Twitter, Investor Demographics, and the Diffusion of Information in Financial Markets

Mahnaz Paydarzarnaghi and David Rakowski[1]

January 25, 2023

Abstract

This research explores the link between information diffusion and social media user characteristics through an analysis of Twitter posts and stock returns. We examine over 9.7 million company-specific Twitter posts and 398,129 Twitter users from 2017 to 2019. We test how stock price reactions to information differ between human and bot (automated accounts) users and how stock price reactions to information are associated with the race, ethnicity, age, and gender of social media users who post that information. We find that posts generated by real people are more strongly associated with information, while posts generated by bots are more associated with a temporary liquidity shock that dissipates within days. We also show that Twitter posts, including images and URLs, impact stock prices more than posts with text. We find that posts generated by white or Hispanic social media users substantially impact stock prices more than other races and ethnicities. In addition, we find that posts by men have a stronger impact on stock returns than posts by women. Finally, we find that the age of a social media user is positively associated with the impact of that user's posts.

KEYWORDS

Asset Pricing, Information Diffusion, Social Media, Twitter, Bots, Human users, Demographic Characteristics

[1] Paydarzarnaghi (email: Mahnaz.paydarzarnaghi@uta.edu), and Rakowski (email: rakowski@uta.edu) are at the University of Texas at Arlington.

Introduction

This research explores the link between information diffusion with Twitter posts and users. It is also known that the speed of information diffusion is important for market efficiency (Lu 2011), especially the information posted on social media ((Peress 2014) and (Rakowski, Shirley, and Stark 2021)). By enormous volume and the rapid speed of information transmission, social media provides a more comprehensive real-time news database than traditional media channels. By existing social bots, the large amount of information can also contain potential noise that might mislead readers (Fan, Talavera, and Tran 2020). The bots can disseminate fake news and manipulate stock markets. According to Fan, Talavera, and Tran (2020), the influence of social media bot activities on stock markets is significant. This research aims also to understand better social media users' roles, both human and bot (automated accounts) users, in the diffusion of information in financial markets. We expect that the human and bots' tweets have an inverse impact on impacts on stock prices, and the magnitude of the influence of human activity is more remarkable than bots.

Furthermore, we test how stock price reactions to information are associated with the demographic characteristics of social media users who post that information. In this study, we examine race, gender, and age characteristics. According to (Yao, Gutter, and Hanna, 2005), culture provides a context in which information is framed, and preferences are formed; their model illustrates the importance that culture, represented by race and ethnic status, has on the financial decision-making process. According to Gutter and Montalto (1999), racial differences in investor behavior may indicate differences in risk tolerance and investment choice or cultural differences in investment behavior. Moreover, since Black and Hispanic investors are more risk averse in their choice of assets than whites ((Badu and Salandro, 1999) and (Yao, Gutter, and Hanna 2005)), we estimate that white social media users have a more substantial impact on stock prices. According to Cueva et al. (2019), psychological research demonstrates that, in areas such as finance, men are more overconfident than women. Thus, the theory predicts that men will trade more excessively than women. Therefore, we expect that men social media users will post information that substantially impacts stock prices. Research shows that investor behavior varies with the age, of investors (Nagy and Obenberger 1994). Because age is known to be negatively associated with noise on social media (Ozimek and Bierhoff 2016), we conjecture that older social media users will post information that has a stronger impact on stock prices.

Social media data allow us to examine the sentiment and perspectives of a larger group of individuals. We collect Twitter posts and users' profile information using Twitter Application Programming Interface (API); We also obtain the Google search volume from Google Trends and stock prices from the Center for Research in Security Prices (CRSP), Wharton Research Data Services (WRDS). In addition, we apply machine learning methods to infer the demographic characteristics of Twitter's users.

This analysis will provide new evidence on the heterogeneity of news sources in terms of how the market interprets news produced by types of social media users. As social media networks displace traditional media as news sources for investors, it is necessary to understand how some social media posts appear as valuable news while others are better characterized as noise. Our novel machine learning approach to social media inferred demographics provides a route to new evidence on these issues. We will first provide evidence of how a particular demographic characteristic, race, gender, and age, are associated with how the market interprets the value of information from social media users.

Our results suggest that Twitter post type, post contents, Tweets' contextual information, Twitter user type, and Twitter users' race, gender, and age affect stock prices. On the one hand, human users reduce stock returns through retweets and replies with a delayed effect. On the other hand, bots increase the stock price through Retweeting and decrease it by tweeting and quoting. We observe that human users provide mostly information on Twitter, while bots provide mainly price pressure. Moreover, Tweet's contextual information can influence stock excess returns. We find that human Hispanic users positively influence the stock price. We also observe that male and female human users' impact is inverse on the stock price. Finally, we notice that stock prices react more strongly to posts by older social media users than younger social media users.

Our research contributes to several streams of research. First and foremost, our paper is the first research on social media users' demographic characteristics in finance literature based on our topic modeling procedure. Secondly, it contributes to the growing literature on the role of social media in financial markets. Third, we investigate and introduce new methods to infer the demographic characteristics of users, like their race, gender, and age. Fourth, several studies focus on the impact of Tweets sentiment on the financial market; however, we focus on the influence of the source of Tweets, bots, contents of tweets, and dissimilar types of Twitter posts.

The important implication of our research is that Twitter should establish policies for enhancing the transparency of tweets posted by bots. In addition, policymakers should monitor social media platforms and prevent them from spreading fake information. Moreover, investors and readers of social media who seek stock news should increase their awareness about the Tweets' content and who the sender is.

This paper is organized as follows: Section 2 provides relevant background information. Section 3 describes the sample and data collection. Section 4 contains hypothesis development and models. Section 5 reports the preliminary results. Finally, we present conclusions in Section 6.

2. LITERATURE

According to Lu (2004), asset pricing models typically assume both that the diffusion of every type of publicly available information takes place instantaneously among all investors and that investors act on the information as soon as it is received. The speed of information diffusion is important for market efficiency. Peress (2014) investigates causal impact of media on trading and price formation by examining national newspaper strikes in several countries. His findings demonstrate that the media contribute to the efficiency of the stock market by improving the dissemination of information among investors and its incorporation into stock prices.

According to Rakowski, Shirley, and Stark (2021), with the advent of social media, sources of information are shifting from curated, top-down providers to a more democratized setting in which an individual can share almost anything using little more than an Internet connection. The emergence of crowd-sourced social media and news platforms has transformed technologies intended for social communications into, among other things, channels for price discovery in the financial markets (Rakowski, Shirley, and Stark 2021).

Microblogging platforms have become an easy and fast way to share and consume information of interest on the Web in real-time. For instance, in recent years, Twitter has emerged as an important source of real-time information exchange platform. It has empowered citizens, companies, marketers to act as content generators, that is, people share information about what they experience, eyewitness, and observe about topics from a wide range of fields such as epidemics, elections, stocks and more (Uddin, Imran, and Sajjad 2014). Twitter activity increases the diffusion of information across a wide range of firms (Rakowski, Shirley, and Stark 2021). According to Rakowski, Shirley, and Stark (2021), twitter is an ideal mechanism to disseminate information about any topic including financial securities. Furthermore, the design, tagging system, and searchability of Twitter allow tweets to be used as a measure of investor attention that is directly tied to a particular stock or topic. In Twitter, users communicate with each other by publishing text-based posts. The popularity and open structure of Twitter have attracted a large number of automated programs, known as bots, which spread spam or malicious contents (Chu et al. 2012).

The attitudes and behaviors of social media users are central to policy-making, commercial prediction tasks and financial market. Changing demographic structure of the U.S. population affects returns through its impact on the risk tolerance of potential investor (Poterba 2001). According to (Yao, Gutter, and Hanna, 2005), culture provides a context in which information is framed, and preferences are formed; their model illustrates the importance that culture, represented by race and ethnic status, has on the financial decision-making process. According to Gutter and Montalto (1999), racial differences in investor behavior may indicate differences in risk tolerance and investment choice or cultural differences in investment behavior. According to Cueva et al. (2019), psychological research demonstrates that, in areas such as finance, men are more overconfident than women. Thus, the theory predicts that men will trade more excessively than women. Research shows that investor behavior varies with the age, of investors (Nagy and Obenberger 1994).

According to Asmussen and Møller (2019), topic modeling enables vast amounts of papers to be reviewed in a transparent, reliable, faster, and reproducible way. We apply topic modeling on the titles and abstracts of 2,024 collections of papers in information systems, computer science, business, finance, economics, and management contexts of the Web of Science database to find what other researchers did on information diffusion topics. We search "information diffusion", "diffusion of information", "spreading of information," "information speed", "speed of information", "diffusion of news", "news diffusion", "news speed", "speed of news," "news distribution," "distribution of news," and "spreading of news" keywords. We use VOSViewer[2] software for constructing and visualizing Co-occurrence networks of the main keywords of the abstract and title of the papers in Figures 1 and 2. Figure 1 depicts five colorful clusters related to information diffusion. For example, previous literature works are related to spreading information and the internet, technology, social media, social network, behavior, efficiency, stock return, game theory, and investor sentiments. Figure 2 displays one of the clusters of keywords, and we can see

[2] van Eck, N. J.; Waltman, L. (2010) VOSViewer: Visualizing Scientific Landscapes

that researchers find an association between information diffusion, social media, fake news, and Twitter.

[Figure 1 inserts here]

[Figure 2 inserts here]

In Figure 3, we apply semantic network analysis to the words of titles and abstracts of all 2,024 papers related to information diffusion. A semantic network, or frame network, represents semantic relations between concepts and keywords in the network. A directed or undirected graph consists of vertices representing semantic relations between concepts, mapping, or connecting semantic fields. The line segment in the network is called Arc. The Arc represents the relation between nodes, and it may be followed to proceed from node to node. The number of lines between nodes indicates the number of relations of the node with other nodes. Figure 3 is the undirected graph representing the main keyword using nodes and the underlying co-occurrences employing connecting edges, in which the larger the number, the greater their relative frequency. The thickness of the line indicates the strength of the relation between the words. The thicker lines indicate there is a stronger association between nodes. For example, social media, information diffusion, and network association are thick and robust. However, we cannot discover any strong association between information diffusion and demographic characteristics of users and investors in previous literature.

[Figure 3 inserts here]

Then we filter the papers into 578 collections of papers in the Web of Science database's business, finance, economics, and management contexts. We have three steps for the topic modeling: pre-processing, topic modeling, and visualization, where the topic model Latent Dirichlet Allocation (LDA) is used. We clean text data of abstracts and titles, tokenize sentences and lemmatize for the pre-processing step. Then, we find the optimal topics using the Elbow method and the LDA Topic Model (Blei 2012). For the last step, we use pyLDAvis for visualization. We can apply coherence or another metric to determine optimum topics. We apply networks of co-occurring words on information diffusion and age in the collection of our abstracts and titles. However, we can see in Figure 4 that there needs to be research on information diffusion and age in the business context. We use pyLDAV, a commonly used and excellent way to visualize information in a topic model. We choose the number of Topic 5 and visualize it by pyLDAvis. The result is provided in Figure 5. The main topic is topic three since its circle is reasonably big and separate from other topics. The top 30 most relevant for topic 3 reveal no age keyword in the previous research on information diffusion. We also apply LDA (Latent Dirichlet Allocation) Topic Model for 20 topics. Figure 6 presents the 20 topics, and there is no age, race, or gender in information diffusion research of prior literature.

[Figure 4 inserts here]

[Figure 5 inserts here]

[Figure 6 inserts here]

Using topic modeling, we find a gap in previous literature about the demographic characteristics of social media users and information diffusion.

3. DATA, MEASURES and SAMPLE CREATION

First, we explain trending stocks and our measures of dependent variables in this section. Second, we describe Twitter, cashtags, Twitter post types, different types of tweet contents, diverse kinds of Tweet entity, source of the Twitter posts, and bot messages. We then explain the Twitter variables measures. Third, we describe the prediction of the demographic characteristics of users and their measures. Fourth, we describe how we collect the Google Trend and the creation of its measure.

3.1 Trending Stocks

Trending stocks are those when a stock is undergoing a significant move compared to its underlying index. The trend can be upward or downwards. We select Trending share since we can have enough daily Twitter posts about them. For this research, we collect tickers of 58 Trending stocks from YahooFinance![3] and focus on these stocks. The list of the tickers and the company names is provided in Table A1.

[Table A1 inserts here]

We collect daily security-level data from the Center for Research in Security Prices database (CRSP) including trading volume, price, and returns. We also obtain the Fama-French three-factor model data from Kenneth R. French's website[4]. We calculate daily value (FAMA and FRENCH 1992) three-Factor Model excess return in the basis points for stock *i*, on day *t* varying values of n.

[3] https://finance.yahoo.com/trending-tickers/

[4] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

3.2 Twitter

According to Cesare, Grant, and Nsoesie (2017), there are several benefits to using social media data. First, social media data provide real-time updates on users' thoughts, feelings, and experiences, allowing researchers to track users' attitudes and behaviors as they emerge. Both the population and individual-level scale of these data create the opportunity to study behaviors that are difficult to assess through traditional means of data collection. Second, because social media posts are unsolicited, users may report opinions and behaviors with greater fidelity than they would in the context of interviews or surveys.

Twitter is an online social networking service that enables users to send short 280-character messages called tweets.[5] Twitter, a social networking site launched in 2006, is one of the most popular social media platforms available today, with 217 million daily active users and 500 million tweets sent daily[6]. As of Nov 22, 2022, Statsita[7], social network Twitter is particularly popular in the United States, whereas of January 2022, the microblogging service had audience reach of 76.9 million users. Statista is a German company specializing in market and consumer data. According to the company, its platform contains more than 1,000,000 statistics on more than 80,000 topics from more than 22,500 sources and 170 different industries and generates a revenue of about €60 million (Wikipedia Contributors, 2019).

[Figure 7 inserts here]

The Pew Research Center measured news consumption on social media by surveying over the period from August 31 to September 7, 2020. Figure 8 presents the percentage of each social media site's users who regularly get news from that particular site. It shows that more than 50% of Twitter users get news regularly on Twitter.

[Figure 8 inserts here]

As of Feb 9, 2021, Statista's website[8] published the global social networks ranked by several users in 2021. It mentions that social networking sites now have 278,414 to 3.6 billion users. According to Statista's website, as of February 2021, most Twitter users are young. Figure 9 shows that 42 percent of Twitter users are younger than 30.

[Figure 9 inserts here]

[5] https://www.statista.com/topics/737/twitter/#topicHeader__wrapper
[6] https://www.omnicoreagency.com/twitter-statistics/
[7] https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/
[8] Statista is a German company specializing in market and consumer data. According to the company, its platform contains more than 1,000,000 statistics on more than 80,000 topics from more than 22,500 sources and 170 different industries and generates a revenue of about €60 million (Wikipedia Contributors, 2019).

According to Rakowski, Shirley, and Stark (2021), with such a large and active user base, Twitter is an ideal mechanism to disseminate information about any topic including financial securities. Furthermore, the design, tagging system, and searchability of Twitter allow tweets to be used as a measure of investor attention that is directly tied to a particular stock or topic.

We collect Twitter posts and users' profile information using Twitter Application Programming Interface (API). We obtain 9,679,647 posts by cashtags of individual securities at daily frequencies from 2017 to 2019. Twitter users place a dollar sign ($) before a ticker, such as $Tsla, relaying that the tweet is about Tesla stock. Figure 10 is an example of an anonymized tweet containing the $Tsla cashtag and the user profile. According to Rakowski, Shirley, and Stark (2021), consumers of information on Twitter can quickly sort information through an enormous amount of data by finding specific companies using cashtags to focus on those tweets about financial securities of interest, helping to reduce the information overload problem.

[Figure 10 inserts here]

A Twitter post can be a Tweet, a Retweet, a Reply, or a Quote. A Retweet is a re-posting of a Tweet. Twitter's Retweet feature helps you and others quickly share that Tweet with all of your followers[9]. A reply is when you respond to another person's Tweet. A Quote feature allows you to Tweet another person's Tweet with your own comment added[10]. People post Tweets, which may contain photos, videos, links, and text[11]. Figures 11,12, and 13 are examples of different types of post Tweets.

[Figure 11 inserts here]

[Figure 12 inserts here]

[Figure 13 inserts here]

Figure 14 summarizes our data collection based on the posts and tweet contents. From 9,679,647 posts, we collected 6,492,648 tweets, 2,261,190 retweets, 584,905 replies, and 340,904 quotes. We have 3,097,999 Tweets, including URLs, 1,284,439 Tweets, including images/videos, and 2,150,895 Tweets, including just text.

[Figure 14 inserts here]

In our data, we collect the Tweet annotations which offer a way to understand contextual information about the Tweet itself. Entity annotations are programmatically defined entities that

[9] https://help.twitter.com/en/using-twitter/retweet-faqs
[10] https://help.twitter.com/en/using-twitter/types-of-tweets
[11] https://help.twitter.com/en/resources/new-user-faq

are nested within the entities field and are reflected as annotations in the payload[12]. The entity annotations can have Person, Place, Product, Organization, and other. Twitter classifies Tweets semantically, meaning that we curate lists of keywords, hashtags, and @handles that are relevant to a given topic. If a Tweet contains the text Twitter has specified, it will be labeled appropriately[13]. According to Finin et al., (2010) , person entities are limited to humans (living, deceased, fictional, deities, ...) identified by name, nickname, or alias. Organization entities are limited to corporations, institutions, government agencies and other groups of people defined by an established organizational structure. Some examples are businesses (Bridgestone Sports Co.), stock ticker symbols (NASDAQ), multinational organizations (European Union), and sports teams (the Yankees). Place entities include names of politically or geographically defined places (cities, provinces, countries, international regions, bodies of water, mountains, etc.). Locations also include man-made structures like airports, highways, streets, and factories. Produce entities include names of products such as Mountain Dew and Mozilla Firefox.

According to (Chu et al. (2012), the popularity and open structure of Twitter have attracted a large number of automated programs, known as bots, which appear to be a double-edged sword to Twitter. Legitimate bots generate a large amount of benign tweets delivering news and updating feeds, while malicious bots spread spam or malicious contents. Chu et al. (2012) state that human tweets are manually posted via the Twitter website, mobile applications (e.g., Twitter), or desktop clients (e.g., TweetDeck); tweeting via such devices requires human participation. Figure 15 shows the distribution of the Tweets from the top 50 tweeting devices.

[Figure 15 inserts here]

We separate tweets based on Tweeting devices and make three different datasets based on the source of Tweets. The first dataset includes Tweets from any Tweeting devices (sources). For the second dataset, we filter Tweets and pick tweets if they come from Twitter Web Client, Twitter Web app, Twitter for iPhone, Twitter for Android, Twitter for iPad, and TweetDeck. According to Chu et al. (2012), our second dataset is human tweets rather than bot messages. Figure 16 shows the distribution of the Tweets for Tweet's source of the second dataset.

[Figure 16 inserts here]

[12] https://developer.twitter.com/en/docs/twitter-api/annotations/overview

[13] https://developer.twitter.com/en/docs/twitter-api/annotations/faq

For the third one, we include all Tweeting devices except Twitter Web Client, Twitter Web app, Twitter for iPhone, Twitter for Android, Twitter for iPad, and TweetDeck. The third dataset includes more bot tweets.

3.3 Demographic Characteristics

Having demographic information of social media users would allow us to make inferences about how attitudes towards securities differ across user demographics. However, the major limitation of social media platforms data is a lack of demographic indicators such as age, race, and gender. To overcome this hurdle, we extract names, profile photos, and surnames of the Twitter users.

For the race and ethnicity, we take the users' names and apply the proposed Bayesian approach of Chang et al. (2010), in which the accuracy of the model is between 0.78 and 0.84. Table A2 exhibits the example of the names most associated with each ethnicity learned by the proposed model of Chang et al. (2010).

[TABLE A2 inserts here]

For age and gender prediction, we download the profile picture of the users and analyze profile pictures by using a face recognition algorithm running on Amazon Web Services cloud to find the age range and gender of users. Figure 17 shows that from 398,129 unique Twitter users, we could download 335,616 profile pictures, and we predict 203047 users' age and gender. We could infer the race of 299,141 users. We also find the gender and age of 203,047 users by using AWS.

[Figure 17 inserts here]

3.4 Google Trends

We also obtain the Google search volume from Google Trend; We search for the stock tickers in Google; if it brings up the stock price or a box with information about the firm, we use the stock ticker in collecting the Google Search Index Volume as the keyword. Otherwise, we add ticker before stock name in searching for it, followed by (Ben-Rephael, Da, and Israelsen 2017). Figure 18 shows an example of selecting keywords for searching in Google Trends. Engelberg and Parsons (2011) calculate Abnormal Google's daily Search Volume (ADSVI) measure calculated as the natural log of the ratio of DSVI on day *t* to the average of DSVI over the previous month where DSVI is Google's daily Search Volume. We follow (Ben-Rephael, Da, and Israelsen 2017) and we assign DSVI on day *t* a score of 0, 1, 2, 3 or 4 if the average is between 80% and 90%, 90% and 94%, 94% and 96%, or greater than 96% of the previous 30 days' daily GSVI, respectively.

[Figure 18 inserts here]

Table 1 displays the names, sources, and brief definitions for all of the variables that appear in our paper. We normalized all independent and control variables.

[TABLE 1 inserts here]

Table 2 reports descriptive statistics of all variables. After merging and making the panel data, we calculate descriptive statistics of Excess Return, Price, Volume, and DSVI. For the demographic variables, We make panel data for each demographic variable for 2017-2019, then calculate descriptive statistics before merging its dataset with other datasets. Table 2 reports that the stocks in our sample, on average, receive 121.68 daily tweets, 50.561 daily retweets, 15.206 daily replies, 10.627 daily quotes, total 15.601 million in Volume, and have a share price of $4.324. we have, on average, 59.180 daily posts including URLs, 27.086 daily posts including images or videos, and 45.682 daily posts including just text. The report shows that, on average, 95.901 daily tweets are labeled as the Organization entity, and 62.770 posts do not have a proper label (labeled as other entity). The sample has 9.142, 7.242, and 9.496 daily Person, place, and product entities, respectively. These statistics appear different from other samples in the literature. For example, in Rakowski, Shirley, and Stark's (2021) paper, the average number of tweets is 9.502 for 1,976 stocks. One reason might be that we collect 59 trending stocks. Another reason is that they calculate one time-series value for each stock within their sample and present the cross-sectional mean. However, we make panel data, then present the descriptive statistics.

[TABLE 2 inserts here]

4. HYPOTHESES and MODELS

We develop our hypotheses in this section and present the models and approach for each hypothesis. By enormous volume and the rapid speed of information transmission, social media provides a more comprehensive real-time news database than traditional media channels. By existing social bots, the large amount of information can also contain potential noise that might mislead readers (Fan, Talavera, and Tran 2020). In addition, the bots can disseminate fake news and manipulate stock markets. According to Fan, Talavera, and Tran (2020), the influence of social media bot activities on stock markets is significant. We want to understand better social media users' roles in the diffusion of information in financial markets, both human and bot (automated accounts) users. The first research question is how is the type of Twitter user (bot or human) associated with the magnitude of the stock price reaction to information on social media? Our first null and alternative hypotheses are:

$H1_0$: The impact of bots and human social media users on the financial market is similar.

$H1_A$: The impact of human activity is stronger than bots.

To test this hypothesis, we filter Tweets and pick tweets if they come from Twitter Web Client, Twitter Web app, Twitter for iPhone, Twitter for Android, Twitter for iPad, and TweetDeck. Panel B of all the models below includes human users' posts, and panel C includes the bot users' posts.

Rakowski, Shirley, and Stark (2021) identify a supply of the information (Twitter attention) and subsequent diffusion of the information (retweet attention) to determine how the different avenues of information diffusion affect prices. They find that the additional spread of information from retweets increases the magnitude of the associated price impact. The data section explains that each Twitter post could be a Tweet, Retweet, quote, or reply. We assert that retweeting, replying, and quoting aid the diffusion of information, and we explore how the different avenues of information diffusion affect prices. Our second research question is how is the type of Twitter posts associated with the magnitude of the stock price reaction to information on social media? Does this impact contrast between human and Bot users (first hypothesis)? Our second null and alternative hypotheses are:

$H2_0$: The type of Twitter post (a tweet, a retweet, a reply, or a quote) does not impact on stock prices.

$H2_A$: The impacts of the spread of information by retweeting, replying, and quoting differ on stock prices.

To test this hypothesis, we follow Rakowski, Shirley, and Stark (2021), we estimate panel regression models of the form:

$$Excess\ Returns_{i,t+n} = \alpha + \beta_1 Tweets_{i,t} + \beta_2 Non\_Tweets_{i,t} + \sum_{j=5}^{m} \beta_j Controls_{i,t} + \varepsilon_{i,t} \quad \textbf{Eq. (1)}$$

$$Excess\ Returns_{i,t+n} = \alpha + \beta_1 Tweets_{i,t} + \beta_2 Retweets_{i,t} + \beta_3 Replies_{i,t} + \beta_4 Quotes_{i,t} + \sum_{j=5}^{m} \beta_j Controls_{i,t} + \varepsilon_{i,t} \quad \textbf{Eq. (2)}$$

In model (1), our variables of interest, $Tweets_{i,t}$, is tally tweets and, $Non\_Tweets_{i,t}$, is total non-tweets (retweets, replies, and quotes) for Stock *i* on day *t*. In model (2), $Tweets_{i,t}$, is tally tweets, $Retweets_{i,t}$, is total retweets, $Replies_{i,t}$, is total replies, and $Quotes_{i,t}$, is total quotes for Stock *i* on day *t*.

Tetlock (2007) finds that measures of media content serve as a proxy for investor sentiment or noninformational trading. According to Vempala et al., (2019), text in social media posts is frequently accompanied by images in order to provide content, supply context, or to express feelings. According to Buffer[14] (2015), tweets with images receive more engagement than tweets without images. Buffer also reports that tweets without links got more retweets, favorites, and replies than tweets with links. Tweets with links tend to receive less engagement than tweets without links. Our third research question is how Twitter post's contents associated with the magnitude of the stock price reaction to information on social media? Does this impact contrast between human and Bot users (first hypothesis)? Our third null and alternative hypotheses are:

$H3_0$: The content of Twitter post does not impact on stock prices.

$H3_A$: Twitter posts including images impact stock prices more than posts with text or links.

To examine this hypothesis, we have models 3 to 6. Each model of 3 to 5 includes URL_posts, Text posts, and Media posts, respectively. Model 6 includes all three types of post content. Our variables of interest, $URL\_Posts_{i,t}$, is all posts including URLs, $Media\_Posts_{i,t}$, is all posts including images/videos, and $Text\_Posts_{i,t}$, is all posts including just text without any media or URLs.

$$Excess\ Returns_{i,t+n} = \alpha + \beta_1 URL\_Posts_{i,t} + \sum_{j=2}^{m} \beta_j Controls_{i,t} + \varepsilon_{i,t} \qquad \textbf{Eq. (3)}$$

$$Excess\ Returns_{i,t+n} = \alpha + \beta_1 Text\_Posts_{i,t} + \sum_{j=2}^{m} \beta_j Controls_{i,t} + \varepsilon_{i,t} \qquad \textbf{Eq. (4)}$$

$$Excess\ Returns_{i,t+n} = \alpha + \beta_1 Media\_Posts_{i,t} + \sum_{j=2}^{m} \beta_j Controls_{i,t} + \varepsilon_{i,t} \qquad \textbf{Eq. (5)}$$

Eq. (6)

$$Excess\ Returns_{i,t+n} = \alpha + \beta_1 URL\_Posts_{i,t} + \beta_2 Text\_Posts_{i,t} + \beta_3 Media\_Posts_{i,t} + \sum_{j=4}^{m} \beta_j Controls_{i,t} + \varepsilon_{i,t}$$

[14] Buffer is a software application for the web and mobile, designed to manage accounts in social networks, by providing the means for a user to schedule posts to Twitter, Facebook, Instagram, Instagram Stories, Pinterest, and LinkedIn, as well as analyze their results and engage with their community. (https://en.wikipedia.org/wiki/Buffer_(application))

We collect the Tweet annotations is one of the key information extraction tasks, which is concerned with identifying names of entities such as people, locations, organizations, and products. According to Finin et al., (2010) , person entities are limited to humans identified by name, nickname, or alias. Organizational entities are limited to corporations, institutions, government agencies, and other groups of people defined by an established organizational structure. As a result, organizations' and person entities' tweets engage in our data since they include the name of firms and people names. The fourth research question is, does the contextual information of Tweet influence stock prices? Our fourth null and alternative hypotheses are:

$H4_0$: The contextual information of Tweet does not impact on stock prices.

$H4_A$: The Person and organization entities impacts are stronger on stock prices.

To assess this hypothesis, we have model 7, where independent variables are $Organization\_entity_{i,t}$, is all Tweets labeled Organization entity, $Person\_entity_{i,t}$, is all Tweets labeled Person entity, $Place\_entity_{i,t}$, is all Tweets labeled Place entity, $Product\_entity_{i,t}$, is all Tweets labeled Product entity, and $Other\_entity_{i,t}$, is all Tweets labeled Other entity.

Eq. (7)

$$Excess\ Returns_{i,t+n} = \alpha + \beta_1 Organization\_entity_{i,t} + \beta_2 Person\_entity_{i,t} + \beta_3 Place\_entity_{i,t} + \beta_4 Product\_entity_{i,t} + \beta_5 Other\_entity_{i,t} + \sum_{j=6}^{m} \beta_j Controls_{i,t} + \varepsilon_{i,t}$$

According to (Yao, Gutter, and Hanna, 2005), culture provides a context in which information is framed, and preferences are formed; their model illustrates the importance that culture, represented by race and ethnic status, has on the financial decision-making process. According to Gutter and Montalto (1999), racial differences in investor behavior may indicate differences in risk tolerance and investment choice or cultural differences in investment behavior. Moreover, since Black and Hispanic investors are more risk averse in their choice of assets than whites ((Badu and Salandro, 1999) and (Yao, Gutter, and Hanna 2005)). The fifth research question is, how is the race of social media users associated with the magnitude of the stock price reaction to information on social media?

Our fifth null and alternative hypotheses are:

$H5_0$: There is no relation between the race of social media users and the magnitude of stock price reaction to information on social media.

$H5_A$: White social media users have a more substantial impact on stock prices than Hispanic and Blacks social media users.

Eq. (8)

$$Excess\ Returns_{i,t+n} = \alpha + \beta_1 White\ Users_{i,t} + \beta_2 API\ Users_{i,t} + \beta_3 Black\ Users_{i,t} + \beta_4 Hispanic\ Users_{i,t} + \sum_{j=5}^{m} \beta_j Controls_{i,t} + \varepsilon_{i,t}$$

For testing the fifth hypothesis, we use model 8, where independent variables are $White\ Users_{i,t}$, all posts by White users, $API\ Users_{i,t}$, all posts by Asian and Pacific Islander users, $Black\ Users_{i,t}$, all posts by Black users and $Hispanic\ Users_{i,t}$, all posts by Hispanic users for Stock *i* on day *t*.

According to Cueva et al. (2019), psychological research demonstrates that, in areas such as finance, men are more overconfident than women. Thus, the theory predicts that men will trade more optimistically than women. The sixth research question is, how is the gender of social media users associated with the magnitude of the stock price reaction to information on social media? Our sixth null and alternative hypotheses are:

$H6_0$: There is no relation between the gender of social media users and the magnitude of stock price reaction to information on social media.

$H6_A$: Men social media users will post information that substantially impacts stock prices.

$$Excess\ Returns_{i,t+n} = \alpha + \beta_1 Male\ Users_{i,t} + \beta_2 Female\ Users_{i,t} + \sum_{j=3}^{m} \beta_j Controls_{i,t} + \varepsilon_{i,t}$$

Eq. (9)

For assessing the sixth hypothesis, we use model 9, where independent variables are $Male\ Users_{i,t}$, all posts by male users and $Female\ Users_{i,t}$, all posts by female users for Stock *i* on day *t*.

Research shows that investor behavior varies with the age, of investors (Nagy and Obenberger 1994). Because age is known to be negatively associated with noise on social media (Ozimek and Bierhoff 2016), we conjecture that older social media users will post information that has a stronger impact on stock prices. The last research question is, how is the age of social media users associated with the magnitude of the stock price reaction to information on social media? Our seventh null and alternative hypotheses are:

$H7_0$: There is no relation between the age of social media users and the magnitude of stock price reaction to information on social media.

H7A: Social media users' age and the magnitude of information diffusion are positively associated. (Stock prices react more strongly to posts by older social media users than younger social media users).

To test this hypothesis, we run models 11 and 12. For model 11, we create a daily index of the imbalance in activity between young (less than 40 years old) and old (more than 39 years old) social media users for each stock in our sample:

$$age\_imbalance_{i,t} = \frac{\sum_{y=1}^{Y} young_{i,t} - \sum_{o=1}^{O} old_{i,t}}{\sum_{y=1}^{Y} young_{i,t} + \sum_{o=1}^{O} old_{i,t}} \quad \textbf{Eq. (10)}$$

where $young_{i,t}$ ($old_{i,t}$) indicates the count of social media posts by young (old) social media users for stock *i* during time *t*.

$$Excess\ Returns_{i,t+n} = \alpha + \beta_1 age\_imbalance_{i,t} + \sum_{j=2}^{m} \beta_j Controls_{i,t} + \varepsilon_{i,t} \quad \textbf{Eq. (11)}$$

For model 12, our variables of interest are, $GenZ\ Users_{i,t}$, all posts by generation Z users (Ages 7-22), $GenY\ Users_{i,t}$, all posts by generation Y users (Ages 23-38), $GenX\ Users_{i,t}$, all posts by generation X users (Ages 39-54), and $Boomers\ Users_{i,t}$), all posts by Boomers' users (Ages 55-73), for Stock *i* on day *t*.

Eq. (12)

$$Excess\ Returns_{i,t+n} = \alpha + \beta_1 GenZ\ Users_{i,t} + \beta_2 GenY\ Users_{i,t} + \beta_3 GenX\ Users_{i,t} + \beta_4 Boomers\ Users_{i,t} + \sum_{j=5}^{m} \beta_j Controls_{i,t} + \varepsilon_{i,t}$$

In all above models, the dependent variable is $Excess\ Returns_{i,t+n}$, daily value Fama and French (1992) three-Factor Model excess return in basis points for stock *i*, on day *t* varying values of *n*. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls on all models for Volume, Price, GDSV and five lags of absolute excess returns (AER). We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs).

5. RESULTS

Table 3 and Table 4 provide results to examine whether the Twitter post type affects stock prices. In Table 3, our variables of interest, $Tweets_{i,t}$, is the daily tally of tweets and, $Non\_Tweets_{i,t}$, is the count of total non-tweets (i.e., retweets, replies, and quotes) for stock *i* on day *t*. We present the results of the full dataset, including posts from any tweeting devices (sources) in Panel A. This sample includes both human users and bots. Columns 1 and 2 of Table 3 show results for day t=0. The coefficient estimate on tweets suggests that a one standard deviation increase in the number of tweets is associated with an increase of 0.1016 bps in stock excess returns. Columns 5 and 6 of Table 3, Panel A, indicate a reversal on day 2. In Panel A, non-tweets do not impact excess returns.

In Panel B of Table 3, we restrict the tweets source to Twitter Web Client, Twitter Web app, Twitter for iPhone, Twitter for Android, Twitter for iPad, and TweetDeck. In this second dataset, we have mostly human posts rather than bot messages. Columns 1 and 2 show results for the day t=0. We again observe a significant increase in stock returns, of 0.1973 bps for a 1 SD increase in Tweets. The reversal for day 2 is now weaker. In the existing literature, a positive coefficient on day zero followed by a reversal in later days, is interpreted as evidence of a liquidity shock (i.e., price pressure). A positive coefficient on day zero with no reversal is interpreted as evidence of new information. The results of Table 3, Panels A and B, therefore, suggest that posts generated by real people are more strongly associated with information (i.e., accompanied by a weak reversal), while posts generated by bots are more associated with a temporary liquidity shock that dissipates in later days (i.e., accompanied by a stronger reversal). In columns 9 and 10, we can see that non-tweets posts are associated with a decrease in excess stock returns of 0.1479 bps, with a greater magnitude than day 0. The results show that non-tweets have a delayed effect.

In panel C of Table 3, we present the results of the third dataset; we include all Tweeting devices except Twitter Web Client, Twitter Web app, Twitter for iPhone, Twitter for Android, Twitter for iPad, and TweetDeck. The third data set likely includes a large proportion of bot posts.

Columns 1 and 2 of Table 3, Panel C, show a positive impact on returns on day t=0, followed by a strong reversal on most following days. Non-tweets display a positive delayed impact on excess returns. Overall, Table 3 shows that, on the one hand, tweets from bots are consistent with price pressure on day t=0 that is followed by a reversal on later days. Non-tweets are consistent with information, but this is processed by the market only with a substantial delay.

[TABLE 3 inserts here]

Since the non-tweets in Table 3 are associated with excess returns, we decompose non-tweets into retweets, replies, and quotes. Model estimates are reported in Table 4. Panel A of Table 4, columns 1 and 2, show results for the day t=0. We observe an increase in stock returns of 0.1017 bps for tweets, followed by insignificant reversal in later days. We can see an increase in excess returns of 0.0355 bps for Retweets day 1. Moreover, we can observe a decrease in stock return of 0.0491 bps for Replies on day 2. In Panel A, quotes do not impact excess returns.

In Panel B, we have mostly human posts rather than bot messages. We again observe a significant increase in stock returns in Tweets for the day t=0. The reversal for day 2 is now weaker compared with Panel A. For the retweets, again an insignificant positive coefficient on day zero with no reversal is interpreted as true information. In columns 9 and 10, we can see that retweets posts are associated with a decrease in excess stock returns, with a reversal respect to day 0. The results show that retweets have a delayed effect. In columns 9 and 10, we can see that replies posts are associated with a decrease in excess stock returns, with a greater magnitude than day 0. The results show that replies have a delayed effect. In Panel B, quotes do not impact excess returns.

Panel C of Table 4 for Bot users, shows a positive impact on returns on day t=0, followed by a strong reversal on most following days for tweets. Retweets display a positive delayed impact on excess returns. For retweets, we can see positive and significant coefficient over days t + 4 through t + 15, which is interpreted as a delayed information reaction. Retweets display a positive delayed impact on excess returns. In Panel C, quotes posts generated by bot users impact excess returns. quotes display a negative delayed impact on excess returns. Based on the results of tables 3 and 4, we reject the first and second null hypotheses, and we conclude that the impacts of the spread of information by retweeting, replying, and quoting differ on stock prices. Moreover, the impact of human activity is more substantial than bots users on the financial market. Real users provide mostly information on Twitter, while bots provide mainly price pressure.

[TABLE 4 inserts here]

For exploring the third null hypothesis, the post's contents do not impact stock excess returns; we have models 3 to 6. First, we examine each post content type separately in models 3 to 5, and the results are provided in Table 5. Then, we have model (6) for including all types of Tweet content. Results are presented in Table 6. In Table 5, panel A, for the entire dataset, we observe a positive coefficient on day 0 for all types of posts (URL, Text, and Media), then a negative coefficient in later days (i.e., a reversal), which is indicating a non-informative liquidity shock on day 0.

In panel B, we observe the significant negative coefficient of 0.019 on URL posts on day 2, indicating that a one standard deviation increase in the number of URL posts by human users is associated with a decrease in excess returns of 0.019 bps. In columns 26, 28, and 30, we can see that all types of posts (URL, Text, and Media) are associated with a decrease in excess stock returns of 0.0846 bps, 0.0785 bps, and 0.079 bps with a greater magnitude than day 0. The results show that all posts (URL, Text, and Media) have a delayed effect on excess return.

Panel C, which is more bot messages than human messages, shows that a 1 SD increase in Media posts is associated with an increase in excess returns equivalent to (0.0452/0.088) 51% of the SD of excess returns. We have reversal on later days (columns 30), a positive coefficient on day 0 then a negative coefficient in later days is interpreted as indicating a non-informative liquidity shock on day 0. We can observe that automated users (bots) are more associated with temporary liquidity shock that dissipates in later days (accompanied by a stronger reversal) by generating posts.

[TABLE 5 inserts here]

We include three types of post contents (URL, Text, and Media) in the model (6). In Table 6, panel A, for the entire dataset, all human and bot messages, in column (10), for URL posts, the negative coefficient is interpreted as a delayed information reaction. Text posts do not impact excess returns. For media posts, the significant positive coefficient of 0.0566 on day 0 is interpreted as "true information" since it is not reversed in later days. The significant positive coefficient of 0.0566 on Tweets indicates that a one standard deviation increases in the number of posts, including media, is associated with an increase in excess returns of 0.0566 bps; if we do a simple annualization, we get 14.15 bps = 0.142% per year.

In panel B, we have primarily human posts rather than bot messages; URL posts display a negative delayed impact on excess returns. Text posts do not impact excess returns in Panel B. In panel C, which has mostly bot messages, we observe a significant increase in stock returns of 0.0488 bps for a 1 SD increase in Media posts; A 1 SD increase in Media posts is associated with an increase in excess returns equivalent to 55.44% of the SD of excess returns. (Table 2 lists the SD of excess daily returns as 0.088, so an increase in excess returns of 0.0488 represents 55.44% of an SD (0.0488/0.088=55.44%).)

Panel C has mostly bot messages; we observe a significant increase in stock returns of 0.0488 bps for a 1 SD increase in Media posts. For Media posts, show a positive impact on returns on day t=0, followed by a strong reversal on most following days. Media posts from bots are consistent with price pressure on day t=0 that is followed by a reversal on later days. We observe a significant decrease in stock returns of 0.041 bps for a 1 SD increase in URL posts. Bot messages impact stock prices with Twitter posts, including images than posts with links and text.

[TABLE 6 inserts here]

Overall, Tables 5 and 6 show that real users provide mostly information, while bots provide mostly price pressure by comparing your panels B and C. In the existing literature, the "Twitter effect", or any news effect, is often characterized as being driven by news or liquidity based on return reversals over time. News (or information) should be associated with returns at t=0, with no reversals later on. Liquidity shocks (price pressure) should also be associated with returns at t=0 or t=1, followed by a return reversal later on. Based on Table 5 and 6 we can reject the third null hypothesis. Twitter posts, including images and links impact stock prices more than posts with text.

In Table 7, we investigate the fourth hypothesis: Tweet's contextual information does not influence stock excess returns. In panel A, for the entire data, we observe a significant decrease in stock return of 0.052 (0.113) bps for Organization_entity (Person_entity) between days t + 4 through t + 15. Therefore, we have a delayed effect for Organization_entity and Person_entity. A negative of coefficient of Place_entity on day 0 is interpreted as "true information" since it is not reversed in later days.

In panel B, for human users, we find a significant decrease in stock prices of 0.063, 0.112, and 0.126 bps for Organization_entity, Person_entity, and Place_entity between days t + 4 through t + 15, respectively. The significant negative coefficient of Organization_entity, Person_entity, and Place_entity on a later day is interpreted as a delayed information reaction. We can see that the magnitude of Place_entity and Person_entity is higher than the Organization entity. For example, a 1 SD increase in Place_entity and Person_entity is associated with a decrease in excess returns equivalent to 143.18% and 127.3% of excess returns, respectively.

In panel C, posts generated by bots show a decrease in stock prices of 0.069 (0.083) bps for Place_entity (Organization_entity). The significant negative coefficient of 0.069 (0.083) on Place_entity (Organization_entity) indicates that a one standard deviation increases in the number of Place_entity (Organization_entity) posts is associated with a decrease in excess returns of 0.069 (0.083) bps. In panel C, compared with Panel B, the Organization entity's magnitude is more remarkable than Place_entity, which is the opposite of Panel B. We also can observe that bots generate Product_entity posts which are more associated with a temporary liquidity shock that dissipates in later days. Overall, We can reject the fourth hypothesis, and Tweet's contextual information can influence stock excess returns.

[TABLE 7 inserts here]

In Table 8, we explore how the race of social media users is associated with the magnitude of the stock price reaction to information on social media is? In panel A, all bot and human posts, we observe that the posts by Hispanic and Asian, and Pacific Islander (API) social media users do not substantially impact stock prices. The significant negative coefficient of 0.054 and 0.048 on White_Users indicate that a one standard deviation increase in the number of posts by White users is associated with a decrease in excess returns of 0.054 bps and 0.048 bps on days 1 and 3, respectively. In column 10, for Black_Users, the significant positive coefficient of 0.09 indicates that one standard deviation increase in posts by Black users is associated with an increase in excess returns of 0.09 bps. Posts by Black users are associated with a delayed market reaction.

In panel B, more human posts than bot messages, we observe that, similar to panel A, the posts by Asian and Pacific Islander (API) social media users do not substantially impact stock prices. On the other hand, for posts generated by Hispanic users, A positive coefficient on day 0 for Hispanic users is interpreted as "true information" since it is not reversed in later days most later days. A one standard deviation increase in the number of Hispanic users is associated with an increase in excess returns of 0.077 bps on day 0 and 0.039 bps on day 3, and 0.168 in column 10. For Black users, a one standard deviation increase in the number of Black users is associated with a decrease in excess returns of 0.037 bps on day 0. We observe the significant negative coefficient of 0.054 and 0.048 on White_Users, indicating that a one standard deviation increase in the number of posts by White users is associated with a decrease in excess returns of 0.054 bps and 0.048 bps on days 1 and 3, respectively.

The results of Panel C, column 10, show that Black and Hispanic users'users' posts have a delayed effect. In Panel C, column 10, we can see that Black users' bot posts are associated with an increase in excess stock returns of 0083 bps, with a greater magnitude than day 0. We can also see in column 10 that Hispanic users' bot posts are associated with a decrease in excess stock returns of 0047 bps, with a greater magnitude than day 0. a positive coefficient on day 0 for API_Users, then an insignificant negative coefficient in the latter days indicates an insignificant non-informative liquidity shock on day 0. In panel C, white users do not impact stock prices. Finally, the results of Table 8 show that we can reject the fifth hypothesis, and white and Hispanic social media users have a more substantial impact on stock prices than others when we have primarily human posts rather than bot messages.

[TABLE 8 inserts here]

Table 9 provides results to investigate whether the gender of social media users is associated with the magnitude of the stock price reaction to information on social media. Variables of the interests are $Male\ Users_{i,t}(Female\ Users_{i,t})$ indicates the count of social media posts by Male (Female) social media users for stock i during time t. In panel A, for the full dataset, we can see a positive coefficient on day 0 for Male users and a negative coefficient in later days (i.e., a reversal), indicating a non-informative liquidity shock on day 0. Male social media users post information that substantially impacts stock prices.

We have more human posts than bot posts in Panel B. the posts by Men impact the magnitude of information diffusion. The significant positive coefficient of 0.162 indicates that a one standard deviation increase in the number of posts by men is associated with an increase in excess returns of 0.162 bps on day 0, equivalent to an annual impact of 40.5 bps. In panel B, men's social media users will post information that substantially impacts stock prices. We can see a positive coefficient on day 0 for Male users and a negative coefficient in later days (i.e., a reversal), indicating a non-informative liquidity shock on day 0. In panel B, the sign of coefficient on day 0 for Female_Users differs from Men_Users. In panel C, Female_Users has significant positive coefficients on days 0 and 1, and negative coefficients in the latter days indicate a non-informative liquidity shock on day 0. We can reject the sixth hypothesis and conclude that men social media users will post information that substantially impacts stock prices.

[TABLE 9 inserts here]

Tables 10 and 11 provide results to examine whether the age of social media users is associated with the magnitude of the stock price reaction to information on social media. For Table 10, the variable of interest is age_imbalance, defined in equation 10. In panel A, for the whole dataset, results show no relationship between the age of social media users and the magnitude of stock price reaction to information on social media for the entire dataset, both human users and bots. In Panel B, we have primarily human posts rather than bot messages. Social media users' age and the magnitude of information diffusion are positively associated on day 2. A significant increase in stock returns of 0.0015 bps for a 1 SD increase in the count of social media posts by old users. A 1 SD increase in posts by older is associated with an increase in excess returns equivalent to 1.7% of the SD of excess returns. (Table 2 lists the SD of excess daily returns as 0.088, so an increase in excess returns of 0.0015 represents 15% of an SD (0.0015/0.088=0.017).) In panel C, results show no association between the age of social media users and stock price reaction.

[TABLE 10 inserts here]

In Table 11, our variables of interests are $GenZ\_Users_{i,t}$, $GenY\ Users_{i,t}$, $GenX\ Users_{i,t}$, and $Boomers\ Users_{i,t}$, the count of social media posts by generation Z users (Ages 7-22), generation Y users (Ages 23-38), generation X users (Ages 39-54), and boomers' users (Ages 55-73), for Stock *i* on day *t*, respectively. In panel A, full dataset, for Generation Y, we have a positive coefficient on day 0 and negative coefficients in later days (i.e., a reversal), it is indicating a non-informative liquidity shock on day 0. The posts by Boomers_Users impact the magnitude of information diffusion. The significant negative coefficient of 0.086 on Boomers_Users indicates that a one standard deviation increase in the number of posts by Boomers is associated with a decrease in excess returns of 0.086 bps on day 0. In column 10, the impact of posts by Boomers, 0.181 bps on excess daily returns is equivalent to an annual impact of 45.25 bps. A negative coefficient on day 0 for Boomers_Users is interpreted as "true information" since it is not reversed in later days.

In Panel B, with more human users than bots, we have a significant positive coefficient for Generation Y on day 0. However, there is no remarkable coefficient in panel C for Generation Y. In panel B, the posts by Boomers_Users impact the magnitude of information diffusion greater than in panel A. A negative coefficient on day 0 for Boomers_Users is interpreted as "true information" since it is not reversed in later days in both panels A and B. The magnitude of the coefficient of Boomers in panel B is more remarkable than in panel A. For example, a 1 SD increase in posts generated by boomers is associated with a decrease in excess returns equivalent to 85.22% of the SD of excess returns. (Table 2 lists the SD of excess daily returns as 0.088, so a decrease in excess returns of 0.075 represents 85.22% of an SD (0.075 /0.088=0.8522).)

In panel C, which has bot messages rather than human messages, we can see that the impact of posts by Boomers disappears. In panel C, Gen Z_Users and Gen Y_Users have significant positive coefficients, and negative coefficients in later days (i.e., a reversal) indicate a non-informative liquidity shock on day 0.

Overall, of the results of Tables 10 and 11, we can reject the seventh hypothesis and conclude that Social media users' age and the magnitude of information diffusion are positively associated. Moreover, the posts generated by older social media users have a more substantial impact on stock prices.

[TABLE 11 inserts here]

6. CONCLUSION

This research examines the link between information diffusion with Twitter posts and users. The speed of information diffusion is essential for market efficiency (Lu, 2011), especially the information posted on social media (Peress, 2014) and (Rakowski, Shirley, and Stark, 2021). We test how stock price reactions to information differ between human and bot (automated accounts) users and how stock price reactions to information are associated with the demographic characteristics of social media users who post that information.

By existing social bots, the large amount of information can also contain potential noise that might mislead readers (Fan, Talavera, and Tran 2020). We find that posts generated by real people are more strongly associated with information, while posts generated by bots are more associated with a temporary liquidity shock that dissipates later. According to (Yao, Gutter, and Hanna, 2005), culture provides a context in which information is framed, and preferences are formed; their model illustrates the importance that culture, represented by race and ethnic status, has on the financial decision-making process. We extract the name and surnames of Twitter users; we infer the race of users by applying the proposed Bayesian approach of Chang et al. (2010). Then, we discover that posts generated by White and Hispanic social media users substantially impact stock prices more than other ethnicities.

For age and gender prediction, we download the users' profile pictures and analyze profile pictures using a face recognition algorithm running on Amazon Web Services cloud to find the age range and gender. According to Cueva et al. (2019), psychological research demonstrates that, in areas such as finance, men are more overconfident than women. Thus, the theory predicts that men will trade more excessively than women. Our results show that men's social media users post information that substantially impacts stock. Research shows that investor behavior varies with the age of investors (Nagy & Obenberger, 1994). Age is known to be negatively associated with noise on social media (Ozimek & Bierhoff, 2016). Finally, we display that social media users' age, and the magnitude of information diffusion are positively associated. Stock prices react more strongly to posts by older social media users than younger social media users.

REFERENCES:

Asmussen, Claus Boye, and Charles Møller. 2019. "Smart Literature Review: A Practical Topic Modelling Approach to Exploratory Literature Review." *Journal of Big Data* 6 (1). https://doi.org/10.1186/s40537-019-0255-7.

Badu, Y. A., Daniels, K. N., & Salandro, D. P. (1999). An empirical analysis of differences in Black and White asset and liability combinations☆. *Financial Services Review*, *8*(3), 129-147.

Ben-Rephael, Azi, Zhi Da, and Ryan D. Israelsen. 2017. "It Depends on Where You Search: Institutional Investor Attention and Underreaction to News." *Review of Financial Studies* 30 (9): 3009–47. https://doi.org/10.1093/rfs/hhx031.

Blei, David M. 2012. "Probabilistic Topic Models." In *Communications of the ACM*, 55:77–84. https://doi.org/10.1145/2133806.2133826.

Cesare, Nina, Christan Grant, and Elaine O Nsoesie. n.d. "Detection of User Demographics on Social Media: A Review of Methods and Recommendations for Best Practices." www.faceplusplus.com.

Chang, J., Rosenn, I., Backstrom, L., & Marlow, C. (2010, May). epluribus: Ethnicity on social networks. *In Fourth International AAAI Conference on Weblogs and Social Media*.

Cueva, C., Iturbe-Ormaetxe, I., Ponti, G., & Tomás, J. (2019). Boys will still be boys: Gender differences in trading activity are not due to differences in (over) confidence. *Journal of Economic Behavior & Organization*, *160*, 100-120.

Chu, Zi, Steven Gianvecchio, Haining Wang, and Sushil Jajodia. 2012a. "Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?" *IEEE Transactions on Dependable and Secure Computing* 9 (6): 811–24. https://doi.org/10.1109/TDSC.2012.75.

Engelberg, Joseph E., and Christopher A. Parsons. 2011. "The Causal Impact of Media in Financial Markets." *Journal of Finance* 66 (1): 67–97. https://doi.org/10.1111/j.1540-6261.2010.01626.x.

FAMA, EUGENE F., and KENNETH R. FRENCH. 1992. "The Cross-Section of Expected Stock Returns." *The Journal of Finance* 47 (2): 427–65. https://doi.org/10.1111/j.1540-6261.1992.tb04398.x.

Fan, Rui, Oleksandr Talavera, and Vu Tran. 2020. "Social Media Bots and Stock Markets." *European Financial Management* 26 (3): 753–77. https://doi.org/10.1111/eufm.12245.

Finin, Tim, Will Murnane, Anand Karandikar, Nicholas Keller, Justin Martineau, and Mark Dredze. n.d. "Annotating Named Entities in Twitter Data with Crowdsourcing." http://bit.ly/dwoQMD.

Gutter, M. S., Fox, J. J., & Montalto, C. P. (1999). Racial differences in investor decision making. *Financial Services Review*, *8*(3), 149-162.

Lu, Hai. 2004. "Information Diffusion and The Boundary of Market Efficiency." University of Southern California.

Nagy, Robert A, and Robert W Obenberger. n.d. "Factors Influencing Individual Investor Behavior."

Ozimek, Phillip, and Hans Werner Bierhoff. 2016. "Facebook Use Depending on Age: The Influence of Social Comparisons." *Computers in Human Behavior* 61: 271–79. https://doi.org/10.1016/j.chb.2016.03.034.

Peress, Joel. 2014a. "The Media and the Diffusion of Information in Financial Markets: Evidence from Newspaper Strikes." *Journal of Finance* 69 (5): 2007–43. https://doi.org/10.1111/jofi.12179.

Poterba, James M. 2001. "Demographic Structure and Asset Returns." *Source: The Review of Economics and Statistics*. Vol. 83.

Rakowski, David, Sara E. Shirley, and Jeffrey R. Stark. 2021. "Twitter Activity, Investor Attention, and the Diffusion of Information." *Financial Management* 50 (1): 3–46. https://doi.org/10.1111/fima.12307.

Tetlock, Paul C. 2007. "Giving Content to Investor Sentiment: The Role of Media in the Stock Market." *Journal of Finance* 62 (3): 1139–68. https://doi.org/10.1111/j.1540-6261.2007.01232.x.

Uddin, Muhammad Moeen, Muhammad Imran, and Hassan Sajjad. 2014. "Understanding Types of Users on Twitter," June. http://arxiv.org/abs/1406.1335.

Vempala, Alakananda, Bloomberg Lp, Daniel Preot¸iuc, and Preot¸iuc-Pietro Bloomberg. n.d. "Categorizing and Inferring the Relationship between the Text and Image of Twitter Posts." https://github.com/.

Yao, Rui, Michael S Gutter, and Sherman D Hanna. n.d. "The Financial Risk Tolerance of Blacks, Hispanics and Whites." http://ssrn.com/abstract=2255105.

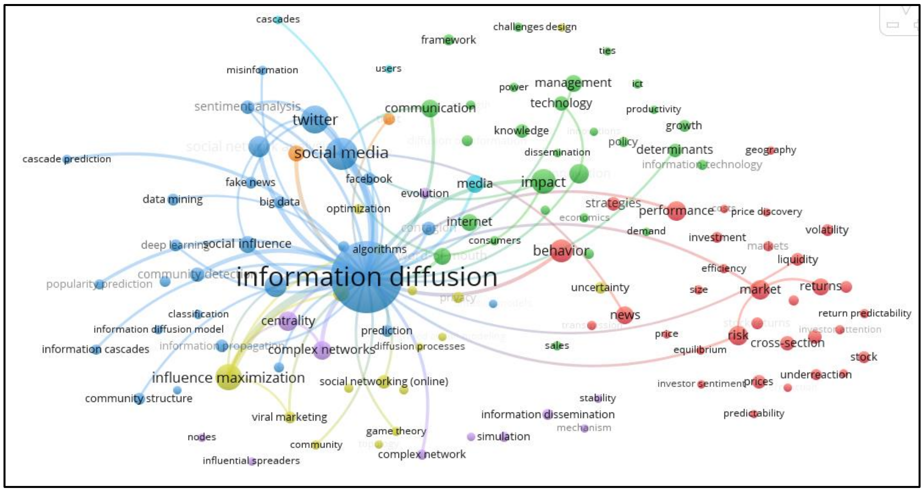

Figure 1: Co-occurrence network visualization

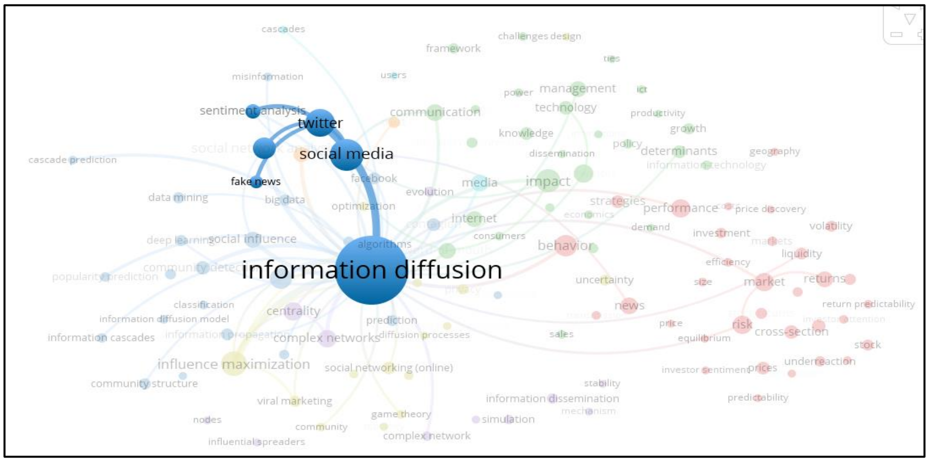

Figure 2: Partial co-occurrence network visualization

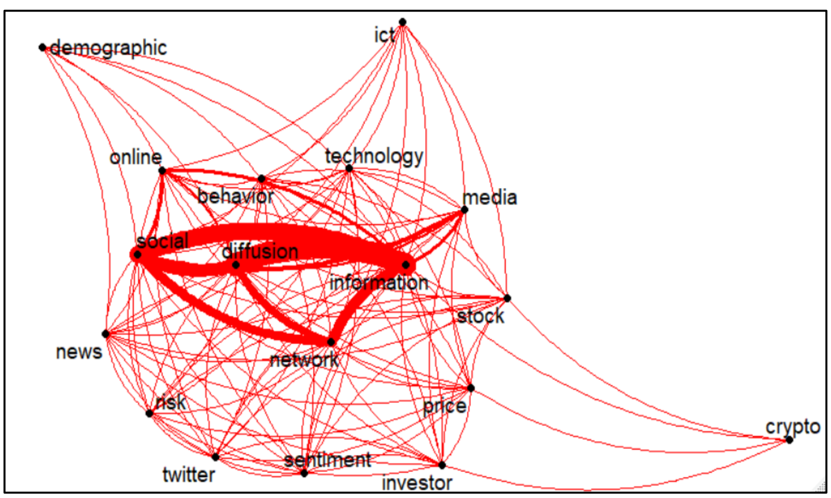

Figure 3: Semantic network analysis

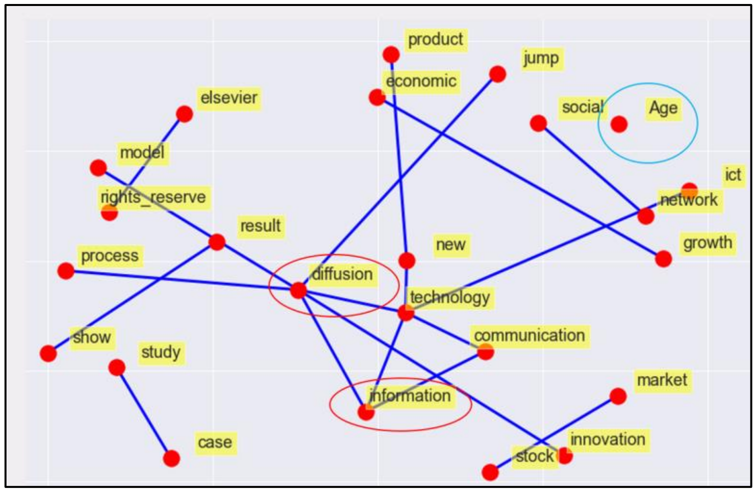

Figure 4: Networks of co-occurring words on information diffusion and age

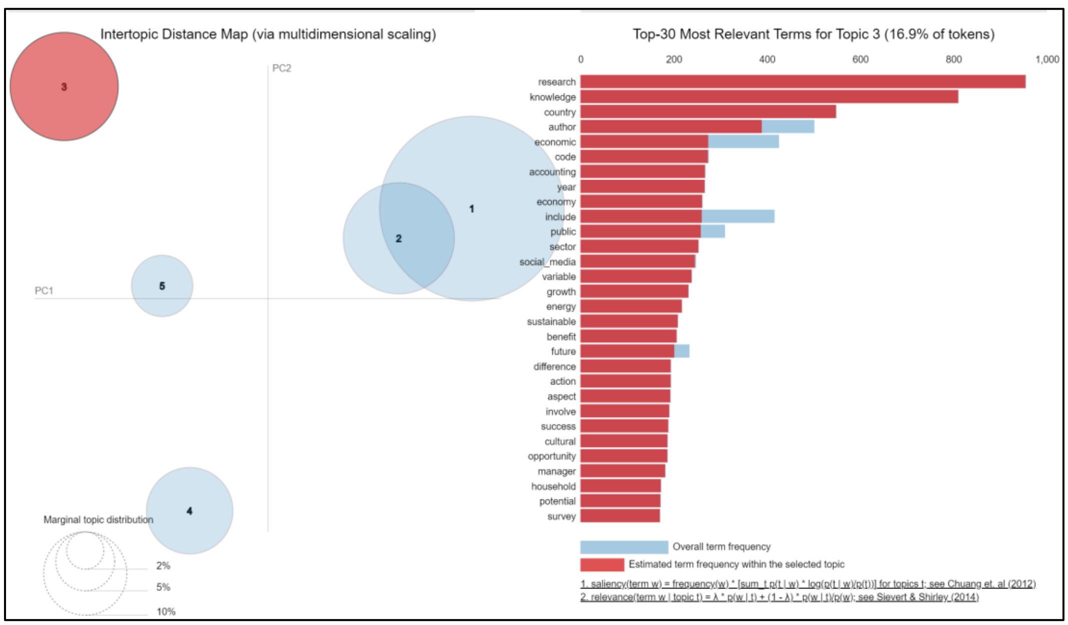

Figure 5: Visualizations by pyLDAvis for 5 topics

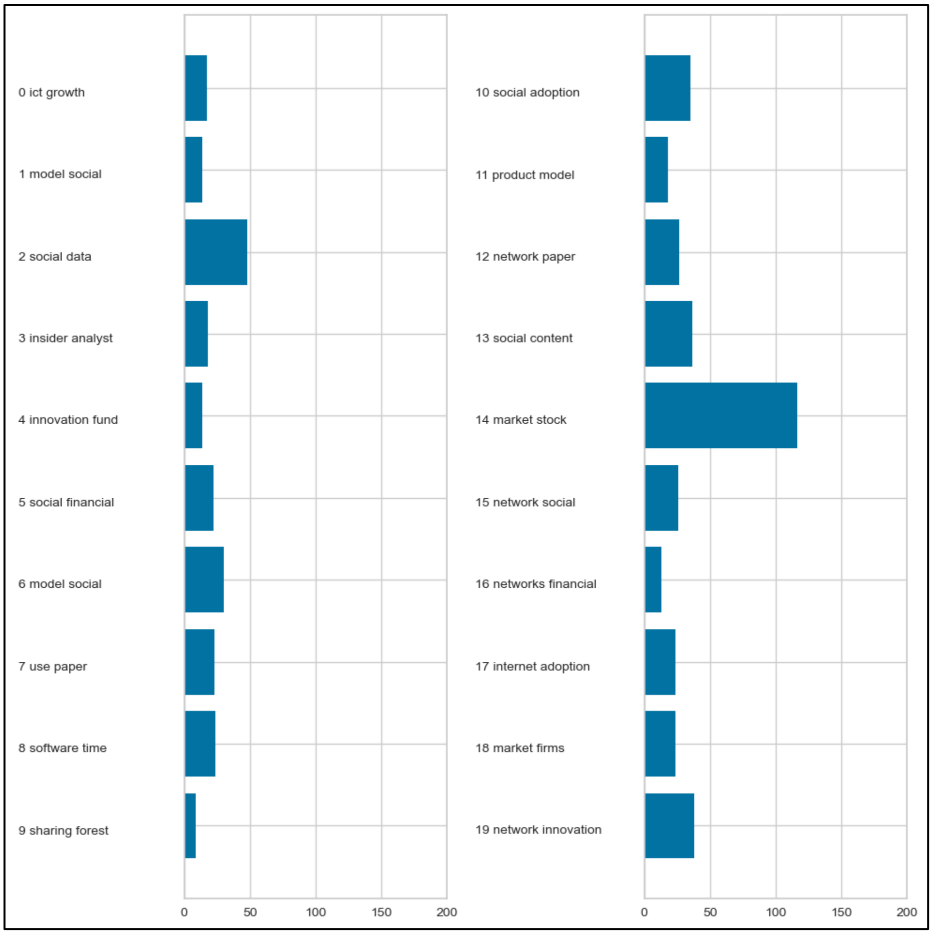

Figure 6: Visualizations for 20 topics

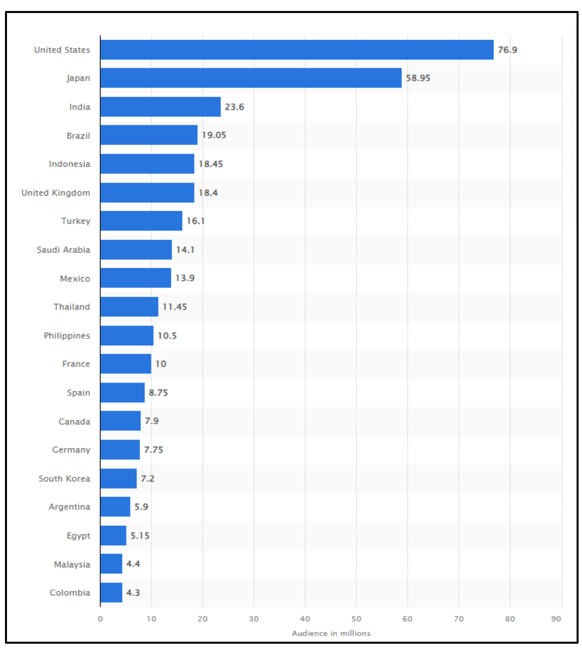

Figure 7: Leading countries based on number of Twitter users as of January 2022 (in millions)

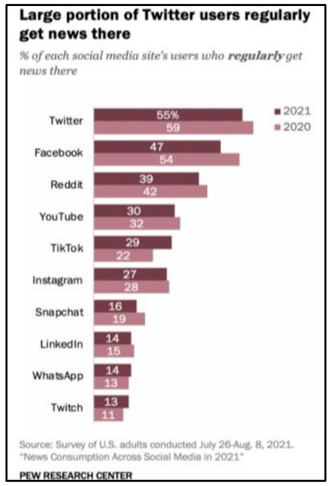

Figure 8: % of each social media site's users who regularly get news there as of Aug 2021

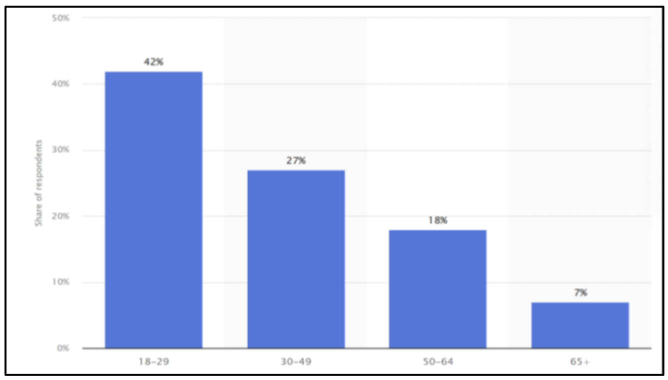

Figure 9: Percentage of U.S. adults who use Twitter as of February 2021, by age group

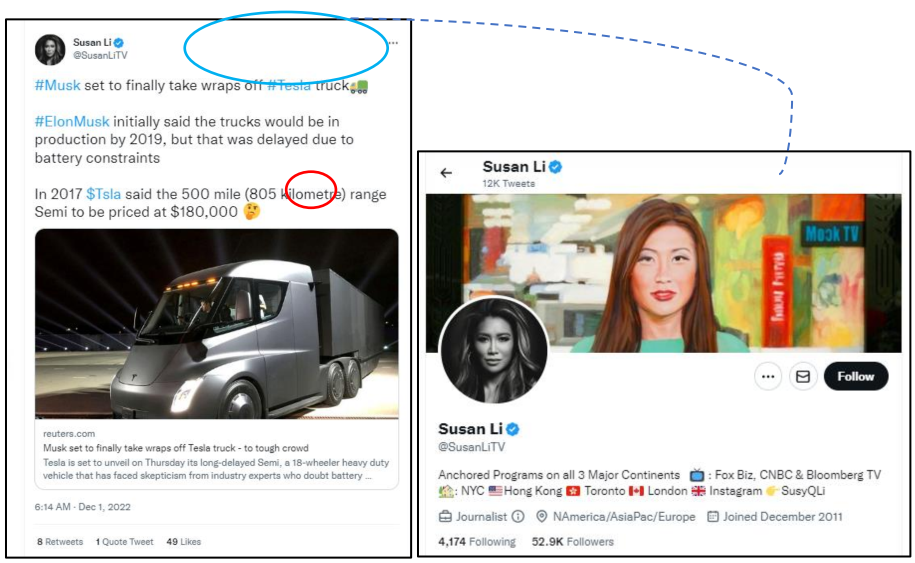

Figure 10: Example of the Tweet and user profile

Figure 11: Example of a Post including URL

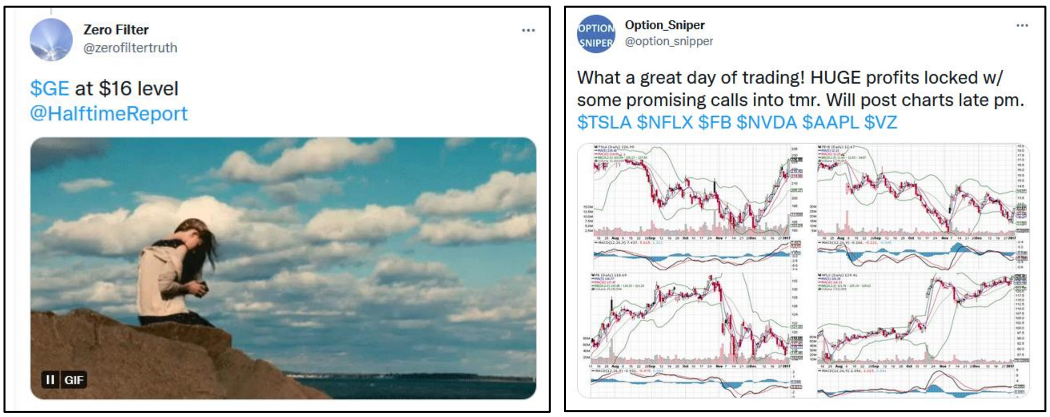

Figure 12: Example of a Post including a media (jpeg, GIF)

Figure 13: Example of a Post without URLs and media

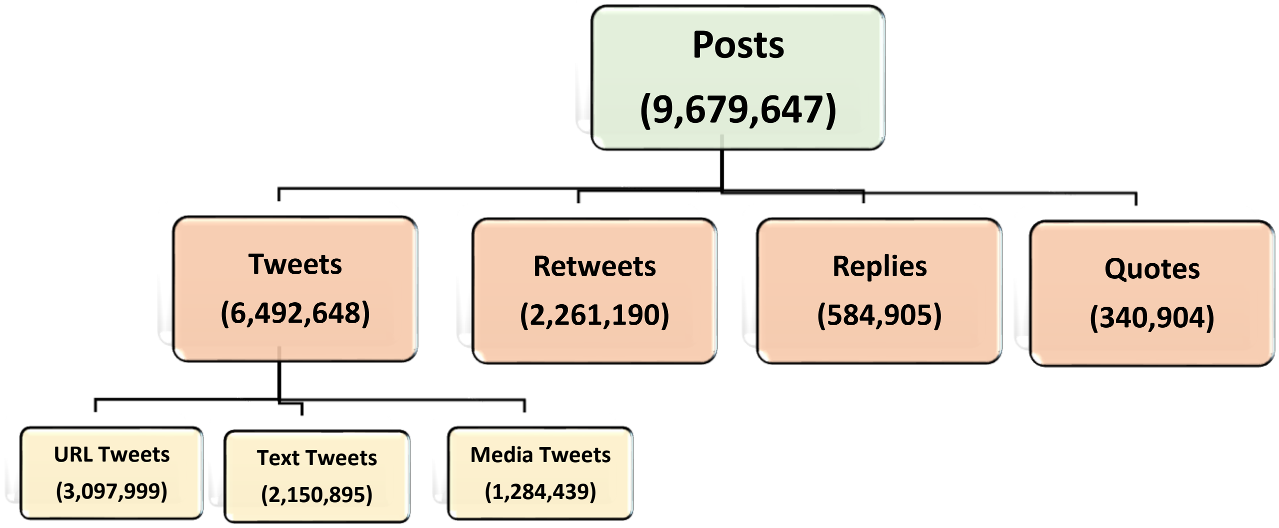

Figure 14: Post Categories and type of contents

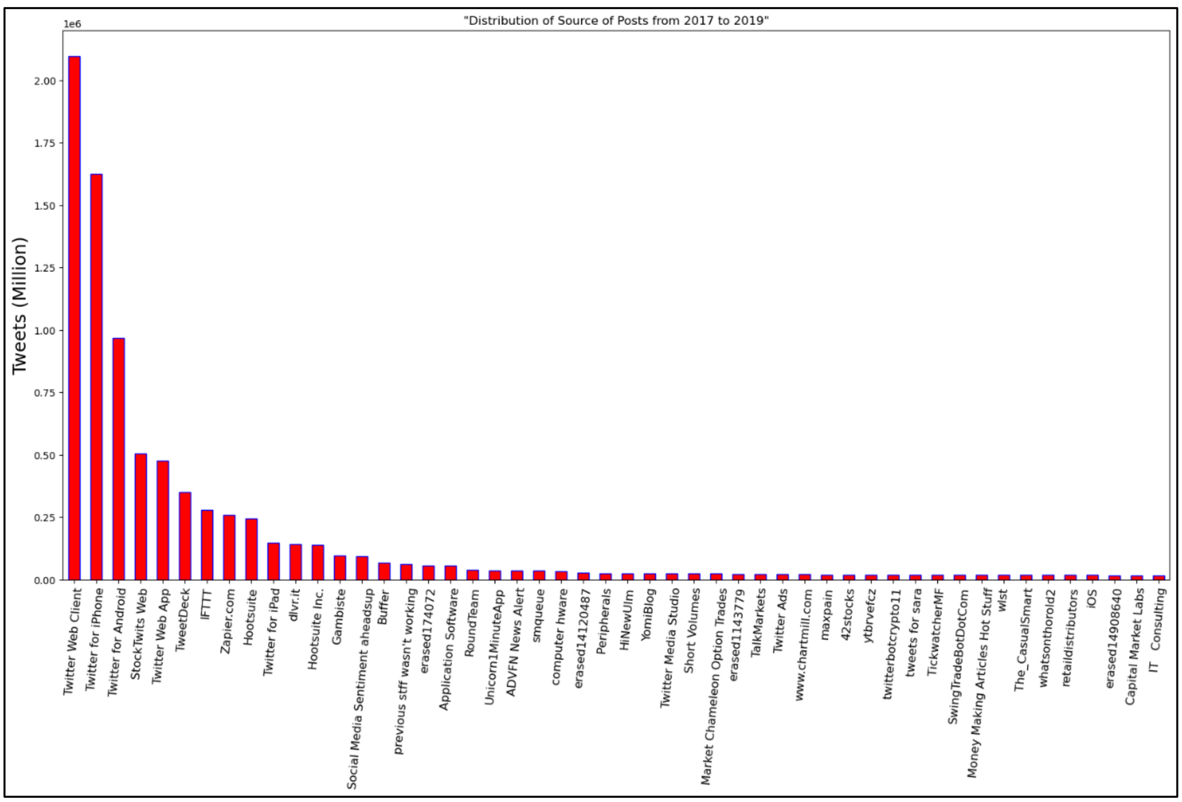

Figure 15: Distribution of the Tweets from tweeting devices (sources).

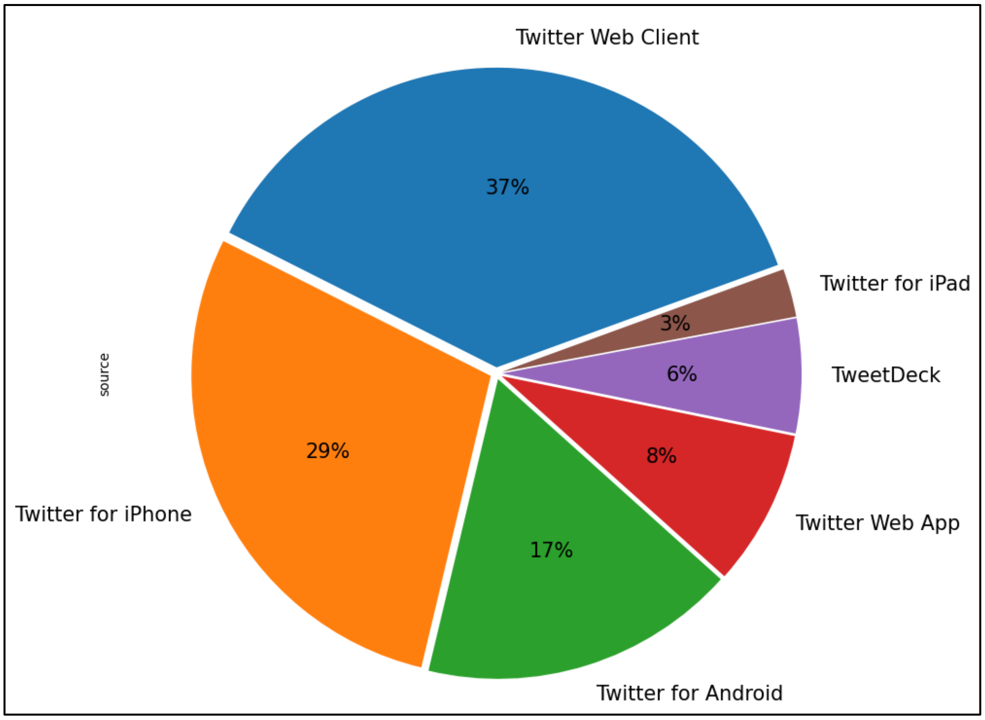

Figure 16: Distribution of the Tweets for Tweet's source of the second dataset

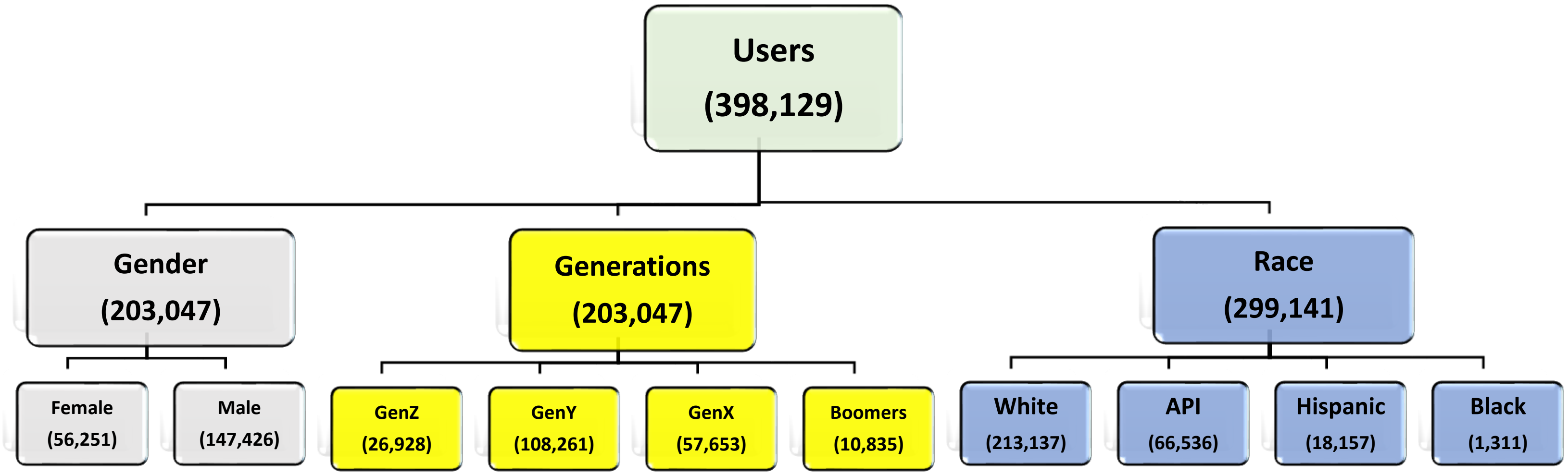

Figure 17: Users and Demographic Characteristics

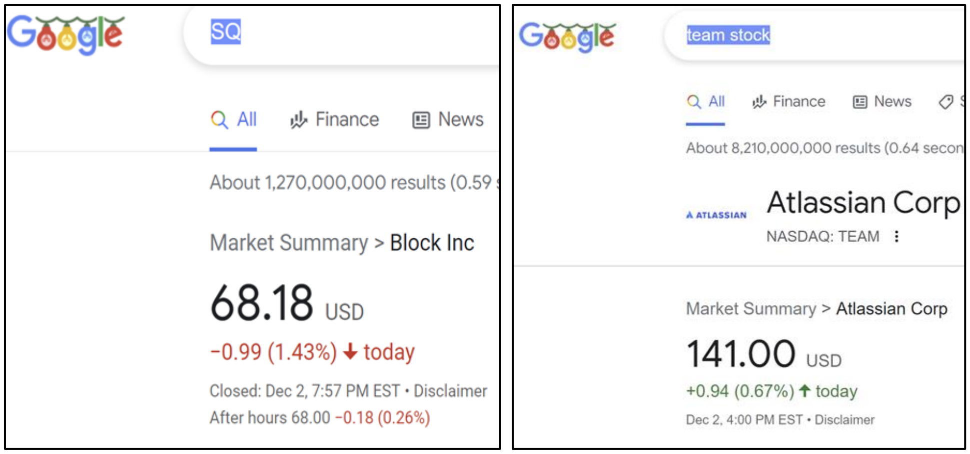

Figure 18: Example of selecting keywords for searching in Google Trends

TABLE A1: Trending Stocks List

No	Ticker	Company Name
1	PYPL	PayPal Holdings, Inc.
2	COIN	Coinbase Global, Inc.
3	SQ	Block, Inc.
4	SBUX	Starbucks Corporation
5	TWLO	Twilio Inc.
6	TEAM	Atlassian Corporation
7	WBD	Warner Bros. Discovery, Inc.
8	CVNA	Carvana Co.
9	NET	Cloudflare, Inc.
10	DASH	DoorDash, Inc.
11	HUDI	Huadi International Group Co., Ltd.
12	HDGE	AdvisorShares Ranger Equity Bear ETF
13	AAPL	Apple Inc.
14	MELI	MercadoLibre, Inc.
15	FIS	Fidelity National Information Services, Inc.
16	NNDM	Nano Dimension Ltd.
17	TLRY	Tilray Brands, Inc.
18	CTRA	Coterra Energy Inc.
19	LSPD.TO	Lightspeed Commerce Inc.
20	TWTR	Twitter, Inc.
21	PTON	Peloton Interactive, Inc.
22	MATIC-USD	Polygon USD
23	EXPE	Expedia Group, Inc.
24	COP	ConocoPhillips
25	DBX	Dropbox, Inc.
26	LNC	Lincoln National Corporation
27	ILMN	Illumina, Inc.
28	BE	Bloom Energy Corporation
29	PBR	Petróleo Brasileiro S.A. - Petrobras
30	AXSM	Axsome Therapeutics, Inc.
31	MSFT	Microsoft Corporation
32	GOOGL, GOOG	Alphabet Inc.
33	AMZN	Amazon.com, Inc.
34	TSLA	Tesla, Inc.
35	XOM	Exxon Mobil Corporation
36	JNJ	Johnson & Johnson
37	WMT	Walmart Inc.
38	JPM	JPMorgan Chase & Co.
39	NVDA	NVIDIA Corporation
40	HD	The Home Depot, Inc.
41	BAC	Bank of America Corporation

No	Ticker	Company Name
42	PFE	Pfizer Inc.
43	KO	The Coca-Cola Company
44	META	Meta Platforms, Inc.
45	MCD	McDonald's Corporation
46	DIS	The Walt Disney Company
47	CSCO	Cisco Systems, Inc.
48	VZ	Verizon Communications Inc.
49	T	AT&T Inc.
50	NFLX	Netflix, Inc.
51	QCOM	QUALCOMM Incorporated
52	INTC	Intel Corporation
53	AMD	Advanced Micro Devices, Inc.
54	BA	The Boeing Company
55	C	Citigroup Inc.
56	GE	General Electric Company
57	ABNB	Airbnb, Inc.
58	F	Ford Motor Company

TABLE A2: The example of the First names most associated with each ethnicity learned by the proposed model of Chang et al. (2010)

Rank	White	Black	Asian / Pacific Islander	Hispanic
1	barb	latoya	rahul	luis
2	conor	latonya	syed	javier
3	peg	deandre	wei	jose
4	deb	lakeisha	minh	jorge
5	kurt	tameka	nguyen	hector
6	colleen	latrice	tuan	yesenia
7	meghan	jermaine	thanh	mayra
8	meaghan	lashonda	sandeep	julio
9	connor	jamaal	phuong	alejandro
10	brendan	lakisha	yi	cesar

TABLE 1: Variable Definitions

Type	Variables	Source	Definitions
Dependent Variable	$Excess\ Returns_{i,t+n}$	CRSP and Kenneth R. French's website	Daily value Fama and French (1992) three-Factor Model excess return in basis points for stock i, on day t varying values of n.
Control Variable	$Volume$	CRSP	Natural log of the number of shares traded daily for a stock in millions.
Control Variable	$Price$	CRSP	Natural log of daily closing price of a stock listed in U.S. dollars.
Control Variable	$DSVI$	Google Trend	DSVI is Google's daily Search Volume; We follow Ben-Rephael et al., (2017) and we assign DSVI on day t a score of 0, 1, 2, 3 or 4 if the average is between 80% and 90%, 90% and 94%, 94% and 96%, or greater than 96% of the previous 30 days' daily GSVI, respectively.
Control Variable	AER	CRSP and Kenneth R. French's website	The absolute value of daily excess returns. We include five lags of AER in our models as control variables.
Independent Variable, Tweet Level	$Tweets_{i,t}$	Twitter	All Tweets for Stock *i* on day *t*.
Independent Variable, Tweet Level	$Non\_Tweets_{i,t}$	Twitter	All non-Tweets (Retweets, replies, and Quotes) for Stock *i* on day *t*.
Independent Variable, Tweet Level	$Retweets_{i,t}$	Twitter	All Retweets for Stock *i* on day *t*.
Independent Variable, Tweet Level	$Replies_{i,t}$	Twitter	All Replies for Stock *i* on day *t*.
Independent Variable, Tweet Level	$Quotes_{i,t}$	Twitter	All Quotes for Stock *i* on day *t*.
Independent Variable, Tweet Level	$URL\_Posts_{i,t}$	Twitter	All posts including URLs for Stock *i* on day *t*.
Independent Variable, Tweet Level	$Text\_Posts_{i,t}$	Twitter	All posts including just text for Stock *i* on day *t*.
Independent Variable, Tweet Level	$Media\_Posts_{i,t}$	Twitter	All posts including media such as image, GIF, videos for Stock *i* on day *t*.
Independent Variable, Tweet Level	$Organization\_entity_{i,t}$	Twitter	All Tweets labeled Organization entity for Stock *i* on day *t*.
Independent Variable, Tweet Level	$Person\_entity_{i,t}$	Twitter	All Tweets labeled Person entity for Stock *i* on day *t*.
Independent Variable, Tweet Level	$Place\_entity_{i,t}$	Twitter	All Tweets labeled Place entity for Stock *i* on day *t*.
Independent Variable, Tweet Level	$Product\_entity_{i,t}$	Twitter	All Tweets labeled Product entity for Stock *i* on day *t*.
Independent Variable, Tweet Level	$Other\_entity_{i,t}$	Twitter	All Tweets labeled Other entity for Stock *i* on day *t*.
Independent Variable, User Level	$Male\ Users_{i,t}$	Machine Learning, AWS, Twitter	All Male users for Stock *i* on day *t*.
Independent Variable, User Level	$Female\ Users_{i,t}$	Machine Learning, AWS, Twitter	All Female users for Stock *i* on day *t*.
Independent Variable, User Level	$Age\_imbalance_{i,t}$	Machine Learning, AWS, Twitter	$= \frac{\sum_{y=1}^{Y} young_{i,t} - \sum_{o=1}^{O} old_{i,t}}{\sum_{y=1}^{Y} young_{i,t} + \sum_{o=1}^{O} old_{i,t}}$
Independent Variable, User Level	$GenZ\ Users_{i,t}$	Machine Learning, AWS, Twitter	All generation Z users (Ages 7-22) for Stock *i* on day *t*.
Independent Variable, User Level	$GenY\ Users_{i,t}$	Machine Learning, AWS, Twitter	All generation Y users (Ages 23-38) for Stock *i* on day *t*.
Independent Variable, User Level	$GenX\ Users_{i,t}$	Machine Learning, AWS, Twitter	All generation X users (Ages 39-54) for Stock *i* on day *t*.
Independent Variable, User Level	$Boomers\ Users_{i,t}$	Machine Learning, AWS, Twitter	All boomers' users (Ages 55-73) for Stock *i* on day *t*.
Independent Variable, User Level	$White\ Users_{i,t}$	Machine Learning, Twitter	All White users for Stock *i* on day *t*.
Independent Variable, User Level	$API\ Users_{i,t}$	Machine Learning, Twitter	All Asian and Pacific Islander users for Stock *i* on day *t*.
Independent Variable, User Level	$Black\ Users_{i,t}$	Machine Learning, Twitter	All Black users for Stock *i* on day *t*.
Independent Variable, User Level	$Hispanic\ Users_{i,t}$	Machine Learning, Twitter	All Hispanic users for Stock *i* on day *t*.

Note. This table displays the names, sources, and brief definitions for all of the variables that appear in our paper. Abbreviations: CRSP, Center for Research in Security Prices. All independent and control variables are standardized.

TABLE 2: Descriptive Statistics

	All Stocks, N= 44				
	Minimum	**Mean**	**Median**	**SD**	**Maximum**
Excess Return (basis points)	-0.996	-0.003	0.001	0.088	1.337
Price (USD)	0.247	4.324	4.331	1.202	7.621
Volume (millions)	5.576	15.601	15.888	1.616	19.599
DSVI	-2.968	0.122	0.695	0.972	3.124
Number of Tweets	1	121.681	56	214.476	9,665
Number of Retweets	1	50.561	14	135.876	5,558
Number of Replies	1	15.206	5	43.146	1,417
Number of Quotes	1	10.627	4	28.530	924
Number of URL Posts	1	59.180	27	112.594	5,011
Number of Text Posts	1	45.682	11	135.461	5,492
Number of Media Posts	1	27.086	9	63.869	2,291
Number of Organization Entity	1	95.901	29	238.587	6,683
Number of Person Entity	1	9.142	3	31.045	2,234
Number of Place Entity	1	7.242	3	19.640	1,059
Number of Product Entity	1	9.496	2	40.397	1,425
Number of other Entity	1	62.770	27	127.848	12,028
Number of Male Users	1	49.234	17	115.599	6,182
Number of Female Users	1	16.652	7	34.220	1,155
Number of Generation Z Users	1	7.854	3	16.134	587
Number of Generation Y Users	1	29.790	10	70.992	3,643
Number of Generation X Users	1	26.730	10	6.797	2,648
Number of Generation Boomers Users	1	7.474	4	13.422	459
Number of White Users	1	126.295	55	252.658	11,640
Number of Asian and Pacific Islander Users	1	26.183	9	61.528	3,050
Number of Black Users	1	7.823	3	15.711	282
Number of Hispanic Users	1	9.369	5	16.160	585

TABLE 3: Panel A

	[t = 0]		[t+1]		[t+2]		[t+3]		[t+4, t+15]	
Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
Tweets	0.1294***	0.1016**	-0.0381	-0.0385	-0.0713**	-0.0642**	-0.037	-0.0358	-0.0029	-0.0504
	(0.0417)	(0.0402)	(0.0258)	(0.0261)	(0.0317)	(0.0325)	(0.0228)	(0.0238)	(0.0646)	(0.065)
Non_Tweets	-0.0207	-0.0174	0.012	0.0102	0.0091	0.0057	0.0078	0.0042	-0.029	-0.0521
	(0.0216)	(0.0207)	(0.0179)	(0.0178)	(0.0147)	(0.0144)	(0.0106)	(0.0105)	(0.0429)	(0.0421)
Price		0.002***		-0.0019***		-0.002***		-0.0021***		-0.0241***
		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0019)
Volume		-0.0011		-0.0002		0.0001		-0.0001		-0.0004
		(0.0011)		(0.0004)		(0.0004)		(0.0003)		(0.001)
DSVI		0.0003*		0.0004*		-0.0001		-0.0002		-0.0006
		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0007)
R-Square	0.1878	0.2005	0.1821	0.1868	0.1842	0.1887	0.1831	0.1873	0.1860	0.2023
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm Fix Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time Fix Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44
Time Series Length	754	754	754	754	754	754	754	754	754	754
Observations	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915

Note. We present the results of the first dataset including Tweets from any Tweeting devices (sources) in Panel A, for model (1). Our variables of interest, $Tweets_{i,t}$, is tally Tweets and, $Non\_Tweets_{i,t}$, is total non-Tweets (retweets, replies, and quotes) for Stock *i* on day *t*. In all models, the dependent variable is $Excess\ Returns_{i,t+n}$, daily value Fama and French (1992) three-Factor Model excess return in basis points for stock *i*, on day *t* varying values of *n*. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls in our models for Volume, Price, GDSV and five lags of absolute excess returns (AER). The results with control variables are presented in even columns. We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs). *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively.

TABLE 3: Panel B

Variable	[t = 0] (1)	(2)	[t+1] (3)	(4)	[t+2] (5)	(6)	[t+3] (7)	(8)	[t+4, t+15] (9)	(10)
Tweets	0.247***	0.1973**	-0.0735	-0.0741	-0.1334*	-0.1199	-0.0768	-0.0721	0.1291	0.1039
	(0.0799)	(0.0781)	(0.0554)	(0.0556)	(0.0734)	(0.0746)	(0.0524)	(0.0549)	(0.14)	(0.1387)
Non_Tweets	-0.0751**	-0.0616**	0.0273	0.0253	0.037	0.0303	0.0263	0.0205	-0.1109**	-0.1479***
	(0.0295)	(0.0272)	(0.0231)	(0.0229)	(0.0237)	(0.0235)	(0.0175)	(0.0176)	(0.0565)	(0.0557)
Price		0.0019***		-0.0019***		-0.002***		-0.0021***		-0.0243***
		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0019)
Volume		-0.001		-0.0002		0.0001		-0.0001		-0.0006
		(0.0011)		(0.0004)		(0.0004)		(0.0003)		(0.001)
DSVI		0.0003*		0.0004*		-0.0001		-0.0002		-0.0006
		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0007)
R-Square	0.1902	0.2020	0.1824	0.1870	0.1849	0.1892	0.1834	0.1875	0.1862	0.2024
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm Fix Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time Fix Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44
Time Series Length	754	754	754	754	754	754	754	754	754	754
Observations	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915

Note. We present the results of the second dataset in panel B, for model (1). we filter Tweets and pick tweets if it comes from Twitter Web Client, Twitter Web app, Twitter for iPhone, Twitter for Android, Twitter for iPad, and TweetDeck. Our variables of interest, $Tweets_{i,t}$, is tally Tweets and, $Non\_Tweets_{i,t}$, is total non-Tweets (retweets, replies, and quotes) for Stock *i* on day *t*. In all models, the dependent variable is $Excess\ Returns_{i,t+n}$, daily value Fama and French (1992) three-Factor Model excess return in basis points for stock *i*, on day *t* varying values of *n*. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls in our models for Volume, Price, GDSV and five lags of absolute excess returns (AER). The results with control variables are presented in even columns. We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs). *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively.

TABLE 3: Panel C

	[t = 0]		[t+1]		[t+2]		[t+3]		[t+4, t+15]	
Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
Tweets	0.0634***	0.0444*	-0.0244*	-0.0246**	-0.0426***	-0.0391**	-0.0178	-0.02*	-0.1498***	-0.2221***
	(0.0245)	(0.0227)	(0.0132)	(0.0125)	(0.0159)	(0.0153)	(0.0112)	(0.011)	(0.0364)	(0.0431)
Non_Tweets	0.0161	0.0115	0.0069	0.0074	-0.0032	-0.0007	-0.0007	0.0008	0.0831***	0.0824***
	(0.0121)	(0.0117)	(0.0077)	(0.0077)	(0.0089)	(0.0085)	(0.0068)	(0.0068)	(0.0188)	(0.019)
Price		0.0021***		-0.002***		-0.0021***		-0.0022***		-0.0242***
		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0019)
Volume		-0.0009		-0.0003		0.0001		-0.0002		-0.0003
		(0.0011)		(0.0003)		(0.0003)		(0.0003)		(0.001)
DSVI		0.0003*		0.0004*		-0.0001		-0.0002		-0.0006
		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0007)
R-Square	0.1845	0.1985	0.1819	0.1866	0.1831	0.1878	0.1828	0.1870	0.1866	0.2028
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm Fix Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time Fix Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44
Time Series Length	754	754	754	754	754	754	754	754	754	754
Observations	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915

Note. We present the results of the third dataset in panel C; for model (1). For the third one, we include all Tweeting devices except Twitter Web Client, Twitter Web app, Twitter for iPhone, Twitter for Android, Twitter for iPad, and TweetDeck. The third dataset includes more bot tweets. Our variables of interest, $Tweets_{i,t}$, is tally Tweets and, $Non\_Tweets_{i,t}$, is total non-Tweets (retweets, replies, and quotes) for Stock *i* on day *t*. In all models, the dependent variable is $Excess\ Returns_{i,t+n}$, daily value Fama and French (1992) three-Factor Model excess return in basis points for stock *i*, on day *t* varying values of *n*. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls in our models for Volume, Price, GDSV and five lags of absolute excess returns (AER). The results with control variables are presented in even columns. We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs). *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively.

TABLE 4: Panel A

	[t = 0]		[t+1]		[t+2]		[t+3]		[t+4, t+15]	
Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
Tweets	0.1333***	0.1071***	-0.0336	-0.0334	-0.0674**	-0.0608*	-0.0338	-0.0328	0.0049	-0.0413
	(0.0413)	(0.0404)	(0.0257)	(0.0265)	(0.0307)	(0.0319)	(0.0221)	(0.0234)	(0.0657)	(0.0666)
Retweets	0.0203	0.0224	0.0331**	0.0355**	0.0146	0.0151	0.0161	0.0147	0.0017	0.0001
	(0.0187)	(0.0179)	(0.0149)	(0.0151)	(0.0127)	(0.0129)	(0.012)	(0.012)	(0.039)	(0.0389)
Replies	-0.0039	-0.0209	-0.0301	-0.0291	-0.0517**	-0.0491**	-0.0353*	-0.0361*	-0.1012	-0.1152*
	(0.0321)	(0.0331)	(0.0251)	(0.0241)	(0.0251)	(0.0245)	(0.0214)	(0.0211)	(0.0688)	(0.0676)
Quotes	-0.0471	-0.029	0.0026	-0.0039	0.0438	0.0366	0.0241	0.0223	0.063	0.0508
	(0.0346)	(0.0349)	(0.0277)	(0.0274)	(0.0269)	(0.027)	(0.0231)	(0.0231)	(0.0849)	(0.0831)
Price		0.002***		-0.002***		-0.002***		-0.0021***		-0.0241***
		(0.0005)		(0.0004)		(0.0005)		(0.0005)		(0.0019)
Volume		-0.0011		-0.0002		0.0001		-0.0001		-0.0004
		(0.0011)		(0.0004)		(0.0004)		(0.0003)		(0.001)
DSVI		0.0003*		0.0004*		-0.0001		-0.0002		-0.0006
		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0007)
R-Square	0.1883	0.2010	0.1824	0.1871	0.1846	0.1890	0.1833	0.1875	0.1861	0.2024
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm Fix Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time Fix Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44
Time Series Length	754	754	754	754	754	754	754	754	754	754
Observations	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915

Note. We present the results of the first dataset including Tweets from any Tweeting devices (sources) in Panel A, for model (2). Our variables of interest, $Tweets_{i,t}$, is tally tweets, $Retweets_{i,t}$, is total retweets, $Replies_{i,t}$, is total replies, and $Quotes_{i,t}$, is total quotes for Stock *i* on day *t*. In all models, the dependent variable is $Excess\ Returns_{i,t+n}$, daily value Fama and French (1992) three-Factor Model excess return in basis points for stock *i*, on day *t* varying values of *n*. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls in our models for Volume, Price, GDSV and five lags of absolute excess returns (AER). The results with control variables are presented in even columns. We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs). *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively.

TABLE 4: Panel B

Variable	[t = 0] (1)	[t = 0] (2)	[t+1] (3)	[t+1] (4)	[t+2] (5)	[t+2] (6)	[t+3] (7)	[t+3] (8)	[t+4, t+15] (9)	[t+4, t+15] (10)
Tweets	0.2815***	0.2334***	-0.061	-0.0595	-0.1293*	-0.1149	-0.0702	-0.0651	0.1582	0.1404
	(0.0802)	(0.0791)	(0.0599)	(0.0616)	(0.0767)	(0.0792)	(0.0548)	(0.0583)	(0.1504)	(0.1491)
Retweets	-0.0032	0.0026	0.0332**	0.0351**	0.0224*	0.021	0.0212*	0.0183	-0.0648*	-0.0785**
	(0.018)	(0.0167)	(0.015)	(0.0151)	(0.0133)	(0.0132)	(0.0123)	(0.0121)	(0.038)	(0.0375)
Replies	-0.0511*	-0.0599**	-0.021	-0.0207	-0.0282	-0.0282	-0.0239	-0.0261	-0.143**	-0.1595**
	(0.0275)	(0.029)	(0.0219)	(0.0213)	(0.0191)	(0.019)	(0.0181)	(0.018)	(0.0645)	(0.0637)
Quotes	-0.0537	-0.0357	0.0061	-0.0001	0.0435	0.0364	0.0263	0.0248	0.0783	0.0632
	(0.033)	(0.034)	(0.0282)	(0.0281)	(0.0269)	(0.0275)	(0.0235)	(0.0238)	(0.0858)	(0.084)
Price		0.0018***		-0.0019***		-0.0019***		-0.0021***		-0.0243***
		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0019)
Volume		-0.0011		-0.0002		0.0001		-0.0001		-0.0007
		(0.0011)		(0.0004)		(0.0004)		(0.0003)		(0.001)
DSVI		0.0003*		0.0004*		-0.0001		-0.0002		-0.0006
		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0007)
R-Square	0.1915	0.2033	0.1825	0.1872	0.1851	0.1894	0.1835	0.1876	0.1863	0.2026
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm Fix Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time Fix Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44
Time Series Length	754	754	754	754	754	754	754	754	754	754
Observations	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915

Note. We present the results of the second data in panel B, for model (2). we filter Tweets and pick tweets if they come from Twitter Web Client, Twitter Web app, Twitter for iPhone, Twitter for Android, Twitter for iPad, and TweetDeck. Our variables of interest, $Tweets_{i,t}$, is tally tweets, $Retweets_{i,t}$, is total retweets, $Replies_{i,t}$, is total replies, and $Quotes_{i,t}$, is total quotes for Stock *i* on day *t*. In all models, the dependent variable is $Excess\ Returns_{i,t+n}$, daily value Fama and French (1992) three-Factor Model excess return in basis points for stock *i*, on day *t* varying values of *n*. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls in our models for Volume, Price, GDSV and five lags of absolute excess returns (AER). The results with control variables are presented in even columns. We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs). *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively.

TABLE 4: Panel C

	[t = 0]		[t+1]		[t+2]		[t+3]		[t+4, t+15]	
Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
Tweets	0.0636**	0.0456*	-0.0211	-0.0212*	-0.0421**	-0.039**	-0.0176	-0.0199*	-0.1289***	-0.1982***
	(0.0259)	(0.0233)	(0.0131)	(0.0125)	(0.0172)	(0.0165)	(0.0118)	(0.0115)	(0.0358)	(0.0411)
Retweets	0.0152	0.0113	0.0079	0.0085	-0.0031	-0.0009	-0.0006	0.0009	0.0876***	0.0888***
	(0.0117)	(0.0111)	(0.0074)	(0.0074)	(0.0087)	(0.0084)	(0.0067)	(0.0068)	(0.0181)	(0.0179)
Replies	0.0065	0.0044	-0.0022	-0.002	-0.009*	-0.0081*	-0.0003	0.0002	0.0093	0.0088
	(0.057)	(0.0057)	(0.0056)	(0.0057)	(0.0048)	(0.0047)	(0.0043)	(0.0043)	(0.0125)	(0.0124)
Quotes	-0.0063	-0.0073	-0.0059	-0.0069	0.01	0.0096	-0.0003	-0.0007	-0.0611**	-0.0716***
	(0.0147)	(0.0143)	(0.0098)	(0.0098)	(0.0094)	(0.0093)	(0.0068)	(0.0069)	(0.0239)	(0.0241)
Price		0.0021***		-0.002***		-0.0021***		-0.0022***		-0.0242***
		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0019)
Volume		-0.0009		-0.0003		0.0001		0.0002		-0.0002
		(0.001)		(0.0003)		(0.0003)		(0.0003)		(0.001)
DSVI		0.0003**		0.0004*		-0.0001		-0.0002		-0.0006
		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0007)
R-Square	0.1846	0.1986	0.1820	0.1866	0.1833	0.1879	0.1828	0.1870	0.1868	0.2031
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm Fix Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time Fix Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44
Time Series Length	754	754	754	754	754	754	754	754	754	754
Observations	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915

Note. We present the results of the third dataset in panel C; for model (2). For the third one, we include all Tweeting devices except Twitter Web Client, Twitter Web app, Twitter for iPhone, Twitter for Android, Twitter for iPad, and TweetDeck. The third dataset includes more bot tweets. Our variables of interest, $Tweets_{i,t}$, is tally tweets, $Retweets_{i,t}$, is total retweets, $Replies_{i,t}$, is total replies, and $Quotes_{i,t}$, is total quotes for Stock *i* on day *t*. In all models, the dependent variable is $Excess\ Returns_{i,t+n}$, daily value Fama and French (1992) three-Factor Model excess return in basis points for stock *i*, on day *t* varying values of *n*. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls in our models for Volume, Price, GDSV and five lags of absolute excess returns (AER). The results with control variables are presented in even columns. We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs). *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively.

TABLE 5: Panel A

	[t = 0]						[t +1]						[t +2]					
Variable	**[1]**	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	**[9]**	**[10]**	**[11]**	**[12]**	**[13]**	**[14]**	**[15]**	**[16]**	**[17]**	**[18]**
URL_Posts	0.05**	0.0313					-0.0189	-0.021					-0.032**	-0.03**				
	(0.02)	(0.0207)					(0.0134)	(0.013)					(0.013)	(0.012)				
Text_Posts			0.0693**	0.0498*					-0.02	-0.022					-0.0363	-0.033		
			(0.0279)	(0.0279)					(0.02)	(0.019)					(0.022)	(0.021)		
Media_Posts					0.0741***	0.0558**					-0.015	-0.015					-0.0324*	-0.0273
					(0.0243)	(0.0245)					(0.016)	(0.016)					(0.018)	(0.017)
Price		0.0022***		0.0021***		0.0021***		-0.002***		-0.002***		-0.002***		-0.0021***		-0.0021***		-0.0021***
		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0005)
Volume		-0.0008		-0.0009		-0.0009		-0.0003		-0.0002		-0.0003		-0.0003		-0.0003		-0.0003
		(0.0011)		(0.0011)		(0.0011)		(0.0003)		(0.0004)		(0.0004)		(0.0003)		(0.0003)		(0.0003)
DSVI		0.0004**		0.0003**		0.0003**		0.0004*		0.0004*		0.0004*		-0.0001*		-0.0001		-0.0001
		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0002)
R-Square	0.1832	0.1979	0.1859	0.1993	0.1868	0.1999	0.1819	0.1866	0.1821	0.1868	0.1820	0.1866	0.1828	0.1875	0.1835	0.1880	0.1833	0.1880
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44	44	44	44	44	44	44	44	44
Time Series	754	754	754	754	754	754	754	754	754	754	754	754	754	754	754	754	754	754
Observations	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915

TABLE 5: Panel A (cont.)

	[t +3]						[t+4, t+15]					
Variable	**[19]**	**[20]**	**[21]**	**[22]**	**[23]**	**[24]**	**[25]**	**[26]**	**[27]**	**[28]**	**[29]**	**[30]**
URL_Posts	-0.011 (0.01)	-0.013 (0.009)					-0.069** (0.031)	-0.151*** (0.034)				
Text_Posts			-0.012 (0.016)	-0.0126 (0.016)					-0.028 (0.043)	-0.0803* (0.042)		
Media_Posts					-0.01 (0.013)	-0.009 (0.012)					-0.032 (0.034)	-0.0723** (0.032)
Price		-0.0022*** (0.0005)		-0.0021*** (0.0005)		-0.0021*** (0.0005)		-0.0243*** (0.0019)		-0.0241*** (0.0019)		-0.0241*** (0.0019)
Volume		-0.0002 (0.0003)		-0.0002 (0.0003)		-0.0002 (0.0003)		-0.0002 (0.001)		-0.0005 (0.001)		-0.0005 (0.001)
DSVI		-0.0002 (0.0002)		-0.0002 (0.0002)		-0.0002 (0.0002)		-0.0006 (0.0007)		-0.0006 (0.0007)		-0.0006 (0.0007)
R-Square	0.1828	0.1869	0.1828	0.1870	0.1828	0.1869	0.1861	0.2026	0.1860	0.2023	0.1860	0.2022
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44	44	44
Time Series	754	754	754	754	754	754	754	754	754	754	754	754
Observations	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915

Note. We present the results of the first dataset including Tweets from any Tweeting devices (sources) in Panel A, for models 3 to 5. Our variables of interest, $URL\_Posts_{i,t}$, is all posts including URLs, $Media\_Posts_{i,t}$, is all posts including images/videos, and $Text\_Posts_{i,t}$, is all posts including just text without any media or URLs. In all models, the dependent variable is $Excess\ Returns_{i,t+n}$, daily value Fama and French (1992) three-Factor Model excess return in basis points for stock i, on day t varying values of n. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls in our models for Volume, Price, GDSV and five lags of absolute excess returns (AER). The results with control variables are presented in even columns. We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs). *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively.

TABLE 5: Panel B

Variable	[t = 0]						[t +1]						[t +2]					
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]
URL_Posts	0.034**	0.0222					-0.0123	-0.0138					-0.021**	-0.019**				
	(0.013)	(0.014)					(0.0102)	(0.0099)					(0.0095)	(0.009)				
Text_Posts			0.0664**	0.0476*					-0.02	-0.021					-0.0354	-0.032		
			(0.0276)	(0.0272)					(0.02)	(0.019)					(0.022)	(0.021)		
Media_Posts					0.0758***	0.0552*					-0.020	-0.0207					-0.0383*	-0.033
					(0.0295)	(0.0245)					(0.019)	(0.0187)					(0.023)	(0.0211)
Price		0.0022***		0.0021***		0.0021***		-0.002***		-0.002***		-0.002***		-0.0021***		-0.0021***		-0.0021***
		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0005)
Volume		-0.0008		-0.0009		-0.0008		-0.0003		-0.0003		-0.0003		-0.0003		-0.0003		-0.0003
		(0.0011)		(0.0011)		(0.0011)		(0.0003)		(0.0004)		(0.0004)		(0.0003)		(0.0003)		(0.0003)
DSVI		0.0004**		0.0003**		0.0003**		0.0004*		0.0004*		0.0004*		-0.0001*		-0.0001		-0.0001
		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0002)
R-Square	0.1834	0.1980	0.1857	0.1992	0.1857	0.1993	0.1819	0.1866	0.1821	0.1867	0.1820	0.1867	0.1829	0.1877	0.1834	0.1880	0.1833	0.1879
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44	44	44	44	44	44	44	44	44
Time Series	754	754	754	754	754	754	754	754	754	754	754	754	754	754	754	754	754	754
Observations	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915

TABLE 5: Panel B (cont.)

Variable	[t +3]						[t+4, t+15]					
	[19]	[20]	[21]	[22]	[23]	[24]	[25]	[26]	[27]	[28]	[29]	[30]
URL_Posts	-0.007	-0.0084					-0.0399	-0.0846***				
	(0.007)	(0.007)					(0.0248)	(0.0241)				
Text_Posts			-0.012	-0.013					-0.028	-0.0785*		
			(0.016)	(0.0156)					(0.043)	(0.042)		
Media_Posts					-0.012	-0.012					-0.037	-0.079**
					(0.016)	(0.0153)					(0.041)	(0.038)
Price		-0.0022***		-0.0022***		-0.0021***		-0.0243***		-0.0241***		-0.0241***
		(0.0005)		(0.0005)		(0.0005)		(0.0019)		(0.0019)		(0.0019)
Volume		-0.0002		-0.0002		-0.0002		-0.0004		-0.0005		-0.0006
		(0.0003)		(0.0003)		(0.0003)		(0.001)		(0.001)		(0.001)
DSVI		-0.0002		-0.0002		-0.0002		-0.0006		-0.0006		-0.0006
		(0.0002)		(0.0002)		(0.0002)		(0.0007)		(0.0007)		(0.0007)
R-Square	0.1828	0.1869	0.1828	0.1870	0.1828	0.1869	0.1861	0.2025	0.1860	0.2023	0.1860	0.2022
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44	44	44
Time Series	754	754	754	754	754	754	754	754	754	754	754	754
Observations	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915

Note. We present the results of the second data in panel B, for models 3 to 5. we filter Tweets and pick tweets if they come from Twitter Web Client, Twitter Web app, Twitter for iPhone, Twitter for Android, Twitter for iPad, and TweetDeck. Our variables of interest, $URL\_Posts_{i,t}$, is all posts including URLs, $Media\_Posts_{i,t}$, is all posts including images/videos, and $Text\_Posts_{i,t}$, is all posts including just text without any media or URLs. In all models, the dependent variable is $Excess\ Returns_{i,t+n}$, daily value Fama and French (1992) three-Factor Model excess return in basis points for stock i, on day t varying values of n. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls in our models for Volume, Price, GDSV and five lags of absolute excess returns (AER). The results with control variables are presented in even columns. We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs). *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively.

TABLE 5: Panel C

	[t = 0]						[t +1]						[t +2]					
Variable	**[1]**	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	**[9]**	**[10]**	**[11]**	**[12]**	**[13]**	**[14]**	**[15]**	**[16]**	**[17]**	**[18]**
URL_Posts	0.0423*	0.0225					-0.0171	-0.0177*					-0.029*	-0.024**				
	(0.0245)	(0.021)					(0.011)	(0.0099)					(0.0149)	(0.0134)				
Text_Posts			0.0552**	0.0381*					-0.0147	-0.0149					-0.0257	-0.0212		
			(0.0209)	(0.0213)					(0.0142)	(0.0139)					(0.0159)	(0.0153)		
Media_Posts					0.057***	0.0452***					-0.001	-0.0001					-0.0143**	-0.0093
					(0.0118)	(0.0115)					(0.008)	(0.0073)					(0.0072)	(0.0065)
Price		0.0022***		0.0021***		0.0021***		-0.002***		-0.002***		-0.002***		-0.0021***		-0.0021***		-0.0021***
		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0005)
Volume		-0.0007		-0.0009		-0.0011		-0.0003		-0.0002		-0.0004		-0.0001		-0.0001		-0.0001
		(0.0011)		(0.0011)		(0.0011)		(0.0003)		(0.0004)		(0.0003)		(0.0003)		(0.0003)		(0.0003)
DSVI		0.0004**		0.0003**		0.0003**		0.0004*		0.0004*		0.0004*		-0.0001		-0.0001		-0.0001
		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0002)
R-Square	0.1825	0.1976	0.1852	0.1988	0.1875	0.2005	0.1818	0.1865	0.1820	0.1866	0.1818	0.1864	0.1826	0.1874	0.1831	0.1877	0.1827	0.1874
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44	44	44	44	44	44	44	44	44
Time Series	754	754	754	754	754	754	754	754	754	754	754	754	754	754	754	754	754	754
Observations	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915

TABLE 5: Panel C (cont.)

Variable	[t +3]						[t+4, t+15]					
	[19]	[20]	[21]	[22]	[23]	[24]	[25]	[26]	[27]	[28]	[29]	[30]
URL_Posts	-0.011 (0.008)	-0.0119 (0.0076)					-0.075*** (0.0278)	-0.161*** (0.0473)				
Text_Posts			-0.0053 (0.0116)	-0.0045 (0.0115)					-0.0166 (0.0313)	-0.0524* (0.03)		
Media_Posts					-0.003 (0.0058)	-0.001 (0.0053)					-0.0153 (0.0158)	-0.0419*** (0.0147)
Price		-0.0022*** (0.0005)		-0.0021*** (0.0005)		-0.0022*** (0.0005)		-0.0243*** (0.0019)		-0.0241*** (0.0019)		-0.0241*** (0.0019)
Volume		-0.0002 (0.0003)		-0.0002 (0.0003)		-0.0002 (0.0003)		-0.0004 (0.001)		-0.0005 (0.001)		-0.0004 (0.001)
DSVI		-0.0002 (0.0002)		-0.0002 (0.0002)		-0.0002 (0.0002)		-0.0006 (0.0007)		-0.0006 (0.0007)		-0.0006 (0.0007)
R-Square	0.1827	0.1869	0.1827	0.1869	0.1827	0.1869	0.1861	0.2023	0.1860	0.2021	0.1860	0.2021
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44	44	44
Time Series	754	754	754	754	754	754	754	754	754	754	754	754
Observations	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915

Note. We present the results of the third dataset in panel C, for models 3 to 5. For the third one, we include all Tweeting devices except Twitter Web Client, Twitter Web app, Twitter for iPhone, Twitter for Android, Twitter for iPad, and TweetDeck. The third dataset includes more bot tweets. Our variables of interest, $URL\_Posts_{i,t}$, is all posts including URLs, $Media\_Posts_{i,t}$, is all posts including images/videos, and $Text\_Posts_{i,t}$, is all posts including just text without any media or URLs. In all models, the dependent variable is $Excess\ Returns_{i,t+n}$, daily value Fama and French (1992) three-Factor Model excess return in basis points for stock i, on day t varying values of n. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls in our models for Volume, Price, GDSV and five lags of absolute excess returns (AER). The results with control variables are presented in even columns. We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs). *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively.

TABLE 6: Panel A

	[t = 0]		[t +1]		[t +2]		[t +3]		[t+4, t+15]	
Variable	**[1]**	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	**[9]**	**[10]**
URL_Posts	-0.0501*	-0.0412	-0.0006	-0.0037	0.008	0.0032	0.001	-0.004	-0.079*	-0.1432***
	(0.0304)	(0.0265)	(0.0192)	(0.0184)	(0.025)	(0.024)	(0.0186)	(0.0183)	(0.046)	(0.0488)
Text_Posts	0.0368	0.0233	-0.0224	-0.024	-0.03	-0.029	-0.0111	-0.0133	0.023	-0.0089
	(0.0392)	(0.0351)	(0.0265)	(0.026)	(0.031)	(0.029)	(0.0241)	(0.0235)	(0.064)	(0.0597)
Media_Posts	0.0687***	0.0566**	0.0033	0.0052	-0.013	-0.007	-0.0016	0.0032	-0.013	0.0002
	(0.0257)	(0.0231)	(0.0125)	(0.013)	(0.010)	(0.001)	(0.011)	(0.0108)	(0.036)	(0.0387)
Price		-0.002***		-0.002***		-0.0021***		-0.0022***		-0.0243***
		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0019)
Volume		-0.0009		-0.0002		-0.0001		-0.0002		-0.0002
		(0.0011)		(0.0003)		(0.0003)		(0.0003)		(0.001)
DSVI		0.0003**		0.0004**		0.0004**		-0.0002		-0.0006
		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0007)
R-Square	0.1874	0.2003	0.1821	0.1868	0.1835	0.1881	0.1828	0.1870	0.1861	0.2026
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44
Time Series	754	754	754	754	754	754	754	754	754	754
Observations	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915

Note. We present the results of the first dataset including Tweets from any Tweeting devices (sources) in Panel A, for model 6. Our variables of interest, $URL\_Posts_{i,t}$, is all posts including URLs, $Media\_Posts_{i,t}$, is all posts including images/videos, and $Text\_Posts_{i,t}$, is all posts including just text without any media or URLs. In all models, the dependent variable is $Excess\ Returns_{i,t+n}$, daily value Fama and French (1992) three-Factor Model excess return in basis points for stock i, on day t varying values of n. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls in our models for Volume, Price, GDSV and five lags of absolute excess returns (AER). The results with control variables are presented in even columns. We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs). *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively.

TABLE 6: Panel B

	[t = 0]		[t +1]		[t +2]		[t +3]		[t+4, t+15]	
Variable	**[1]**	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	**[9]**	**[10]**
URL_Posts	-0.0489	-0.0377	0.0054	0.0027	0.0155	0.0097	0.0057	0.0018	-0.0572	-0.0881**
	(0.0308)	(0.0273)	(0.0199)	(0.0194)	(0.0266)	(0.0256)	(0.0198)	(0.0198)	(0.0499)	(0.0443)
Text_Posts	0.0736	0.0529	-0.0181	-0.0188	-0.0351	-0.0303	-0.0141	-0.0139	0.0368	0.0068
	(0.0473)	(0.0419)	(0.0312)	(0.0302)	(0.0381)	(0.0366)	(0.0294)	(0.0289)	(0.0718)	(0.0629)
Media_Posts	0.0567**	0.0445*	-0.0091	-0.0068	-0.0217	-0.0158	-0.0051	-0.0013	-0.0155	-0.0021
	(0.0281)	(0.0255)	(0.0155)	(0.0157)	(0.016)	(0.0157)	(0.0148)	(0.0146)	(0.0439)	(0.0436)
Price		0.002***		-0.002***		-0.0021***		-0.0021***		-0.0243***
		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0019)
Volume		-0.0008		-0.0003		-0.0001		-0.0002		-0.0004
		(0.0011)		(0.0003)		(0.0003)		(0.0003)		(0.001)
DSVI		0.0003**		0.0004*		-0.0001		-0.0002		-0.0006
		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0007)
R-Square	0.1870	0.1999	0.1821	0.1868	0.1836	0.1881	0.1828	0.1870	0.1861	0.2025
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44
Time Series	754	754	754	754	754	754	754	754	754	754
Observations	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915

Note. We present the results of the second data in panel B, for model (6). we filter Tweets and pick tweets if they come from Twitter Web Client, Twitter Web app, Twitter for iPhone, Twitter for Android, Twitter for iPad, and TweetDeck. Our variables of interest, $URL\_Posts_{i,t}$, is all posts including URLs, $Media\_Posts_{i,t}$, is all posts including images/videos, and $Text\_Posts_{i,t}$, is all posts including just text without any media or URLs. In all models, the dependent variable is $Excess\ Returns_{i,t+n}$, daily value Fama and French (1992) three-Factor Model excess return in basis points for stock i, on day t varying values of n. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls in our models for Volume, Price, GDSV and five lags of absolute excess returns (AER). The results with control variables are presented in even columns. We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs). *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively.

TABLE 6: Panel C

	[t = 0]		[t +1]		[t +2]		[t +3]		[t+4, t+15]	
Variable	**[1]**	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	**[9]**	**[10]**
URL_Posts	-0.0486***	-0.041**	-0.0127	-0.0146	-0.0048	-0.0087	-0.0068	-0.0115	-0.0694**	-0.1296***
	(0.0187)	(0.0163)	(0.0106)	(0.0102)	(0.014)	(0.014)	(0.0099)	(0.0094)	(0.03)	(0.0481)
Text_Posts	0.0151	0.0052	-0.0246	-0.0239	-0.0241	-0.0221	-0.0043	-0.0051	0.0001	-0.0193
	(0.0246)	(0.0238)	(0.0164)	(0.016)	(0.0199)	(0.019)	(0.0145)	(0.0144)	(0.0408)	(0.0387)
Media_Posts	0.0569***	0.0488***	0.0141**	0.0141**	-0.0009	0.003	0.0003	0.0033	-0.0042	-0.0125
	(0.0101)	(0.009)	(0.0062)	(0.0061)	(0.0052)	(0.0052)	(0.0055)	(0.0054)	(0.0181)	(0.0181)
Price		0.002***		-0.002***		-0.0021***		-0.0022***		-0.0242***
		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0019)
Volume		-0.001		-0.0003		-0.0001		-0.0002		-0.0002
		(0.001)		(0.0003)		(0.0003)		(0.0003)		(0.001)
DSVI		0.0003**		0.0004*		-0.0001		-0.0002		-0.0006
		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0007)
R-Square	0.1879	0.2008	0.1822	0.1868	0.1831	0.1877	0.1827	0.1869	0.1861	0.2023
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44
Time Series	754	754	754	754	754	754	754	754	754	754
Observations	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915

Note. We present the results of the third dataset in panel C, for model (6). For the third one, we include all Tweeting devices except Twitter Web Client, Twitter Web app, Twitter for iPhone, Twitter for Android, Twitter for iPad, and TweetDeck. The third dataset includes more bot tweets. Our variables of interest, $URL\_Posts_{i,t}$, is all posts including URLs, $Media\_Posts_{i,t}$, is all posts including images/videos, and $Text\_Posts_{i,t}$, is all posts including just text without any media or URLs. In all models, the dependent variable is $Excess\ Returns_{i,t+n}$,daily value Fama and French (1992) three-Factor Model excess return in basis points for stock i, on day t varying values of n. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls in our models for Volume, Price, GDSV and five lags of absolute excess returns (AER). The results with control variables are presented in even columns. We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs). *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively.

TABLE 7: Panel A

	[t = 0]		[t +1]		[t +2]		[t +3]		[t+4, t+15]	
Variable	**[1]**	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	**[9]**	**[10]**
Organization_entity	0.019	0.016	0.013	0.012	0.008	0.008	-0.002	-0.002	-0.036	-0.052**
	(0.015)	(0.015)	(0.012)	(0.012)	(0.011)	(0.01)	(0.011)	(0.01)	(0.024)	(0.025)
Person_entity	-0.066*	-0.059*	-0.016	-0.018	-0.004	-0.007	-0.001	-0.004	-0.096**	-0.113**
	(0.036)	(0.036)	(0.018)	(0.018)	(0.011)	(0.01)	(0.019)	(0.02)	(0.038)	(0.048)
Place_entity	-0.03**	-0.029**	-0.014	-0.015	-0.004	-0.006	0.002	-0.001	-0.116***	-0.13***
	(0.014)	(0.014)	(0.009)	(0.009)	(0.007)	(0.007)	(0.007)	(0.007)	(0.027)	(0.028)
Product_entity	-0.04***	-0.032**	0.005	0.007	0.003	0.003	0.001	0.003	-0.008	0.012
	(0.013)	(0.015)	(0.022)	(0.023)	(0.014)	(0.014)	(0.012)	(0.013)	(0.045)	(0.052)
Other_entity	0.291***	0.239*	-0.095	-0.1	-0.191*	-0.184	-0.068	-0.074	0.276	0.187
	(0.111)	(0.125)	(0.092)	(0.093)	(0.114)	(0.119)	(0.12)	(0.128)	(0.191)	(0.233)
Price		0.0019***		-0.0019***		-0.0019***		-0.0021***		-0.0244***
		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0019)
Volume		-0.0011		-0.0002		0.0001		-0.0001		-0.0005
		(0.001)		(0.0004)		(0.0003)		(0.0004)		(0.001)
DSVI		0.0003**		0.0004*		-0.0001		-0.0002		-0.0006
		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0007)
R-Square	0.1926	0.2038	0.1830	0.1876	0.1866	0.1909	0.1833	0.1875	0.1871	0.2032
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44
Time Series	754	754	754	754	754	754	754	754	754	754
Observations	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915

Note. We present the results of the first dataset including Tweets from any Tweeting devices (sources) in Panel A, for model 7. Our variables of interest, $Organization\_entity_{i,t}$, is all Tweets labeled Organization entity, $Person\_entity_{i,t}$, is all Tweets labeled Person entity, $Place\_entity_{i,t}$, is all Tweets labeled Place entity, $Product\_entity_{i,t}$, is all Tweets labeled Product entity, and $Other\_entity_{i,t}$, is all Tweets labeled Other entity. In all models, the dependent variable is $Excess\ Returns_{i,t+n}$, daily value Fama and French (1992) three-Factor Model excess return in basis points for stock i, on day t varying values of n. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls in our models for Volume, Price, GDSV and five lags of absolute excess returns (AER). The results with control variables are presented in even columns. We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs). *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively.

TABLE 7: Panel B

	[t = 0]		[t +1]		[t +2]		[t +3]		[t+4, t+15]	
Variable	**[1]**	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	**[9]**	**[10]**
Organization_entity	0.017	0.014	0.009	0.008	0.01	0.01	-0.003	-0.003	-0.042	-0.063**
	(0.016)	(0.017)	(0.014)	(0.014)	(0.012)	(0.012)	(0.012)	(0.012)	(0.027)	(0.028)
Person_entity	-0.068*	-0.062	-0.014	-0.015	-0.005	-0.007	-0.003	-0.004	-0.107**	-0.112**
	(0.041)	(0.039)	(0.018)	(0.019)	(0.01)	(0.01)	(0.022)	(0.023)	(0.045)	(0.051)
Place_entity	-0.026*	-0.026*	-0.012	-0.013	-0.005	-0.007	0.005	0.003	-0.115***	-0.126***
	(0.013)	(0.013)	(0.008)	(0.009)	(0.007)	(0.007)	(0.008)	(0.008)	(0.029)	(0.03)
Product_entity	-0.015	-0.01	0.026	0.028	-0.005	-0.004	0.005	0.006	-0.007	0.011
	(0.011)	(0.012)	(0.018)	(0.018)	(0.007)	(0.008)	(0.008)	(0.008)	(0.031)	(0.032)
Other_entity	0.32***	0.264*	-0.107	-0.111	-0.211*	-0.205	-0.08	-0.085	0.316	0.236
	(0.105)	(0.125)	(0.105)	(0.107)	(0.127)	(0.133)	(0.146)	(0.155)	(0.197)	(0.237)
Price		0.0019***		-0.0019***		-0.0019***		-0.0021***		-0.0244***
		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0019)
Volume		-0.001		-0.0002		0.0001		0.0001		-0.0006
		(0.0011)		(0.0004)		(0.0003)		(0.0004)		(0.001)
DSVI		0.0003**		0.0004*		-0.0001		-0.0002		-0.0006
		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0007)
R-Square	0.1930	0.2042	0.1832	0.1878	0.1869	0.1911	0.1834	0.1876	0.1871	0.2032
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44
Time Series	754	754	754	754	754	754	754	754	754	754
Observations	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915

Note. We present the results of the second data in panel B, for model 7. we filter Tweets and pick tweets if they come from Twitter Web Client, Twitter Web app, Twitter for iPhone, Twitter for Android, Twitter for iPad, and TweetDeck. Our variables of interest, $Organization\_entity_{i,t}$, is all Tweets labeled Organization entity, $Person\_entity_{i,t}$, is all Tweets labeled Person entity, $Place\_entity_{i,t}$, is all Tweets labeled Place entity, $Product\_entity_{i,t}$, is all Tweets labeled Product entity, and $Other\_entity_{i,t}$, is all Tweets labeled Other entity. In all models, the dependent variable is $Excess\ Returns_{i,t+n}$, daily value Fama and French (1992) three-Factor Model excess return in basis points for stock i, on day t varying values of n. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls in our models for Volume, Price, GDSV and five lags of absolute excess returns (AER). The results with control variables are presented in even columns. We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs). *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively.

TABLE 7: Panel C

	[t = 0]		[t +1]		[t +2]		[t +3]		[t+4, t+15]	
Variable	**[1]**	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	**[9]**	**[10]**
Organization_entity	0.034	0.028	0.021	0.022	0.004	0.007	-0.003	-0.001	-0.07*	-0.083**
	(0.025)	(0.024)	(0.018)	(0.018)	(0.016)	(0.015)	(0.013)	(0.013)	(0.039)	(0.04)
Person_entity	-0.025***	-0.018*	-0.004	-0.008	0.004	0.002	0.009	0.005	0.003	-0.034
	(0.009)	(0.01)	(0.008)	(0.009)	(0.007)	(0.007)	(0.006)	(0.006)	(0.022)	(0.023)
Place_entity	-0.029*	-0.028*	-0.011	-0.012	0.002	0.001	-0.015	-0.016*	-0.058**	-0.069**
	(0.015)	(0.016)	(0.013)	(0.013)	(0.009)	(0.009)	(0.009)	(0.009)	(0.027)	(0.028)
Product_entity	-0.071***	-0.054**	-0.011	-0.011	0.018	0.014	0.001	0.001	0.035	0.08
	(0.024)	(0.024)	(0.016)	(0.017)	(0.017)	(0.018)	(0.013)	(0.014)	(0.086)	(0.107)
Other_entity	0.096*	0.071	-0.031	-0.032	-0.066*	-0.062	-0.016	-0.017	0.052	-0.002
	(0.051)	(0.049)	(0.029)	(0.028)	(0.038)	(0.038)	(0.029)	(0.029)	(0.064)	(0.063)
Price		0.0021***		-0.002***		-0.0021***		-0.0021***		-0.0244***
		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0019)
Volume		-0.001		-0.0002		0.0001		-0.0001		-0.0006
		(0.0011)		(0.0004)		(0.0004)		(0.0003)		(0.001)
DSVI		0.0003**		0.0004*		-0.0001		-0.0002		-0.0006
		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0007)
R-Square	0.1867	0.1998	0.1823	0.1869	0.1843	0.1887	0.1829	0.1871	0.1862	0.2024
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44
Time Series	754	754	754	754	754	754	754	754	754	754
Observations	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915

Note. We present the results of the third dataset in panel C, for model (7). For the third one, we include all Tweeting devices except Twitter Web Client, Twitter Web app, Twitter for iPhone, Twitter for Android, Twitter for iPad, and TweetDeck. The third dataset includes more bot tweets. Our variables of interest, $Organization\_entity_{i,t}$, is all Tweets labeled Organization entity, $Person\_entity_{i,t}$, is all Tweets labeled Person entity, $Place\_entity_{i,t}$, is all Tweets labeled Place entity, $Product\_entity_{i,t}$, is all Tweets labeled Product entity, and $Other\_entity_{i,t}$, is all Tweets labeled Other entity. In all models, the dependent variable is $Excess\ Returns_{i,t+n}$, daily value Fama and French (1992) three-Factor Model excess return in basis points for stock i, on day t varying values of n. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls in our models for Volume, Price, GDSV and five lags of absolute excess returns (AER). The results with control variables are presented in even columns. We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs). *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively.

TABLE 8: Panel A

	[t = 0]		[t +1]		[t +2]		[t +3]		[t+4, t+15]	
Variable	**[1]**	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	**[9]**	**[10]**
White_Users	-0.011	-0.025	-0.049**	-0.054**	-0.034	-0.034	-0.047**	-0.048**	-0.05	-0.115*
	(0.033)	(0.034)	(0.024)	(0.024)	(0.025)	(0.025)	(0.021)	(0.021)	(0.069)	(0.069)
API_Users	0.102*	0.095*	0.02	0.019	-0.036	-0.034	-0.006	-0.006	-0.015	-0.014
	(0.053)	(0.049)	(0.039)	(0.038)	(0.045)	(0.044)	(0.032)	(0.032)	(0.097)	(0.094)
Black_Users	-0.005	-0.008	0.003	0.004	0.005	0.007	-0.0008	0.001	0.079***	0.09***
	(0.007)	(0.007)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.015)	(0.016)
Hispanic_Users	0.014	0.011	0.007	0.01	0.009	0.011	0.024	0.024	0.01	0.006
	(0.024)	(0.022)	(0.019)	(0.019)	(0.018)	(0.017)	(0.015)	(0.015)	(0.048)	(0.047)
Price		0.0022***		-0.002***		-0.0021***		-0.0022***		-0.0244***
		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0019)
Volume		-0.0007		-0.0003		-0.0001		-0.0002		-0.0006
		(0.0011)		(0.0004)		(0.0003)		(0.0003)		(0.001)
DSVI		0.0004**		0.0004*		-0.0001		-0.0002		-0.0006
		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0007)
R-Square	0.1872	0.2003	0.1821	0.1868	0.1840	0.1886	0.1832	0.1873	0.1867	0.2031
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44
Time Series	754	754	754	754	754	754	754	754	754	754
Observations	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915

Note. We present the results of the first dataset including Tweets from any Tweeting devices (sources) in Panel A, for model 8. Our independent variables are $White\ Users_{i,t}$, $API\ Users_{i,t}$, $Black\ Users_{i,t}$, and $Hispanic\ Users_{i,t}$ indicates the count of social media posts by White, Asian and Pacific Islander (API), Black and Hispanic social media users for stock *i* during time *t*. In all models, the dependent variable is $Excess\ Returns_{i,t+n}$, daily value Fama and French (1992) three-Factor Model excess return in basis points for stock i, on day t varying values of n. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls in our models for Volume, Price, GDSV and five lags of absolute excess returns (AER). The results with control variables are presented in even columns. We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs). *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively.

TABLE 8: Panel B

	[t = 0]		[t +1]		[t +2]		[t +3]		[t+4, t+15]	
Variable	**[1]**	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	**[9]**	**[10]**
White_Users	0.011	-0.019	-0.109**	-0.119**	-0.073	-0.074	-0.12***	-0.122***	-0.082	-0.179
	(0.069)	(0.071)	(0.051)	(0.05)	(0.055)	(0.055)	(0.045)	(0.045)	(0.146)	(0.147)
API_Users	0.032	0.046	0.061*	0.062*	0.022	0.021	0.051*	0.049*	-0.071	-0.061
	(0.045)	(0.041)	(0.037)	(0.036)	(0.032)	(0.032)	(0.03)	(0.029)	(0.101)	(0.102)
Black_Users	-0.038***	-0.037***	0.01	0.012	0.006	0.007	0.0004	0.003	-0.041	-0.021
	(0.013)	(0.013)	(0.011)	(0.011)	(0.011)	(0.011)	(0.01)	(0.01)	(0.03)	(0.03)
Hispanic_Users	0.089***	0.077***	0.014	0.018	-0.017	-0.012	0.038**	0.039**	0.159**	0.168**
	(0.034)	(0.029)	(0.026)	(0.026)	(0.02)	(0.02)	(0.019)	(0.019)	(0.072)	(0.069)
Price		0.0021***		-0.002***		-0.0021***		-0.0022***		-0.0243***
		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0019)
Volume		-0.0007		-0.0003		-0.0001		-0.0002		-0.0007
		(0.0011)		(0.0004)		(0.0003)		(0.0003)		(0.0009)
DSVI		0.0004**		0.0004*		-0.0001		-0.0002		-0.0006
		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0007)
R-Square	0.1881	0.2011	0.1824	0.1871	0.1841	0.1886	0.1834	0.1876	0.1863	0.2024
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44
Time Series	754	754	754	754	754	754	754	754	754	754
Observations	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915

Note. We present the results of the second data in panel B, for model 8. we filter Tweets and pick tweets if they come from Twitter Web Client, Twitter Web app, Twitter for iPhone, Twitter for Android, Twitter for iPad, and TweetDeck. Our independent variables are $White\ Users_{i,t}$, $API\ Users_{i,t}$, $Black\ Users_{i,t}$, and $Hispanic\ Users_{i,t}$ indicates the count of social media posts by White, Asian and Pacific Islander (API), Black and Hispanic social media users for stock *i* during time *t*. In all models, the dependent variable is $Excess\ Returns_{i,t+n}$, daily value Fama and French (1992) three-Factor Model excess return in basis points for stock i, on day t varying values of n. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls in our models for Volume, Price, GDSV and five lags of absolute excess returns (AER). The results with control variables are presented in even columns. We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs). *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively.

TABLE 8: Panel C

	[t = 0]		[t +1]		[t +2]		[t +3]		[t+4, t+15]	
Variable	**[1]**	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	**[9]**	**[10]**
White_Users	-0.003 (0.015)	-0.007 (0.014)	-0.009 (0.011)	-0.01 (0.01)	-0.008 (0.011)	-0.007 (0.01)	0.001 (0.009)	-0.0005 (0.009)	-0.01 (0.031)	-0.046 (0.029)
API_Users	0.09*** (0.03)	0.073*** (0.027)	-0.004 (0.02)	-0.004 (0.019)	-0.042* (0.026)	-0.037 (0.025)	-0.021 (0.018)	-0.019 (0.018)	0.005 (0.045)	-0.016 (0.041)
Black_Users	-0.001 (0.006)	-0.004 (0.006)	0.002 (0.004)	0.003 (0.004)	0.003 (0.004)	0.005 (0.004)	-0.001 (0.004)	0.0001 (0.004)	0.076*** (0.014)	0.083*** (0.015)
Hispanic_Users	-0.006 (0.008)	-0.006 (0.007)	-0.001 (0.005)	-0.001 (0.005)	0.004 (0.006)	0.004 (0.006)	0.002 (0.005)	0.001 (0.005)	-0.036** (0.015)	-0.047*** (0.014)
Price		0.0022*** (0.0005)		-0.002*** (0.0005)		-0.0021*** (0.0005)		-0.0022*** (0.0005)		-0.0242*** (0.0019)
Volume		-0.0006 (0.0011)		-0.0003 (0.0004)		-0.0001 (0.0003)		-0.0002 (0.0003)		-0.0008 (0.0009)
DSVI		0.0004** (0.0002)		0.0004* (0.0002)		-0.0001 (0.0002)		-0.0002 (0.0002)		-0.0006 (0.0007)
R-Square	0.1877	0.2005	0.1819	0.1865	0.1838	0.1883	0.1830	0.1871	0.1869	0.2034
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44
Time Series	754	754	754	754	754	754	754	754	754	754
Observations	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915

Note. We present the results of the third dataset in panel C, for model (8). For the third one, we include all Tweeting devices except Twitter Web Client, Twitter Web app, Twitter for iPhone, Twitter for Android, Twitter for iPad, and TweetDeck. The third dataset includes more bot tweets. Our independent variables are $White\ Users_{i,t}$, $API\ Users_{i,t}$, $Black\ Users_{i,t}$, and $Hispanic\ Users_{i,t}$ indicates the count of social media posts by White, Asian and Pacific Islander (API), Black and Hispanic social media users for stock *i* during time *t*. In all models, the dependent variable is $Excess\ Returns_{i,t+n}$, daily value Fama and French (1992) three-Factor Model excess return in basis points for stock i, on day t varying values of n. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls in our models for Volume, Price, GDSV and five lags of absolute excess returns (AER). The results with control variables are presented in even columns. We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs). *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively

TABLE 9: Panel A

	[t = 0]		[t +1]		[t +2]		[t +3]		[t+4, t+15]	
Variable	**[1]**	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	**[9]**	**[10]**
Male_Users	0.181**	0.145*	-0.065	-0.067	-0.135*	-0.128*	-0.028	-0.029	0.015	-0.062
	(0.084)	(0.081)	(0.056)	(0.055)	(0.073)	(0.072)	(0.056)	(0.056)	(0.145)	(0.143)
Female_Users	-0.049	-0.042	0.032	0.033	0.052*	0.051*	0.001	0.0004	-0.039	-0.032
	(0.033)	(0.031)	(0.026)	(0.025)	(0.03)	(0.029)	(0.022)	(0.022)	(0.061)	(0.06)
Price		0.0022***		-0.002***		-0.0021***		-0.0022***		-0.0242***
		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0019)
Volume		-0.0006		-0.0003		-0.0001		-0.0002		-0.0007
		(0.0011)		(0.0004)		(0.0003)		(0.0003)		(0.0009)
DSVI		0.0004**		0.0004*		-0.0001		-0.0002		-0.0006
		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0007)
R-Square	0.1880	0.2007	0.1822	0.1868	0.1847	0.1891	0.1830	0.1872	0.1860	0.2022
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44
Time Series	754	754	754	754	754	754	754	754	754	754
Observations	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915

Note. We present the results of the first dataset including Tweets from any Tweeting devices (sources) in Panel A, for model 9. Our independent variables are $Male\ Users_{i,t}$, all posts by male users and $Female\ Users_{i,t}$, all posts by female users for Stock *i* on day *t*. In all models, the dependent variable is $Excess\ Returns_{i,t+n}$, daily value Fama and French (1992) three-Factor Model excess return in basis points for stock i, on day t varying values of n. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls in our models for Volume, Price, GDSV and five lags of absolute excess returns (AER). The results with control variables are presented in even columns. We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs). *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively.

TABLE 9: Panel B

	[t = 0]		[t +1]		[t +2]		[t +3]		[t+4, t+15]	
Variable	**[1]**	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	**[9]**	**[10]**
Male_Users	0.2**	0.162**	-0.049	-0.05	-0.13*	-0.122*	-0.025	-0.025	-0.031	-0.096
	(0.079)	(0.077)	(0.054)	(0.053)	(0.069)	(0.068)	(0.053)	(0.053)	(0.141)	(0.139)
Female_Users	-0.068**	-0.059**	0.02	0.019	0.05*	0.047*	-0.002	-0.004	-0.009	-0.01
	(0.029)	(0.028)	(0.023)	(0.023)	(0.027)	(0.026)	(0.02)	(0.02)	(0.057)	(0.055)
Price		0.0022***		-0.002***		-0.0021***		-0.0022***		-0.0242***
		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0019)
Volume		-0.0006		-0.0003		-0.0001		-0.0002		-0.0007
		(0.0011)		(0.0004)		(0.0003)		(0.0003)		(0.0009)
DSVI		0.0004**		0.0004*		-0.0001		-0.0002		-0.0006
		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0007)
R-Square	0.1881	0.2008	0.1821	0.1867	0.1846	0.1891	0.1830	0.1872	0.1860	0.2022
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44
Time Series	754	754	754	754	754	754	754	754	754	754
Observations	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915

Note. We present the results of the second data in panel B, for model 9. we filter Tweets and pick tweets if they come from Twitter Web Client, Twitter Web app, Twitter for iPhone, Twitter for Android, Twitter for iPad, and TweetDeck. Our independent variables are $Male\ Users_{i,t}$, all posts by male users and $Female\ Users_{i,t}$, all posts by female users for Stock *i* on day *t*. In all models, the dependent variable is $Excess\ Returns_{i,t+n}$, daily value Fama and French (1992) three-Factor Model excess return in basis points for stock i, on day t varying values of n. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls in our models for Volume, Price, GDSV and five lags of absolute excess returns (AER). The results with control variables are presented in even columns. We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs). *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively.

TABLE 9: Panel C

	[t = 0]		[t +1]		[t +2]		[t +3]		[t+4, t+15]	
Variable	**[1]**	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	**[9]**	**[10]**
Male_Users	0.051*	0.04	-0.024	-0.025	-0.04	-0.037	-0.014	-0.015	0.059	0.008
	(0.03)	(0.028)	(0.02)	(0.02)	(0.025)	(0.025)	(0.018)	(0.018)	(0.041)	(0.039)
Female_Users	0.044***	0.039***	0.02**	0.022**	-0.002	0.00001	0.002	0.004	-0.052**	-0.045**
	(0.01)	(0.01)	(0.009)	(0.009)	(0.006)	(0.006)	(0.007)	(0.006)	(0.02)	(0.02)
Price		0.0022***		-0.002***		-0.0021***		-0.0022***		-0.0242***
		(0.0005)		(0.0005)		(0.0005)		(0.0005)		(0.0019)
Volume		-0.0007		-0.0003		-0.0001		-0.0002		-0.0008
		(0.0011)		(0.0004)		(0.0003)		(0.0003)		(0.0009)
DSVI		0.0004**		0.0004*		-0.0001		-0.0002		-0.0006
		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0007)
R-Square	0.1891	0.2018	0.1821	0.1868	0.1836	0.1881	0.1828	0.1870	0.1862	0.2021
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44
Time Series	754	754	754	754	754	754	754	754	754	754
Observations	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915

Note. We present the results of the third dataset in panel C, for model (9). For the third one, we include all Tweeting devices except Twitter Web Client, Twitter Web app, Twitter for iPhone, Twitter for Android, Twitter for iPad, and TweetDeck. The third dataset includes more bot tweets. Our independent variables are $Male\ Users_{i,t}$, all posts by male users and $Female\ Users_{i,t}$, all posts by female users for Stock *i* on day *t*. In all models, the dependent variable is $Excess\ Returns_{i,t+n}$, daily value Fama and French (1992) three-Factor Model excess return in basis points for stock i, on day t varying values of n. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls in our models for Volume, Price, GDSV and five lags of absolute excess returns (AER). The results with control variables are presented in even columns. We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs). *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively.

TABLE 10: Panel A

	[t = 0]		[t +1]		[t +2]		[t +3]		[t+4, t+15]	
Variable	**[1]**	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	**[9]**	**[10]**
Age_imbalance	0.0008 (0.0008)	0.0002 (0.0007)	0.0005 (0.0009)	0.0007 (0.0009)	0.0003 (0.0009)	0.0006 (0.0009)	-0.0004 (0.0009)	-0.0002 (0.0008)	-0.0025 (0.003)	-0.001 (0.003)
Price		0.0023*** (0.0006)		-0.0024*** (0.0006)		-0.0027*** (0.0005)		-0.0025*** (0.0005)		-0.0289*** (0.0019)
Volume		-0.0011 (0.0012)		-0.0003 (0.0004)		-0.0002 (0.0003)		-0.0004 (0.0003)		-0.0003 (0.001)
DSVI		0.0003 (0.0002)		0.0004* (0.0002)		-0.0001 (0.0002)		-0.0003 (0.0002)		-0.0009 (0.0007)
R-Square	0.1983	0.2203	0.1992	0.1992	0.20	0.2057	0.2043	0.2093	0.2043	0.2252
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44
Time Series	750	750	750	750	750	750	750	750	750	750
Observations	26,503	26,503	26,503	26,503	26,503	26,503	26,503	26,503	26,503	26,503

Note. We present the results of the first dataset including Tweets from any Tweeting devices (sources) in Panel A, for model 11. Our variable of interest, Age_imbalance$_{i,t}$, create a daily index of the imbalance in activity between young (less than 39 years old) and old (more than 38 years old) social media users for each stock *i* on day *t*. In all models, the dependent variable is $Excess\ Returns_{i,t+n}$, daily value Fama and French (1992) three-Factor Model excess return in basis points for stock i, on day t varying values of n. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls in our models for Volume, Price, GDSV and five lags of absolute excess returns (AER). The results with control variables are presented in even columns. We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs). *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively.

TABLE 10: Panel B

Variable	[t = 0] [1]	[2]	[t +1] [3]	[4]	[t +2] [5]	[6]	[t +3] [7]	[8]	[t+4, t+15] [9]	[10]
Age_imbalance	0.0004	0.0001	-0.0002	-0.0002	0.0015**	0.0015**	-0.0001	-0.0001	0.0007	-0.0001
	(0.0006)	(0.0006)	(0.0009)	(0.0009)	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0025)	(0.0025)
Price		0.0023***		-0.0028***		-0.0031***		-0.0025***		-0.0294***
		(0.0006)		(0.0006)		(0.0006)		(0.0005)		(0.0019)
Volume		-0.0011		-0.0006		-0.0002		-0.0004		-0.0009
		(0.0012)		(0.0004)		(0.0003)		(0.0003)		(0.001)
DSVI		0.0003		0.0005*		-0.0001		-0.0002		-0.0009
		(0.0002)		(0.0002)		(0.0002)		(0.0002)		(0.0007)
R-Square	0.2001	0.2222	0.1951	0.2005	0.2053	0.2114	0.2232	0.2284	0.2158	0.2374
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44
Time Series	750	750	750	750	750	750	750	750	750	750
Observations	26,503	26,503	26,503	26,503	26,503	26,503	26,503	26,503	26,503	26,503

Note. We present the results of the second data in panel B, for model 11. we filter Tweets and pick tweets if they come from Twitter Web Client, Twitter Web app, Twitter for iPhone, Twitter for Android, Twitter for iPad, and TweetDeck.Our variable of interest, $\text{Age\_imbalance}_{i,t}$, create a daily index of the imbalance in activity between young (less than 39 years old) and old (more than 38 years old) social media users for each stock *i* on day *t*. In all models, the dependent variable is $Excess\ Returns_{i,t+n}$, daily value Fama and French (1992) three-Factor Model excess return in basis points for stock i, on day t varying values of n. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls in our models for Volume, Price, GDSV and five lags of absolute excess returns (AER). The results with control variables are presented in even columns. We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs). *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively.

TABLE 10: Panel C

	[t = 0]		[t +1]		[t +2]		[t +3]		[t+4, t+15]	
Variable	**[1]**	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	**[9]**	**[10]**
Age_imbalance	0.0007 (0.0006)	0.0004 (0.0005)	0.0005 (0.0005)	0.0006 (0.0005)	0.0001 (0.0005)	0.0003 (0.0005)	-0.0006 (0.0005)	-0.0004 (0.0005)	-0.002 (0.0018)	0.0001 (0.0018)
Price		0.0025*** (0.0006)		-0.0026*** (0.0006)		-0.0028*** (0.0006)		-0.0025*** (0.0006)		-0.0301*** (0.0019)
Volume		-0.0013 (0.0014)		-0.0005 (0.0004)		-0.0003 (0.0003)		-0.0004 (0.0003)		-0.0015 (0.001)
DSVI		0.0003 (0.0002)		0.0005* (0.0003)		-0.0002 (0.0002)		-0.0004** (0.0002)		-0.0009 (0.0007)
R-Square	0.2035	0.2278	0.2015	0.2065	0.2064	0.2120	0.2141	0.2187	0.2132	0.2373
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44
Time Series	750	750	750	750	750	750	750	750	750	750
Observations	26,503	26,503	26,503	26,503	26,503	26,503	26,503	26,503	26,503	26,503

Note. We present the results of the third dataset in panel C, for model (11). For the third one, we include all Tweeting devices except Twitter Web Client, Twitter Web app, Twitter for iPhone, Twitter for Android, Twitter for iPad, and TweetDeck. The third dataset includes more bot tweets. Our variable of interest, $\text{Age\_imbalance}_{i,t}$, create a daily index of the imbalance in activity between young (less than 39 years old) and old (more than 38 years old) social media users for each stock *i* on day *t*. In all models, the dependent variable is $Excess\ Returns_{i,t+n}$, daily value Fama and French (1992) three-Factor Model excess return in basis points for stock i, on day t varying values of n. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls in our models for Volume, Price, GDSV and five lags of absolute excess returns (AER). The results with control variables are presented in even columns. We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs). *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively.

TABLE 11: Panel A

	[t = 0]		[t +1]		[t +2]		[t +3]		[t+4, t+15]	
Variable	**[1]**	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	**[9]**	**[10]**
GenZ_Users	-0.01 (0.02)	-0.005 (0.02)	0.005 (0.015)	0.008 (0.015)	0.016 (0.016)	0.015 (0.017)	-0.014 (0.018)	-0.015 (0.018)	-0.002 (0.043)	0.014 (0.045)
GenY_Users	0.117** (0.052)	0.106** (0.046)	-0.012 (0.042)	-0.015 (0.041)	-0.043 (0.042)	-0.042 (0.041)	0.008 (0.039)	0.007 (0.039)	0.029 (0.105)	-0.002 (0.101)
GenX_Users	0.116** (0.046)	0.084 (0.056)	0.008 (0.036)	0.008 (0.037)	-0.045 (0.035)	-0.037 (0.034)	-0.014 (0.031)	-0.01 (0.032)	0.127 (0.097)	0.098 (0.097)
Boomers_Users	-0.097*** (0.024)	-0.086*** (0.022)	-0.017 (0.019)	-0.018 (0.019)	0.01 (0.021)	0.007 (0.02)	-0.004 (0.017)	-0.007 (0.017)	-0.162*** (0.053)	-0.181*** (0.052)
Price		0.002*** (0.0005)		-0.002*** (0.0005)		-0.0021*** (0.0005)		-0.0021*** (0.0005)		-0.0242*** (0.0019)
Volume		-0.0009 (0.0011)		-0.0003 (0.0004)		-0.0001 (0.0003)		-0.0001 (0.0003)		-0.0003 (0.001)
DSVI		0.0003** (0.0002)		0.0004* (0.0002)		-0.0001 (0.0002)		-0.0002 (0.0002)		-0.0006 (0.0007)
R-Square	0.1889	0.2015	0.1820	0.1866	0.1841	0.1885	0.1830	0.1872	0.1863	0.2026
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44
Time Series	754	754	754	754	754	754	754	754	754	754
Observations	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915

Note. We present the results of the first dataset including Tweets from any Tweeting devices (sources) in Panel A, for model 12. Our variables of interest, $GenZ\_Users_{i,t}$, $GenY\ Users_{i,t}$, $GenX\ Users_{i,t}$, and $Boomers\ Users_{i,t}$, the count of social media posts by generation Z users (Ages 7-22), generation Y users (Ages 23-38), generation X users (Ages 39-54), and boomers' users (Ages 55-73), for Stock i on day t, respectively. In all models, the dependent variable is $Excess\ Returns_{i,t+n}$, daily value Fama and French (1992) three-Factor Model excess return in basis points for stock i, on day t varying values of n. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls in our models for Volume, Price, GDSV and five lags of absolute excess returns (AER). The results with control variables are presented in even columns. We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs). *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively.

TABLE 11: Panel B

	[t = 0]		[t +1]		[t +2]		[t +3]		[t+4, t+15]	
Variable	**[1]**	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	**[9]**	**[10]**
GenZ_Users	-0.036 (0.023)	-0.032 (0.022)	0.003 (0.015)	0.004 (0.015)	0.017 (0.016)	0.015 (0.016)	-0.011 (0.018)	-0.013 (0.017)	0.01 (0.043)	0.009 (0.045)
GenY_Users	0.143*** (0.048)	0.131*** (0.044)	-0.009 (0.04)	-0.011 (0.04)	-0.051 (0.037)	-0.05 (0.037)	0.003 (0.037)	0.002 (0.037)	0.036 (0.098)	0.011 (0.094)
GenX_Users	0.099* (0.052)	0.073 (0.055)	0.006 (0.042)	0.005 (0.042)	-0.042 (0.041)	-0.035 (0.04)	-0.022 (0.035)	-0.019 (0.036)	0.139 (0.103)	0.118 (0.1)
Boomers_Users	-0.081*** (0.027)	-0.075*** (0.024)	-0.017 (0.02)	-0.017 (0.019)	0.013 (0.024)	0.012 (0.023)	0.003 (0.018)	0.002 (0.018)	-0.183*** (0.054)	-0.196*** (0.052)
Price		0.0021*** (0.0005)		-0.002*** (0.0005)		-0.0021*** (0.0005)		-0.0021*** (0.0005)		-0.0242*** (0.0019)
Volume		-0.0008 (0.0011)		-0.0003 (0.0004)		-0.0001 (0.0003)		-0.0002 (0.0003)		-0.0003 (0.001)
DSVI		0.0003** (0.0002)		0.0004* (0.0002)		-0.0001 (0.0002)		-0.0002 (0.0002)		-0.0006 (0.0007)
R-Square	0.1885	0.2013	0.1820	0.1867	0.1840	0.1885	0.1830	0.1872	0.1865	0.2028
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44
Time Series	754	754	754	754	754	754	754	754	754	754
Observations	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915

Note: We present the results of the second data in panel B, for model 12, we filter Tweets and pick tweets if they come from Twitter Web Client, Twitter Web app, Twitter for iPhone, Twitter for Android, Twitter for iPad, and TweetDeck. Our variables of interest, $GenZ\_Users_{i,t}$, $GenY\ Users_{i,t}$, $GenX\ Users_{i,t}$, and $Boomers\ Users_{i,t}$, the count of social media posts by generation Z users (Ages 7-22), generation Y users (Ages 23-38), generation X users (Ages 39-54), and boomers' users (Ages 55-73), for Stock i on day t, respectively. In all models, the dependent variable is $Excess\ Returns_{i,t+n}$, daily value Fama and French (1992) three-Factor Model excess return in basis points for stock i, on day t varying values of n. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls in our models for Volume, Price, GDSV and five lags of absolute excess returns (AER). The results with control variables are presented in even columns. We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs). *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively.

TABLE 11: Panel C

	[t = 0]		[t +1]		[t +2]		[t +3]		[t+4, t+15]	
Variable	**[1]**	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	**[9]**	**[10]**
GenZ_Users	0.1196*** (0.032)	0.12*** (0.03)	0.009 (0.025)	0.014 (0.025)	-0.006 (0.018)	-0.002 (0.018)	-0.018 (0.019)	-0.01 (0.019)	-0.055 (0.067)	0.015 (0.066)
GenY_Users	0.0203 (0.0203)	0.013 (0.018)	-0.004 (0.017)	-0.005 (0.017)	-0.007 (0.015)	-0.006 (0.015)	0.001 (0.011)	0.001 (0.011)	-0.019 (0.033)	-0.036 (0.032)
GenX_Users	0.0624*** (0.0167)	0.049** (0.023)	-0.0002 (0.013)	0.002 (0.013)	-0.028*** (0.011)	-0.025** (0.011)	-0.002 (0.01)	-0.001 (0.01)	0.015 (0.034)	-0.018 (0.035)
Boomers_Users	-0.0053 (0.01)	-0.003 (0.01)	-0.0004 (0.009)	-0.001 (0.009)	-0.01 (0.009)	-0.012 (0.023)	-0.009 (0.007)	-0.011 (0.007)	0.026 (0.019)	0.012 (0.019)
Price		0.0021*** (0.0005)		-0.002*** (0.0005)		-0.0021*** (0.0005)		-0.0022*** (0.0005)		-0.0242*** (0.0019)
Volume		-0.0012 (0.001)		-0.0004 (0.0004)		0.0001 (0.0004)		-0.0002 (0.0003)		-0.0006 (0.001)
DSVI		0.0003** (0.0002)		0.0004* (0.0002)		-0.0001 (0.0002)		-0.0002 (0.0002)		-0.0006 (0.0007)
R-Square	0.1895	0.2022	0.1818	0.1864	0.1837	0.1882	0.1829	0.1870	0.1861	0.2020
Five lags AER	NO	Yes	NO	Yes	NO	Yes	NO	Yes	NO	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cross Sections	44	44	44	44	44	44	44	44	44	44
Time Series	754	754	754	754	754	754	754	754	754	754
Observations	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915	28,915

Note: We present the results of the third dataset in panel C, for model (12). For the third one, we include all Tweeting devices except Twitter Web Client, Twitter Web app, Twitter for iPhone, Twitter for Android, Twitter for iPad, and TweetDeck. The third dataset includes more bot tweets. Our variables of interest, $GenZ\_Users_{i,t}$, $GenY\ Users_{i,t}$, $GenX\ Users_{i,t}$, and $Boomers\ Users_{i,t}$, the count of social media posts by generation Z users (Ages 7-22), generation Y users (Ages 23-38), generation X users (Ages 39-54), and boomers' users (Ages 55-73), for Stock i on day t, respectively. In all models, the dependent variable is $Excess\ Returns_{i,t+n}$, daily value Fama and French (1992) three-Factor Model excess return in basis points for stock i, on day t varying values of n. We calculate this measure of return over days t = 0, t + 1, t + 2, and t + 3 and the cumulative return over days t + 4 through t + 15. We include controls in our models for Volume, Price, GDSV and five lags of absolute excess returns (AER). The results with control variables are presented in even columns. We define all variables in Table 1. We cluster standard errors across each day and stock and include day and stock-level fixed effects (FEs). *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively.