DECISION SCIENCES
A JOURNAL OF THE DECISION SCIENCES INSTITUTE

Decision Sciences
Volume 48 Number 3
June 2017

Trading on Twitter: Using Social Media Sentiment to Predict Stock Returns

Hong Kee Sul
Wharton Research Data Service, The Wharton School, University of Pennsylvania Philadelphia, PA 19104, e-mail: hongkee@wharton.upenn.edu

Alan R. Dennis
Operations and Decision Technologies Department, Kelley School of Business Indiana University, Bloomington, IN 47405, e-mail: ardennis@indiana.edu

Lingyao (Ivy) Yuan[†]
Department of Supply Chain and Information Systems, College of Business, Iowa State University, Ames, IA 50011, e-mail: lyuan@iastate.edu

ABSTRACT

Decision making is often based on the rational assessment of information, but recent research shows that emotional sentiment also plays an important role, especially for investment decision making. Emotional sentiment about a firm's stock that spreads rapidly through social media is more likely to be incorporated quickly into stock prices (e.g., on the same trading day it was expressed), while sentiment that spreads slowly takes longer to be incorporated into stock prices and thus is more likely to predict stock prices on future days. We analyzed the cumulative sentiment (positive and negative) in 2.5 million Twitter postings about individual S&P 500 firms and compared this to the stock returns of those firms. Our results show that the sentiment in tweets about a specific firm from users with less than 171 followers (the median in our sample) had a significant impact on the stock's returns on the next trading day, the next 10 days, and the next 20 days. Interestingly, sentiment in tweets from users with fewer than 171 followers that were *not* retweeted had the greatest impact on future stock returns. A trading strategy based on these findings produced meaningful economic gains on the order of an 11–15% annual return. [Submitted: December 4, 2014. Revised: April 1, 2016. Accepted: April 12, 2016.]

Subject Areas: Twitter, Emotion, Sentiment, Stock returns, and S&P500.

INTRODUCTION

Almost 75% of adult Internet users use social media, and this percentage is increasing (Pew-Research, 2014). Twitter is one of the most popular social media

[†]Corresponding author.

platforms in the world. Not only has the number of people using social media increased dramatically, so too has the amount of use. In 2015, there were about 300 million Twitter users worldwide, who sent an average of 500 million tweets per day ("About Twitter, Inc.," 2015). Users have integrated social media into many aspects of their daily life (Ellison, 2007), including investment decision making (Oh & Sheng, 2011). Numerous professional and amateur investors and analysts use Twitter to post news articles, and opinions, often providing information and comments more frequently than the professional news media (Sprenger, Tumasjan, Sandner, & Welpe, 2014).

Stock returns, or the profits from trading stocks, are influenced by many factors. Along with fundamental factors and transaction costs, investor sentiment also plays an important role in influencing stock return (Baker & Wurgler, 2007). Market sentiment can be expressed in many ways. The development of social media provides a new meaningful channel for users to share information and their personal feelings. As such, it also serves as a convenient method to capture market sentiment.

Prior research has studied whether the emotional content of tweets can be used to predict stock returns. Bollen, Mao, and Zeng (2011b) assessed the emotional state (calm, alert, sure, vital, kind, and happy) in 10 million tweets that were not related to the stock market. They found that the amount of one state, "calm," was significantly positively correlated with changes in the Dow Jones Industrial average (DJIA) several days later; in other words, when there was a great deal of "calm" in tweets on a given day, the DJIA tended to rise over the following days. Oh and Sheng (2011) examined 200,000 tweets from StockTwits that focused on specific stocks and classified each tweet as "bullish," "bearish," or "neutral" to create a "bullishness" index for each stock. They found the 5-day rolling average of the bullishness index was useful in predicting stock price movements. Sprenger et al. (2014) also used machine learning to create a different bullishness index that they too found to be predictive of stock returns several days later. Smailović, Grčar, Lavrač, and Žnidaršič (2014) used machine learning to examine sentiment (i.e., positive emotion) in tweets and found it to be predictive of stock returns several days later. Risius, Akolk, and Beck (2015) examined emotional states (happiness, affection, satisfaction, fear, anger, depression, contempt) and positive and negative sentiment, and found negative sentiment and "depression" to predict stock returns on the following day.

These findings are promising in suggesting that the emotional state and sentiment in tweets can be used to predict stock returns, but there are still many unanswered questions. Although empirical research has shown that certain emotional states and sentiments in tweets can predict stock price movements, there is a lack of theory to explain why they influence stock returns days later. We argue that the Gradual Information Diffusion model (Hong & Stein, 1999) is useful in understanding how tweets are linked to future stock returns. Under this theoretical perspective, information (in our case, sentiment) influences stock prices as it spreads through the investing public. Sentiment that spreads quickly has an immediate influence on prices, while sentiment that spreads slowly has a slower effect. Sentiment that spreads slowly opens the door for a trading strategy that capitalizes on the stock returns from slowly rising or falling prices.

We analyzed almost 2 years' worth of data collected from Twitter and linked it to the average daily stock returns of firms in the S&P 500. Our results show that the sentiment in tweets about specific firms was significantly related to stock returns on subsequent days. Tweets from individuals with fewer followers had a stronger impact on future returns than tweets from those with many followers, because their tweets took longer to spread. Likewise, tweets that were *not* retweeted took longer to spread and were linked to greater future stock returns.

PRIOR RESEARCH AND THEORY

Information and Stock Return Prediction

Whether stock returns can be predicted has long been a debate. Based on the Efficient Market Hypothesis (EMH), early research argued that stock returns are random and cannot be predicted (Eppen & Fama, 1969; Dockery & Kavussanos, 1996). Research shows that new information, especially news, is a major factor influencing stock returns and quickly leads to stock price changes (Malkiel & Fama, 1970; Hong, Lim, & Stein, 2000; Qian & Rasheed, 2007). Under EMH, positive or negative news (e.g., merger, terrorist attack) is quickly factored into a stock price within minutes, so there is little opportunity to profit, unless of course, one has insider knowledge of an event before it occurs. Mass media outlets play an important role in disseminating information to a broad audience, especially individual investors (Fang & Peress, 2009). This suggests that information contained in social media such as Twitter, which also reaches a broad audience, may be linked to stock returns (Bollen et al., 2011a; Oh & Sheng, 2011; Smailović et al., 2014; Sprenger et al., 2014; Risius et al., 2015).

EMH assumes that information travels quickly and that investors are rational and capable of understanding the full implications of the information they receive (Hong & Stein, 1999). An alternative view is the Gradual Information Flow (GIF) model of Hong and Stein (1999), although the term GIF did not emerge until later (Hong & Stein, 2007). GIF argues that the modern world differs in two important ways from that assumed by EMH. First, some information is private, known only to some investors, and this information diffuses more slowly than public information. Second, investors have cognitive limitations and biases that limit their ability to fully process all implications of the information they receive. In general, investors are either news followers who use fundamental firm information to make investment decisions or momentum traders who use past changes in stock prices to make investment decisions. Both act under bounded rationality, and because they focus primarily on the information relevant to their investing style, they overlook other types of information they receive, and thus prices do not respond to new information as quickly as EMH would predict.

GIF predicts that the speed of information diffusion through the investing public influences how quickly stock prices change in response to new information. Under reasonably efficient markets, information diffuses rapidly among the investing public and is quickly incorporated into stock prices (Hong & Stein, 1999). Conversely, if information diffuses more slowly, it will take longer for that information to be fully incorporated into their prices, and thus there may be opportunities to profit from information before it is fully incorporated into prices (Hong et al., 2000). Under GIF, information should spread rapidly for stocks covered by the mass media but more slowly for stocks not covered by the media. Research shows that stocks not covered by the mass media earn significantly higher future returns than stocks that are covered, after controlling for risk characteristics (Merton, 1987; Fang & Peress, 2009), suggesting that the speed of information diffusion is important in understanding how information may be used to predict future stock prices and thus the returns that can be made by investing.

Information Diffusion in Social Media

Twitter is a social media platform in which users post short text messages of up to 140 characters, called tweets. Anyone can open a Twitter account and begin sending tweets. Users can subscribe to or “follow” other users, and the followers are notified immediately when a user tweets. Many Twitter users have few followers, while commentators, journalists, and celebrities have thousands or more. The median number of followers has gradually increased over time and was about 100 in 2014 (Liu, Kilman-Silver, & Mislove, 2014). Most users follow more people than they have followers; the median number of users followed has gradually increased over time and was around 140 in 2014 (Liu et al., 2014).

During the past several years, Twitter has drawn interests of researchers from multiple disciplines. Current research on Twitter includes several streams. One stream is its impact on information diffusion and supporting communication/collaboration (Honey & Herring, 2009) in many different contexts. Using Twitter during a talk show decreased the psychological distance between the host and his/her audience (Larsson, 2013). In the context of education, Twitter is a potential learning tool in classrooms (Dhir, Buragga, & Boreqqah, 2013). Twitter has become an important tool to spread information during natural disasters and social crises (Sakaki, Okazaki, & Matsuo, 2010; Oh, Agrawal, & Rao, 2013).

Another research stream using Twitter is designing and developing network analysis techniques and algorithms. The abundant data exchanged on Twitter every minute provide researchers, especially those in computer science, the opportunity to observe the social network change. Other related techniques, such as text mining and data mining techniques, also became more refined by studying Twitter data. A third stream is using Twitter to predict individual behavior. Using opinion mining tools and sentiment analysis techniques, researchers are able to predict election results (Tumasjan, Sprenger, Sandner, & Welpe, 2010), hospital-associated mortality (Daley et al., 1988), and heart disease in middle-aged and older persons (Gordon, Castelli, Hjortland, Kannel, & Dawber, 1977).

Due to its popularity, the investment community has adopted Twitter. This community uses the convention of tagging stock-related tweets with a dollar sign ($) followed by the firm’s stock ticker symbol. For example, an individual tweeting about PepsiCo would include $PEP in the tweet. A sample tweet from our data: “$PEP has been strong all day. And who doesn’t love those Frito-Lay snacks? Be honest."

Any Twitter user can send a tweet and include a stock ticker with a dollar sign to indicate that he or she thinks the tweet contains financial information.

Depending upon how many followers that user has, that information may reach a few users, many users, or even tens of thousands of users. Other users can "retweet" the information to their followers so that the information in the original tweet will spread throughout a broad audience of Twitter users—and to non-Twitter users if some users choose to spread the information using other media.

In his seminal work on networks, Barabási (2002) shows just how interconnected we are. Twitter is a directed graph network in that connections are directional. I receive information from users I follow, but they do not receive information from me unless they follow me back. The speed at which information spreads through such a network depends upon how many followers a user has (Barabási, 2002). The general formula for the number of hops it takes to reach any other node in a network is $d = \text{Log}\, N / \log k$, where N is the total number of nodes in the network, and k is the average number of connections per node (i.e., followers) (Barabási, 2002). The number of active traders is on the order of 10 million depending upon how one defines active (trade-IQ, 2011). It is difficult to estimate the average number of followers in this community; links in networks typically follow a power law distribution—not a normal distribution (Barabási, 2002)—and our data were no different, so we use the median of 171. Using these data, we see that it takes about 3.1 hops for information from one node to reach any other node. Of course, Twitter is not the only mechanism through which information is spread. Individual investors can talk with or e-mail other investors. Most people know 200–5,000 people by name (Barabási, 2002), which suggests that we are three to four hops away from anyone else on the planet (Barabási, 2002).

The speed of information diffusion influences whether the information is quickly incorporated in stock prices or takes longer—perhaps days—to be fully disseminated and incorporated into stock prices (Hong & Stein, 1999). Empirical studies of networks show that the number of connections is not randomly distributed (Barabási, 2002). In every network, there are hubs, individuals who have substantially more connections than the average (Barabási, 2002). These hubs are often opinion leaders who facilitate the rapid diffusion of information through the network (Barabási, 2002). Some professional analysts routinely tweet information and thus have a large number of followers; Jim Cramer of CNBC's *Mad Money*, for example, has over 650,000 followers. These individuals are the hubs in the diffusion of investment information, reaching a significant proportion of investors in one hop.

If a Twitter user is a hub (i.e., has many followers), the information he or she tweets will spread more quickly than if the same information is tweeted by a user who has few followers (Barabási, 2002). Tweets from a hub not only reach more people in a single hop, but also tend to be more influential (Barabási, 2002). If a Twitter user has many followers, any information he or she tweets will be quickly disseminated, and stock prices should quickly change to incorporate that information that same day, and there should be little or no effect on stock returns on future days. For example, during the day on July 7, 2011, Twitter account howardlindzon (which has about 200,000 followers) tweeted, "Looks like Howard will be adding more $aapl on a good close. He's predictable." In this tweet of 14 words, two were positive, making it a positive tweet on Apple. Apple had closed the previous day at $351.76, and rose to close at $357.20 that day.

Conversely, if a Twitter user has few followers, information should be slower to disseminate because it will take more hops to reach a critical mass of investors and because it will be less influential than tweets from a hub (Barabási, 2002). Therefore, there is more likely to be a relationship between that information and stock returns on future days because it will take longer for that information to reach many investors and be incorporated into stock prices. For example, during the day on August 8, 2011, several tweets came from multiple twitter accounts with fewer than 100 followers. User ibshakey tweeted, "Still love $AAPL. Continue to love gold. The US in general, not so much." User cronked tweeted, "After that last quarter, how can you not buy $AAPL here? There are some bright spots out there. Not all is lost.," and drewmethey tweeted, "$AAPL I have to buy Apple here. It's just too cheap!!!! Can't resist." Each tweet contained two positive words, making all three a positive tweet on Apple, Inc. From the previous close price of $373.62, the stock price fell to $353.21 at the end of the day. However, on the next day and over the next 10 days, the stock price gradually rose, closing a month later, on September 8, at $384.14.

Sentiment and Contagion

Much of the investment information shared using traditional media and social media is facts and opinions, but individual behavior is not only the outcome of rational decision making. Emotions triggered by these facts and opinions can also influence decisions (Bechara & Damasio, 2005). Twitter provides a good environment to foster the sharing of emotion (Bollen, Pepe, & Mao, 2011b) because the length of each tweet is restricted to 140 characters. The limitation on length of tweets encourages users to be brief and get to the point (Oh & Sheng, 2011). Thus, a short message can provide a focused and more intense trigger for the receiver.

Individual moods, emotions, and other affects are influenced by both internal factors and external factors. Internal factors include personality, individual competency, and so on. External factors include experiences, and information the individual receives. Different affects have different impacts on individuals (Frijda, 1994). Affects can be broad and vague or acute and specific. Affects may have a long-term influence; their effects can also be short term.

Emotion, as one type of affect, has the characteristics of having a clear trigger and a short but more intense effect (Frijda, 1994). Emotion is a subjective feeling related, triggered by a stimulus such as an event, an object, or information in one's environment. Once the stimulus conditions, the stimulus itself, or the supporting cognition, perceptions, or other triggers are no longer active, the emotion will disappear. Emotion can be highly contagious (Schoenewolf, 1990; Hatfield, Cacioppo, & Rapson, 1993).

There are many ways to conceptualize the way emotion is expressed, but two dominant approaches have emerged (Russell, 1980, 2003; Calvo & Kim, 2013). The classic approach, used by Bollen et al. (2011a) and Risius et al. (2015), is to consider specific emotional states, such as joy, anger, sadness, etc. The other approach, used by Smailović et al. (2014), is the dimensional model in which emotional affect is conceptualized as having two dimensions: valence (positive or negative) and arousal (high or low) (Osgood, Suci, & Tannenbaum, 1957; Russell, 1980, 2003; Cacioppo, Petty, Losch, & Kim, 1986); some authors also include a third dimension of dominance (Bradley & Lang, 1994). Neither approach is more or less "correct" (Calvo & Kim, 2013), and it is straightforward to map emotional states onto the dimensional model (Russell, 1980, 2003; Bradley & Lang, 1994). For example, the "calm" emotional state studied by Bollen et al. would be considered neutral valence and low arousal (Russell, 2003).

Both models are commonly used, although Calvo and Kim (2013) conclude that researchers in natural language processing are more likely to use the emotional states model, while researchers in psychology are more likely to use the dimensional model. When natural language processing researchers use the dimensional model, they commonly focus on only the valence dimension, which they term "sentiment" (e.g., Wiebe, Wilson, & Cardie, 2005; Abbasi & Chen, 2008; Oh & Sheng, 2011; Smailović et al., 2014). In this study, we focus on sentiment, as has been common in financial research (e.g., Tetlock, 2007).

Sentiment affects decision making (Bakamitsos, 2006). According to Construal Level Theory (CLT), positive and negative sentiment may have different effects (Liberman & Trope, 1998; Bar-Anan, Liberman, & Trope, 2006; Fujita, Trope, Liberman, & Levin-Sagi, 2006). Positive sentiment increases abstract construal, that is, the adoption of abstract, future goals, while negative sentiment triggers a focus on immediate and proximal concerns and reduces the adoption of abstract future goals (Liberman & Trope, 1998; Eyal, Liberman, Trope, & Walther, 2004; Bar-Anan et al., 2006; Fujita et al., 2006; Labroo & Patrick, 2009). Positive sentiment is more likely to induce individuals to make a decision than negative sentiment, which tends to slow the decision process (Qiu & Yeung, 2008). Positive sentiment also may induce an individual to act on a decision (Frijda, 1994). Positive sentiment can increase consumers' impulse to buy in the context of electronic commerce (Parboteeah, Valacich, & Wells, 2009) but increase an individual's resistance to temptation in other contexts (Fedorikhin & Patrick, 2010). Thus, positive sentiment and negative sentiment are more than the opposite ends of the same dimension; they can trigger different behaviors. For this reason, researchers often have measured them separately to capture their full effects (e.g., Tetlock, Saar-Tsechanksky, & Macskassy, 2008; Risius et al., 2015).

Sentiment is contagious (Schoenewolf, 1990; Hatfield et al., 1993). Social contagion is "the tendency to automatically mimic and synchronize expressions, vocalizations, postures, and movements with those of another person's and to converge emotionally" (Hatfield et al., 1993). Contagion happens implicitly and explicitly (Singer & Lamm, 2009). Sentiment is expressed through facial impressions, physical gestures, vocal tones, and written words. When individuals exchange messages via written text, photos, audio, or even video on social media, the message sender's sentiment also is exchanged. Thus, tweets and the sentiment they contain have the potential to influence the receiver's behavior (Risius et al., 2015).

Hypotheses

We argue that the sentiment contained in social media tweets will have a direct effect on stock returns in a manner similar to the effects that professional news media have on stock returns. Positive sentiment should be associated with positive returns and negative sentiment should be associated with negative returns (Tetlock et al., 2008).

The effects of sentiment spread in the same manner in which information spreads through the network. Thus, the speed of diffusion is important. If a tweet about a specific firm is sent by a hub (a user who has many followers), the sentiment it contains will spread faster than the sentiment sent by a user who has few followers because more individuals will see it immediately, and it will be more influential (Barabási, 2002). Sentiment that is spread more quickly will be incorporated into prices faster, so that it will have an effect on returns sooner (Hong & Stein, 1999; Tetlock et al., 2008). Its effects are more likely to be seen on the same trading day on which it was tweeted. Thus, it will have little effect on stock returns on future days, because its effects are immediately incorporated into prices (Hong & Stein, 1999; Tetlock et al., 2008).

In contrast, sentiment contained in tweets from those with fewer followers will take longer to disseminate because fewer people will see them on the first hop, and it will be less influential (Barabási, 2002). Thus, the number of followers affects the speed of sentiment diffusion. This less visible sentiment from users with few followers is spread more slowly and will take longer to affect stock prices (Hong & Stein, 1999). Therefore, it will have a larger effect on stock returns on future trading days (Hong & Stein, 1999; Tetlock et al., 2008). Thus:

> *H1: The sentiment in tweets about a specific firm sent by individuals with few followers is directly related to stock returns on future trading days.*

This sentiment diffusion process will also be affected by the extent to which tweets are "retweeted"—that is, whether an individual who receives a tweet resends it to his or her followers. A study of 37 billion public tweets found that the percentage of retweets has increased over time: about 5% in 2010, 10% in 2011, 20% in 2012, and 25% in 2013 (Liu et al., 2014).

Individuals retweet for a variety of reasons. The most common reasons are because they believe the tweet's information would be of interest to their followers or to express support for the original tweeter (Macskassy & Michelson, 2011; Liu et al., 2014). In the investing context where tweets are deliberately tagged with the $ and ticker symbol, we theorize that most tweets are retweeted because the sender believes they have potential information for other investors.

Retweets affect the diffusion process. Retweeting someone else's tweet is a deliberate signal that the user believes the tweet would be of interest to his or her followers. Retweeting spreads the sentiment in the tweet faster than if the tweet was not retweeted and makes the tweet more influential because now two people advocate for its content, not one. Sentiment in tweets that are retweeted will be more quickly incorporated in stock price, so it will have *less* of an effect on returns on future trading days. Therefore, it is the combination of few followers and *not* being retweeted that leads to the greatest stock returns on future days because the sentiment in these tweets will take the longest to diffuse through the network. Therefore, we hypothesize:

H2: The sentiment in tweets about a specific firm sent by individuals with few followers that are not retweeted is directly related to stock returns on future trading days.

METHODOLOGY

Financial Data

To ensure sufficient reliability of Twitter data, we focused only on firms that are part of the S&P 500. Financial data, including the closing price of each stock in the S&P 500, were obtained from Compustat, Center for Research in Security Prices (CRSP), Institutional Brokers' Estimate System (IBES), and Kenneth French's Web site (Rai, Patnayakuni, & Seth, 2006). The sample period is from March 2011 to January 2013.

Twitter Data

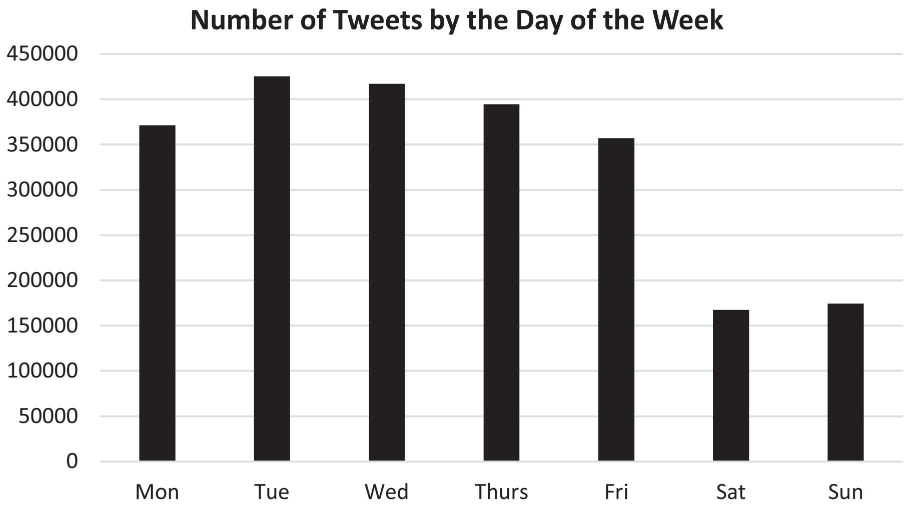
This study used data collected from Twitter. The focus of this article is on whether the sentiment in tweets about an individual firm can predict stock returns. Thus, it is important to match tweets to specific firms. The convention in Twitter is to precede the stock ticker symbol with a dollar sign ($) to indicate that a tweet contains investment information about a firm. We collected all public tweets that contained the relevant $ symbol with an S&P 500 stock ticker from Twitter using a developer account. We retrieved 3,475,428 tweets during the sample time period. Of all the tweets, 16.02% were retweets. We excluded all the tweets that contained more than one ticker symbol because we could not be sure if the information in the tweet pertained to one firm or all firms equally. For example, a tweet like "I also like long $AAPL @347.40 ... and short $RIMM @62.70" would be excluded from the analysis. This produced a final sample of 2,503,385 tweets. An inspection of 500 randomly selected tweets found no tweets from the firm itself. Figure 1 shows the distribution of the tweets by days of the week.

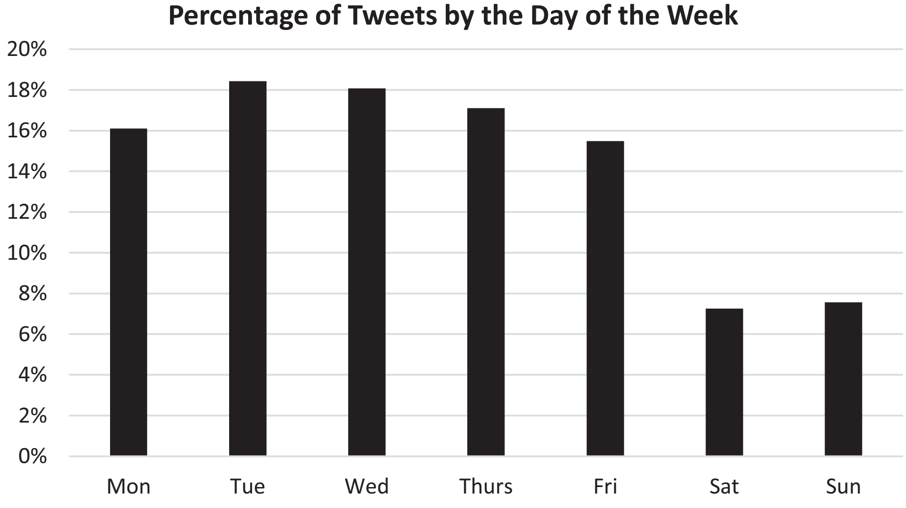
The Sentiment in Tweets

There are many approaches to sentiment analysis (Feldman, 2013). We used the word analysis strategy. Each word in a tweet was matched to a dictionary of terms to determine its sentiment. We used the Harvard-IV dictionary (Jorgenson & Vu, 2005), which is a commonly used source for word classification in the financial content analysis of popular press articles and Web news sites, used, for example, by Tetlock (2007), Tetlock et al. (2008), and Da, Engelberg, and Gao (2011). There are other dictionaries that could be used (e.g., the financial dictionary of Loughran and McDonald (2011)), but these dictionaries are designed for the analysis of legal and financial documents which contain formal English (e.g., 10K filings), not the slang version of English used in Twitter.

CLT argues that positive and negative sentiment may have different effects (Liberman & Trope, 1998; Bar-Anan et al., 2006; Fujita et al., 2006), so it is important to track both positive and negative sentiment because they may have different effects. Empirical research on stock returns has shown that sometimes

Figure 1: Number of tweets (top) and percentage distribution (bottom) by the day of the week.





positive sentiment has an impact (Smailović et al., 2014), while other times negative sentiment has an impact (Risius et al., 2015).

We counted all words in the tweets that had the "NEG" tag in the Harvard-IV dictionary as words that conveyed a negative sentiment. We counted "POS" tagged words as words conveying a positive sentiment. Although this approach has been widely used in prior research (Tetlock et al., 2008), it is an imperfect measure of sentiment, because it cannot detect subtle meanings in English, such as sarcasm or semantic word groups that combine positive and negative words (e.g., "not bad":

Xie et al., 2015). Likewise, we did not include emoticons in our analysis, so this is another limitation (Xie et al., 2015). An analysis of 100,000 randomly selected tweets from our sample found 430 to contain emoticons (i.e., less than 1%).

We used three separate measures to better model sentiment, as has been done in prior research (Tetlock et al., 2008). If our measures do not accurately capture sentiment, then we are less likely to find a significant relationship, so this approach is a more conservative test of the relationship between sentiment and stock returns than human analysis of the tweets, which would be effectively impossible given our sample size of over 2.5 million tweets. Table 1 shows descriptive statistics about the tweets.

Because we are using daily stock returns as our dependent variable, we combined all tweets for each firm on a given day. Daily returns are defined as close-to-close daily returns, so we match day t return with firm level Twitter content on day t up to the market close time of 4 p.m. New York's time. Any tweet that was posted after 4 p.m. was treated as day $t + 1$. Following Tetlock et al. (2008), we used three variables to measure sentiment. Sentiment is measured as following, where P, N, and T are the daily aggregate number of positive, negative, and total words for each day for a given firm.

$$\text{Sentiment} = \begin{cases} \text{neg1} \equiv \frac{N}{T} \\ \text{pos1} \equiv \frac{P-N}{P+N} \\ \text{pos2} \equiv \log\left(\frac{1+P}{1+N}\right) \end{cases}. \tag{1}$$

Conceptually, neg is the ratio of the amount of negative sentiment to the total communication (positive, negative, and neither). Pos1 is a normalized ratio (on a –1 to +1 scale) of the overall positive or negative sentiment expressed (omitting words with no sentiment). Pos2 is an unstandardized ratio of positive to negative sentiment, but log adjusted to capture the potential for diminishing marginal effects. All three measures may produce similar results, but we included all three for greater insight. Descriptive statistics can be found in Table 2.

Analysis

To answer the question of whether social media have sentiment information that can predict future returns, we examine whether the speed of information dissemination (i) reflected by the number of followers and (ii) retweet history is associated with future returns.

To answer the first research question, we test the following equation:

$$CAR^i_{t,\ t+n} = \alpha + \beta_0 \text{sentiment}_{u_t}^{\ i} + \beta_1 \text{sentiment}_{o_t}^{\ i} + \gamma CV^i_t + \epsilon^i_t, \tag{2}$$

where $CAR^i_{t,t+n}$ is the cumulative abnormal return about firm i from day t +1 to day t +n; $\text{sentiment}_{u_t}^{\ i}$ is the sentiment about firm i on day t expressed in tweets from users with a number of followers at or under a given threshold; $\text{sentiment}_{o_t}^{\ i}$ is the sentiment about firm i on day t expressed in tweets from users with a number of followers over a given threshold; and CV are five control variables, as described below.

Table 1: Descriptive statistics of tweets.

	March 2011	April 2011	May 2011	June 2011	July 2011	August 2011	September 2011	October 2011	November 2011	December 2011	January 2012	February 2012
Total stock tickers	497	497	498	497	498	498	498	498	498	498	498	497
Total number of tweets	126,926	216,303	197,978	160,119	189,655	244,763	205,893	215,702	199,471	171,531	190,144	189,079
Average number of tweets per firm	255.38	435.22	397.55	322.17	380.83	491.49	413.44	433.14	400.54	344.44	381.82	380.44
Average number of words per tweet for all firms	15.299	15.615	15.437	15.358	15.161	15.130	14.804	14.953	14.662	15.267	15.160	14.735
Average number of positive words per tweet for all firms	0.400	0.433	0.401	0.359	0.261	0.331	0.289	0.322	0.309	0.322	0.349	0.336
Average number of negative words per tweet for all firms	0.245	0.264	0.247	0.223	0.254	0.220	0.204	0.217	0.195	0.203	0.229	0.207
Average percentage of positive words for all firms	0.026	0.027	0.026	0.023	0.023	0.022	0.019	0.021	0.021	0.021	0.023	0.023
Average percentage of negative words for all firms	0.016	0.017	0.016	0.014	0.016	0.014	0.014	0.014	0.013	0.013	0.015	0.013
SD % positive words for all firms	0.012	0.013	0.011	0.012	0.012	0.012	0.013	0.012	0.012	0.011	0.012	0.011
SD % negative words for all firms	0.009	0.010	0.010	0.008	0.010	0.008	0.009	0.008	0.009	0.008	0.008	0.008

Table 2: Descriptive statistics of the firm/trading day data.

	N	Mean	Variance	*SD*	Min	25th Percentile	50th Percentile	75th Percentile	Max
Pos1 Sentiment $_{i,t}$	119,727	0.254	0.474	0.688	–1.000	–0.111	0.333	1.000	1.000
Pos 2 Sentiment $_{i,t}$	119,727	0.347	0.704	0.839	–4.500	–0.182	0.452	0.860	5.226
Neg Sentiment $_{i,t}$	119,727	0.040	0.002	0.042	0.000	0.000	0.035	0.060	0.667
Surprise $_{i,t}$	119,727	0.001	0.001	0.031	–2.080	0.000	0.000	0.000	3.950
Control $2_{i,t\text{-}30,t\text{-}2}$	112,938	–0.005	0.008	0.092	–1.010	–0.052	0.000	0.047	0.652
Control 1 $_{i,t\text{-}1}$	112,858	0.000	0.000	0.019	–0.673	–0.008	0.000	0.008	0.532
Upgrade $_{i,t}$	119,727	0.023	0.029	0.171	0.000	0.000	0.000	0.000	9
Downgrade $_{i,t}$	119,727	0.021	0.023	0.151	0.000	0.000	0.000	0.000	5

Notes: Surprise i,t: Earnings surprise, relative to median analyst estimate.
Control 1 i,t-30,t-2: Past returns, cumulative abnormal return from the [–30, –2] trading window.
Control 2 i,t-1: The abnormal return on the prior trading day.
Upgrade i,t: The number of financial analyst upgrades for company i on day t.
Downgrade i,t: The number of financial analyst downgrades for company i on day t.

Table 3: Definitions of variables.

AR_t^i	The AR_t^i is the abnormal return of firm i on date t, adjusted using the size and book-to-market matched characteristic portfolio's return. $$AR_t^i = R_t^i - Pfo_t^i,$$ where $R_t^i = \ln(r_t^i + 1)$, and $Pfo_t^i = \frac{1}{n} \sum_{j=1}^{n} w_j(R_t^j)$. Note that r_t^i is the daily holding period return (ret) in the CRSP daily stock database (CRSP.DSF) including and w_j is the value weight of the jth firm in the portfolio of firms $j = 1 \sim n$, such that $\sum_{j=1}^{n} w_j = 1$. The size and book-to-market characteristic portfolio was formed using the 30th and 70th NYSE book-to-market percentiles and the median NYSE market equity.
$CAR_{t,\ t+n}^i$	The $CAR_{t,\ t+n}^i$ is future cumulative abnormal return of firm i on date t, the dependent variable in our regressions. It is the summation of the abnormal returns of the next n days starting day $t+1$. $$CAR_{t,t+n}^i = \sum_{j=1}^{n}\left(AR_{t+j}^i\right)$$
$Control1_{i,t}$	The $Control1_{i,t}$ is the abnormal return of firm i for the date $t-1$. $Control1_{i,t} = AR_{t-1}^i$
$Control2_{i,t}$	The $Control2_{i,t}$ is the cumulative abnormal return, or the summation of the abnormal return, of firm i for the previous 30 days excluding the abnormal return of the previous date. $$Control2_{i,t} = \sum_{j=2}^{30}\left(AR_{t-j}^i\right)$$
Earnings Surprise$_{i,t}$	An earnings surprise is calculated for each firm i on each earnings announcement date. The Earnings Surprise$_{i,t}$ is calculated as the difference between the actual EPS (actual) and the median EPS (medest) from the IBES summary statistics database (IBES.STATSUM).
$Upgrade_{i,t}$ $Downgrade_{i,t}$	An upgrade/downgrade is recorded as 1 if an analyst increased/decreased the IBES recommendation code (ireccd) from IBES recommendation detail database (IBES.RECDDET). The $Upgrade_{i,t}$, $Downgrade_{i,t}$ is the summation of the number of upgrade/downgrade for all analysts for firm i on the same date.

Equation (2) examines future abnormal returns for days 1 to n after the tweets were made. Table 3 shows how we calculated $CAR_{t,t+n}^i$. We have chosen to use three time periods: next day returns (i.e., $n = 1$), next-day-to-10th-day returns (i.e., $n = 10$), and next-day-to-20th-day returns for a longer view (i.e., $n = 20$). These are trading days, so 10 days is approximately 2 weeks, and 20 days is approximately one month. These time periods are consistent with prior research (e.g., Tetlock et al., 2008; Fang & Peress, 2009; Chen, De, Hu, & Hwang, 2014).

Table 4a: Correlations among the sentiment variables used in H1.

	Pos1 Over	Pos2 Over	Neg Over	Pos1 Under	Pos2 Under
Pos2 Over	.905				
Neg Over	–.800	–.741			
Pos1 Under	.253	.253	–.193		
Pos2 Under	.242	.276	–.196	.905	
Neg Under	–.191	–.193	.199	–.811	–.766

Notes: All correlations are significant at $p < .001$. Cells in gray are correlations among variables used in the same regression analysis.

We used five control variables. Stock returns exhibit autocorrelation so we included two control variables to control for autocorrelation: control1 and control2 capture the abnormal returns on the day before (i.e., $t - 1$) and the cumulative return over the prior 30 days, respectively (Tetlock et al., 2008; Chen et al., 2014) (Table 3). The third control variable is earnings surprise, calculated as the actual earnings per share for a given firm announced on a given day minus the median analyst earnings per share prediction, where the median analyst prediction is the "Median Estimate" from IBES Summary. The last two control variables were the upgrades and downgrades on the company from professional stock analysts as control variables because upgrades and downgrades can influence stock returns (Chen et al., 2014). We counted the number of upgrades and downgrades on the specific firm's stock on the same trading day as the tweets and included these numbers as controls. As is commonly done in financial research, we obtained analyst recommendations from the IBES, categorized each change in recommendation as either an upgrade or downgrade, and counted the number of each on each trading day (Chen et al., 2014) (Table 3).

To examine H1, the impact of the number of followers, we need to test whether β_0 is positive when sentiment is pos1 or pos2 and test whether β_0 is negative when sentiment is neg for the different trading periods. We split the tweets into two groups based on the number of followers of the tweeters, those with many followers and those with few followers. The question is, what is "many" and "few?" The median number of followers in our sample was 171, so we selected this as the break point for assigning tweets into groups of users with few followers and many followers. The sample size was 48,538 because we can analyze the data only when there are tweets from individuals both over and under the threshold on the same day for the same firm.

To examine H2, the combined impact of number of followers and retweets, we divided the tweets into four groups: many followers and retweeted; many followers and not retweeted; few followers and retweeted; and few followers and not retweeted. We used the same break point (171) as the threshold for assigning tweets into groups with few followers and many followers. The sample size was 8,245 because we can analyze data only when there are tweets in all four groups on the same day for the same firm.

Tables 4a and (b) shows the correlations among the sentiment variables. If the sentiment in tweets from those with few and many followers are highly correlated, multicollinearity could bias the results. The correlations indicate little risk due to

Table 4b: Correlations among the sentiment variables used in H2.

	1	2	3	4	5	6	7	8	9	10	11
1. Pos1 over not retweeted											
2. Pos1 over retweeted	**.189**										
3. Pos1 under not retweeted	**.103**	**.067**									
4. Pos1 under retweeted	**.078**	**.209**	**.102**								
5. Pos2 over not retweeted	**.927**	**.235**	**.125**	**.117**							
6. Pos2 over retweeted	**.174**	**.910**	**.072**	**.265**	**.239**						
7. Pos2 under not retweeted	**.096**	**.077**	**.915**	**.115**	**.131**	**.091**					
8. Pos2 under retweeted	**.074**	**.232**	**.101**	**.906**	**.121**	**.337**	**.123**				
9. Neg over not retweeted	**–.793**	**–.151**	**–.085**	**–.063**	**–.750**	**–.141**	**–.080**	**–.062**			
10. Neg over retweeted	**–.145**	**–.436**	**–.007**	**–.074**	**–.142**	**–.405**	**.021**	**–.087**	**.151**		
11. Neg under not retweeted	**–.069**	–.003	**–.495**	–.001	**–.050**	.002	**–.436**	**–.002**	**.066**	**.144**	
12. Neg under retweeted	**–.067**	**–.073**	.010	**–.411**	**–.059**	**–.090**	**.041**	**–.383**	**.065**	**.304**	**.199**

Note: All correlations shown in **bold** are significant at $p < .001$. Cells in gray are correlations among variables used in the same regression analysis.

multicollinearity, with the highest correlation between variables in the same model being less than .30. We included Variance Inflation Factors in all analyses and found that most were less than 1.2, and none exceeded 2.0, indicating that it is highly unlikely that our data suffer from multicollinearity.

One of the issues with large data sets is that the traditional approach of using *p* values can be misleading because the large sample size means that any relationship is likely to be significant (Lin, Lucas, & Shmueli, 2013). Lin et al. (2013) offer several strategies for the analysis of large data sets. We adopt three of their recommendations, plus a fourth traditionally used in the analysis of large sample stock return data. First, we present confidence intervals for the size of effects. Second, we conduct a series of robustness checks using alternate models to see the extent to which our results are dependent on the specific models we use. Third, we examine the predictive ability of the models by comparing them to a controls-only model using symmetric mean absolute percent error (SMAPE), the absolute value of the difference between the predicted and actual divided by the mean of the absolute value of the predicted and the absolute value of the actual (Armstrong, 1985; Makridakis, 1993; Tofallis, 2015). Finally, one of the strongest tests of the practical significance of models used to predict stock returns is a trading strategy analysis—a test of whether an investor who builds a trading strategy using the results would experience a profit after accounting for trading costs (Tetlock et al., 2008).

RESULTS

Impact of Followers

We begin with H1, which argues that future abnormal returns would be directly related to the sentiment in tweets from users with few followers. Table 5 shows that the beta coefficients are significant and in the hypothesized direction for all three measures (pos1, pos2, and neg) for all three time periods (next day, next-to-10th-day, and next-to-20th-day), except for next day returns for neg, which is in the hypothesized direction but not significant.

The adjusted R^2 for these analyses are in a similar range to those in other studies of stock returns, such as (Chordia, Roll, & Subrahmanyam, 2002; Tetlock, 2007; Tetlock et al., 2008; Bollen et al., 2011a; Chen et al., 2014). SMAPE values for the models with only the five control variables for next day returns, next-to-10th-day, and next-to-20th-day are 1.9231, 1.8685, and 1.8210, respectively. The SMAPE values for all nine models in Table 5 are below the SMAPE values for the controls-only models, indicating they are a better fit. Based on these results (8 of 9 hypothesis tests significant, R^2 equivalent to R^2 in prior research, and lower SMAPE), we conclude that H1 is supported.

Combined Impact of Followers and Retweets

H2 argues that future abnormal returns would be directly related to the sentiment in tweets from users with few followers that are not retweeted. Table 6 shows that the beta coefficients on the sentiment in tweets from those with few followers that were not retweeted are significant and in the hypothesized direction for all

Table 5: Regression results of emotional sentiment on abnormal returns by number of followers.

(a) Pos1	Next Day	Next-to-10th-Day	Next-to-20th-Day
Sentiment_Over $_{i,t}$	–0.140	1.020*	1.103
Sentiment_Under $_{i,t}$	0.388**	1.512***	2.582***
Control 1 $_{i,t\text{-}1}$	0.004	0.017	–0.011
Control 2$_{i,t\text{-}30,t\text{-}2}$	0.001	0.013***	0.022***
Surprise $_{i,t}$	0.010***	0.004	0.001
Upgrade $_{i,t}$	0.002***	0.000	–0.001
Downgrade $_{i,t}$	–0.002***	0.000	–0.003
Intercept	0.000***	–0.003***	–0.006***
SMAPE	1.9208	1.8662	1.8200
Adj. R^2	0.001	0.001	0.001

(b) Pos2	Next Day	Next-to-10th-Day	Next-to-20th-Day
Sentiment_Over $_{i,t}$	–0.105	0.618	0.133
Sentiment_Under $_{i,t}$	0.349**	1.405***	2.674***
Control 1 $_{i,t\text{-}1}$	0.004	0.017	–0.010
Control 2$_{i,t\text{-}30,t\text{-}2}$	0.001	0.012***	0.022***
Surprise $_{i,t}$	0.010***	0.004	0.001
Upgrade $_{i,t}$	0.002***	0.000	–0.001
Downgrade $_{i,t}$	–0.002***	0.000	–0.003
Intercept	0.000***	–0.004***	–0.007***
SMAPE	1.9190	1.8648	1.8187
Adj. R^2	0.002	0.001	0.002

(c) Neg	Next Day	Next-to-10th-Day	Next-to-20th-Day
Sentiment_Over $_{i,t}$	2.313	–13.140	–6.829
Sentiment_Under $_{i,t}$	–3.265	–22.735***	–52.175***
Control 1 $_{i,t\text{-}1}$	0.004	0.018	–0.010
Control 2$_{i,t\text{-}30,t\text{-}2}$	0.001*	0.013***	0.022***
Surprise $_{i,t}$	0.010***	0.004	0.001
Upgrade $_{i,t}$	0.002***	0.000	–0.001
Downgrade $_{i,t}$	–0.002***	0.000	–0.003
Intercept	0.000***	–0.001**	–0.003***
SMAPE	1.8200	1.8187	1.8201
Adj. R^2	0.001	0.001	0.002

Notes: The coefficients are multiplied by 1,000.
*$p \leq .05$, ** $p \leq .01$, *** $p \leq .001$.

three measures (pos1, pos2, and neg) for all three time periods (next day, next-to-10th-day, and next-to-20th-day), except for next-to-10th day returns for neg which is in the hypothesized direction but not significant. Table 7 presents confidence intervals for the betas.

The adjusted R^2 for these analyses are equivalent to or substantially higher (by an order of magnitude—i.e., 1,000%) than adjusted R^2 in prior studies (Chordia

Table 6: Regression results of emotional sentiment and retweeting on abnormal returns by number of followers.

(a) Pos1	Next Day	Next-to-10th-Day	Next-to-20th-Day
Sentiment_OverNo $_{i,t}$	–0.531	2.043	1.871
Sentiment_OverRe $_{i,t}$	0.162	–0.407	–1.000
Sentiment_UnderNo $_{i,t}$	0.917*	3.598**	5.850**
Sentiment_UnderRe $_{i,t}$	0.468	–0.037	–0.122
Control 1 $_{i,t\text{-}1}$	0.020*	0.006*	0.036
Control 2$_{i,t\text{-}30,t\text{-}2}$	0.003	0.043***	0.082***
Surprise $_{i,t}$	0.008**	0.000	0.007
Upgrade $_{i,t}$	0.002**	–0.002	–0.005
Downgrade $_{i,t}$	–0.001	–0.005*	0.004
Intercept	–0.001***	–0.005***	–0.009***
SMAPE	0.2219	0.2148	0.2172
Adj. R^2	0.002	0.008	0.013

(b) Pos2	Next Day	Next-to-10th-Day	Next-to-20th-Day
Sentiment_OverNo $_{i,t}$	–0.126	0.778	–0.116
Sentiment_OverRe $_{i,t}$	–0.327	–0.412	–1.214
Sentiment_UnderNo $_{i,t}$	0.831**	3.353***	4.649***
Sentiment_UnderRe $_{i,t}$	0.522	0.221	0.901
Control 1 $_{i,t\text{-}1}$	0.020*	0.064	0.036
Control 2$_{i,t\text{-}30,t\text{-}2}$	0.004	0.043***	0.008***
Surprise $_{i,t}$	0.008*	0.000	0.007
Upgrade $_{i,t}$	0.002	–0.002	–0.005
Downgrade $_{i,t}$	–0.001	0.005	–0.004
Intercept	–0.001***	–0.005***	–0.009***
SMAPE	0.2217	0.2146	0.2172
Adj. R^2	0.003	0.009	0.013

(c) Neg	Next Day	Next-to-10th-Day	Next-to-20th-Day
Sentiment_OverNo $_{i,t}$	3.008	–57.433	–79.755
Sentiment_OverRe $_{i,t}$	8.017	10.249	25.681
Sentiment_UnderNo $_{i,t}$	–14.274*	–40.149	–67.144*
Sentiment_UnderRe $_{i,t}$	–7.962	3.638	2.038
Control 1 $_{i,t\text{-}1}$	0.020*	0.064*	0.035
Control 2$_{i,t\text{-}30,t\text{-}2}$	0.003	0.004***	0.082***
Surprise $_{i,t}$	0.009*	0.000	0.007
Upgrade $_{i,t}$	0.002*	–0.002	–0.005
Downgrade $_{i,t}$	–0.001	0.005*	0.004
Intercept	0.000	0.000	–0.003
SMAPE	0.2219	0.2148	0.2172
Adj. R^2	0.002	0.008	0.012

Notes: The coefficients are multiplied by 1,000.
*$p \leq .05$, ** $p \leq .01$, *** $p \leq .001$.

Table 7: Confidence intervals for beta for sentiment in non-retweeted tweets from individuals with 171 or fewer followers.

(a) Pos1	Next Day	Next-to-10th-Day	Next-to-20th-Day
$Sentiment\_UnderNo_{i,t}$	0.053 to 1.781	0.946 to 6.250	2.048 to 9.652
(b) Pos2	Next Day	Next-to-10th-Day	Next-to-20th-Day
$Sentiment\_UnderNo_{i,t}$	0.213 to 1.449	1.459 to 5.247	1.932 to 7.366
(c) Neg	Next Day	Next-to-10th-Day	Next-to-20th-Day
$Sentiment\_UnderNo_{i,t}$	–0.005 to –28.543	3.635 to –83.933	–4.361 to –129.927

et al., 2002; Tetlock, 2007; Tetlock et al., 2008; Bollen et al., 2011a; Chen et al., 2014). SMAPE values for the models with only the five control variables for next day returns, next-to-10th-day, and next-to-20th-day are 1.8476, 1.7763, and 1.7042, respectively. The SMAPE values for all nine models in Table 6 are below the SMAPE values for their matching controls-only models, indicating they are a better fit. Based on these results (eight of nine hypothesis tests significant, R^2 equivalent to or higher than R^2 in prior research, and lower SMAPE), we conclude that H2 is supported.

These results support our arguments that the speed of sentiment diffusion affects future returns. When sentiment spreads the slowest (i.e., tweets sent by those with fewer than the median number of followers [171] that are not retweeted), it affects stock returns on future trading days.

Robustness Checks

We conducted several robustness checks. We ran separate analyses for H1 and H2 treating missing values as zero emotion (which produced sample sizes of 83,891) and found the same pattern of results.

We conducted a separate analysis for H1 using a different split in the number of followers. We used 1,000 followers as the threshold between many and few. A user with 1,000 followers falls in the top 4% of all Twitter users ("About Twitter, Inc.," 2014), so they are what Barabási (2002) would call hubs. The Twitter users with over 1,000 followers were typically well-known media or analysts, such as CNN, ABC, WSJ, CNBC, Fox News, *Fortune Magazine*, and Jim Cramer, who is a writer, TV show host, and co-founder of *TheStreet.com*. This produced the analysis of "few" and "many" followers, with a 96–4% split. The split using 1,000 followers followed the same pattern as with the median split (Table A1 in the Appendix). We conclude that this hypothesis is robust to the choice of threshold for the number of followers. For this hypothesis, as long as one is not a hub (i.e., the top 4% of all Twitter users), it takes days before sentiment in your tweets spreads.

We conducted a similar analysis for H2 using the 1,000 follower threshold (96–4% split). The sample size here was 9,014. The pattern here was different; only three of the nine hypothesis tests were significant (Table A2 in the Appendix). We conclude that this hypothesis is not robust to the choice of threshold for what "few" followers means. If we consider retweeting behavior, then the threshold

Table 8: Annualized returns from a trading strategy using sentiment in non-retweeted tweets from individuals with 171 or fewer followers.

	Holding Period		
	1 Day	10 Days	20 Days
Without trading costs	11.44%	17.91%	12.59%
With trading costs	–28.57%	15.65%	11.41%

number to identify "few" followers must be such that we do not consider only hubs and nonhubs in the network.

We examined the effects of the sentiment contained in tweets only from those with many followers, omitting the sentiment in tweets from those with few followers. Four of the nine hypothesis tests were significant (Table A3 in the Appendix), but the SMAPE values are not improved. We conclude that the sentiment in tweets from those with many followers is not consistently related to future returns.

If the sentiment in tweets does affect the market, we are likely to see an effect on the same day by users with many followers. For example, if Jim Cramer tweets positively or negatively about a specific stock then its price should move quickly. Table A4 in the Appendix shows the same day effects (i.e., $t = 0$); there is a significant same day effect for the sentiment in tweets from users with many followers both with and without considering the sentiment in tweets from those with few followers. All three betas on the sentiment from those with many followers are greater than the corresponding betas for those with few followers, with significance of $p < .001$. The SMAPE values are lower than those of the corresponding controls-only models.

Effectiveness of a Trading Strategy

One important question is whether these results can be used to build a profitable trading strategy. We theorized, and the empirical results show that the tweets from users with few followers that are *not* retweeted lead to the greatest abnormal returns. We followed the approach of Tetlock et al. (2008) and constructed two equally weighted portfolios, one long, one short. At the close of each trading day, we analyzed the sentiment in that day's tweets about specific firms using pos2 (we choose pos2 because it takes into account both positive and negative sentiment). We purchase firms in the top 10% and short sell the firms in the bottom 10%. Not all firms receive tweets each day, so the number of firms varies from day to day. We used three different holding periods (1 day, 10 days, and 20 days), and at the end of the holding period, we close out our long and short positions. Because we are simultaneously taking long and short positions, there is no need to consider market return as a control; any rise or fall in the market as a whole is controlled for by the simultaneous long and short positions.

Table 8 presents the annualized returns of the trading strategy with and without trading costs for the three different holding periods. Following Tetlock et al. (2008), we assume round-trip trading costs of 10 basis points (i.e., the total cost to buy and sell). The 1-day holding period produces positive returns, but because the strategy executes trades every day, the returns become negative after including trading costs. The trading strategies using 10- and 20-day holding periods produce significant positive returns, both before and after trading costs. These returns compare favorably with those in Tetlock et al. (2008), who found trading strategies using sentiment in news stories produced returns of 23.17% before trading costs and –2.71% after trading costs (i.e., a loss). In other words, the results in Table 8 show that a trading strategy with a 10- or 20-day holding period that balances long and short positions results in meaningful positive returns.

DISCUSSION

We theorized that the sentiment in tweets is related to stock returns of individual stocks based on how fast that sentiment spreads through the market. The sentiment in tweets sent by users with few followers, which diffuses more slowly than sentiment in tweets sent by those with many followers, is significant in predicting the firm's stock returns one trading day, 10 trading days, and 20 trading days after the tweets were posted. The sentiment in tweets from those with few followers that were not retweeted had the strongest effect on returns on future days, as this sentiment takes the longest to spread through the market. A trading strategy with a 10- or 20-day holding period built on these factors shows meaningful annual returns.

We argue that a social contagion process is at work. Tweets spread positive or negative sentiment about a stock through the market and can influence prices, and thus the returns from trading those stocks. Sentiment can spread quickly; for example, network hubs like Jim Cramer can send a tweet that has positive or negative sentiment about a stock and his 650,000 followers immediately see its sentiment. If these followers act, the stock price can respond very quickly. Thus, we conclude that Twitter users with many followers have a market impact similar to traditional news media; the impact of the sentiment in their tweets disseminates rapidly and is quickly incorporated into stock price. However, there are no significant stock returns on future trading days and thus it is difficult to profit from a trading strategy based on them. In contrast, if a user with few followers sends a tweet, few people will see it, and even if they quickly act on its sentiment, the small number of trades will have little immediate impact on the stock price. Over time, however, as the sentiment diffuses through the market, the sentiment will gradually affect the stock price. The sentiment in tweets from users with fewer followers had a stronger impact on stock returns the next trading day and over the next 10–20 days compared to tweets from users with many followers. Because the change is gradual, there is an opportunity to profit from this as a trading strategy. The diffusion is slower for tweets that are not retweeted, and thus they have the greatest impact on returns on future trading days and offer the greatest opportunities to profit from a trading strategy based on this.

Our results offer similar conclusions to other research based on the GIF model of information diffusion. GIF argues that markets are generally efficient at the macro level, but if we examine them at a micro level, we see that it is possible to uncover situations in which markets are not perfectly efficient, because human behavior is not perfectly efficient at spreading information. Studies in line with GIF have suggested that it takes longer for information about stocks that are not routinely covered by the mass media to be absorbed into their price (Merton, 1987; Fang & Peress, 2009). Thus, it is possible to use information to predict future stock returns for firms with systematically slower information diffusion, even after controlling for risk characteristics (Merton, 1987; Fang & Peress, 2009). Thus, we believe the GIF model provides a good theoretical foundation for understanding why Twitter sentiment can be used to predict future stock returns in some cases but not in others.

There exist at least two possible theoretical explanations for how the speed of information diffusion influences stock returns. The first is that the sentiment of tweets "causes" changes in stock prices. Individuals post tweets when they believe they have useful comments about an individual stock. These comments may have facts as well as an underlying sentiment. Sentiment is highly contagious (Schoenewolf, 1990; Hatfield et al., 1993), and it influences how investors make buy/sell decisions as the sentiment spreads through the public. A cumulative positive sentiment triggers positive thoughts about the company and leads to a purchase decision, raising the stock price. A cumulative negative sentiment induces negative thoughts and thus leads to sell decisions, decreasing the stock price. The rate at which this sentiment spreads through the market is influenced by the number of followers of the sender and whether the tweet is retweeted, so that sentiment in tweets from those with few followers that are not retweeted takes a longer time to spread and thus takes longer to influence stock prices; this leads to significant stock returns on future trading days.

A second possible explanation is that tweets "reflect" the underlying information that influences individual stock returns. In this case, it is not the sentiment of the tweets themselves that influence stock returns, but rather the tweets reflect how investors feel about the stock and are a leading indicator of their buy/sell decisions. Investors planning to buy a stock have positive sentiment about the stock and communicate this sentiment in their tweets. Likewise, investors planning to sell a stock communicate negative sentiment in their tweets.

We believe that the first explanation, that the sentiment of tweets causes stock price changes, best explains our findings because the number of followers and whether the tweet was retweeted were significantly related to stock price changes. Two assumptions would have to hold for the explanation that tweets reflect information to be viable. First, investors would have to tweet their information days before acting on it, which is illogical; no rational investor would share information likely to affect prices before acting on it. Second, the underlying information would need to spread through the social network in the same manner as the tweets and retweets but via a different mechanism in order for us to find the relationships we did. This is a less likely explanation than the simpler explanation that it is the tweets themselves that influence behavior.

CLT argues that positive and negative sentiment may have different effects (Liberman & Trope, 1998; Bar-Anan et al., 2006; Fujita et al., 2006). Previous research has shown that both positive and negative sentiment in tweets can affect stock returns, but no study has found both to have effects (Smailović et al., 2014;

Risius et al., 2015). Interestingly, we found both positive and negative sentiment to directly affect returns.

Investors make investment decisions using a variety of information sources, with Twitter being just one of many possible sources. The economic magnitude of the relationships in our study is moderate to high (Tetlock et al., 2008). The trading analysis showed positive annual returns for 10- and 20-day holding periods after considering trading costs. The economic significance of these effects is meaningful.

Limitations

This study also has several limitations. We only studied firms in the S&P 500. We have no empirical data to argue that our results apply or do not apply to smaller firms or firms traded in other markets that are not covered by the S&P 500. We studied one specific time period in the life of the market, so it could be that the market conditions that led to our findings no longer apply. Likewise, we studied the same time period in the life of Twitter, and because Twitter behavior changes over time (Liu et al., 2014), it may be that Twitter users behave differently now, and the behaviors we observed no longer occur.

The fundamental theory underlying our research is the GIF model. Our results are driven by the speed of diffusion of sentiment, so one important theoretical limitation is if this sentiment is based on already widely diffused fundamental information (e.g., a rise in oil prices that could negatively influence transportation stocks) then this sentiment is likely to have little effect on stock returns, because investors have already acted and prices have already changed. We did not examine the extent to which the sentiment in the tweets we analyzed was based on already disseminated fundamental information, so this is an avenue for future research.

Another potential limitation is homophily, the possibility that individuals similar to each other tend to post similar Tweets (Aral, Muchnik, & Sundararajan, 2009; Shalizi & Thomas, 2011). Under this argument, the changes in stock prices are not due to social contagion but are because people similar to each other in the number of followers use similar trading strategies. This is possible, but we view this as less likely than social contagion because it could only be true if trading behaviors were related to the number of followers and the retweet history of tweets. This is possible, but requires additional, somewhat convoluted theorizing to link the number of Twitter followers and retweeting history to trading behaviors. While homophily is useful in understanding some tweeting behaviors, it is often not as powerful as other theoretical models (Macskassy & Michelson, 2011). So, using Occam's razor, we conclude that the social contagion of sentiment is a better explanation for our results.

Implications for Research

Despite these limitations, we believe that these results have implications for future research. Our work builds on recent research showing that the "calmness" or "depression" in tweets (Bollen et al., 2011a; Risius et al., 2015), their "bullishness" or "beariness" (Oh & Sheng, 2011; Sprenger et al., 2014), and their sentiment (Smailović et al., 2014; Risius et al., 2015) can be useful in predicting stock returns. We use social contagion based on the GIF model as underlying theory and show that factors which influence the speed of sentiment diffusion (number of followers, retweeting) significantly affect the stock returns on future trading days. We offered two possible explanations for the theoretical mechanism that links the sentiment in tweets to future stock returns. We need more research to better understand the underlying theoretical mechanism that links sentiment to stock returns.

Our research shows that *who* sends the tweets is an important factor in explaining how the sentiment in tweets affects stock returns. Ironically, users with many followers (i.e., those with more than the median number of 171) have *no* significant influence on stock returns on the next trading day or subsequent days. The sentiment in their tweets is quickly incorporated into stock prices leading to no future returns. In contrast, sentiment expressed by Twitter users who have few followers—and thus diffuses slowly—has significant and meaningful impacts on stock returns on future trading days. We used a simple analysis that divided users into two groups, over and under the median number of followers in our data set. We believe that this calls for more research into who expresses the sentiment in tweets and how this can be used to explain stock price movements and better predict stock returns. The number of connections in a social network typically follows a power law distribution (Barabási, 2002), so an analysis that uses more than two categories to better capture this distribution may better model the speed of sentiment diffusion and provide additional insight.

Our research also shows that what happens to the tweet *after* it is sent has a significant impact on stock returns. Tweets that are retweeted have a faster impact on stock prices and thus do not predict stock returns on future trading days, whereas tweets that are not retweeted can predict future returns. We believe this calls for more research into retweeting behavior. For example, how do investors react to the tweets they receive that are and are not retweeted? The most common reason for retweeting is because the sender believes the tweet's information would be of interest to their followers (Macskassy & Michelson, 2011). Do retweeted tweets appear more important and thus get more attention, so they are more likely to influence behavior?

Prior studies examining how emotion is linked to stock returns have used different approaches to measuring it, including emotional states (e.g., calm, happiness, depression) and sentiment (e.g., positive emotional valence, "bullishness") (Calvo & Kim, 2013). We used the Harvard IV Psychological Dictionary to assess the positive and negative sentiment expressed in the tweets. There are many dictionaries designed to categorize words based on sentiment, such as Loughran and McDonald Financial Sentiment Dictionaries (Loughran & McDonald, 2011). We did not include emoticons in our analysis, which could be examined in future research.

We used the formulas of Tetlock et al. (2008) to build three different measures of sentiment (pos1, pos2, neg) that provided essentially the same conclusions (with some minor differences among them). There are many other formulas and machine learning techniques that can be used to develop sentiment metrics that are more sophisticated (e.g., Oh & Sheng, 2011; Smailović et al., 2014; Sprenger et al., 2014). One key challenge in sentiment analysis is understanding semantics in groups of words. A Twitter post may have both positive and negative terms, so if one considers the semantic rules of groups of words, the meaning may become clearer (e.g., "this is not bad.") (Xie et al., 2015). Additional research is needed that uses different, more sophisticated, analysis strategies to better understand if different approaches are better at predicting stock returns on future trading days.

In this study, we use Twitter as the social media platform. There are many other social media platforms that may also provide insights into future stock returns. We hope our work can spawn future research on this topic. What are the impacts of Facebook, LinkedIn, or other Web media?

Finally, we examined the impact of sentiment on stock returns at a daily level. Future research could use market microstructural data to examine how emotional state and sentiment impact markets in real time.

Implications for Practice

We believe that this study has two implications for practice. The first is providing guidance to investment decision making. Our results show that a trading strategy built on the analysis of the sentiment in tweets from users with few followers that are not retweeted produces significant positive returns after considering trading costs. Tweets are available publically and can be retrieved using Twitter development accounts, so this may be an investable trading strategy. Combining this with a focus on firms that have little coverage from the traditional media may also increase returns (Merton, 1987; Fang & Peress, 2009).

A second implication is that firms should carefully monitor how they use Twitter. Most firms manage formal financial information that could impact stock prices because there are numerous financial regulations in place. Because the sentiment of tweets is linked to future stock prices, firms need to monitor the sentiment in their tweets in addition to the "rational" information they contain.

CONCLUSION

We found that the sentiment in social media postings can predict stock returns on future trading days. Tweets from users with few followers (i.e., less than the median of 171 followers) that were not retweeted had an impact on future returns 10 and 20 days later, while those from users with many followers and those that were retweeted had no impact on future returns. The findings are consistent with our hypothesis that sentiment that is diffused slowly takes longer to be incorporated into prices, while sentiment that is diffused faster will be quickly incorporated into prices and thus will have little association with returns on future days.

REFERENCES

Abbasi, A., & Chen, H. (2008). CyberGate: A design framework and system for text analysis of computer-mediated communication. *MIS Quarterly*, *32*(4), 811–837.

About Twitter, Inc. (2015). Accessed March 14, 2016. Available at: https://about.twitter.com/company

Aral, S., Muchnik, L., & Sundararajan, A., (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, *106*(51), 21544–21549.

Armstrong, J. S. (1985). *Long range forecasting: From crystal ball to computer* (2nd ed.). New York, NY: Wiley.

Bakamitsos, G. A. (2006). A cue alone or a probe to think? The dual role of affect in product evaluations. *Journal of Consumer Research*, *33*(3), 403–412.

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, *21*(2), 129–151.

Bar-Anan, Y., Liberman, N., & Trope, Y. (2006). The association between psychological distance and construal level: Evidence from an implicit association test. *Journal of Experimental Psychology-General*, *135*(4), 609–622.

Barabási, A. L. (2002). *Linked*. Cambridge, MA: Persues Publishing.

Bechara, A., & Damasio, A. R. (2005). The somatic marker hypothesis: A neural theory of economic decision. *Games and Economic Behavior*, *52*(2), 336–372.

Bollen, J., Mao, H., & Zeng, X. (2011a). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8.

Bollen, J., Pepe, A., & Mao, H. (2011b). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, Barcelona.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59.

Cacioppo, J. T., Petty, R. E., Losch, M. E., & Kim, H. S. (1986). Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *Journal of Personality and Social Psychology*, *50*(2), 260–268.

Calvo, R. A., & Kim, S. M. (2013). Emotions in text: Dimensional and categorical models. *Computational Intelligence*, *29*(3), 527–543.

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

Chordia, T., Roll, R., & Subrahmanyam, A. (2002). Order imbalance, liquidity, and market returns. *Journal of Financial economics*, *65*(1), 111–130.

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

Daley, J., Jencks, S., Draper, D., Lenhart, G., Thomas, N., & Walker, J. (1988). Predicting hospital-associated mortality for Medicare patients. *JAMA: The Journal of the American Medical Association*, *260*(24), 3617–3624.

Dhir, A., Buragga, K., & Boreqqah, A. A. (2013). Tweeters on campus: Twitter a learning tool in classroom? *Journal of Universal Computer Science*, *19*(5), 672–691.

Dockery, E., & Kavussanos, M. G. (1996). Testing the efficient market hypothesis using panel data, with application to the Athens stock market. *Applied Economics Letters*, *3*(2), 121–123.

Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, *13*(1), 210–230.

Eppen, G. D., & Fama, E. F. (1969). Cash balance and simple dynamic portfolio problems with proportional costs. *International Economic Review*, *10*(2), 119–133.

Eyal, T., Liberman, N., Trope, Y., & Walther, E. (2004). The pros and cons of temporally near and distant action. *Journal Personality and Social Psychology*, *86*(6), 781–795.

Fang, L., & Peress, J. (2009). Media coverage and the cross-section of stock returns. *The Journal of Finance*, *64*(5), 2023–2052.

Fedorikhin, A., & Patrick, V. M. (2010). Positive mood and resistance to temptation: The interfering influence of elevated arousal. *Journal of Consumer Research*, *37*(4), 698–711.

Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, *56*(4), 82–89.

Frijda, N. H. (1994). Varieties of affect: Emotions and episodes, moods, and sentiments. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions*. Oxford, UK: Oxford University, 197–202.

Fujita, K., Trope, Y., Liberman, N., & Levin-Sagi, M. (2006). Construal levels and self-control. *Journal of Personality and Social Psychology*, *90*(3), 351–367.

Gordon, T., Castelli, W. P., Hjortland, M. C., Kannel, W. B., & Dawber, T. R. (1977). Predicting coronary heart disease in middle-aged and older persons. *JAMA: The Journal of the American Medical Association*, *238*(6), 497–499.

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1993). Emotional contagion. *Current Directions in Psychological Science*, *2*(3), 96–99.

Honey, C., & Herring, S. C. (2009). Beyond microblogging: Conversation and collaboration via Twitter. *Proceedings of the 42nd Hawaii International Conference on System Science*, Waikoloa.

Hong, H., Lim, T., & Stein, J. C. (2000). Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *The Journal of Finance*, *55*(1), 265–295.

Hong, H., & Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance*, *54*(6), 2143–2184.

Hong, H., & Stein, J. C. (2007). Disagreement and the stock market. *Journal of Economic Perspectives*, *21*(2), 109–128.

Jorgenson, D. W., & Vu, K. (2005). Information technology and the World economy. *The Scandinavian Journal of Economics*, *107*(4), 631–650.

Labroo, A. A., & Patrick, V. M. (2009). Psychological distancing: Why happiness helps you see the big picture. *Journal of Consumer Research*, *35*(5), 800–809.

Larsson, A. O. (2013). Tweeting the viewer—Use of Twitter in a talk show context. *Journal of Broadcasting & Electronic Media*, *57*(2), 135–152.

Liberman, N., & Trope, Y. (1998). The role of feasibility and desirability considerations in near and distant future decisions: A test of temporal construal theory. *Journal of Personality and Social Psychology*, *75*(1), 5–18.

Lin, M., Lucas, H. C. Jr., & Shmueli, G. (2013). Too big to fail: Large samples and the p-value problem. *Information Systems Research*, *24*(4), 906–917.

Liu, Y., Kilman-Silver, C., & Mislove, A. (2014). The tweets they are a-changin': Evolution of Twitter users and behavior. *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, Ann Arbor, MI.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35–65.

Macskassy, S. A., & Michelson, M. (2011). "Why do people retweet? Anti-homophily wins the day!" *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, Barcelona.

Makridakis, S. (1993). Accuracy measures: Theoertical and practical concerns. *International Journal of Forecasting*, *9*(4), 527–529.

Malkiel, B. G., & Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417.

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *The Journal of Finance*, *42*(3), 483–510.

Oh, O., Agrawal, M., & Rao, H. R. (2013). Community intelligence and social media services: A rumor theoretic analysis of tweets during social crises. *MIS Quarterly*, *37*(2), 407–426.

Oh, O., & Sheng, O. (2011). Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement. *Proceedings of the International Conference on Information Systems*, Shanghai.

Osgood, C., Suci, G., & Tannenbaum, P. (1957). *The measurement of meaning*. Urbana, IL: University of Illinois.

Parboteeah, D. V., Valacich, J. S., & Wells, J. D. (2009). The influence of website characteristics on a consumer's urge to buy impulsively. *Information Systems Research*, *20*(1), 60–78.

Pew-Research. (2014). Social media use over time, accessed November 11, 2014, available at http://www.pewinternet.org/data-trend/social-media/social-media-use-all-users/.

Qian, B., & Rasheed, K., (2007). Stock market prediction with multiple classifiers. *Applied Intelligence*, *26*(1), 25–33.

Qiu, C., & Yeung, C. W. (2008). Mood and comparative judgment: Does mood influence everything and finally nothing? *Journal of Consumer Research*, *34*(5), 657–669.

Rai, A., Patnayakuni, R., & Seth, N. (2006). Firm performance impacts of digitally enabled supply chain integration capabilities. *MIS Quarterly*, *30*(2), 225–246.

Risius, M., Akolk, F., & Beck, R. (2015). Differential emotions and the stock market—The case of company-specific trading. *European Conference on*

Information Systems, Completed Research Papers, Munster, Germany, Paper 147.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172.

Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes Twitter users: Real-time event detection by social sensors. *Proceedings of the 19th International Conference on the World Wide Web*. Raleigh, NC.

Schoenewolf, G. (1990). Emotional contagion: Behavioral induction in individuals and groups. *Modern Psychoanalysis*, *15*(1), 49–61.

Shalizi, C. R., & Thomas, A. C. (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods & Research*, *40*(2), 211–239.

Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences*, *1156*(1), 81–96.

Smailović, J., Grčar, M., Lavrač, N., & Žnidaršič, M. (2014). Stream-based active learning for sentiment analysis in the financial domain. *Information Sciences*, *285*(32), 181–203.

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

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, *62*(3), 1139–1168.

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

Tofallis, C. (2015). A better measure of relative prediction accuracy for model selection and model estimation, *Journal of the Operational Research Society*, *66*(8), 1352–1362.

trade-IQ. (2011). The U.S. active trader market; Report Preview, accessed December 8, 2015, available at http://tradeiq.blogspot.com/2012/02/us-active-trader-market-report-preview.html.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M., (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, Washington, DC.

Wiebe, J., Wilson, T., & Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, *39*(2-3), 165–210.

Xie, Y., Chen, Z., Zhang, L. K., Cheng, Y., Honbo, D., & Agrawal, A. (2015).et al. MuSES: Multilingual sentiment elictitation system for social media data, *IEEE Intelligent Systems*, *29*(4), 34–42.

Appendix

Table A1: Regression results of emotional sentiment on abnormal returns by number of followers using a breakpoint of 1,000 followers (a 96–4% split). The SMAPE values for the controls-only models are 1.9182, 1.8694, and 1.8116.

(a) Pos1	Next Day	Next-to-10th-Day	Next-to-20th-Day
Sentiment_Over $_{i,t}$	–0.193	0.428	0.088
Sentiment_Under $_{i,t}$	0.384**	1.248**	1.871**
Control 1 $_{i,t\text{-}1}$	0.006	0.022	0.007
Control 2$_{i,t\text{-}30,t\text{-}2}$	0.002*	0.016***	0.027***
Surprise $_{i,t}$	0.010***	0.006	0.005
Upgrade $_{i,t}$	0.002***	0.000	–0.001
Downgrade $_{i,t}$	–0.002***	0.000	–0.002
Intercept	0.000***	–0.003***	–0.006***
SMAPE	1.9146	1.8594	1.8113
Adj. R^2	0.002	0.001	0.001

(b) Pos2	Next Day	Next-to-10th-Day	Next-to-20th-Day
Sentiment_Over $_{i,t}$	–0.164	0.114	–0.625
Sentiment_Under $_{i,t}$	0.301**	1.145***	1.988***
Control 1 $_{i,t\text{-}1}$	0.006	0.022	0.007
Control 2$_{i,t\text{-}30,t\text{-}2}$	0.002*	0.016***	0.027***
Surprise $_{i,t}$	0.010***	0.006	0.005
Upgrade $_{i,t}$	0.002***	0.000	–0.001
Downgrade $_{i,t}$	–0.002**	0.000	–0.002
Intercept	0.000***	–0.003***	–0.006***
SMAPE	1.9142	1.8589	1.8107
Adj. R^2	0.002	0.001	0.001

(c) Neg	Next Day	Next-to-10th-Day	Next-to-20th-Day
Sentiment_Over $_{i,t}$	1.885	–4.643	–1.104
Sentiment_Under $_{i,t}$	–2.922	–20.715**	–14.126***
Control 1 $_{i,t\text{-}1}$	0.006	0.023*	0.026***
Control 2$_{i,t\text{-}30,t\text{-}2}$	0.002*	0.016***	0.007
Surprise $_{i,t}$	0.010***	0.006	0.005
Upgrade $_{i,t}$	0.002***	0.000	–0.001
Downgrade $_{i,t}$	–0.002***	0.000	–0.002
Intercept	0.000*	–0.002***	–0.004***
SMAPE	1.9170	1.8597	1.8112
Adj. R^2	0.002	0.001	0.001

Notes: The coefficients are multiplied by 1,000.
$^*p \leq .05$, $^{**}p \leq .01$, $^{***}p \leq .001$.

Table A2: Regression results of emotional sentiment and retweeting on abnormal returns by number of followers using a breakpoint of 1,000 followers (a 96–4% split). The SMAPE values for the controls-only models are 1.8577, 1.7660, and 1.7168, respectively.

(a) Pos1	Next Day	Next-to-10th-Day	Next-to-20th-Day
Sentiment_OverNo $_{i,t}$	–1.021	2.814	0.672
Sentiment_OverRe $_{i,t}$	0.146	–0.481	0.309
Sentiment_UnderNo $_{i,t}$	1.185*	1.240	3.737
Sentiment_UnderRe $_{i,t}$	0.526	0.998	0.922
Control 1 $_{i,t\text{-}1}$	–0.003	0.058*	0.037
Control 2$_{i,t\text{-}30,t\text{-}2}$	0.008***	0.047***	0.082***
Surprise $_{i,t}$	0.007*	0.001	–0.005
Upgrade $_{i,t}$	0.002	–0.002	–0.005
Downgrade $_{i,t}$	–0.002***	0.003	0.002
Intercept	–0.001**	–0.005***	–0.008***
SMAPE	0.3253	.3145	0.3158
Adj. R^2	0.004	0.009	0.012

(b) Pos2	Next Day	Next-to-10th-Day	Next-to-20th-Day
Sentiment_OverNo $_{i,t}$	–0.655	1.079	–0.745
Sentiment_OverRe $_{i,t}$	–0.165	–1.127	–0.604
Sentiment_UnderNo $_{i,t}$	0.896*	1.444	3.643*
Sentiment_UnderRe $_{i,t}$	0.384	1.186	0.856
Control 1 $_{i,t\text{-}1}$	–0.003	0.059*	0.038
Control 2$_{i,t\text{-}30,t\text{-}2}$	0.008***	0.047***	0.083***
Surprise $_{i,t}$	0.008*	0.001	–0.005
Upgrade $_{i,t}$	0.002	–0.002	–0.005
Downgrade $_{i,t}$	–0.002***	0.003	0.002
Intercept	–0.001**	–0.005***	–0.008
SMAPE	0.3252	0.3144	0.3156
Adj. R^2	0.004	0.009	0.012

(c) Neg	Next Day	Next-to-10th-Day	Next-to-20th-Day
Sentiment_OverNo $_{i,t}$	6.373	–59.201*	–39.345
Sentiment_OverRe $_{i,t}$	4.147	32.940	29.771
Sentiment_UnderNo $_{i,t}$	–13.255	–27.298	–65.048
Sentiment_UnderRe $_{i,t}$	–8.617	–21.707	–22.808
Control 1 $_{i,t\text{-}1}$	–0.003	0.059*	0.037
Control 2$_{i,t\text{-}30,t\text{-}2}$	0.008***	0.046***	0.082***
Surprise $_{i,t}$	0.008*	0.001	-0.005
Upgrade $_{i,t}$	0.002	–0.002	–0.005
Downgrade $_{i,t}$	–0.002***	0.003	0.002
Intercept	0.000	0.000	–0.002
SMAPE	0.3257	0.3145	0.3157
Adj. R^2	0.004	0.009	0.012

Notes: The coefficients are multiplied by 1,000.
*$p \le .05$, ** $p \le .01$, *** $p \le .001$.

Table A3: Regression results of emotional sentiment on abnormal returns using only those with more than 171 followers. The SMAPE values for the controls-only models are .9678, .9399, and .9216.

(a) Pos1	Next Day	Next-to-10th-Day	Next-to-20th-Day
Sentiment_Over $_{i,t}$	–0.056	0.832**	0.952*
Control 1 $_{i,t\text{-}1}$	0.002	0.011	-0.014
Control 2$_{i,t\text{-}30,t\text{-}2}$	0.001	0.007***	0.013***
Surprise $_{i,t}$	0.012***	0.012*	0.014
Upgrade $_{i,t}$	0.002***	0.001	–0.001
Downgrade $_{i,t}$	–0.002***	0.001	–0.004**
Intercept	0.000***	–0.003***	–0.006***
SMAPE	0.9677	0.9399	0.9216
Adj. R^2	0.001	0.000	0.001

(b) Pos2	Next Day	Next-to-10th-Day	Next-to-20th-Day
Sentiment_Over $_{i,t}$	–0.041	0.587*	0.395
Control 1 $_{i,t\text{-}1}$	0.002	0.011	–0.010
Control 2$_{i,t\text{-}30,t\text{-}2}$	0.001	0.012***	0.013***
Surprise $_{i,t}$	0.012***	0.004	0.014
Upgrade $_{i,t}$	0.002***	0.000	-0.001
Downgrade $_{i,t}$	–0.002***	0.000	–0.004***
Intercept	0.000***	–0.003***	–0.004***
SMAPE	0.9677	0.9399	0.9216
Adj. R^2	0.001	0.000	0.001

(c) Neg	Next Day	Next-to-10th-Day	Next-to-20th-Day
Sentiment_Over $_{i,t}$	0.149	–12.808**	–9.950
Control 1 $_{i,t\text{-}1}$	0.001	0.018	–0.013
Control 2$_{i,t\text{-}30,t\text{-}2}$	0.001	0.013***	0.013***
Surprise $_{i,t}$	0.012***	0.004	0.014
Upgrade $_{i,t}$	0.001***	0.000	–0.001
Downgrade $_{i,t}$	–0.002***	0.000	–0.003
Intercept	0.000***	–0.001**	–0.003**
SMAPE	0.9678	0.9399	0.9216
Adj. R^2	0.001	0.001	0.001

Note: The coefficients are multiplied by 1,000.
*$p \le .05$, ** $p \le .01$, *** $p \le .001$.

Table A4: Regression results of emotional sentiment on abnormal returns by number of followers for the same trading day. The SMAPE values for the controls-only model are 1.7935 and .9286, respectively.

(a) Pos1	Both Over and Under	Over Only
Sentiment_Over$_{i,t}$	1.635***	1.135***
Sentiment_Under $_{i,t}$	0.508**	
Control 1 $_{i,t\text{-}1}$	–0.003	–0.003
Control 2$_{i,t\text{-}30,t\text{-}2}$	0.001	0.001
Surprise $_{i,t}$	0.027***	0.022***
Upgrade $_{i,t}$	0.011***	0.010***
Downgrade $_{i,t}$	–0.017***	–0.015***
Intercept	0.000***	0.000***
SMAPE	1.7409	0.9018
Adj. R^2	0.036	0.027

(b) Pos2	Both Over and Under	Over Only
Sentiment_Over $_{i,t}$	1.627***	1.318***
Sentiment_Under $_{i,t}$	0.506***	
Control 1 $_{i,t\text{-}1}$	–0.004	–0.004
Control 2$_{i,t\text{-}30,t\text{-}2}$	0.001	0.001
Surprise $_{i,t}$	0.027***	0.022***
Upgrade $_{i,t}$	0.011***	0.009***
Downgrade $_{i,t}$	–0.017***	–0.015***
Intercept	0.000***	0.000***
SMAPE	1.7206	0.8910
Adj. R^2	0.038	0.028

(c) Neg	Both Over and Under	Over Only
Sentiment_Over $_{i,t}$	–23.415***	–16.503***
Sentiment_Under $_{i,t}$	–7.596**	
Control 1 $_{i,t\text{-}1}$	–0.002	–0.003
Control 2$_{i,t\text{-}30,t\text{-}2}$	0.001	0.001
Surprise $_{i,t}$	0.027***	0.023***
Upgrade $_{i,t}$	0.011***	0.009***
Downgrade $_{i,t}$	–0.017***	–0.015***
Intercept	0.000***	0.000***
SMAPE	1.7495	0.9086
Adj. R^2	0.035	0.027

Notes: The coefficients are multiplied by 1,000.
*$p \leq .05$, ** $p \leq .01$, *** $p \leq .001$.

Hong Kee Sul is a research support director at Wharton Research Data Services, The Wharton School, University of Pennsylvania. In 2015, he graduated with a PhD in Finance from Indiana University, Kelley School of Business. Dr. Sul received a Master of Science degree from Seoul National University and a Bachelor of Science degree from Korean Advanced Institute of Science and Technology. His work has been presented in several conferences, including the 2016 American Economic Association Annual Meetings, 2014 Financial Management Association Annual Meeting Doctoral Consortium, and the 47th Annual Hawaii International Conference on System Sciences.

Alan Dennis is a professor of information systems and holds the John T. Chambers Chair of Internet Systems in the Kelley School of Business at Indiana University. He is a Fellow of the AIS and has written more than 150 research papers. His research focuses on three main themes: the use of computer technologies to support team creativity and decision making; IS for the subconscious; and the use of the Internet to improve education. He is the AIS Vice President of Conferences, the editor in Chief of *Foundations and Trends in Information Systems*, and co-editor in Chief of *AIS Transactions on Replication Research*. Prof. Dennis also has written four books, two on data communications and networking, and two on systems analysis and design. His most recent start-up company is NameInsights.com, which uses big data and analytics to help parents pick baby names.

Lingyao (Ivy) Yuan is an assistant professor of information systems of College of Business at Iowa State University. Her research interests include the impact of noncognition behavior and decision making, especially the impact of emotion, on computer-mediated communication, decision making, and collaboration. She has conducted research in the fields of electronic commerce and social media. She has been published in the *Group Decision and Negotiation* as well as several conferences, including *47*th *Annual Hawaii International Conference on System Sciences*, and *2013 INFORMS Annual Meeting*. She received a Master of Science in Information Technology from University of North Carolina Charlotte in 2011 and Bachelors of Management Information Systems from University of International Business and Economics in 2009.