Social media bots and stock markets

Article

Accepted Version

Fan, R., Talavera, O. and Tran, V. ORCID: https://orcid.org/0000-0001-9561-8118 (2020) Social media bots and stock markets. European Financial Management, 26 (3). pp. 753-777. ISSN 1468-036X doi: 10.1111/eufm.12245 Available at https://centaur.reading.ac.uk/88903/

To link to this article DOI: http://dx.doi.org/10.1111/eufm.12245

Publisher: Wiley

Social media bots and stock markets

Rui Fan[*], Oleksandr Talavera[**], and Vu Tran[* †]

This version: 08th October 2019

Abstract

This study examines the link between information spread by social media bots and stock trading. Based on a large sample of tweets mentioning 55 companies in the FTSE 100 composites, we find significant relations between bot tweets and stock returns, volatility, and trading volume at both daily and intraday levels. These results are also confirmed by an event study of stock response following abnormal increases in tweets volume. The findings are robust to various specifications including controlling for traditional news channel, alternative measures of volatility, information flows in pre-trading hours, and different measures of sentiment.

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

JEL classification: G12, G14, L86

---

[*] School of Management, Swansea University, Fabian Way, Swansea, SA1 8EN, UK

[**] Birmingham Business School, University of Birmingham, Edgbaston Park Road, Birmingham, B15 2TY, UK

[†] Corresponding author: v.tran@swansea.ac.uk

We thank John Doukas (the Editor) and an anonymous referee for their valuable suggestions. We also would like to thank participants of Swansea research seminar, 2018 Royal Economic Society Annual Conference, and Financial Management Association 2019 European Conference for helpful comments and suggestions. Any remaining errors are of our own. The standard disclaimer applies.

Social media bots and stock markets

This version: 08th October 2019

Abstract

This study examines the link between information spread by social media bots and stock trading. Based on a large sample of tweets mentioning 55 companies in the FTSE 100 composites, we find significant relations between bot tweets and stock returns, volatility, and trading volume at both daily and intraday levels. These results are also confirmed by an event study of stock response following abnormal increases in tweets volume. The findings are robust to various specifications including controlling for traditional news channel, alternative measures of volatility, information flows in pre-trading hours, and different measures of sentiment.

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

JEL classification: G12, G14, L86

1. Introduction

Recent years have witnessed the increasing importance of social media platforms as alternative information sources. Given the enormous volume and the rapid speed of information transmission, social media provides a more comprehensive real-time news database compared to traditional media channels (e.g. Zhang et al., 2011). However, this large amount of instant information can also contain potential noise that might mislead readers. These concerns are more critical given the recent rise of social bots, cybots, and social media farms (e.g. Ferrara et al., 2016), which could be weaponized to disseminate fake news and manipulate stock markets (Forbes, 2017).

Indeed, the influence of social media bot activities on stock markets is not negligible. In our data, there are cases in which a sudden increase in the volume of automated (bot) tweets is associated with significant changes in stock returns. For example, on 1st May 2017, there was an upsurge in the number of 'bot' tweets with positive sentiment for Pearson from 18 to 4,349. On the same day, Pearson stock price increased by 1.01%, compared to a 0.63% rise of FTSE 100. Another case is dated on 29th March 2017. The number of bot-created tweets containing the keyword 'Barclays' increased from 5 to 14,668 (all with negative sentiment) and Barclays stock price decreased by 0.35%. In the following week, Barclays' shares lost over 5%. These concerns and stylized observations lead to a question of whether there is an empirically justified link between information spread by automated social media accounts and stock markets.

Our paper is related to a few strands of literature. First, this paper contributes to the recent literature on social media and the stock market. Multiple studies have suggested that stock market participation and Twitter use are positively correlated (e.g. Bonaparte and Kumar, 2013). Tweets can also be used to forecast aggregate market indexes and individual stock performance (e.g. Zhang et al., 2011; Sprenger et al., 2014a, b). A few papers investigate the link between social media and stock market manipulation (e.g. Renault, 2017b; Al Nasseri et al., 2015). However, these studies do not directly consider the fact that not all messages posted on social media are created by humans. Some Twitter users are bots, automated computer algorithms that are designed to pump intended information into public domains. Given that Twitter bots' tweets are autonomously created and spread, they can potentially contain helpful information, noise, and even unreliable information.[1] Thus, it is reasonable to expect differences in the effects of bot tweets and human tweets on stock markets. This has not yet been considered in the existing literature.

In addition, most existing literature investigates the influence of information spread in professional investing platform and/or social media on daily stock prices (e.g. Sprenger et al., 2014b; Rakowski et al., 2018). However, the use of daily data might not always capture the feature of swift information flows in social media. There are a few recent studies (e.g. Renault, 2017a; Behrendt and Schmidt, 2018) examining the impacts of social media information on intraday stock prices. Renault (2017a) finds that online investor sentiment derived from messages posted on StockTwits can help predict intraday stock returns, while Behrendt and Schmidt (2018) do not find economically meaningful co-movement between intraday volatility and stock-related Twitter information. Our study complements and contributes to this strand of literature in two ways. To the best of our knowledge, this work is the first attempt to investigate the link between stock markets and social media content created by automated accounts. Further, the unique dataset allows us to account for the near-instant information flows in social media that might have a different effect on the stock market versus information from other sources that is typically spread at lower speed.

[1] Several studies have found evidence for the disproportional role of Twitter bots in spreading low-credibility content (e.g. Shao et al., 2018). Additionally, some studies (e.g. Kogan et al., 2018) have examined the impacts of fake news on stock market, while our paper focuses on the information spread by social media bots rather than fake information in general.

Second, this study makes a multidisciplinary contribution to the field of social media bots, politics, and the stock market. Gorodnichenko et al. (2018) detect spillover effects from bots to human activities on social media during political events such as the 2016 Brexit Referendum and 2016 US Presidential Election. However, it is difficult to evaluate the relationship between bots and political outcomes because the latter are not observable in real time. We contribute to this strand of literature by examining the link between real-world outcomes and information spread by Twitter bots. The evidence from financial markets could also be extrapolated to political opinions or other real-life events that could be associated with social media information flows.

Our data come from three sources. First, we have a unique and comprehensive Twitter dataset.[2] About 69.76 million tweets are collected from the Twitter Streaming application programing interface (API) from August 2015 to July 2018. These tweets contain the names of FTSE 100 firms. All Twitter messages have information about the content of the tweets as well as users' metadata such as username and ID, date, location, and friend and follower counts. After excluding less frequently tweeted firms, our final data cover 55 firms with an average number of daily tweets of 100 or more during the sample period.

Second, daily stock prices, volume, and bid-ask spread for the sampled companies are obtained from Datastream. Intraday data are also employed: 5-min stock prices and volume are collected from Tickdatamarket between August 2015 and July 2018.[3] Finally, traditional news data are hand collected from the Financial Times.[4] Thus, there are 37,674 firm-days for daily data and 1,408,538 observations for 5-min intraday data.

[2] Some prior studies use secondary data of aggregate tweet sentiment and other features. For example, Behrendt and Schmidt (2018) use 1-min Twitter count and sentiment data from Bloomberg. Cathcart et al. (2019) extract a measure of media tone from the Thomson Reuters News Analytics database. However, in this study, primary tweet data are collected and allow us to experiment with different measures of size and sentiment of information flows.

[3] Source: www.tickdatamarket.com

[4] Source: www.ft.com. Financial Times is widely recognized as a leading news outlet at least for UK companies

Our main results suggest that Twitter information is related to stock returns, volatility, and trading volume of individual stocks regardless of whether daily or high-frequency data are used. Both human and bot tweets have significant relations with stock features, and the magnitude of the impact of human tweets is larger. Moreover, the volume and sentiment of messages are also significantly associated with stock trading indicators. These results are robust to multiple alternative specifications such as tweets during pre-trading hours, lagged tweet features, as well as alternative measures of volatility, volume, bid-ask spread, and the sentiment of Twitter messages. Further, we use an event study to detect abnormal increases in the volume of tweets and bot-tweets and examine stock responses following these increases. The results show significant associations between tweeting activities and stock volatility as well as trading volume. We also detect a bounce-back pattern of stock prices following increases in the volume of negative retweets.

Our findings raise some important implications. First, the transparency of tweets posted by bots on social media should be enhanced. Moreover, policy makers and regulators should establish a code of practice to monitor social media providers to prevent the spread of fake information. Second, there is a need for resources to mitigate the potential problems arising from a lack of social media literacy because people could be misled by false information. Finally, our study suggests the potential benefits of data availability for researchers. The large amount of data from social media should be made available for investigations and studies. These data in turn play an important role in monitoring (ab)usage of social media networks.

The rest of 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

---

and UK financial news. All relevant news on sampled companies should be covered due to competition between news outlets.

methodology. Section 5 presents the empirical results. Section 6 summarizes the findings and concludes.

2. Related literature

2.1 News, stock message boards, Twitter information intermediaries, and social media bots

Several studies have investigated the relationship between news, information spread in social media platforms such as Twitter, and stock performance. For instance, Dougal et al. (2012) find a causal relation between the Wall Street Journal columnists and Dow Jones Industrial Average daily returns. Local media coverage of earnings announcements of S&P 500 index firms can help to forecast local stock trading (Engelberg and Parsons, 2011). Chen et al. (2014) show that articles and commentaries on a popular online forum, seekingalpha.com, predict future stock returns and earnings surprises. Based on high-frequency data of UK stocks, Gross-Klussmann and Hautsch (2011) find significant reactions of returns, volatility, trading volume, and bid-ask spread in response to news announcements collected from Reuters NewsScope Sentiment Engine.

Multiple researchers have directed their attention to the large amount of qualitative user-generated information from online stock forums, and their findings have been mixed. Wysocki (1998) shows that the posting volume on Yahoo! message boards can forecast next day trading volume and returns of the related stocks. Contrarily, Tumarkin and Whitelaw (2001) 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 a Naive Bayes algorithm to study information from both Yahoo! Finance and Raging Bull for 45 companies in the Dow Jones Industrial Average and the Dow Jones Internet Index. They acknowledge that the volume of such messages can help forecast volatility while the effect on stock returns is statistically significant but economically small. In contrast, Behrendt and Schmidt (2018) argue that stocks-related tweets are not useful to assess and forecast intraday volatility for individual-level stocks. Some other studies have considered the sentiment embedded in messages on stock message boards such as StockTwits, and find the significant link between sentiment and stock index returns (e.g. Das and Chen, 2007; Renault, 2017a).

Since social media platforms such as Twitter provide alternative sources of information for investors, there are several studies examining how information spread in social media could affect investors' beliefs. For example, emotional tweet percentage is linked to three major US stock market indicators, namely the Dow Jones Industrial Average, NASDAQ, and S&P 500 (e.g. Zhang et al., 2011). Moreover, Sprenger et al. (2014a) find significant reactions in an S&P 500 company's stock prices to unusually high tweets volume about that company. Sprenger et al. (2014b) show that there is a strong relationship between Twitter messages sentiment, volume, and individual stock returns, trading volume, and volatility.[5]

However, these studies have not yet accounted for the fact that some messages on social media are posted by bots, automated computer algorithms that could spread information and potentially mislead the general public. This concern has led to an increasing number of studies in computer science documenting the existence of social media bots and their influence. Stukal et al. (2017) acknowledge the use of bots to spread news stories in Russia between 2014 and 2015. Kollanyi et al. (2016) find that bots influenced public opinion during the 2016 EU Referendum and 2016 US Presidential Election. Overall, these studies reveal that bots can imitate humans and are useful in disseminating information. More importantly, human and bot accounts interact with each other, and the former is more likely to retweet/share information

[5] Similarly, Ranco et al. (2015) illustrate 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. Renault (2017b) demonstrate that fraudsters could use social media to manipulate stock prices of small capitalization firms.

than the latter (e.g. Ferrara et al., 2016). Motivated by the gap in finance literature and the findings in computer science literature, we aim to explore the effects of information spread by Twitter bots on the stock market in this paper.

2.2. Psychological bias, the spread of information, and impacts of news among and on investors

An emerging strand of literature argues that the spread of information could affect public opinions during political events. DeMarzo et al. (2003) theoretically show that repeated signals convert multi-dimensional to uni-dimensional (dis)agreements that are in turn effective at polarizing public opinions. Gorodnichenko et al. (2018) empirically reveal that the spread of information by Twitter bots polarized political beliefs during the 2016 Brexit Referendum and 2016 US Presidential Election.

Another strand of relevant literature focuses on how various channels of information dissemination could differently influence investors' behavior. 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 due to limited attention problems. Dimpfl and Jank (2016) find a strong association between Dow Jones index volatility and the volume of retail investors' search queries for its name. Further, trading decisions can be influenced via numerous channels including word of mouth or epidemic types of transmission (e.g. Hong et al., 2005).

There are two main groups of research into the mechanism via which media coverage and public information impact stock markets. The information view states that media coverage can lower the cost of information acquisition and reduce information asymmetry between firms and investors (e.g. Tetlock, 2010; Bushee et al., 2010; Blankespoor et al., 2014). The salience view argues that more media coverage might bring the company more investor attention and investments. For instance, Da et al. (2011) find that increasing attention, proxied by a search volume index, forecasts high stock prices in the next two weeks and big first-day returns after IPOs. Solomon et al. (2012) state that fund holdings with stocks recently covered in the press attract more investments than fund holdings with stocks not featured in the media.

We conjecture that there might be a third view on the mechanism via which public information influences investors' decisions. Since there is vast amount of information in social media, and due to limited attention and costly information acquisitions (e.g. Barber and Odean, 2008), social media (bots) could exacerbate noise trading. Traditional asset pricing rules out the impact of irrationality based on an argument that competition among rational arbitrageurs would eliminate irrational beliefs and noise. However, under certain circumstances, noise trading could dominate the market. In practice, arbitrageurs who enforce the no-arbitrage principle and uphold market efficiency face significant risks and constraints (e.g. Pontiff, 2006). In particular, rational arbitrageurs would require a noise-trader risk premium and could be deterred by the unpredictability of noise traders' misperceptions because such collective misperceptions could induce substantial losses (e.g. DeLong et al., 1990).

3. Tweet features and bot identifications

3.1. Tweet data collection and cleaning

This study uses Twitter Streaming API to collect data.[6] API can be treated as an interface between users and the system. The interface passes the inquiries raised by users to the system and then returns the responses to the users. We make requests to collect tweets with a FTSE 100 company name and have a random sample of all tweets containing any FTSE 100 company

[6] API can collect real-time tweets with pre-determined characters. More information can be found at https://developer.twitter.com/en/docs/tweets/filterrealtime/guides/powertrack_rules_and_filtering, accessed on 28 June, 2018.

names. The collected information includes the tweets' content, metadata, username and ID, date, location, friend, and follower counts. Approximately 69.76 million tweets containing names of FTSE 100 companies are collected from 1st August 2015 to 31st July 2018.[7] Daily stock data for FTSE 100 companies are acquired from Datastream from 1st January 2014 to 31st July 2018 to estimate abnormal measures of stock returns. The 5-min stock data are obtained from Tickdatamarket between August 2015 and July 2018.

An extra cleaning process is applied to make sure that our Twitter data capture the sound information flows in social media. In particular, we cannot obtain satisfactory Twitter data for some companies (e.g. 'Aberdeen Asset Management'), probably because Twitter users are not likely to post long company names. Therefore, we exclude all companies with fewer than 100 average daily tweets. This leaves us with a final sample of 55 companies (full list in Table A1 the online Appendix).

Following Kollanyi et al. (2016), tweets are cleaned in three steps. First, special characters in tweets are deleted such as link tokens (starting with 'http', 'https', 'www'), hashtag tokens (starting with '#'), and user identifier tokens (starting with '@') from the tweet messages. Second, all tweets containing only links or URLs are deleted. Finally, all non-English tweets are excluded.

3.2. The 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 using TextBlob. TextBlob is a text-processing tool in Python that returns a polarity score for each tweet posting.[8, 9] The polarity

[7] FTSE 100 composites as of 1st January 2014 are used.

[8] Refer to Loria (2018) for more details about TextBlob.

[9] The correlation between the sentiment using TextBlob and Renault (2017a) social media lexicon is 0.7 for the sampled companies. This result is conditional on that the sentiment is detected (at least one word is defined as either positive or negative).

scores range from -1 to 1: a negative (positive) score indicates that the sentiment is negative (positive) while 0 suggests that the sentiment is neutral. Both PatternAnalyzer and NaiveBayesAnalyzer in TextBlob are employed to perform sentiment analysis, and the same sentiment score is obtained for each tweet posting. Table A3 in the online Appendix gives examples of how the polarity scores of tweets in our sample are obtained using TextBlob.

3.3. Humans vs. automated bots Twitter accounts

There is no perfect procedure to recognize bot tweets because bots/cybots can imitate human behavior (Haustein et al., 2016). That being said, several studies (Chu et al., 2010; Cook et al., 2014) have discussed the criteria that can help to separate human and bot accounts. First, human users are more likely to tweet on weekdays and during the day; the tweeting time of bot accounts does not follow these patterns. Moreover, Haustein et al. (2016) document that the daily average number of posts is about five for bot accounts and two for human accounts. Finally, Lee et al. (2010) find that bot agents are more likely to constantly post similar messages whereas human accounts do not. Based on these, we identify a Twitter account as a potential bot agent if any one of the following three criteria about suspicious tweeting activities is met: (i) at least five tweets during abnormal tweeting times, i.e., from 0:00 to 6:00 am[10]; (ii) more than ten tweets are posted a day; and (iii) repeating the same tweet content three times or more on one day. A Twitter account is classified as a bot account if we record suspicious activities on more than 50% of active days during the sampled period. For example, if an account has tweeting activities for ten days, then it is detected as a bot account if this account is flagged as

[10] Our time is always London time, which is Greenwich Mean Time (GMT) in the winter and British Summer Time (BST) in the summer.

having suspicious bot behavior on more than five days.[11] To cross-check the validity of our bot detection, we employ Botometer (previously BotOrNot). This is an online social media bot detection tool developed by researchers from Indiana University and Northeastern University.[12] Our analyses based on these techniques are consistent with each other.

A reader might be interested in what type of information bots actually spread. To provide a rough answer to this question, we extract the most frequent words from the content of bot tweets. The mentioned firms are categorized into industry groups, namely manufacturing, non-financial services, or financial services industries (See Table 1). In general, bot accounts that mention manufacturing and non-financial services firms are likely to post advertisement related to products and services. These bots also mention words related to firm events: 'deal', 'launch', 'event', and 'merger'. When bots mention financial services firms, they are more likely to mention financial terms such as 'rate', 'profit', 'hold', 'keep', 'target', 'buy', or 'loss'.

3.4. Original tweets and retweets

Original tweets, which are posted for the first time, are separated from their retweets. First, 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. This is the username of the Twitter account from which the tweet was retweeted. Consequently, the original tweets and their retweets can be identified.

---

[11] We try alternative threshold values for every criterion and identify an account as bot agent if tweeting behavior satisfies all criteria on more than half of the days and for three or more days, and we obtain similar results.

[12] More information can be found in Davis et al. (2016). A threshold of 0.5 is used to detect automated accounts.

4. Aggregation of tweet information and estimation framework

4.1. Tweet posting aggregation

All daily tweets are aggregated to examine the relationship between all tweet postings and stock price changes on a daily basis. We explore the relationship between market features (stock returns, trading volume, volatility, and bid-ask spread) and tweet features (positiveness, message volume, and agreement). Similar to Antweiler and Frank (2004), our aggregate sentiment measure is given as:

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

where $M_t^{positive}$ and $M_t^{negative}$ are the counts of positive and negative tweets on day *t*. The tweet message volume is the natural logarithm of the count of all tweet messages containing one sample company name on day *t*. Tweet agreement is defined as follows:

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

If all tweet messages are positive or negative, then the agreement among all messages equals 1. Tweet and stock feature data are available for most company-day observations. Finally, consistent with London Stock Exchange's 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. Tweets posted during weekends are pooled to the following Mondays.

4.2. Stock indicators

Following prior literature (e.g. Antweiler and Frank, 2004), we measure abnormal return ($AR_{i,t}$) as the difference between the log-return ($R_{i,t}$) and the expected return $E(R_{i,t})$ on a given day. Expected return is calculated using the market model. Following previous literature, a

100-day estimation period starting 110 days before the relevant date is used.[13]

Two measures of volatility are used in our analysis. First, Parkinson (1980) daily volatility, which is based on intraday high and low stock prices ($S_{t,high}$, $S_{t,low}$), is estimated as follows:

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

Second, an abnormal change in the volatility measure equals volatility today minus the average volatility over the past 100 trading days (i.e., [-110, -10]). Similarly, trading volume is defined as the natural logarithm of the number of shares traded on a given day. A similar measure of abnormal changes in trading volume as above is used. Finally, the bid-ask spread is the logarithm of the difference between the bid and ask quotes. It is measured by basis points, and we make an identical transformation to obtain abnormal changes in the bid-ask spread.

4.3. Summary statistics

Table 2 presents the descriptive statistics of market and tweet features during the sampling period. The average number of daily tweets is 1,852, and the standard deviation is around 4,901 tweet postings per day. A large number of Twitter messages per company per day indicates that our sample comprises a sound information flow. Meanwhile, the average number of tweets generated by automated bot accounts is significantly smaller, about 97 messages. Notably, tweets generated by bots are more negative than tweets generated by humans. The positiveness measure from bot-tweets is 0.5810, which is smaller than the counterpart from all tweet messages, 1.0615. There are also significant correlations between trading volume, volatility,

[13] Similarly, an alternative 1-year estimation period yields quantitatively similar results as shown in Table D11 in the online Appendix. There is a caveat against very long estimation periods since stock characteristics such as beta are not time invariant.

bid-ask spread, and tweet features.[14]

Figure 1 depicts weekly aggregate features from human-posted and bot-posted tweets on the 55 sampled firms. During the week commencing 12th September 2016, there were fewer tweet activities (< 300,000 tweet messages). The highest number of tweets was around 870,000 in the week beginning February 2016 and the week ending January 2017. Furthermore, we observe a significant correlation between human and bot tweet volume (Figure 1). Sentiment measures from human-posted and bot-posted tweets are also correlated but to a much lesser extent (Figure 1).

4.4. Empirical specification

The baseline regression is as follows:

$$y_{i,t} = \alpha + \beta_1 Positiveness_{i,t} + \beta_2 Message_{i,t} + \beta_3 Agreement_{i,t} + \delta_1 TNews_{i,t} + \delta_2 R_{i,t-1} + \delta_3 R_t^{FTSE} + u_i + \varepsilon_{i,t} \quad (3)$$

where i stands for firm and t denotes time, and $y_{i,t}$ are the individual stock returns, volatility, trading volume, or bid-ask spread.

Return is calculated as the log return, and the market-model abnormal return controls for the stock's systematic risk. The market return is the FTSE 100 index return.

Normalized trading volume is the natural logarithm of the number of shares traded. Volatility is the Parkinson (1980) intraday volatility and bid-ask spread is the logarithm of the difference between bid and ask quotes scaled by 10,000.

We also perform regressions with three more measures of volatility, volume, and bid-ask spread: abnormal volatility (volume/bid-ask spread) is the volatility (volume/bid-ask spread)

[14] See the online Appendix for detailed results.

minus the average from a 100-day estimation period starting 110 days before the relevant date.

The $Positiveness_{i,t}$ variable is defined using equation (1) to examine the impact of sentiment on stock market features. Tweets created between 4:30 pm on the previous trading day, when the London Stock Exchange closes, and 4:29 pm on the current day are used. Tweets posted after the market closes can only affect stocks on the next day. Consistent with prior literature (e.g. Ranco et al., 2015), a positive (negative) coefficient of positiveness in explaining returns (volatility, trading volume) is expected. $Message_{i,t}$ is defined as the natural logarithm of the number of tweets between 4:30 pm on the previous trading day and 4:29 pm today. In line with Antweiler and Frank (2004), we anticipate a positive coefficient of message to explain volatility and trading volume but not returns. $Agreement_{i,t}$ describes the extent that tweets agree with or are different from each other, i.e., similar or very different number of positive versus negative tweets. Compatible with Sprenger et al. (2014b), negative coefficients of agreement to explain returns, volatility, and trading volume are anticipated.

We control for a number of relevant factors. Following Chen et al. (2014), we include the number of traditional news ($TNews_{i,t}$) related to firm *i* during time period *t*, proxied by news on Financial Times.[15] Other control variables include lagged return ($R_{i,t-1}$), market return (FTSE 100 return) R_i^{FTSE}, and firm individual fixed effects u_i. Finally, $\varepsilon_{i,t}$ is the error term.[16]

5. Results discussion

5.1. Relation of tweet and market indicators

Table 3 reports fixed effects panel regressions for two return measures, volatility, volume,

[15] Robustness checks controlling for sentiment of traditional news provide quantitatively similar results, as reported in Tables D12 to D14 in the online Appendix.

[16] See a full list of definition of variables in Table A2 in the online Appendix.

and bid-ask spread as the dependent variables. The key independent variables are the three aggregate features (positiveness, volume, and agreement) from all tweets containing a sampled firm's name. In line with Chen et al. (2014), we find a significant link between aggregate sentiment in messages and stock returns, and the magnitude of this effect (0.0218) is economically large because the mean return and market model return are 0.0166 and -0.0056 (Table 2). There is no significant relationship between Twitter messages volume and stock returns and bid-ask spread.[17] This is different from Wysocki (1998), who documents a significant relationship between the message volume of stock message boards and stock returns. The results reinforce our initial concerns that—in contrast to information on professional stock forums, information flows in social media potentially contain significant noise.

There are strongly significant correlations of tweet indicators with volatility and trading volume. The coefficients of message volume are positive and statistically significant in explaining variation in volatility and trading volume, a 1% increase in tweet volume is associated with a 0.07% increase in trading volume. We also detect statistically significant relations between positiveness and volatility, trading volume, and between agreement measure and trading volume. This is consistent with our expectation and the prior literature (e.g. Sprenger et al., 2014b). However, the magnitude of these coefficients is much smaller than those of the tweet message measure.

Next, we split tweet-based measures into human-posted and bot-posted. The independent variables are Twitter features aggregated from human and bot tweets. Table 4 shows significant relations between sentiment of both human and bot tweets and stock returns. This effect is economically meaningful. This effect is economically meaningful: 1% increase in human (bot) tweets positiveness measure of sentiment is associated with 0.0206% (0.0165%) increase in

[17] Robustness investigations based on Fama-MacBeth regressions yield quantitatively similar results, as shown in Table D15 in the online Appendix.

returns. Positive (negative) human and bot tweets are associated with increases (decreases) in the two return measures. Moreover, there are strong associations between human and bot tweet features and, volatility and trading volume. The coefficients of both human and bot message volume are significantly positive in explaining the variation in volatility and trading volume. However, there are very few significant relations between Twitter features and bid-ask spread.

In our further investigation, we employ the interaction terms between human and bot tweet characters as additional independent variables. We find that bot tweet message volume could reinforce the human tweet message volume in explaining the changes in volatility and trading volume (Table 5). The marginal effects of human and bot tweet messages on volatility and trading volume are also illustrated in Figure 2. A higher current level of human (bot) tweets on a given stock leads to a stronger impact of the bot (human) tweets. For example, if the current level of human tweets is at the 10th (90th) percentile level, then a 1% increase in bot messages leads to 0.03% (over 0.04%) increase in trading volume. The average number of bot tweets during the sampled period is below 97 (see Table 2). It is therefore affordable to increase these tweets more than a few percentage points. This gives practical implications for market participants and regulators that, a few hundred bot (human) tweets posted at a right moment could potentially spark an increase of several percentage points in trading volume.

Tables 4 and 5 show that information extracted from both human and bot tweets is associated with significant changes in all stock indicators, and the impacts of human tweets are larger. This result implies that investors are likely to treat the repeated tweets, which are mainly spread by bots, as noise and thus less likely to act on this information. Nonetheless, there are significant increases in volatility and trading volume associated with bot-tweets. This is in support of Enikolopov et al. (2018), who argue that non-institutional investors tend to be affected by tweet postings. Furthermore, they might only change the trading volume but not the returns. Also, our measure of volatility (i.e., the high-low range) can be interpreted as a measure of ex-post disagreement among market participants on a given day. Increased volatility is associated with bot-tweets (reported in Tables 4 and 5). This supports the argument that tweets posted by bots are more likely to cause disagreement among investors' views on a stock's fundamental value and hence its fair price.

5.2. Intraday analysis

The real-time nature of Twitter suggests that investigations using intraday data might provide further insights. Table 6 presents regression results based on 5-min stock prices from August 2015 to July 2018. There are significant relationships between tweets and stock returns, volatility, and trading volume. The impact on returns and volatility are economically significant. A 1% increase in tweet volume is associated with a 0.11% decrease and a 1.2% increase in returns and volatility respectively. These indicate that the information from tweets is disseminated in the stock market mainly through disagreement among market participants.

In Table 7, we decompose tweets posted by humans and bots and estimate the intraday regression using 5-min data between August 2015 and July 2018. The results are similar to those in Table 6. We find a significant relationship between tweets and trading volume, and statistically significant and economically meaningful coefficients of tweet features on returns and ex-post volatility. Our results are consistent with prior literature based on high-frequency data. Gross-Klussmann and Hautsch (2011) find that news sentiment indicators can forecast future stock returns, volatility, trading volume, and bid-ask spread when using 20-s prices of UK stocks. Moreover, Renault (2017a) shows that online investor sentiment could help predict intraday stock index returns when grouping news sentiment into half-hour intervals. Thus, a trading strategy based on intraday sentiment-driven noise trading to buy (sell) in the last half-hour trading interval and sell (buy) before the market closes is proposed. Similarly, our findings also suggest profitable trading strategies based on 5-min trading intervals.

Table 8 shows the results of regressing lagged tweet features on market features based on 5-min data. The results confirm significant relations between information in tweets and bot-tweets with stock returns, volatility, and trading volume. The effects of tweets on returns and volatility are economically significant. For example, a 1% increase in human tweet volume is associated with a 0.17% decrease in returns. These results provide further evidence that the sentiment (i.e., positiveness and agreement) and volume of tweets are associated with stock returns, trading volume, and disagreement among market participants (i.e., ex-post volatility when tweet features are collected from the previous 5-min intervals). Our analysis is useful for investors to establish profitable trading strategies at high-frequency intervals, and for regulators to closely monitor the use of Twitter to disseminate information about the stock market.

The intraday analysis results indicate that information spread by Twitter bot accounts is related to volatility and trading volume of individual stocks, which is consistent with our daily analysis. In line with our daily analysis results, we note a significant relationship between information disseminated by Twitter bots and stock returns when 5-min stock prices are used.

5.3. Event study

We also use an event study to identify abnormal increases in volume of tweets and bot tweets and examine the impacts of the 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 volume of tweets or bot-tweets is larger than 100%; and 3) the absolute increase in volume of tweets or bot-tweets is greater than 500. Figure C1 in the online Appendix 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 most of the sampling period. The highest number of events per day is 29, which occurred during the week commencing 23rd January 2017.

Panel A of Table 9 reports changes in returns following different types of events. There is no significant relationship on the event days (0) when there are abnormal increases in the volume of positive tweets. In contrast, stock prices decrease significantly when there are abnormal increases in negative tweet volume. All changes are reported in percentages. Hence, the magnitude of reduction is 0.38%. However, we detect 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. If we focus on tweets posted by bots, then we find no significant relationship on the event days (0) when the tweet volume has abnormal increases, but there are significant negative returns in the following week (1, 5) after abnormal increases in positive tweet volume. Panel B of Table 9 reports equivalent event study results while returns are market-model abnormal returns. 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. The effect of returns is economically relevant.

Panel C of Table 9 presents the responses of an ex-post volatility measure to abnormal increases in tweet volume. There are strongly significant and positive associations for all time windows. This implies that abnormal increases in the volume of both positive and negative tweets, and tweets posted by bots could deepen the ex-post disagreement among market participants. Panel D of Table 9 shows the relationship between normalized trading volume and abnormal increases in tweet volume. There are statistically significant and economically meaningful associations in trading volume during all time windows after abnormal increases in tweet volume regardless of the type of tweets: positive tweets, negative tweets, or bot tweets.[18]

[18] Following abnormal increases in the volume of all tweets, trading volume increases over 9.1%+17.9%=27.0% on average among all events in the next week.

Panel E of Table 9 reports the changes in bid-ask spread, an inversed measure of liquidity, after abnormal increases in tweet volume. Again, we find statistically and economically significant associations in the bid-ask spread.[19] Human tweets are separated into positive and negative tweets, and correlations are similar to the positive and negative bot tweets in Table 9. All tweets are further decomposed into original tweets and retweets, and we find that the bounce-back pattern of stock prices is related to abnormal increases in the volume of negative retweets but not original tweets.[20]

Overall, this event study suggests that information embedded in tweets is strongly linked to market participants' disagreement (i.e., ex-post volatility, trading volume, and liquidity); and there is a bounce-back pattern of stock prices only in relation to negative tweets. Again, these findings support the argument that tweets are more likely to affect non-institutional investors.

5.4. Robustness checks

It is possible that Twitter bots respond to changes in market conditions instead of causing them.[21] To partially mitigate this potential issue, we conduct 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 8:00 am today as pre-trading, and between 8:00 am and 4:30 pm today as trading. Contemporaneous regressions using tweet features collected during pre-trading hours yield similar results (Table 10). There are a few significant associations between tweet indicators and stock returns or bid-ask spread, but we find statistically significant relations between tweet features (positiveness, message volume, and agreement) and volatility

[19] After the event (abnormal increases in tweets volume), the mean change of bid-ask spread decreases by 15.1+52.9=68.0 basis points the following week.

[20] See Tables C1 to C3 in the online Appendix for more details.

[21] Intuitively, this potential econometric issue is not very plausible because bot accounts holders have less incentives to simply react to market conditions. If they do react, then market conditions are exacerbated. Our findings of the association between bot tweets and stock performance remain highly relevant.

and trading volume.

To account for persistence of volatility, GARCH-type models are employed instead of the measure based on intraday high-low range. This setup also allows asymmetry between positive and negative shocks (Glosten et al., 2003). The results of GARCH (1, 1) and GJR-GARCH (1, 1) models are quantitatively similar to the main findings (reported in Tables D1 and D2 in the online Appendix).[22, 23] Across all models, the coefficients of bot message are strongly significant and positive. Increases in bot tweets are linked to heightened volatility after controlling for the long memory characteristics of volatility. Positiveness of bot tweets reduces volatility, implying that more negative bot tweets are associated with higher volatility. There is also evidence that bot and human tweets interact and reinforce each other. The interaction seems to be stronger during pre-trading periods.

Further robustness checks are conducted to investigate the lagged relations between tweet and market features, and the 1-day lagged tweet features are regressed on market features. Similar results are obtained and reported in Table D3 in the online Appendix. Specifically, there is no statistically significant association between lagged positiveness and stock returns. Contrary to Antweiler and Frank (2004), we detect few statistically significant relations between message volume and stock returns. This might be caused by the differences in datasets and subjects in our study. Investors might be more thoughtful when assessing the information content of Twitter messages compared to professional stock forums. Importantly, there are still statistically significant associations between 1-day lagged positiveness, message volume, and volatility, trading volume. We also use alternative measures of volatility (volume and bid-ask

[22] One-step ahead estimation from GARCH and GJR models is used as dependent variables for volatility regressions, and this further mitigates the backward causality issue, i.e., market volatility causes changes in tweet features.

[23] Hansen and Lunde (2005) show that the GARCH (1, 1) model outperforms within a class of 330 GARCH-type models.

spread) and obtain similar results (Table D4 in the online Appendix).

Finally, two alternative aggregate measures of sentiment are employed as robustness checks. Following Brown and Cliff (2005), a positive-negative spread is calculated as follows:

$$Spread_t = \frac{M_t^{positive} - M_t^{negative}}{Total\ tweets_t} \tag{4}$$

where $M_t^{positive}$ and $M_t^{negative}$ are the counts of positive and negative tweets on day t.

We also employ the standardized percentages of positive/negative tweets as used in Tetlock et al. (2008), and negativeness is given as follows:

$$Negativeness_t = \frac{Neg_t - \mu_{Neg}}{\sigma_{Neg}} \tag{5}$$

here, $Neg_t = \frac{M_t^{negative}}{Total\ tweeets_t}$, and μ_{Neg}, σ_{Neg} are the mean and standard deviation from each company's empirical distribution of Neg_t.

Both robustness measures yield quantitatively similar results, as reported in Tables D5-D10 in the online Appendix. These results collectively confirm that both human and bot tweets are influential with strong evidence of interaction/reinforcement. The effect is particularly robust regarding trading volume and volatility.

6. Conclusions

Social media has become a popular platform for information sharing and acquisition. However, its convenience and popularity also come with threats. Recent literature and events have intensified the focus 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. US Intelligence Committee, 2018). Scientific evidence of bots' associations with actual outcomes, however, is limited given the fact that political beliefs are hard to measure at a reasonable frequency. This paper employs characteristics of FTSE 100 composites as an alternative measure of actual outcomes. Specifically, we investigate whether the volume and sentiment of tweets/bot-tweets are associated with abnormal changes in stock indicators using a sample of 55 companies in the FTSE 100 composites during the period from August 2015 to July 2018.

Based on the daily frequency, we find significant relations between tweets/bot-tweets and stock returns, volatility, and trading volume. This indicates that information embedded in social media can help to forecast certain stock features. In addition, there is evidence of significant interaction between the volume of tweets posted by humans and bots in affecting stock market. Given the fact that humans are more likely to retweet bots (e.g. Ferrara et al., 2016), this finding suggests a seeding role of bots in magnifying noise trading or any potential market manipulation. The 5-min stock prices are also used to perform an intraday analysis. There are associations between tweets (bot-tweets) and stock returns, volatility, and trading volume, which are consistent with the daily data analysis. Our results are robust to multiple alternative specifications, such as tweets during pre-trading hours, lagged tweets, GARCH and GJR volatility, and alternative tweets sentiment measures.

Following existing literature on message volume and sentiment (e.g. Wysocki, 1998; Chen et al., 2014), an event study is conducted to identify stock responses following abnormal increases in tweets and bot-tweets volume. The results show that information embedded in tweets is strongly linked to market participants' disagreement (i.e., ex-post volatility, trading volume and liquidity). There is a bounce-back pattern of stock prices when there are increases in the volume of negative retweets.

Our findings collectively reveal evidence of social media bots' relations with real outcomes. Particularly, abuses of social media bots could provoke instability (i.e., to intensify public heterogeneous beliefs and opinions polarization). Compared to institutional investors, small investors are more sensitive to such information flows.[24] Our findings have several implications. First, to prevent the potential spread of fake information, this paper emphasizes the transparency of information posted on social media and a proper code of practice for social media. Besides, there are needs to raise social media literacy and attention on the uses and abuses of social media bots. Finally, our study suggests the potential benefits of making social media data available for future research.

[24] See Ben-Rephael et al. (2017) for a more detailed discussion about institutional and individual investors and stock market.

References

Al Nasseri, A., Tucker, A., & De Cesare, S. (2015). Quantifying StockTwits semantic terms' trading behavior in financial markets: An effective application of decision tree algorithms. *Expert Systems with Applications, 42(23)*, 9192-9210.

Antweiler, W., & Frank, M.Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *Journal of Finance, 59(3)*, 1259-1294.

Barber, B., & 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(2)*, 785-818.

Behrendt, S., & Schmidt, A. (2018). The Twitter myth revisited: Intraday investor sentiment, Twitter activity and individual level stock return volatility. *Journal of Banking and Finance, 96*, 355-367.

Ben-Rephael, A., Da, Z., & Israelsen, R.D. (2017). It depends on where you search: Institutional investor attention and underreaction to news. *Review of Financial Studies, 30(9)*, 3009-3047.

Blankespoor, E., Miller, G.S., & White, H.D. (2014). The role of dissemination in market liquidity: Evidence from firm's use of Twitter. *Accounting Review, 89(1)*, 79-112.

Bonaparte, Y., & Kumar, A. (2013). Political activism, information costs, and stock market participation. *Journal of Financial Economics, 107(3)*, 760-786.

Brown, G.W., & Cliff, M.T. (2005). Investor sentiment and asset valuation. *Journal of Business, 78(2)*, 405-440.

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

Cathcart, L., Gotthelf, N.M., Uhl, M., & Shi, Y. (2019). News sentiment and sovereign credit risk. *European Financial Management*, forthcoming.

Chen, H., De, P., Hu, Y., & Hwang, B-H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies, 27(5),* 1367-1403.

Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2010). Who is tweeting on Twitter: human, bot, or cyborg? In *Proceedings of the 26th Annual Computer Security Applications Conference*, 21-30, ACM.

Cook, D.M., Waugh, B., Abdipanah, M., Hashemi, O., & Rahman, S.A. (2014). Twitter deception and influence: Issues of identity, slacktivism, and puppetry. *Journal of Information Warfare*, *13(1)*, 58-IV.

Da, Z., Engelberg, J.E., & Gao, P. (2011). In search of attention. *Journal of Finance, 66(5),* 1461-1499.

Das, S., & Chen, M. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science, 53(9),* 1375-1388.

Davis, C.A., Varol, O., Ferrara, E., Flammini, A., & Menczer, F. (2016). BotOrNot: A system to evaluate social bots. Available at https://arxiv.org/pdf/1602.00975.pdf.

DeLong, J.B., Shleifer, A., Summers, L.H., & Waldmann, R.J. (1990). Noise trader risk in financial markets. *Journal of Political Economy, 98(4),* 703-738.

DeMarzo, P.M., Vayanos, D., & Zwiebel, J. (2003). Persuasion bias, social influence, and unidimensional opinions. *Quarterly Journal of Economics, 118(3),* 909-968.

Dimpfl, T., & Jank, S. (2016). Can internet search queries help to predict stock market volatility? *European Financial Management*, *22(2),* 171-192.

Dougal, C., Engelberg, J.E., Garcia, D., & Parsons, C.A. (2012). Journalists and the stock market. *Review of Financial Studies*, *25(3)*, 639-679.

Engelberg, J.E., & Parsons, C.A. (2011). The causal impact of media in financial markets. *Journal of Finance*, *66(1)*, 67-97.

Enikolopov, R., Petrova, M., & Sonin, K. (2018). Social media and corruption. *American Economic Journal: Applied Economics*, *10(1)*, 150-174.

Ferrara, E., Varol, O., Davis, C.A., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, *59(7)*, 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, 2018).

Glosten, L.R., Jagannathan, R., & Runkle, D.E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance, 48(5)*, 1779-1801.

Gorodnichenko, Y., Pham, T., & Talavera, O. (2018). Social media, sentiment and public opinions: Evidence from #Brexit and #USElection. *NBER Working paper* 24631.

Gross-Klussmann, A., & 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(2)*, 321-340.

Hansen, P.R., & Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH (1, 1)? *Journal of Applied Econometrics*, *20(7)*, 873-889.

Haustein, S., Bowman, T.D., Holmberg, K., Tsou, A., Sugimoto, C.R., & Larivière, V. (2016). Tweets as impact indicators: Examining the implications of automated "bot" accounts on Twitter. *Journal of the Association for Information Science and Technology, 67(1)*, 232-238.

Hirshleifer, D., & Teoh, S.H. (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics*, *36(1-3)*, 337-386.

Hong, H., Kubik, J.D., & Stein, J.C. (2005). The neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers. *Journal of Finance*, *60(6)*, 2801-2824.

Kogan, S., Moskowitz, T.J., & Niessner, M. (2018). Fake news: Evidence from financial markets. *Working paper*, *Massachusetts Institute of Technology*, *Yale University and AQR Capital Management*.

Kollanyi, B., Howard, P.N., & Woolley, S.C. (2016). Bots and automation over Twitter during the US Election. *Data memo* 2016.4. Oxford, UK: Project on computational propaganda.

Lee, K., Caverlee, J., & Webb, S. (2010). Uncovering social spammers: social honeypots + machine learning. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 435-442, ACM.

Loria, S. (2018). TextBlob Documentation. (Online) Available at: https://media.readthedocs.org/pdf/textblob/dev/textblob.pdf (Accessed 14 September 2018).

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of Business, 53(1)*, 61-65.

Pontiff, J. (2006). Costly arbitrage and the myth of idiosyncratic risk. *Journal of Accounting and Economics*, *42*, 35–52.

Rakowski, D.A., Shirley, S.E., & Stark, J.R. (2018). Is all that Twitters gold? Social media attention, stock returns, and the spread of information. *Working paper*, *University of Texas at Arlington and Middle Tennessee State University*.

Ranco, G., Aleksovski, D., Caldarelli, G., Grcar, M., & Mozetic, I. (2015). The effects of Twitter sentiment on stock price returns. *Plos One*, *10(9)*, 1-21.

Renault, T. (2017a). Intraday online investor sentiment and return patterns in the US stock market. *Journal of Banking and Finance, 84*, 25-40.

Renault, T. (2017b). Market manipulation and suspicious stock recommendations on social media. *Working paper*, *IESEG School of Management*.

Shao, C., Ciampaglia, G.L., Varol, O., Yang, K.C., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature Communications*, *9*, 1-9.

Solomon, D.H., Soltes, E., & Sosyura, D. (2014). Winners in the spotlight: media coverage of fund holdings as a driver of flows. *Journal of Financial Economics*, *113(1)*, 53-72.

Sprenger, T.O., Sandner, P.G., Tumasjan, A., & Welpe, I.M. (2014a). News or noise? Using

Twitter to identify and understand company-specific news flow. *Journal of Business Finance and Accounting, 41*, 791-830.

Sprenger, T.O., Tumasjan, A., Sandner, P.G., & Welpe, I.M. (2014b). Tweets and trades: The information content of stock microblogs. *European Financial Management*, *20(5),* 926-957.

Stukal, D., Sanovich, S., Bonneau, R., & Tucker, J.A. (2017). Detecting bots on Russian political Twitter. *Big Data*, *5(4)*, 310-324.

Tetlock, P.C., Saar-Tsechansky, M., & Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *Journal of Finance*, *63(3)*, 1437-1467.

Tetlock, P.C. (2010). Does public financial news resolve asymmetric information? *Review of Financial Studies*, *23(9)*, 3520-3557.

Tumarkin, R., & Whitelaw, R.F. (2001). News or noise? Internet postings and stock prices. *Financial Analysts Journal, 57(3),* 41-51.

US 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.D. (1998). Cheap talk on the web: The determinants of postings on stock message boards. *Working paper 98025, University of Michigan Business School*.

Zhang, X., Fuehres, H., & Gloor, P.A. (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.

Table 1: Most mentioned words in bot tweet activity spikes

This table reports most mentioned words in large number bot tweet activities by categorizing the mentioned firms into three industries, i.e. manufacturing, non-financial services and financial services. Words are lemmatized before counting. 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 100.

Industry	Most frequently mentioned words
Manufacturing	bag, check, handbag, preview, job, exhibit, fashion, new, leather, deal, men, event, launch, authentic, beauty
Non-Financial Services	jump, live, count, day, tour, say, idea, get, merger, deal, talk, new, ready, come, shop
Financial Services	read, rate, bank, profit, hold, keep, price, headquarter, target, analyst, buy, loss, move, financial, stay

Table 2: Descriptive statistics

This table reports summary statistics. The sample covers 55 of FTSE 100 composites from 1st August 2015 to 31st July 2018. Returns are log returns, market-model abnormal return is based on market model that reflects stock's idiosyncratic risk, market return is FTSE 100 index return. Volume is the number of shares traded, and we calculate Parkinson (1980) intraday volatility using daily high and low prices. Traditional news is the number of relevant news on Financial Times. Positiveness is given as $Positiveness_t = \ln\left(\frac{1+M_t^{positive}}{1+M_t^{negative}}\right)$, where $M_t^{positive}$ and $M_t^{negative}$ are the counts of positive and negative tweets on day t, and Agreement is $Agreement_t = 1 - \sqrt{1-\left(\frac{M_t^{positive}-M_t^{negative}}{M_t^{positive}+M_t^{negative}}\right)^2}$. Human (bots) tweet features are based on human (bots) tweets.

	Mean	Std. Dev.	No. of Obs.
Return (%)	0.0166	1.8184	37,674
Mrk-model return (%)	-0.0056	1.5262	37,674
Trading (log) volume	8.3334	1.4613	37,674
Abnormal trading (log) volume	0.0026	0.4355	37,674
Volatility i.e. high-low range (%)	1.3210	0.9504	37,674
Abnormal volatility (%)	-0.0030	0.8756	37,674
Bid-ask spread (bpts)	5.6131	17.0815	37,674
Abnormal bid-ask spread (bpts)	0.0760	17.0179	37,674
Traditional News	0.0608	0.2726	37,674
No. of tweets	1851.8089	4900.9761	37,674
Positiveness	1.0615	1.0018	37,674
Agreement	0.2550	0.2815	37,067
No. of tweets generated by humans	1755.3040	4709.4215	37,674
Positiveness of human tweets	1.0613	0.9926	37,674
Agreement of human tweets	0.2655	0.2920	36,977
No. of tweets generated by bots	96.5049	359.4532	37,674
Positiveness of bot tweets	0.5810	1.1130	37,674
Agreement of bot tweets	0.3864	0.4411	37,067

Table 3: Regressions for all tweets

This table reports fixed-effects regressions of stock indicators by tweet characteristics. Dependent variables are (log) return, market-model return, volatility i.e. Parkinson (1980) intraday high-low range, (normalized) trading volume, and bid-ask spread. Main independent variables are aggregate Twitter characters based on tweets collected from 4:30 pm day -1 to 4:29 pm day 0. Message is the natural logarithm of the number of tweets, Positiveness is given as $Positiveness_t = \ln\left(\frac{1+M_t^{positive}}{1+M_t^{negative}}\right)$, where $M_t^{positive}$ and $M_t^{negative}$ are the counts of positive and negative tweets on day t, and Agreement is $Agreement_t = 1 - \sqrt{1-\left(\frac{M_t^{positive}-M_t^{negative}}{M_t^{positive}+M_t^{negative}}\right)^2}$. Traditional news is the number of relevant news on Financial Times. T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 10%, 5%, 1% significance, respectively.

	Return	Mrk-model return	Volatility	Volume	Bid-Ask
	(1)	(2)	(3)	(4)	(5)
Positiveness	0.0218***	0.0218**	-0.0474***	-0.0065***	-0.0041
	(2.77)	(2.35)	(-5.26)	(-2.90)	(-1.19)
Message	-0.0220	-0.0214	0.1630***	0.0670***	-0.0001
	(-1.64)	(-1.36)	(10.34)	(17.47)	(-0.01)
Agreement	-0.0032	-0.0038	-0.0088	-0.0104***	0.0071
	(-0.48)	(-0.48)	(-1.18)	(-4.60)	(0.83)
Traditional News	-0.0022	-0.0003	0.0947***	0.0307***	-0.0031
	(-0.27)	(-0.04)	(13.90)	(17.36)	(-0.69)
Lagged Return	0.0205*	0.0205	-0.0538***	-0.0135***	0.0083
	(1.85)	(1.53)	(-3.78)	(-6.19)	(1.28)
FTSE100 Return	0.5293***	0.0283***	-0.0899***	-0.0182***	-0.0044
	(79.67)	(3.67)	(-11.09)	(-10.89)	(-1.00)
Observations	37,619	37,619	37,619	37,619	37,619
R^2	0.282	0.003	0.233	0.918	0.014

Table 4: Decomposition of human vs. bot tweets

This table reports fixed-effects regressions of stock indicators by tweet characteristics. Dependent variables are (log) return, market-model return, volatility i.e. Parkinson (1980) intraday high-low range, (normalized) trading volume, and bid-ask spread. Main independent variables are either human-originated or bot-originated Twitter characters based on tweets collected from 4:30 pm day -1 to 4:29 pm day 0. Human (bots) tweet features are based on human (bots) tweets. Message is the natural logarithm of the number of tweets, Positiveness is given as $Positiveness_t = \ln\left(\frac{1+M_t^{positive}}{1+M_t^{negative}}\right)$, where $M_t^{positive}$ and $M_t^{negative}$ are the counts of positive and negative tweets on day t, and Agreement is $Agreement_t = 1 - \sqrt{1-\left(\frac{M_t^{positive}-M_t^{negative}}{M_t^{positive}+M_t^{negative}}\right)^2}$. Traditional news is the number of relevant news on Financial Times. T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 10%, 5%, 1% significance, respectively.

	(1) Return	(2) Mrk-model return	(3) Volatility	(4) Volume	(5) Bid-Ask
Human Positiveness	0.0206***	0.0213**	-0.0429***	-0.0067***	-0.0015
	(2.65)	(2.33)	(-4.91)	(-3.06)	(-0.38)
Human Message	-0.0340	-0.0332	0.2640***	0.0851***	0.0082
	(-1.62)	(-1.34)	(11.47)	(16.19)	(0.77)
Human Agreement	-0.0032	-0.0038	0.0010	-0.0079***	0.0076
	(-0.45)	(-0.45)	(0.13)	(-3.43)	(0.94)
Bots Positiveness	0.0165***	0.0146**	-0.0346***	-0.0047**	-0.0087**
	(2.70)	(2.03)	(-5.56)	(-2.34)	(-2.20)
Bots Message	0.0043	0.0067	0.0941***	0.0430***	0.0141
	(0.47)	(0.63)	(10.35)	(13.55)	(1.22)
Bots Agreement	-0.0066	-0.0081	0.0003	-0.0026	-0.0082*
	(-1.23)	(-1.30)	(0.05)	(-1.50)	(-1.80)
Traditional News	-0.0013	0.0005	0.0819***	0.0272***	-0.0039
	(-0.16)	(0.05)	(12.15)	(15.45)	(-0.85)
Lagged Return	0.0210*	0.0212	-0.0533***	-0.0132***	0.0080
	(1.88)	(1.57)	(-3.75)	(-6.15)	(1.22)
FTSE100 Return	0.5295***	0.0279***	-0.0903***	-0.0186***	-0.0044
	(79.34)	(3.60)	(-11.13)	(-11.02)	(-1.03)
Observations	36,925	36,925	36,925	36,925	36,925
R^2	0.282	0.003	0.242	0.919	0.014

Table 5: Human bot interactions

This table reports fixed-effects regressions of stock indicators by tweet characteristics. Dependent variables are (log) return, market-model return, volatility, (normalized) trading volume, and bid-ask spread. Main independent variables are aggregate Twitter characters based on tweets collected from 4:30 pm day -1 to 4:29 pm day 0. Human (bots) tweet features are based on human (bots) tweets. Message is the natural logarithm of the number of tweets, Positiveness is given as $Positiveness_t = \ln\left(\frac{1+M_t^{positive}}{1+M_t^{negative}}\right)$, where $M_t^{positive}$ and $M_t^{negative}$ are the counts of positive and negative tweets on day t, and Agreement is $Agreement_t = 1 - \sqrt{1-\left(\frac{M_t^{positive}-M_t^{negative}}{M_t^{positive}+M_t^{negative}}\right)^2}$. The 'Hu × Bo ...' are interaction terms between Human and Bot tweets characteristics. Traditional news is the number of relevant news on Financial Times. T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 10%, 5%, 1% significance, respectively.

	(1)	(2)	(3)	(4)	(5)
	Return	Mrk-model return	Volatility	Volume	Bid-Ask
Human Positiveness	0.0208**	0.0225**	-0.0342***	-0.0057**	-0.0003
	(2.51)	(2.31)	(-3.64)	(-2.49)	(-0.06)
Human Message	-0.0315	-0.0209	0.3806***	0.0980***	0.0128
	(-0.97)	(-0.55)	(11.63)	(12.36)	(0.85)
Human Agreement	-0.0061	-0.0071	-0.0046	-0.0083***	0.0111
	(-0.80)	(-0.78)	(-0.55)	(-3.20)	(0.83)
Bots Positiveness	0.0188**	0.0181*	-0.0356***	-0.0045	-0.0080*
	(2.04)	(1.68)	(-3.85)	(-1.64)	(-1.69)
Bots Message	0.0048	0.0107	0.1330***	0.0474***	0.0163
	(0.40)	(0.74)	(11.23)	(12.48)	(1.61)
Bots Agreement	-0.0104	-0.0112	0.0087	-0.0014	-0.0030
	(-1.45)	(-1.33)	(1.19)	(-0.60)	(-0.61)
Hu Pos. × Bot Pos.	-0.0034	-0.0059	-0.0040	-0.0008	-0.0014
	(-0.39)	(-0.57)	(-0.44)	(-0.29)	(-0.31)
Hu Mess. × Bot Mess.	0.0026	0.0131	0.1286***	0.0142**	0.0046
	(0.12)	(0.50)	(5.85)	(2.26)	(0.44)
Hu Agree. × Bot Agree.	0.0071	0.0069	-0.0050	-0.0010	-0.0084
	(1.04)	(0.85)	(-0.73)	(-0.40)	(-0.78)
Traditional News	-0.0013	0.0004	0.0806***	0.0270***	-0.0040
	(-0.17)	(0.04)	(11.98)	(15.37)	(-0.86)
Lagged Return	0.0210*	0.0212	-0.0535***	-0.0132***	0.0080
	(1.88)	(1.57)	(-3.77)	(-6.17)	(1.22)
FTSE100 Return	0.5295***	0.0278***	-0.0904***	-0.0186***	-0.0044
	(79.34)	(3.59)	(-11.15)	(-11.03)	(-1.04)
Observations	36,925	36,925	36,925	36,925	36,925
R^2	0.282	0.004	0.244	0.919	0.014

Table 6: Intraday regressions

This table reports fixed-effects regressions of stock indicators by tweet characteristics. Dependent variables are (log) return, market-model return, volatility in basis points, and (normalized) trading volume at the end of each 5-minute interval. Main independent variables are aggregated Twitter characters based on tweets collected during that 5-minute interval. Tweets before trading hours i.e. after 4:30 pm day -1 and before 8:00 am day 0 are pooled to the first 5-minute interval of day 0. Message is the natural logarithm of the number of tweets, Positiveness is given as $Positiveness_t = \ln\left(\frac{1+M_t^{positive}}{1+M_t^{negative}}\right)$, where $M_t^{positive}$ and $M_t^{negative}$ are the counts of positive and negative tweets on day t, and Agreement is $Agreement_t = 1 - \sqrt{1 - \left(\frac{M_t^{positive} - M_t^{negative}}{M_t^{positive} + M_t^{negative}}\right)^2}$. Traditional news is the number of relevant news on Financial Times. T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 10%, 5%, 1% significance, respectively.

	(1) Return	(2) Mrk-model return	(3) Volatility	(4) Volume
Positiveness	0.0584	0.0579	-0.2702***	-0.0129***
	(1.60)	(1.59)	(-4.55)	(-11.97)
Message	-0.1114*	-0.1070*	1.2147***	0.1132***
	(-1.76)	(-1.70)	(16.96)	(74.06)
Agreement	0.0065	0.0017	-0.0504	-0.0238***
	(0.08)	(0.02)	(-0.50)	(-9.19)
Traditional News	-0.4472	-0.4448	5.3028***	0.3018***
	(-0.28)	(-0.28)	(6.41)	(6.04)
Lagged Return	-0.4449	-0.4449	-0.0020	-0.0000
	(-1.44)	(-1.44)	(-1.23)	(-0.10)
FTSE100 Return	0.9345***	-0.0269	-0.5667***	-0.0352***
	(2.87)	(-0.08)	(-10.04)	(-37.18)
Observations	1,408,538	1,408,538	1,408,538	1,408,538
R^2	0.199	0.199	0.004	0.694

Table 7: Decomposition of human vs. bot tweets at intraday frequency

This table reports fixed-effects regressions of stock indicators by tweet characteristics. Dependent variables are (log) return, market-model return, volatility in basis points, and (normalized) trading volume at the end of each 5-minute interval. Main independent variables are aggregated Twitter characters based on tweets collected during that 5-minute interval. Tweets before trading hours i.e. after 4:30 pm day -1 and before 8:00 am day 0 are pooled to the first 5-minute interval of day 0. Human (bots) tweet features are based on human (bots) tweets. Message is the natural logarithm of the number of tweets, Positiveness is given as $Positiveness_t = \ln\left(\frac{1+M_t^{positive}}{1+M_t^{negative}}\right)$, where $M_t^{positive}$ and $M_t^{negative}$ are the counts of positive and negative tweets on day t, and Agreement is $Agreement_t = 1 - \sqrt{1-\left(\frac{M_t^{positive}-M_t^{negative}}{M_t^{positive}+M_t^{negative}}\right)^2}$. Traditional news is the number of relevant news on Financial Times. T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 10%, 5%, 1% significance, respectively.

	(1)	(2)	(3)	(4)
	Return	Mrk-model return	Volatility	Volume
Human Positiveness	0.0648*	0.0645*	-0.2576***	-0.0122***
	(1.66)	(1.66)	(-4.36)	(-11.20)
Human Message	-0.1074*	-0.1040*	1.1727***	0.1100***
	(-1.72)	(-1.66)	(16.21)	(70.90)
Human Agreement	0.0295	0.0249	-0.0838	-0.0266***
	(0.36)	(0.30)	(-0.82)	(-10.24)
Bots Positiveness	-0.1032	-0.1073	-0.0985	-0.0060*
	(-0.99)	(-1.03)	(-1.49)	(-1.89)
Bots Message	0.1219	0.1267	0.3020**	0.0176***
	(0.87)	(0.90)	(2.52)	(7.99)
Bots Agreement	-0.2436	-0.2409	0.1119	0.0227***
	(-0.75)	(-0.74)	(0.50)	(6.24)
Traditional News	-0.5511	-0.5484	5.2484***	0.2994***
	(-0.34)	(-0.34)	(6.29)	(5.95)
Lagged Return	-0.4458	-0.4458	-0.0021	-0.0000
	(-1.44)	(-1.44)	(-1.24)	(-0.31)
FTSE100 Return	0.9352***	-0.0249	-0.5717***	-0.0354***
	(2.86)	(-0.08)	(-9.89)	(-36.95)
Observations	1,375,955	1,375,955	1,375,955	1,375,955
R^2	0.199	0.199	0.004	0.695

Table 8: Regressions for 5-minute lagged tweets

This table reports fixed-effects regressions of stock indicators by tweet characteristics. Dependent variables are (log) return, market-model return, volatility in basis points, and (normalized) trading volume at the end of each 5-minute interval. Main independent variables are aggregate Twitter characters based on tweets collected during the previous 5-minute interval. Tweets before trading hours i.e. after 4:30 pm day -1 and before 8:00 am day 0 are pooled to the first 5-minute interval of day 0. Human (bots) tweet features are based on human (bots) tweets. Message is the natural logarithm of the number of tweets, Positiveness is given as $Positiveness_t = \ln\left(\frac{1+M_t^{positive}}{1+M_t^{negative}}\right)$, where $M_t^{positive}$ and $M_t^{negative}$ are the counts of positive and negative tweets on day t, and Agreement is $Agreement_t = 1 - \sqrt{1-\left(\frac{M_t^{positive}-M_t^{negative}}{M_t^{positive}+M_t^{negative}}\right)^2}$. Traditional news is the number of relevant news on Financial Times. T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 10%, 5%, 1% significance, respectively.

	(1) Return	(2) Mrkt-model return	(3) Volatility	(4) Volume
Lagged Human Positiveness	0.0157	0.0155	-0.1779***	-0.0106***
	(0.43)	(0.42)	(-3.27)	(-9.76)
Lagged Human Message	-0.1746***	-0.1710***	1.3971***	0.1162***
	(-2.81)	(-2.75)	(22.35)	(75.74)
Lagged Human Agreement	-0.0243	-0.0290	-0.0758	-0.0269***
	(-0.36)	(-0.43)	(-1.26)	(-10.38)
Lagged Bots Positiveness	0.0726	0.0675	-0.2018***	-0.0050
	(0.84)	(0.78)	(-4.27)	(-1.57)
Lagged Bots Message	0.0267	0.0313	0.3184***	0.0205***
	(0.21)	(0.25)	(2.80)	(9.36)
Lagged Bots Agreement	0.1814	0.1856	-0.0612	0.0216***
	(0.81)	(0.83)	(-0.45)	(5.94)
Traditional News	0.0622	0.0665	6.1710***	0.3920***
	(0.04)	(0.04)	(7.27)	(7.92)
Lagged Return	-0.4456	-0.4456	-0.0021	0.0000
	(-1.43)	(-1.43)	(-1.24)	(0.11)
Lagged FTSE100 Return	0.9158***	-0.0430	-0.6001***	-0.0355***
	(2.91)	(-0.14)	(-10.64)	(-36.95)
Observations	1,375,936	1,375,936	1,375,936	1,375,936
R^2	0.200	0.200	0.004	0.695

Table 9: 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 spread in responses to abnormal increases in the number 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 10%, 5%, 1% significance, respectively.

Time windows	All	Positive	Negative	Bots	Bot pos.	Bot neg.
Panel A: Response of returns						
[0]	-0.0030	-0.1098	-0.3820***	-0.0309	0.0473	-0.2414*
[1]	0.0929*	0.1476**	0.1024	0.0151	0.0303	0.0290
[1,5]	0.0943	0.0664	0.2865*	0.2405	-0.4009**	0.3466*
Obs.	1218	721	522	695	328	181
Panel B: Response of market-model abnormal returns						
[0]	0.0562	0.0170	-0.2477**	0.0398	0.0729	-0.1330*
[1]	0.0425	0.0985*	0.0950	-0.0244	0.0352	0.0507
[1,5]	0.0507	0.0625	0.1286	0.1638	-0.4794**	0.2143*
Obs.	1218	721	522	695	328	181
Panel C: Response of volatility						
[0]	0.2948***	0.2533***	0.3664***	0.1607**	0.0419*	0.0253
[1]	0.1626***	0.1403***	0.2262***	0.0765**	0.0111*	0.0548
[1,5]	0.3558***	0.3069***	0.3633***	0.3916**	0.3348**	0.1152
Obs.	1218	721	522	695	328	181
Panel D: Response of (normalized) trading volume						
[0]	9.1225***	6.3627***	7.5173***	-0.6154	-3.8387*	-6.5174**
[1]	9.6686***	7.2143***	11.1803***	5.5388***	2.8529	1.4550
[1,5]	17.9111***	13.7167***	23.9366***	16.8697**	11.8529*	-3.9375
Obs.	1218	721	522	695	328	181
Panel E: Response of bid-ask spread						
[0]	-15.1186***	-14.5438***	-15.0848***	-13.9371**	-10.3607***	-10.8067***
[1]	-14.6694***	-15.4248***	-14.2154***	-12.7374*	-5.2121**	-7.7433***
[1,5]	-52.9291***	-53.4368***	-55.8268***	-42.8992*	-32.6067***	-34.0654***
Obs.	1218	721	522	695	328	181

Table 10: Pre-trading hour regressions

This table reports fixed-effects regressions of stock indicators by tweet characteristics. Dependent variables are (log) return, market-model return, volatility i.e. Parkinson (1980) intraday high-low range, (normalized) trading volume, and bid-ask spread. Main independent variables are aggregate Twitter characters based on tweets collected from 4:30 pm day -1 to 8:00 am day 0. Human (bots) tweet features are based on human (bots) tweets. Message is the natural logarithm of the number of tweets, Positiveness is given as $Positiveness_t = \ln\left(\frac{1+M_t^{positive}}{1+M_t^{negative}}\right)$, where $M_t^{positive}$ and $M_t^{negative}$ are the counts of positive and negative tweets on day t, and Agreement is $Agreement_t = 1 - \sqrt{1-\left(\frac{M_t^{positive}-M_t^{negative}}{M_t^{positive}+M_t^{negative}}\right)^2}$. Traditional news is the number of relevant news on Financial Times. T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 10%, 5%, 1% significance, respectively.

	(1) Return	(2) Mrk-model return	(3) Volatility	(4) Volume	(5) Bid-Ask
Human Positiveness	0.0154**	0.0183**	-0.0289***	-0.0049**	-0.0005
	(2.06)	(2.06)	(-3.48)	(-2.31)	(-0.10)
Human Message	-0.0249	-0.0236	0.1858***	0.0692***	0.0152
	(-1.31)	(-1.06)	(8.54)	(13.38)	(1.17)
Human Agreement	-0.0085	-0.0116	-0.0017	-0.0080***	0.0035
	(-1.19)	(-1.36)	(-0.22)	(-3.51)	(0.83)
Bots Positiveness	0.0131**	0.0106	-0.0365***	-0.0059***	-0.0050
	(2.09)	(1.44)	(-5.68)	(-2.93)	(-1.60)
Bots Message	0.0031	0.0027	0.0655***	0.0294***	0.0031
	(0.33)	(0.24)	(6.80)	(9.33)	(0.57)
Bots Agreement	-0.0112*	-0.0128*	0.0087	-0.0000	-0.0059*
	(-1.94)	(-1.88)	(1.46)	(-0.02)	(-1.77)
Traditional News	-0.0001	0.0021	0.0923***	0.0301***	-0.0039
	(-0.02)	(0.22)	(13.50)	(16.94)	(-0.83)
Lagged Return	0.0214*	0.0217	-0.0527***	-0.0131***	0.0080
	(1.89)	(1.59)	(-3.64)	(-5.93)	(1.22)
FTSE100 Return	0.5311***	0.0292***	-0.0914***	-0.0188***	-0.0045
	(78.50)	(3.71)	(-11.08)	(-10.95)	(-1.07)
Observations	36,157	36,157	36,157	36,157	36,157
R^2	0.284	0.003	0.236	0.918	0.014

Figure 1: Weekly aggregate measures of human and bot tweets

This figure describes weekly aggregate measures of tweets posted by human and bot on the 55 sampled firms from FTSE 100 composites from August 2015 to July 2018. Message is log-number of tweets. The left (right) y-axis is for human (bot) message. $Positiveness_t = \ln\left(\frac{1+M_t^{positive}}{1+M_t^{negative}}\right)$, and $Agreement_t = 1 - \sqrt{1-\left(\frac{M_t^{positive}-M_t^{negative}}{M_t^{positive}+M_t^{negative}}\right)^2}$. Tetlock et al. (2008) Negativeness is standardized negative percentage. During the week commencing 12th September 2016, there are less tweet activities, i.e. the number of tweet messages is lower than 300,000. The highest number of tweets is around 870,000 in the week beginning February 2016 and the week ending January 2017.

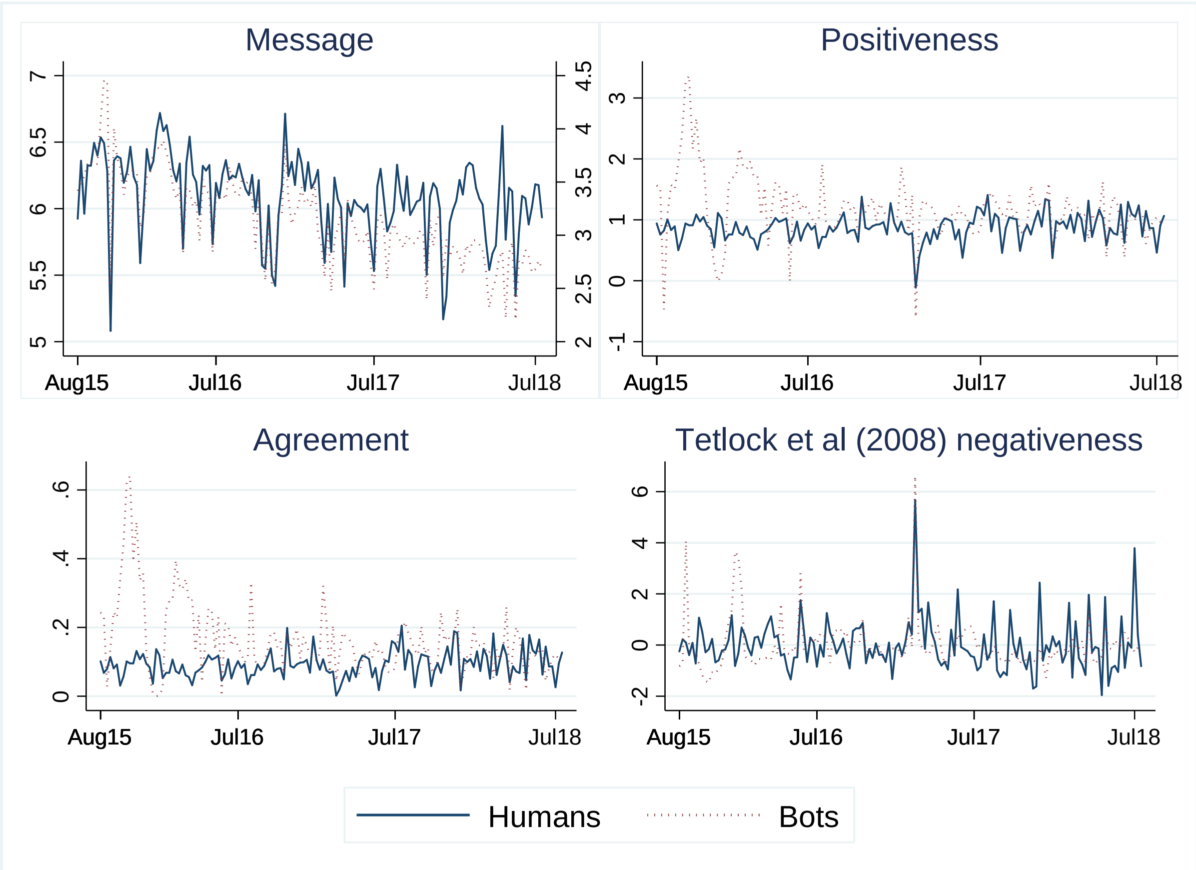
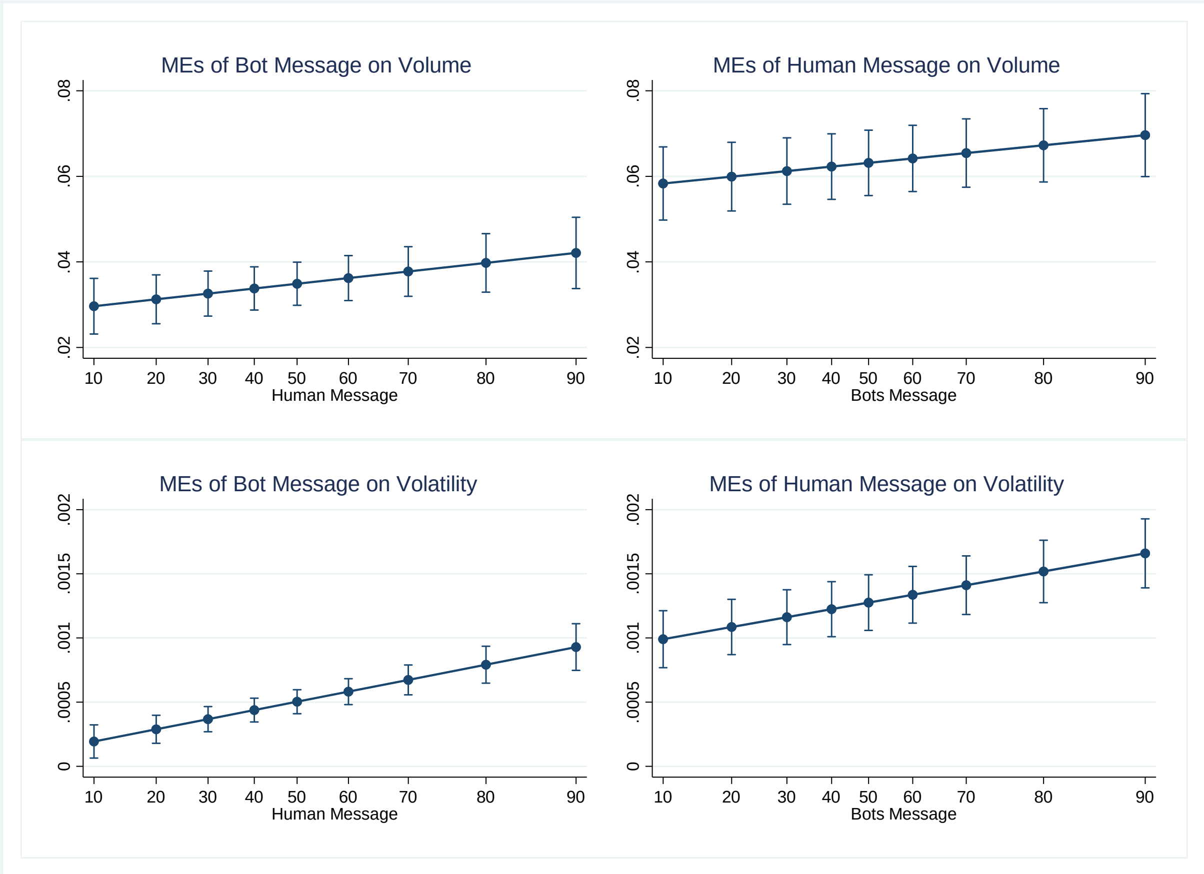


Figure 2: Marginal effects of human and bot messages

This figure shows the marginal effects of human and bot tweets on stocks trading volume (volatility) with different levels of human and bot messages volume. Y axis is the marginal effects of bot (human) messages on either volume or volatility, and x axis denotes percentiles of distribution of human (bot) messages measure. Vertical lines denote 95% confidence intervals.



Online appendix

Table of contents

Appendix A: Textual information

Table A1: Sampled companies

This table lists names of the sampled companies.

	Firm name		Firm name
1	3I Group	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: Definition of variables

This table gives a list of the definition of variables used in the analysis.

Variable	Definition	Measure
	Aggregate Twitter features	
$Positiveness_{i,t}$	Variable for the aggregate sentiment of tweets containing company *i* on day *t*	Natural logarithm of the number of positive tweets containing company *i* on day *t* over the number of corresponding negative tweets
$Message_{i,t}$	Variable for the Twitter volume for company *i* on day *t*	Natural logarithm of the number of all tweets containing company *i* on day *t*
$Agreement_{i,t}$	Variable for the aggregate agreement of tweets containing company *i* on day *t*	1 - square root of [1 – square of (the difference between the number of positive and negative tweets over the total number of positive and negative tweets)]
$Traditional\ News_{i,t}$	Variable for traditional news channels	Number of articles related to firm *i* on day *t* on Financial Times
	Stock indicators	
$Return_{i,t}$	Return for stock *i* on day *t*	Natural logarithm of the return
$Mrk\text{-}model\ return_{i,t}$	Market-model abnormal return	Difference between the log-return of stock *i* on day *t* and an estimate of expected return using the market model over the period from *t-110* to *t-10*
$Volatility_{i,t}$	Intraday volatility of stock prices	Natural logarithm of (the highest price divided by the lowest price of stock *i* on day *t*), adjusted by $\frac{1}{2\sqrt{2}}$
$Volume_{i,t}$	Trading volume	Natural logarithm of the number of stocks *i* traded on day *t*
$Bid\text{-}ask_{i,t}$	Bid-ask spread (in basis points)	Logarithm of the difference between bid and ask quotes for stock *i* on day *t*, scaled by 10,000

Table A3: Examples of polarity score of tweets given by TextBlob

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

Text	Sentiment score
Ivans having a terrible time as Glencore earning slump	-1
HSBC 2015 profit slumps disappointing china market	-0.6
Why does Sainsburys want to buy Argos	0
investment 3 reasons why carnival corporation ccl is a great momentum	0.8
excellent news just in HSBC one of the world's largest bank employing 10s of 000s here will stay headquartered	1

Appendix B: 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 spread. Human (bots) tweet features are based on human (bots) tweets. Message is the natural logarithm of the number of tweets, Positiveness is given as $Positiveness_t = \ln\left(\frac{1+M_t^{positive}}{1+M_t^{negative}}\right)$, where $M_t^{positive}$ and $M_t^{negative}$ are the counts of positive and negative tweets on day t, and Agreement is $Agreement_t = 1 - \sqrt{1 - \left(\frac{M_t^{positive} - M_t^{negative}}{M_t^{positive} + M_t^{negative}}\right)^2}$. Traditional news is the number of relevant news on Financial Times. * denotes correlations that are significantly different from 0 at the 1% significance level.

	Return	Log (volume)	Volatility	Bid-Ask	Human Message	Human Positiveness	Human Agreement	Bots Message	Bots Positiveness
Log (volume)	-0.0227*								
Volatility	-0.1211*	0.2337*							
Bid-Ask	0.0012	-0.0434*	0.0079						
Human Message	-0.0057	0.1688*	0.0779*	0.0003					
Human Positiveness	0.0207*	-0.0967*	-0.0666*	0.0045	0.0073				
Human Agreement	0.0147*	-0.1581*	-0.0748*	0.0067	-0.4066*	0.5843*			
Bots Message	-0.0029	0.1540*	0.1021*	-0.0135*	0.7333*	-0.0054	-0.2738*		
Bots Positiveness	0.0109	0.0683*	-0.0123	-0.012	0.2970*	0.1923*	-0.0118	0.4064*	
Bots Agreement	0.0045	0.0160*	0.0057	-0.0167*	-0.0427*	0.1406*	0.0601*	0.0571*	0.3692*
Traditional News	-0.0033	0.1665*	0.1049*	-0.0125	0.0937*	-0.0342*	-0.0510*	0.0889*	0.0304*

Appendix C: Event study

Figure C1: 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.

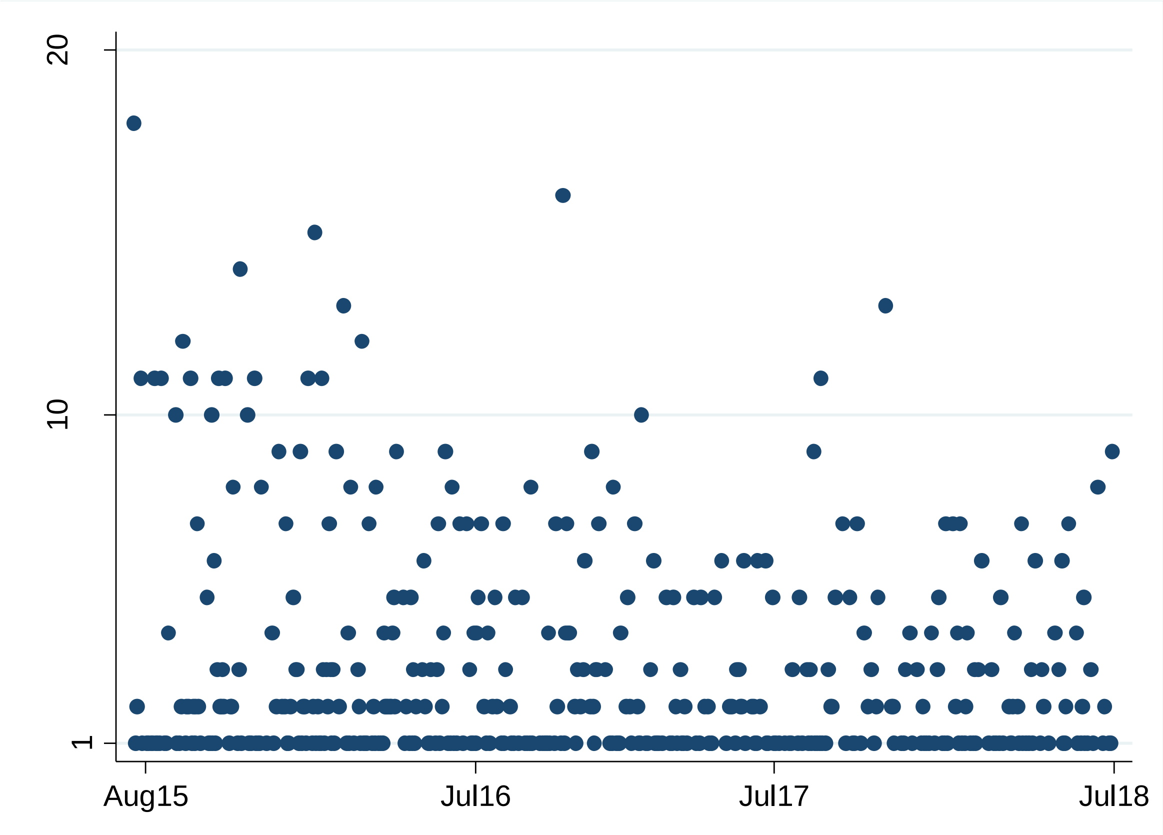


Table C1: Event study – Response of returns

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 10%, 5%, 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]	-0.0030	-0.1098	-0.3820***	-0.0309	0.0473	-0.2414*	-0.0151	-0.1450	-0.3934***
[1]	0.0929*	0.1476**	0.1024	0.0151	0.0303	0.0290	0.1056	0.0972*	0.0939
[1,5]	0.0943	0.0664	0.2865*	0.2405	-0.4009**	0.3466*	0.0979	0.0804	0.3010*
Obs.	1218	721	522	695	328	181	1188	718	519
Responses to abnormal increases in original tweets									
[0]	-0.0437	-0.0845	-0.4103**	-0.0159	-0.0081	-0.4097*	-0.0787	-0.0729	-0.4904**
[1]	0.0416	0.0628	0.2891**	0.0174	-0.0378	0.2795*	0.1004	0.0111	0.2286*
[1,5]	0.1439	0.0246	-0.0461	0.2189	-0.1773	0.7190**	0.0229	0.0336	-0.0762
Obs.	974	469	341	665	301	120	910	453	315
Responses to abnormal increases in retweets									
[0]	-0.0697	-0.1935*	-0.1807	0.0509	0.1483	-0.1872	-0.0814	-0.2093*	-0.1676
[1]	0.0920	0.1367**	0.1475	0.1047	0.0486	-0.1608	0.1063	0.1365**	0.1446
[1,5]	0.2372**	0.2621**	0.4972***	-0.0752	-0.0587	-0.0683	0.2274	0.1939	0.4888***
Obs.	873	563	399	267	129	45	869	560	396
Panel B: Response of market-model abnormal returns									
Responses to abnormal increases in tweets									
[0]	0.0562	0.0170	-0.2477**	0.0398	0.0729	-0.1330*	0.0567	-0.0010	-0.2499**
[1]	0.0425	0.0985*	0.0950	-0.0244	0.0352	0.0507	0.0573	0.0558	0.0802
[1,5]	0.0507	0.0625	0.1286	0.1638	-0.4794**	0.2143*	0.0731	0.0506	0.1599
Obs.	1218	721	522	695	328	181	1188	718	519
Responses to abnormal increases in original tweets									
[0]	0.1069	0.0863	-0.1848	0.0599	0.0619	-0.1979**	0.0813	0.0911	-0.2569
[1]	-0.0131	0.0137	0.2289**	-0.0199	-0.0036	0.2475**	0.0297	0.0071	0.1867*
[1,5]	0.0746	-0.0208	-0.1222	0.1840	-0.3654**	0.5623**	-0.0035	-0.0056	-0.0974
Obs.	974	469	341	665	301	120	910	453	315
Responses to abnormal increases in retweets									
[0]	-0.0211	-0.0736	-0.1176	0.0636	0.1615	-0.1625	-0.0339	-0.0916	-0.1072
[1]	0.0652	0.0907	0.1310	0.0225	-0.0121	-0.0735	0.0825	0.0852	0.1272
[1,5]	0.1248	0.1521	0.3321**	-0.0776	-0.2886*	-0.5535*	0.1256	0.1270	0.3317**
Obs.	873	563	399	267	129	45	869	560	396

Table C2: Event study – Response of volatility and trading volume

This table reports the decomposition of volatility and (normalized) trading volume 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 10%, 5%, 1% significance, respectively.

Time windows	All	Positive	Negative	Bots	Bot pos.	Bot neg.	Human	Human pos.	Human neg.
Panel A: Response of volatility									
Responses to abnormal increases in tweets									
[0]	0.2948***	0.2533***	0.3664***	0.1607**	0.0419*	0.0253	0.2892	0.2312***	0.3611***
[1]	0.1626***	0.1403***	0.2262***	0.0765**	0.0111*	0.0548	0.1670	0.1515***	0.2199***
[1,5]	0.3558***	0.3069***	0.3633***	0.3916**	0.3348**	0.1152	0.3602	0.2788***	0.3338***
Obs.	1218	721	522	695	328	181	1188	718	519
Responses to abnormal increases in original tweets									
[0]	0.4329***	0.3315***	0.3636***	0.1055	0.0912*	0.0200	0.4688	0.3292***	0.4207***
[1]	0.2274***	0.1999***	0.2348***	0.0817	0.0678*	0.1177*	0.2384	0.2250***	0.2777***
[1,5]	0.5696***	0.5310***	0.4463***	0.3749	0.4249***	0.1209	0.5618	0.4734***	0.5535***
Obs.	974	469	341	665	301	120	910	453	315
Responses to abnormal increases in retweets									
[0]	0.1857***	0.1188**	0.2475***	-0.0589	-0.0123	0.2262	0.1869	0.1035**	0.2531***
[1]	0.1446***	0.0846**	0.2247***	-0.0236	-0.0050	0.0830	0.1468	0.0885***	0.2275***
[1,5]	0.2888***	0.2255**	0.3514**	-0.1287	-0.0178	0.0401	0.2975	0.2523**	0.3489**
Obs.	873	563	399	267	129	45	869	560	396
Panel B: Response of (normalized) trading volume									
Responses to abnormal increases in tweets									
[0]	9.1225***	6.3627***	7.5173***	-0.6154	-3.8387*	-6.5174**	9.1613	5.9871***	7.2406***
[1]	9.6686***	7.2143***	11.1803***	5.5388***	2.8529	1.4550	9.5194	7.0642***	10.5213***
[1,5]	17.9111***	13.7167***	23.9366***	16.8697**	11.8529*	-3.9375	17.8745	11.0679**	24.2224***
Obs.	1218	721	522	695	328	181	1188	718	519
Responses to abnormal increases in original tweets									
[0]	12.9208***	6.5964***	4.7994*	-1.6898	-3.8824*	-3.6356	14.5736	7.1038***	5.7998**
[1]	13.1173***	8.5612***	13.6725***	7.3412	6.8460***	6.4731*	14.1471	9.6962***	15.2918***
[1,5]	26.9189***	21.2776***	22.1404***	20.5997	23.1817***	3.7549	28.1391	19.5932***	25.6950***
Obs.	974	469	341	665	301	120	910	453	315
Responses to abnormal increases in retweets									
[0]	3.8963**	4.9069**	7.3725***	-5.3645	-1.7490	0.4087	4.3064	4.5787**	7.7202***
[1]	6.0892***	5.5938***	10.4086***	-0.0230	1.4290	-4.4835	6.7802	5.6728***	10.4510***
[1,5]	14.1659***	10.9526**	26.4060***	-10.9239	-9.3281	-29.0807*	15.3796	11.5705**	26.0623***
Obs.	873	563	399	267	129	45	869	560	396

Table C3: Event study – Response of bid-ask spread

This table reports the decomposition of cumulative changes in bid-ask spread 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 10%, 5%, 1% significance, respectively.

Time windows	All	Positive	Negative	Bots	Bot pos.	Bot neg.	Human	Human pos.	Human neg.
Response of bid-ask spread									
Responses to abnormal increases in tweets									
[0]	-15.1186***	-14.5438***	-15.0848***	-13.9371**	-10.3607***	-10.8067***	-15.3488	-15.6338***	-15.1950***
[1]	-14.6694***	-15.4248***	-14.2154***	-12.7374*	-5.2121**	-7.7433***	-15.0629	-16.1590***	-14.5728***
[1,5]	-52.9291***	-53.4368***	-55.8268***	-42.8992*	-32.6067***	-34.0654***	-54.6402	-56.0207***	-55.4746***
Obs.	1218	721	522	695	328	181	1188	718	519
Responses to abnormal increases in original tweets									
[0]	-12.8980***	-8.3171***	-7.3179***	-13.7383	-6.8517***	-7.6324**	-12.0182	-9.0906***	-7.5614***
[1]	-12.9300***	-8.2405***	-8.1158***	-12.6109	-3.6379*	-9.1111***	-11.5670	-8.4329***	-6.7627***
[1,5]	-44.0075***	-26.0203***	-27.4392***	-42.5304	-15.0391**	-27.4572***	-42.1854	-30.1440***	-25.0329***
Obs.	974	469	341	665	301	120	910	453	315
Responses to abnormal increases in retweets									
[0]	-15.4451***	-16.3414***	-16.6174***	-16.3799	-16.3839***	-14.0473**	-15.2573	-16.1229***	-16.6007***
[1]	-15.9117***	-17.8400***	-14.9649***	-13.0061	-13.4213***	-4.5937	-15.7423	-17.9572***	-15.2808***
[1,5]	-62.0199***	-61.0678***	-67.5662***	-71.2719	-61.1448***	-55.8358***	-60.6125	-62.4888***	-68.6385***
Obs.	873	563	399	267	129	45	869	560	396

Appendix D: Robustness checks

Table D1: Robustness check of volatility measures

This table reports fixed-effects regressions of volatility measures by tweet characteristics. Dependent variables are one-day ahead GARCH (1, 1) volatility for columns (1) - (2) and GJR (1, 1) volatility for columns (3) - (4). Main independent variables are human (bots) tweet characters based on human (bots) tweets collected from 4:30 pm day -1 to 4:29 pm day 0. Message is the natural logarithm of the number of tweets, Positiveness is given as $Positiveness_t = \ln\left(\frac{1+M_t^{positive}}{1+M_t^{negative}}\right)$, where $M_t^{positive}$ and $M_t^{negative}$ are the counts of positive and negative tweets on day t, and Agreement is $Agreement_t = 1 - \sqrt{1-\left(\frac{M_t^{positive}-M_t^{negative}}{M_t^{positive}+M_t^{negative}}\right)^2}$. Traditional news is the number of relevant news on Financial Times. T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 10%, 5%, 1% significance, respectively.

	(1) GARCH volatility	(2) GARCH volatility	(3) GJR volatility	(4) GJR volatility
Human Positiveness	-0.0045	0.0107	-0.0094	0.0125
	(-0.41)	(0.89)	(-0.87)	(1.07)
Human Message	0.2394***	0.2814***	0.2203***	0.3159***
	(7.92)	(6.73)	(7.14)	(6.86)
Human Agreement	-0.0282**	-0.0407**	-0.0183*	-0.0418**
	(-2.33)	(-2.14)	(-1.67)	(-2.30)
Bots Positiveness	-0.0432***	-0.0147	-0.0482***	-0.0128
	(-5.43)	(-1.41)	(-5.94)	(-1.21)
Bots Message	0.1369***	0.1527***	0.1466***	0.1766***
	(10.96)	(11.83)	(11.13)	(12.41)
Bots Agreement	-0.0084	-0.0151	-0.0065	-0.0199
	(-0.89)	(-1.20)	(-0.70)	(-1.59)
Traditional News	0.0125	0.0119	0.0088	0.0076
	(1.27)	(1.21)	(0.99)	(0.87)
Hu Pos. × Bot Pos.		-0.0505***		-0.0619***
		(-4.41)		(-5.23)
Hu Mess. × Bot Mess.		0.0390*		0.0936***
		(1.68)		(3.78)
Hu Agree. × Bot Agree.		0.0232		0.0347**
		(1.32)		(1.99)
Observations	15,267	15,267	14,151	14,151
R^2	0.482	0.483	0.488	0.490

Table D2: Robustness check of volatility measures for pre-trading hours tweets

This table reports fixed-effects regressions of volatility measures by tweet characteristics. Dependent variables are one-day ahead GARCH (1, 1) volatility for columns (1) - (2) and GJR(1,1) volatility for columns (3) - (4). Main independent variables are human (bots) tweet characters based on human (bots) tweets collected from 4:30 pm day -1 to 8:00 am day 0. Message is the natural logarithm of the number of tweets, Positiveness is given as $Positiveness_t = \ln\left(\frac{1+M_t^{positive}}{1+M_t^{negative}}\right)$, where $M_t^{positive}$ and $M_t^{negative}$ are the counts of positive and negative tweets on day t, and Agreement is $Agreement_t = 1 - \sqrt{1-\left(\frac{M_t^{positive}-M_t^{negative}}{M_t^{positive}+M_t^{negative}}\right)^2}$. Traditional news is the number of relevant news on Financial Times. T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 10%, 5%, 1% significance, respectively.

	(1) GARCH volatility	(2) GARCH volatility	(3) GJR volatility	(4) GJR volatility
Human Positiveness	-0.0281***	-0.0211***	-0.0269***	-0.0202***
	(-4.88)	(-3.50)	(-4.48)	(-3.22)
Human Message	0.1205***	0.1908***	0.1150***	0.1817***
	(6.44)	(5.87)	(6.00)	(5.39)
Human Agreement	-0.0051	-0.0080	-0.0069	-0.0105
	(-0.91)	(-1.29)	(-1.17)	(-1.64)
Bots Positiveness	-0.0353***	-0.0281***	-0.0346***	-0.0264***
	(-6.46)	(-3.61)	(-6.42)	(-3.39)
Bots Message	0.1178***	0.1426***	0.1303***	0.1522***
	(14.76)	(13.37)	(15.41)	(13.67)
Bots Agreement	-0.0001	0.0048	-0.0023	0.0011
	(-0.02)	(0.74)	(-0.47)	(0.18)
Traditional News	0.0154***	0.0146***	0.0151**	0.0144**
	(2.79)	(2.66)	(2.56)	(2.45)
Hu Pos. × Bot Pos.		-0.0144**		-0.0154**
		(-2.01)		(-2.11)
Hu Mess. × Bot Mess.		0.0760***		0.0715***
		(3.70)		(3.36)
Hu Agree. × Bot Agree.		-0.0015		0.0011
		(-0.25)		(0.19)
Observations	36,248	36,248	36,257	36,257
R^2	0.491	0.491	0.467	0.467

Table D3: Lagged tweets regressions

This table reports fixed-effects regressions of stock indicators by tweet characteristics. Dependent variables are (log) return, market-model return, volatility i.e. Parkinson (1980) intraday high-low range, abnormal change in volatility, (normalized) trading volume, abnormal change in trading volume, bid-ask spread and abnormal change in bid-ask spread. Main independent variables are aggregate Twitter characters based on tweets collected from 4:30 pm day -2 to 4:29 pm day -1. Human (bots) tweet features are based on human (bots) tweets. Message is the natural logarithm of the number of tweets, Positiveness is given as $Positiveness_t = \ln\left(\frac{1+M_t^{positive}}{1+M_t^{negative}}\right)$, where $M_t^{positive}$ and $M_t^{negative}$ are the counts of positive and negative tweets on day t, and Agreement is $Agreement_t = 1 - \sqrt{1-\left(\frac{M_t^{positive}-M_t^{negative}}{M_t^{positive}+M_t^{negative}}\right)^2}$. Traditional news is the number of relevant news on Financial Times. T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 10%, 5%, 1% significance, respectively.

	(1) Return	(2) Mrk-model return	(3) Volatility	(4) Volume	(5) Bid-Ask
Lagged Human Positiveness	0.0049	0.0004	-0.0281***	-0.0074***	-0.0123
	(0.62)	(0.05)	(-3.70)	(-3.33)	(-1.35)
Lagged Human Message	0.0141	0.0101	0.1344***	0.0592***	-0.0047
	(0.80)	(0.57)	(7.91)	(12.77)	(-0.22)
Lagged Human Agreement	-0.0090	-0.0076	0.0009	-0.0051**	0.0031
	(-1.16)	(-0.96)	(0.12)	(-2.22)	(0.41)
Lagged Bots Positiveness	0.0063	0.0049	-0.0362***	-0.0052***	-0.0080*
	(0.86)	(0.67)	(-5.61)	(-2.58)	(-1.73)
Lagged Bots Message	0.0216**	0.0076	0.0549***	0.0344***	0.0169
	(2.05)	(0.71)	(5.73)	(10.86)	(0.87)
Lagged Bots Agreement	-0.0098	-0.0104	0.0052	-0.0021	-0.0055
	(-1.51)	(-1.57)	(0.89)	(-1.18)	(-1.22)
Traditional News	-0.0012	-0.0025	0.0978***	0.0321***	-0.0037
	(-0.13)	(-0.26)	(14.27)	(18.03)	(-0.82)
Lagged Return	0.0257*	0.0266	-0.0405**	-0.0032	0.0132
	(1.66)	(1.59)	(-2.30)	(-1.18)	(1.10)
Lagged FTSE100 Return	-0.0164*	-0.0110	-0.0202**	-0.0186***	-0.0091
	(-1.71)	(-1.14)	(-2.27)	(-8.58)	(-0.83)
Observations	36,871	36,871	36,871	36,871	36,871
R^2	0.007	0.002	0.226	0.918	0.014

Table D4: Robustness checks for alternative measures of stock indicators

This table reports fixed-effects regressions of stock indicators by tweet characteristics. Dependent variables are (log) return, market-model return, volatility i.e. Parkinson (1980) intraday high-low range, abnormal change in volatility, (normalized) trading volume, abnormal change in trading volume, bid-ask spread and abnormal change in bid-ask spread. Main independent variables are aggregate Twitter characters based on tweets collected from 4:30 pm day -1 to 4:29 pm day 0. Message is the natural logarithm of the number of tweets, Positiveness is given as $Positiveness_t = \ln\left(\frac{1+M_t^{positive}}{1+M_t^{negative}}\right)$, where $M_t^{positive}$ and $M_t^{negative}$ are the counts of positive and negative tweets on day t, and Agreement is $Agreement_t = 1 - \sqrt{1 - \left(\frac{M_t^{positive} - M_t^{negative}}{M_t^{positive} + M_t^{negative}}\right)^2}$. Traditional news is the number of relevant news on Financial Times. T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 10%, 5%, 1% significance, respectively.

	(1) Return	(2) Mrk-model return	(3) Volatility	(4) Abnormal Volatility	(5) Volume	(6) Abnormal Volume	(7) Bid-Ask	(8) Abnormal Bid-Ask
Positiveness	0.0218***	0.0218**	-0.0474***	-0.0481***	-0.0065***	-0.0117	-0.0041	-0.0049
	(2.77)	(2.35)	(-5.26)	(-4.88)	(-2.90)	(-1.58)	(-1.19)	(-1.40)
Message	-0.0220	-0.0214	0.1630***	0.1783***	0.0670***	0.2184***	-0.0001	-0.0038
	(-1.64)	(-1.36)	(10.34)	(10.24)	(17.47)	(17.12)	(-0.01)	(-0.37)
Agreement	-0.0032	-0.0038	-0.0088	-0.0085	-0.0104***	-0.0282***	0.0071	0.0080
	(-0.48)	(-0.48)	(-1.18)	(-1.04)	(-4.60)	(-3.81)	(0.83)	(0.94)
Traditional News	-0.0022	-0.0003	0.0947***	0.1011***	0.0307***	0.0999***	-0.0031	-0.0024
	(-0.27)	(-0.04)	(13.90)	(13.55)	(17.36)	(16.70)	(-0.69)	(-0.54)
Lagged Return	0.0205*	0.0205	-0.0538***	-0.0661***	-0.0135***	-0.0516***	0.0083	0.0081
	(1.85)	(1.53)	(-3.78)	(-4.25)	(-6.19)	(-6.92)	(1.28)	(1.25)
FTSE100 Return	0.5293***	0.0283***	-0.0899***	-0.1016***	-0.0182***	-0.0616***	-0.0044	-0.0044
	(79.67)	(3.67)	(-11.09)	(-11.36)	(-10.89)	(-10.94)	(-1.00)	(-1.01)
Observations	37,619	37,619	37,619	37,619	37,619	37,619	37,619	37,619
R^2	0.282	0.003	0.233	0.071	0.918	0.087	0.014	0.003

Table D5: Regression using bull-bear spread from Brown and Cliff (2005)

This table reports fixed-effects regressions of stock indicators by tweet characteristics. Dependent variables are (log) return, market-model return, volatility i.e. Parkinson (1980) intraday high-low range, (normalized) trading volume, and bid-ask spread. Main independent variables are aggregate Twitter characters based on tweets collected from 4:30 pm day -1 to 4:29 pm day 0. Message is the natural logarithm of the number of tweets, Sentiment spread is given as $Spread_t = \frac{M_t^{positive} - M_t^{negative}}{Total\ tweets_t}$, where $M_t^{positive}$ and $M_t^{negative}$ are the counts of positive and negative tweets on day t, and Agreement is $Agreement_t = 1 - \sqrt{1 - \left(\frac{M_t^{positive} - M_t^{negative}}{M_t^{positive} + M_t^{negative}}\right)^2}$. Traditional news is the number of relevant news on Financial Times. T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 10%, 5%, 1% significance, respectively.

	Return	Mrk-model return	Volatility	Volume	Bid-Ask
	(1)	(2)	(3)	(4)	(5)
Sentiment Spread	0.0154**	0.0164*	-0.0366***	-0.0071***	-0.0045*
	(2.13)	(1.91)	(-4.08)	(-3.13)	(-1.74)
Message	-0.0236	-0.0239	0.2819***	0.1020***	0.0127
	(-1.20)	(-1.03)	(13.46)	(20.37)	(1.37)
Agreement	0.0040	0.0028	-0.0164***	-0.0093***	0.0067
	(0.71)	(0.42)	(-2.64)	(-4.70)	(0.70)
Traditional News	-0.0014	0.0006	0.0837***	0.0278***	-0.0038
	(-0.17)	(0.06)	(12.45)	(15.82)	(-0.83)
Lagged Return	0.0208*	0.0209	-0.0535***	-0.0132***	0.0079
	(1.87)	(1.55)	(-3.76)	(-6.13)	(1.22)
FTSE100 Return	0.5300***	0.0281***	-0.0895***	-0.0183***	-0.0043
	(79.63)	(3.64)	(-11.05)	(-10.93)	(-1.02)
Observations	37223	37223	37223	37223	37223
R^2	0.283	0.003	0.239	0.918	0.014

Table D6: Decomposition of human vs. bot tweets

This table reports fixed-effects regressions of stock indicators by tweet characteristics. Dependent variables are (log) return, market-model return, volatility i.e. Parkinson (1980) intraday high-low range, (normalized) trading volume, and bid-ask spread. Main independent variables are either human-originated or bot-originated Twitter characters based on tweets collected from 4:30 pm day -1 to 4:29 pm day 0. Message is the natural logarithm of the number of tweets, Sentiment spread is given as $Spread_t = \frac{M_t^{positive} - M_t^{negative}}{Total\ tweets_t}$,where $M_t^{positive}$ and $M_t^{negative}$ are the counts of positive and negative tweets on day t, and Agreement is $Agreement_t = 1 - \sqrt{1 - \left(\frac{M_t^{positive} - M_t^{negative}}{M_t^{positive} + M_t^{negative}}\right)^2}$. Traditional news is the number of relevant news on Financial Times. T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 10%, 5%, 1% significance.

	(1) Return	(2) Mrk-model return	(3) Volatility	(4) Volume	(5) Bid-Ask
Human Sentiment Spread	0.0135*	0.0148	-0.0309***	-0.0062***	0.0040
	(1.76)	(1.63)	(-3.32)	(-2.61)	(0.72)
Human Message	-0.0312	-0.0321	0.2655***	0.0912***	0.0133
	(-1.46)	(-1.28)	(11.45)	(16.51)	(1.10)
Human Agreement	0.0079	0.0074	-0.0179***	-0.0110***	0.0078
	(1.27)	(0.99)	(-2.68)	(-5.07)	(0.75)
Bots Sentiment Spread	0.0161***	0.0167**	-0.0296***	-0.0078***	-0.0036*
	(2.79)	(2.43)	(-4.64)	(-3.80)	(-1.68)
Bots Message	0.0042	0.0065	0.0828***	0.0407***	0.0104
	(0.47)	(0.62)	(9.01)	(12.65)	(1.10)
Bots Agreement	-0.0059	-0.0078	-0.0004	-0.0018	-0.0081*
	(-1.11)	(-1.24)	(-0.08)	(-1.02)	(-1.73)
Traditional News	-0.0009	0.0013	0.0827***	0.0281***	-0.0043
	(-0.10)	(0.12)	(11.71)	(14.98)	(-0.79)
Lagged Return	0.0218*	0.0224	-0.0525***	-0.0128***	0.0085
	(1.87)	(1.58)	(-3.54)	(-5.61)	(1.15)
FTSE100 Return	0.5288***	0.0319***	-0.0907***	-0.0192***	-0.0071
	(75.48)	(3.93)	(-10.70)	(-10.73)	(-1.47)
Observations	33474	33474	33474	33474	33474
R^2	0.282	0.004	0.247	0.917	0.016

Table D7: Human bot interactions

This table reports fixed-effects regressions of stock indicators by tweet characteristics. Dependent variables are (log) return, market-model return, volatility, (normalized) trading volume, and bid-ask spread. Main independent variables are aggregate Twitter characters based on tweets collected from 4:30 pm day -1 to 4:29 pm day 0. Human (bots) tweet features are based on human (bots) tweets. Message is the natural logarithm of the number of tweets, Sentiment spread is given as $Spread_t = \frac{M_t^{positive} - M_t^{negative}}{Total\ tweets_t}$, where $M_t^{positive}$ and $M_t^{negative}$ are the counts of positive and negative tweets on day t, and Agreement is $Agreement_t = 1 - \sqrt{1 - \left(\frac{M_t^{positive} - M_t^{negative}}{M_t^{positive} + M_t^{negative}}\right)^2}$. The 'Hu × Bo ...' are interaction terms between Human and Bot tweets characteristics. Traditional news is the number of relevant news on Financial Times. T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 10%, 5%, 1% significance, respectively.

	(1) Return	(2) Mrk-model return	(3) Volatility	(4) Volume	(5) Bid-Ask
Human Sentiment Spread	0.0151*	0.0166*	-0.0303***	-0.0060**	0.0045
	(1.88)	(1.74)	(-3.07)	(-2.50)	(0.74)
Human Message	-0.0208	-0.0111	0.3676***	0.0967***	0.0167
	(-0.62)	(-0.29)	(11.03)	(11.57)	(1.14)
Human Agreement	0.0053	0.0049	-0.0200**	-0.0117***	0.0152
	(0.70)	(0.55)	(-2.57)	(-4.36)	(0.77)
Bots Sentiment Spread	0.0209***	0.0216**	-0.0364***	-0.0075***	-0.0031
	(2.80)	(2.45)	(-4.28)	(-2.73)	(-1.26)
Bots Message	0.0064	0.0112	0.1063***	0.0419***	0.0117
	(0.58)	(0.87)	(9.82)	(11.61)	(1.33)
Bots Agreement	-0.0081	-0.0093	0.0030	-0.0022	-0.0004
	(-1.12)	(-1.09)	(0.41)	(-0.94)	(-0.07)
Hu Sent. × Bot Sent.	-0.0117	-0.0119	0.0139	-0.0010	-0.0007
	(-1.33)	(-1.15)	(1.31)	(-0.25)	(-0.17)
Hu Mess. × Bot Mess.	0.0101	0.0209	0.1058***	0.0156*	0.0028
	(0.47)	(0.82)	(5.09)	(1.90)	(0.31)
Hu Agree. × Bot Agree.	0.0050	0.0041	-0.0034	0.0011	-0.0137
	(0.69)	(0.48)	(-0.47)	(0.37)	(-0.81)
Traditional News	-0.0010	0.0011	0.0818***	0.0280***	-0.0043
	(-0.12)	(0.11)	(11.60)	(14.96)	(-0.80)
Lagged Return	0.0219*	0.0224	-0.0526***	-0.0128***	0.0085
	(1.87)	(1.59)	(-3.56)	(-5.62)	(1.15)
FTSE100 Return	0.5288***	0.0319***	-0.0908***	-0.0192***	-0.0070
	(75.47)	(3.92)	(-10.71)	(-10.74)	(-1.47)
Observations	33474	33474	33474	33474	33474
R^2	0.282	0.004	0.248	0.917	0.016

Table D8: Regression using standardized negative percentage from Tetlock et al. (2008)

This table reports fixed-effects regressions of stock indicators by tweet characteristics. Dependent variables are (log) return, market-model return, volatility i.e. Parkinson (1980) intraday high-low range, (normalized) trading volume, and bid-ask spread. Main independent variables are aggregate Twitter characters based on tweets collected from 4:30 pm day -1 to 4:29 pm day 0. Message is the natural logarithm of the number of tweets, Negativeness is given as $Negativeness_t = \frac{Neg_t - \mu_{Neg}}{\sigma_{Neg}}$, where $Neg_t = \frac{M_t^{negative}}{Total\ tweeets_t}$, and μ_{Neg}, σ_{Neg} are mean and standard deviation from each company's empirical distribution of Neg_t, and Agreement is $Agreement_t = 1 - \sqrt{1 - \left(\frac{M_t^{positive} - M_t^{negative}}{M_t^{positive} + M_t^{negative}}\right)^2}$. Traditional news is the number of relevant news on Financial Times. T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 10%, 5%, 1% significance, respectively.

	Return	Mrk-model return	Volatility	Volume	Bid-Ask
	(1)	(2)	(3)	(4)	(5)
Negativeness	-0.0141**	-0.0158**	0.0132*	0.0015	0.0026
	(-2.40)	(-2.28)	(1.93)	(0.87)	(0.96)
Message	-0.0232	-0.0234	0.2818***	0.1020***	0.0127
	(-1.18)	(-1.01)	(13.48)	(20.37)	(1.35)
Agreement	0.0038	0.0022	-0.0256***	-0.0116***	0.0061
	(0.68)	(0.34)	(-4.30)	(-6.01)	(0.57)
Traditional News	-0.0014	0.0005	0.0839***	0.0278***	-0.0038
	(-0.17)	(0.06)	(12.48)	(15.85)	(-0.83)
Lagged Return	0.0208*	0.0209	-0.0534***	-0.0132***	0.0079
	(1.87)	(1.55)	(-3.75)	(-6.12)	(1.22)
FTSE100 Return	0.5300***	0.0281***	-0.0895***	-0.0183***	-0.0043
	(79.67)	(3.64)	(-11.04)	(-10.92)	(-1.02)
Observations	37223	37223	37223	37223	37223
R^2	0.283	0.003	0.239	0.918	0.014

Table D9: Decomposition of human vs. bot tweets

This table reports fixed-effects regressions of stock indicators by tweet characteristics. Dependent variables are (log) return, market-model return, volatility i.e. Parkinson (1980) intraday high-low range, (normalized) trading volume, and bid-ask spread. Main independent variables are either human-originated or bot-originated Twitter characters based on tweets collected from 4:30 pm day -1 to 4:29 pm day 0. Human (bots) tweet features are based on human (bots) tweets. Message is the natural logarithm of the number of tweets, Negativeness is given as $Negativeness_t = \frac{Neg_t - \mu_{Neg}}{\sigma_{Neg}}$, where $Neg_t = \frac{M_t^{negative}}{Total\ tweeets_t}$, and μ_{Neg}, σ_{Neg} are mean and standard deviation from each company's empirical distribution of Neg_t, and Agreement is $Agreement_t = 1 - \sqrt{1 - \left(\frac{M_t^{positive} - M_t^{negative}}{M_t^{positive} + M_t^{negative}}\right)^2}$. Traditional news is the number of relevant news on Financial Times. T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 10%, 5%, 1% significance, respectively.

	(1) Return	(2) Mrk-model return	(3) Volatility	(4) Volume	(5) Bid-Ask
Human Negativeness	-0.0116*	-0.0122*	0.0150**	0.0028	0.0006
	(-1.86)	(-1.66)	(2.12)	(1.51)	(0.17)
Human Message	-0.0305	-0.0313	0.2656***	0.0913***	0.0129
	(-1.42)	(-1.25)	(11.45)	(16.51)	(1.05)
Human Agreement	0.0081	0.0077	-0.0253***	-0.0129***	0.0099
	(1.30)	(1.04)	(-3.96)	(-6.06)	(0.74)
Bots Negativeness	-0.0114**	-0.0137**	0.0032	-0.0025	0.0021
	(-2.43)	(-2.45)	(0.64)	(-1.56)	(0.77)
Bots Message	0.0049	0.0072	0.0802***	0.0398***	0.0101
	(0.55)	(0.69)	(8.62)	(12.35)	(1.08)
Bots Agreement	-0.0021	-0.0038	-0.0078	-0.0037**	-0.0088*
	(-0.43)	(-0.65)	(-1.49)	(-2.18)	(-1.84)
Traditional News	-0.0010	0.0012	0.0829***	0.0281***	-0.0043
	(-0.11)	(0.12)	(11.75)	(15.03)	(-0.80)
Lagged Return	0.0218*	0.0223	-0.0524***	-0.0128***	0.0085
	(1.86)	(1.58)	(-3.54)	(-5.60)	(1.15)
FTSE100 Return	0.5288***	0.0320***	-0.0907***	-0.0192***	-0.0071
	(75.47)	(3.93)	(-10.68)	(-10.71)	(-1.47)
Observations	33474	33474	33474	33474	33474
R^2	0.282	0.004	0.246	0.917	0.016

Table D10: Human bot interactions

This table reports fixed-effects regressions of stock indicators by tweet characteristics. Dependent variables are (log) return, market-model return, volatility, (normalized) trading volume, and bid-ask spread. Main independent variables are aggregate Twitter characters based on tweets collected from 4:30 pm day -1 to 4:29 pm day 0. Human (bots) tweet features are based on human (bots) tweets. Message is the natural logarithm of the number of tweets, Negativeness is given as $Negativeness_t = \frac{Neg_t - \mu_{Neg}}{\sigma_{Neg}}$, where $Neg_t = \frac{M_t^{negative}}{Total\ tweeets_t}$, and μ_{Neg}, σ_{Neg} are mean and standard deviation from each company's empirical distribution of Neg_t, and Agreement is $Agreement_t = 1 - \sqrt{1 - \left(\frac{M_t^{positive} - M_t^{negative}}{M_t^{positive} + M_t^{negative}}\right)^2}$. The 'Hu × Bo ...' are interaction terms between Human and Bot tweets characteristics. Traditional news is the number of relevant news on Financial Times. T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 10%, 5%, 1% significance, respectively.

	(1) Return	(2) Mrk-model return	(3) Volatility	(4) Volume	(5) Bid-Ask
Human Negativeness	-0.0120*	-0.0127*	0.0134*	0.0030	0.0008
	(-1.87)	(-1.68)	(1.91)	(1.59)	(0.22)
Human Message	-0.0172	-0.0069	0.3680***	0.0972***	0.0154
	(-0.51)	(-0.18)	(11.07)	(11.60)	(0.94)
Human Agreement	0.0045	0.0041	-0.0263***	-0.0132***	0.0178
	(0.59)	(0.46)	(-3.47)	(-4.94)	(0.78)
Bots Negativeness	-0.0118**	-0.0140**	0.0030	-0.0023	0.0027
	(-2.53)	(-2.53)	(0.61)	(-1.44)	(0.97)
Bots Message	0.0077	0.0126	0.1037***	0.0411***	0.0113
	(0.70)	(0.97)	(9.48)	(11.36)	(1.33)
Bots Agreement	-0.0050	-0.0061	-0.0035	-0.0039*	-0.0009
	(-0.72)	(-0.75)	(-0.49)	(-1.67)	(-0.14)
Hu Neg. × Bot Neg.	0.0004	-0.0002	-0.0030	-0.0013	-0.0014
	(0.09)	(-0.04)	(-0.53)	(-0.73)	(-0.64)
Hu Mess. × Bot Mess.	0.0139	0.0252	0.1046***	0.0058	0.0017
	(0.64)	(0.98)	(5.01)	(0.93)	(0.17)
Hu Agree. × Bot Agree.	0.0056	0.0049	-0.0042	0.0006	-0.0140
	(0.78)	(0.58)	(-0.59)	(0.21)	(-0.84)
Traditional News	-0.0011	0.0010	0.0820***	0.0281***	-0.0043
	(-0.13)	(0.10)	(11.64)	(15.01)	(-0.80)
Lagged Return	0.0218*	0.0223	-0.0525***	-0.0128***	0.0085
	(1.87)	(1.58)	(-3.55)	(-5.60)	(1.14)
FTSE100 Return	0.5288***	0.0319***	-0.0908***	-0.0192***	-0.0070
	(75.46)	(3.93)	(-10.69)	(-10.72)	(-1.47)
Observations	33474	33474	33474	33474	33474
R^2	0.282	0.004	0.247	0.917	0.016

Table D11: Regression of market model return based on prior 250 days

This table reports fixed-effects regressions of market model return by tweet characteristics. Dependent variables is market-model return. Parameters of the market model are estimated based on the [-260, -10] time window. Main independent variables are aggregate Twitter characters based on tweets collected from 4:30 pm day -1 to 4:29 pm day 0. Human (bots) tweet features are based on human (bots) tweets. Message is the natural logarithm of the number of tweets, Positiveness is given as $Positiveness_t = \ln\left(\frac{1+M_t^{positive}}{1+M_t^{negative}}\right)$, where $M_t^{positive}$ and $M_t^{negative}$ are the counts of positive and negative tweets on day t, and Agreement is $Agreement_t = 1 - \sqrt{1-\left(\frac{M_t^{positive}-M_t^{negative}}{M_t^{positive}+M_t^{negative}}\right)^2}$. The 'Hu × Bo ...' are interaction terms between Human and Bot tweets characteristics. Traditional news is the number of relevant news on Financial Times. T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 10%, 5%, 1% significance, respectively.

	(1)	(2)	(3)
Positiveness	0.0232**		
	(2.50)		
Message	-0.0179		
	(-1.15)		
Agreement	-0.0053		
	(-0.67)		
Human Positiveness		0.0231**	0.0212**
		(2.53)	(2.46)
Human Message		-0.0295	-0.0094
		(-1.20)	(-0.68)
Human Agreement		-0.0052	-0.0052
		(-0.62)	(-0.63)
Bots Positiveness		0.0127*	0.0153*
		(1.77)	(1.64)
Bots Message		0.0077	0.0040
		(0.72)	(0.50)
Bots Agreement		-0.0084	-0.0100
		(-1.34)	(-1.24)
Hu × Bot Pos.			-0.0081
			(-0.81)
Hu × Bot Mess.			0.0009
			(0.06)
Hu × Bot Agree.			0.0046
			(0.59)
Traditional News	0.0002	0.0011	0.0009
	(0.02)	(0.11)	(0.10)
Lagged Return	0.0194	0.0201	0.0201
	(1.46)	(1.49)	(1.49)
Market Return	0.0212***	0.0207***	0.0207***
	(2.75)	(2.67)	(2.67)
Observations	37619	36925	36925
R^2	0.003	0.003	0.000

Table D12: Regressions for all tweets

This table reports fixed-effects regressions of stock indicators by tweet characteristics. Dependent variables are (log) return, market-model return, volatility i.e. Parkinson (1980) intraday high-low range, (normalized) trading volume, and bid-ask spread. Main independent variables are aggregate Twitter characters based on tweets collected from 4:30 pm day -1 to 4:29 pm day 0. Message is the natural logarithm of the number of tweets, Positiveness is given as $Positiveness_t = \ln\left(\frac{1+M_t^{positive}}{1+M_t^{negative}}\right)$, where $M_t^{positive}$ and $M_t^{negative}$ are the counts of positive and negative tweets on day t, and Agreement is $Agreement_t = 1 - \sqrt{1-\left(\frac{M_t^{positive}-M_t^{negative}}{M_t^{positive}+M_t^{negative}}\right)^2}$. No. of traditional news is the number of relevant news on Financial Times. Negativeness in news is percentage of negative words in traditional news using McDonald and Loughran dictionary. T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 10%, 5%, 1% significance, respectively.

	(1) Return	(2) Mrk-model return	(3) Volatility	(4) Volume	(5) Bid-Ask
Positiveness	0.0215***	0.0214**	-0.0473***	-0.0065***	-0.0042
	(2.72)	(2.30)	(-5.25)	(-2.89)	(-1.19)
Message	-0.0215	-0.0208	0.1629***	0.0670***	-0.0001
	(-1.61)	(-1.32)	(10.34)	(17.46)	(-0.01)
Agreement	-0.0032	-0.0039	-0.0088	-0.0104***	0.0071
	(-0.48)	(-0.48)	(-1.18)	(-4.60)	(0.83)
No. of Traditional News	0.0186	0.0239*	0.0922***	0.0294***	-0.0007
	(1.54)	(1.68)	(9.49)	(10.73)	(-0.24)
Negativeness in News	-0.0270**	-0.0317**	0.0032	0.0017	-0.0032
	(-2.45)	(-2.43)	(0.34)	(0.63)	(-0.78)
Lagged Return	0.0200*	0.0198	-0.0538***	-0.0135***	0.0082
	(1.80)	(1.48)	(-3.78)	(-6.18)	(1.28)
FTSE100 Return	0.5292***	0.0282***	-0.0899***	-0.0182***	-0.0044
	(79.70)	(3.66)	(-11.09)	(-10.89)	(-1.01)
Observations	37619	37619	37619	37619	37619
R^2	0.282	0.004	0.233	0.918	0.014

Table D13: Decomposition of human vs. bot tweets

This table reports fixed-effects regressions of stock indicators by tweet characteristics. Dependent variables are (log) return, market-model return, volatility i.e. Parkinson (1980) intraday high-low range, (normalized) trading volume, and bid-ask spread. Main independent variables are either human-originated or bot-originated Twitter characters based on tweets collected from 4:30 pm day -1 to 4:29 pm day 0. Human (bots) tweet features are based on human (bots) tweets. Message is the natural logarithm of the number of tweets, Positiveness is given as $Positiveness_t = \ln\left(\frac{1+M_t^{positive}}{1+M_t^{negative}}\right)$, where $M_t^{positive}$ and $M_t^{negative}$ are the counts of positive and negative tweets on day t, and Agreement is $Agreement_t = 1 - \sqrt{1-\left(\frac{M_t^{positive}-M_t^{negative}}{M_t^{positive}+M_t^{negative}}\right)^2}$. No. of traditional news is the number of relevant news on Financial Times. Negativeness in news is percentage of negative words in traditional news using McDonald and Loughran dictionary. T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 10%, 5%, 1% significance, respectively.

	(1) Return	(2) Mrk-model return	(3) Volatility	(4) Volume	(5) Bid-Ask
Human Positiveness	0.0203***	0.0210**	-0.0429***	-0.0067***	-0.0016
	(2.61)	(2.28)	(-4.91)	(-3.05)	(-0.39)
Human Message	-0.0334	-0.0324	0.2639***	0.0850***	0.0083
	(-1.59)	(-1.32)	(11.48)	(16.18)	(0.78)
Human Agreement	-0.0032	-0.0038	0.0010	-0.0079***	0.0076
	(-0.45)	(-0.45)	(0.13)	(-3.43)	(0.94)
Bots Positiveness	0.0164***	0.0144**	-0.0346***	-0.0047**	-0.0087**
	(2.68)	(2.00)	(-5.56)	(-2.34)	(-2.20)
Bots Message	0.0041	0.0065	0.0941***	0.0430***	0.0140
	(0.45)	(0.61)	(10.35)	(13.55)	(1.22)
Bots Agreement	-0.0066	-0.0081	0.0003	-0.0026	-0.0082*
	(-1.23)	(-1.30)	(0.05)	(-1.50)	(-1.80)
No. of Traditional News	0.0176	0.0226	0.0798***	0.0261***	-0.0012
	(1.43)	(1.55)	(8.26)	(9.53)	(-0.43)
Negativeness in News	-0.0247**	-0.0288**	0.0027	0.0014	-0.0035
	(-2.21)	(-2.17)	(0.29)	(0.51)	(-0.86)
Lagged Return	0.0205*	0.0206	-0.0533***	-0.0132***	0.0079
	(1.83)	(1.53)	(-3.75)	(-6.14)	(1.22)
FTSE100 Return	0.5294***	0.0277***	-0.0903***	-0.0186***	-0.0044
	(79.37)	(3.59)	(-11.13)	(-11.01)	(-1.03)
Observations	36925	36925	36925	36925	36925
R^2	0.283	0.004	0.242	0.919	0.014

Table D14: Human bot interactions

This table reports fixed-effects regressions of stock indicators by tweet characteristics. Dependent variables are (log) return, market-model return, volatility, (normalized) trading volume, and bid-ask spread. Main independent variables are aggregate Twitter characters based on tweets collected from 4:30 pm day -1 to 4:29 pm day 0. Human (bots) tweet features are based on human (bots) tweets. Message is the natural logarithm of the number of tweets, Positiveness is given as $Positiveness_t = \ln\left(\frac{1+M_t^{positive}}{1+M_t^{negative}}\right)$, where $M_t^{positive}$ and $M_t^{negative}$ are the counts of positive and negative tweets on day t, and Agreement is $Agreement_t = 1 - \sqrt{1-\left(\frac{M_t^{positive}-M_t^{negative}}{M_t^{positive}+M_t^{negative}}\right)^2}$. The 'Hu × Bo ...' are interaction terms between Human and Bot tweets characteristics. No. of traditional news is the number of relevant news on Financial Times. Negativeness in news is percentage of negative words in traditional news using McDonald and Loughran dictionary. T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 10%, 5%, 1% significance, respectively.

	(1) Return	(2) Mrk-model return	(3) Volatility	(4) Volume	(5) Bid-Ask
Human Positiveness	0.0204**	0.0221**	-0.0342***	-0.0057**	-0.0004
	(2.47)	(2.27)	(-3.63)	(-2.49)	(-0.07)
Human Message	-0.0309	-0.0202	0.3805***	0.0979***	0.0129
	(-0.95)	(-0.53)	(11.64)	(12.36)	(0.85)
Human Agreement	-0.0062	-0.0071	-0.0046	-0.0083***	0.0111
	(-0.81)	(-0.78)	(-0.55)	(-3.20)	(0.83)
Bots Positiveness	0.0186**	0.0180*	-0.0355***	-0.0045	-0.0080*
	(2.03)	(1.67)	(-3.85)	(-1.64)	(-1.69)
Bots Message	0.0046	0.0104	0.1330***	0.0474***	0.0163
	(0.38)	(0.72)	(11.23)	(12.49)	(1.60)
Bots Agreement	-0.0104	-0.0113	0.0087	-0.0014	-0.0030
	(-1.46)	(-1.34)	(1.19)	(-0.60)	(-0.61)
Hu Pos. × Bot Pos.	-0.0035	-0.0059	-0.0040	-0.0008	-0.0014
	(-0.40)	(-0.58)	(-0.44)	(-0.29)	(-0.31)
Hu Mess. × Bot Mess.	0.0025	0.0130	0.1286***	0.0142**	0.0046
	(0.11)	(0.49)	(5.85)	(2.26)	(0.44)
Hu Agree. × Bot Agree.	0.0072	0.0070	-0.0050	-0.0010	-0.0084
	(1.04)	(0.86)	(-0.73)	(-0.40)	(-0.77)
No. of Traditional News	0.0176	0.0225	0.0784***	0.0259***	-0.0013
	(1.43)	(1.54)	(8.12)	(9.47)	(-0.45)
Negativeness in News	-0.0247**	-0.0288**	0.0028	0.0014	-0.0035
	(-2.21)	(-2.17)	(0.30)	(0.52)	(-0.85)
Lagged Return	0.0205*	0.0206	-0.0534***	-0.0132***	0.0079
	(1.83)	(1.53)	(-3.76)	(-6.15)	(1.22)
FTSE100 Return	0.5294***	0.0277***	-0.0904***	-0.0186***	-0.0044
	(79.36)	(3.58)	(-11.15)	(-11.02)	(-1.04)
Observations	36925	36925	36925	36925	36925
R^2	0.283	0.004	0.244	0.919	0.014

Table D15: Fama-MacBeth regressions for all tweets

This table reports fixed-effects regressions of stock indicators by tweet characteristics. Dependent variables are (log) return, market-model return, volatility i.e. Parkinson (1980) intraday high-low range, (normalized) trading volume, and bid-ask spread. Main independent variables are aggregate Twitter characters based on tweets collected from 4:30 pm day -1 to 4:29 pm day 0. Message is the natural logarithm of the number of tweets, Positiveness is given as $Positiveness_t = \ln\left(\frac{1+M_t^{positive}}{1+M_t^{negative}}\right)$, where $M_t^{positive}$ and $M_t^{negative}$ are the counts of positive and negative tweets on day t, and Agreement is $Agreement_t = 1 - \sqrt{1-\left(\frac{M_t^{positive}-M_t^{negative}}{M_t^{positive}+M_t^{negative}}\right)^2}$. Traditional news is the number of relevant news on Financial Times. Lagged return and market return are included as control variables but not reported for brevity. T-statistics based on Huber-White robust standard errors are reported in parentheses. *, **, *** denote 10%, 5%, 1% significance, respectively.

	(1) Return	(2) Mrk-model return	(3) Volatility	(4) Volume	(5) Bid-Ask
Positiveness	0.0177**	0.0174**	-0.0595***	-0.0373***	-0.0023
	(2.43)	(2.03)	(-7.42)	(-5.68)	(-0.40)
Message	-0.0021	-0.0035	0.0491***	0.1518***	0.0076
	(-0.42)	(-0.61)	(9.73)	(41.15)	(1.49)
Agreement	-0.0020	-0.0044	-0.0034	-0.0839***	0.0097
	(-0.30)	(-0.56)	(-0.47)	(-10.26)	(0.98)
Traditional news	-0.0070	-0.0043	0.0836***	0.1431***	-0.0154***
	(-0.81)	(-0.42)	(10.12)	(21.45)	(-2.85)
Observations	37619	37619	37619	37619	37618
R^2	0.092	0.092	0.104	0.115	0.070