Trading on Twitter:
The Financial Information Content of Emotion in Social Media

Hong Keel Sul
Indiana University
hsul@indiana.edu

Alan R. Dennis
Indiana University
ardennis@indiana.edu

Lingyao (Ivy) Yuan
Indiana University
yuanl@indiana.edu

Abstract

We collected data from Twitter posts about firms in the S&P 500 and analyzed their cumulative emotional valence (i.e., whether the posts contained an overall positive or negative emotional sentiment). We compared this to the average daily stock market returns of firms in the S&P 500. Our results show that the cumulative emotional valence (positive or negative) of Twitter tweets about a specific firm was significantly related to that firm's stock returns. The emotional valence of tweets from users with many followers (more than the median) had a stronger impact on same day returns, as emotion was quickly disseminated and incorporated into stock prices. In contrast, the emotional valence of tweets from users with few followers had a stronger impact on future stock returns (10-day returns).

1. Introduction

Social network sites have attracted millions of users and are now a meaningful channel in which consumers share information and ideas. In 2012, 63% of Internet users used social media at least once a month and this percentage is increasing [40]. Twitter is among Facebook, LinkedIn, YouTube, Pinterest and Instagram as one of the most popular social media platforms in the world. In early 2013, there were about 500 million Twitter users worldwide who sent an average of 400 million tweets per day [41].

Users have integrated social media into many aspects of their daily life [15]. Because of its advantages over traditional media in terms of reach, frequency, usability, immediacy and permanence, more and more industries use social media to distribute information [1]. Numerous professional and amateur investors and analysts have begun using Twitter to post news articles, and opinions, often more frequently than the professional news media.

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 are a major factor in providing valuable implicit or explicit information for making fast and advantageous decisions [4]. Twitter provides a good environment to foster the sharing of emotion [6]. The length of each tweet is restricted to 140 words. Emotion, as one type of affect, has the characteristics of having a clear trigger and having a short but more intense effect on the individuals [21]. The limitation on length of twits encourages users to convey one type of emotion per message. The short message can provide a focused and perhaps intense trigger to the information receiver. Twitter also supports both bi-directional and single direction relationships. A user can have friends they "listen" to, followers they "talk' to, and some who are both friends and followers. Its flexibility can support complex social structures.

Prior research has studied whether the emotional content of tweets can be used to predict stock returns. Bollen [5] used text processing techniques to abstract emotion level from about 10 million tweets over a nine month period. They compared the extent to which six emotions were expressed in the tweets (calm, alert, sure, vital, kind, and happy) to the closing values of the Dow Jones Industrial average (DJIA) on subsequent days. They found that one emotion, "calm," was significantly positively correlated (r=.013) with changes in the DJIA two days later and five days later (r=.036). In other words, when there was a great deal of "calm" emotion in tweets on a given day, the DJIA tended to rise over the following two to five days. While these correlations (and R^2 of .0002 to .001) may seem small, they are comparable to effects seen in other research on the impacts of information on future stock returns [9, 10, 40, 41].

These findings are promising in suggesting that the emotion in tweets can be used to predict stock price changes, but there are still many unanswered questions. First, the prior research [5] looked at all tweets (not just those related to investing) and compared them to the return of one stock market index (DJIA). One important question is whether the emotion expressed in tweets related to a specific firm can be used to predict the future price of that one firm's stock. Second, only one emotion ("calm") was found to be significantly correlated with stock index return. Thus another question is whether other emotions or the overall pattern of positive or negative emotion in stock-related tweets can be used to predict future returns. Third, there is a lack of theory to explain why "calm" emotions influence stock index returns days later. We need a better understanding of how emotions are related to stock returns.

We analyzed data collected from Twitter, during the period from March to October of 2011 and linked it to the average daily stock market returns of firms in the S&P 500. Our results show that the emotional valence (positive or negative) of tweets about specific firms was significantly related to stock returns on subsequent days. Interestingly, tweets from those with fewer followers had a stronger impact on future prices than tweets from those with many followers.

2. Prior Theory

2.1. Information and Stock Return Prediction

Whether stock market prices can be predicted has long been a debate. Based on the Efficient Market Hypothesis (EMH), earlier research argued that stock market prices are random and cannot be predicted [14, 16]. However, recent research has found that new information, especially news, is a major factor in predicting stock market prices [31, 35]. Mass media outlets play in important role in disseminating information to a broad audience, especially individual investors [18]. This suggests that new information extracted from social media such as Twitter, which also reaches a broad audience, may be useful in predicting stock market prices [5].

According to the Gradual-Information-Diffusion model, investors are typically 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 [26]. Both act under bounded rationality, and the interaction between these two creates both under-reaction and overreaction to new information.

This model 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 more quickly incorporated into stock prices [26],[27]. 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 price [26],[27]. Information should spread rapidly for stocks covered by the mass media but more slowly for stocks are not covered by the media. Research shows that stock not covered by the mass media earn significantly higher future returns than stocks that are covered, after controlling for risk characteristics [18, 32], suggesting that the speed at which information spreads across the investing public is important in understanding why and how that information may be used to predict future stock prices.

2.2 Information 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 also subscribe to (or "follow") other users and the followers are notified immediately when a user tweets. Many Twitter users have few followers; others (often celebrities) have millions. In early 2013, the average number of followers was about 200 [10].

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 in information diffusion and supporting communication/collaboration [25] in many different contexts. Using Twitter during a talk show decreased the psychological distance between the host and his/her audience [29]. In the context of education, Twitter is also proven to be a potential learning tool in classrooms [13]. Twitter has become an important tool to spread information during natural disasters and social crises [33, 38].

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 [42], hospital-associated mortality [12], and heart disease in middle-aged and older persons [23].

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.[1] For example, an individual tweeting about Apple would include $AAPL in the tweet. A sample tweet might say something like: "$PEP has been strong all day. And who doesn't love those Frito-Lay snacks? Be honest" This tweet is related to Pepsico, Inc, whose NYSE ticker is "PEP."

---

[1] Stocktwits.com claims that they created the notation of using $ symbol as the prefix to the stock ticker.

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 millions 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 such as email.

The speed at which information spreads through a broad audience depends upon how many followers a user has. If a Twitter users has many followers, the information should spread more quickly than if the user has few followers. 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.

As we argued above, the speed of information diffusions influences whether the information is quickly incorporated in stock prices or takes longer – perhaps days – to be fully disseminated and incorporated into stock prices. If a Twitter user has many followers, any information he or she tweets should be quickly disseminated, and prices should quickly change to incorporate that information. In this case, there should be little or no relationship between information and price changes on future days because the information will be reflected in the stock's price on the same day it was tweeted. Conversely, if a Twitter user has few followers, information should be slow to disseminate, so there is more likely to be a relationship between that information and stock price changes on future days because it will take longer for that information to reach more investors. Therefore, future stock price predictability tweets should be related to whether the user who tweeted it has few or many followers.

2.3 Emotion

Individual moods, emotions, and other affect are influenced by both internal factors and external factors. Internal factors include personality, emotion related to individual competency, and so on. External factors include experiences, and information the individual receives. Different affects have different impact on individuals [21]. 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 [21]. 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 [24, 39].

There are many ways to conceptualize emotion. The classic approach, used by Bollen [5], is to consider specific emotions such as joy, anger, sadness, etc. Modern research has reconceptualized emotion as having two dimensions, valence (positive or negative) and arousal (high or low) [7, 8, 37]. In this study, we are interested in how positive and negative emotion affects stock returns. Thus, we focus only valence. As an aside, we note that the "calm" emotion studied by Bollen is somewhat similar to neutral arousal and neutral valence.

Emotion, especially positive emotion, has been well studied in social psychology and marketing. Multiple social psychology theories emphasize the effect of emotion. According to Construal Level Theory (CLT), positive mood increases abstract construal, that is, the adoption of abstract, future goals. Negative mood triggers a focus on immediate and proximal concerns and reduces the adoption of abstract, future goals [3, 17, 22, 28, 30].

Positive emotion affects decision making [2]. Individuals are more likely to be influenced by emotion during the formation of the first evaluation and less likely to be influenced in subsequent evaluations [36]. Positive emotion also has a strong likelihood to cause action [21]. Positive emotion is more likely to individuals to make a choice compared with negative emotion [36]. Positive emotion can increase consumers' impulse to buy in the context of electronic commerce [34], but increase an individual's resistance to temptation in other contexts [19].

2.4 Hypotheses

We argue that the emotional valence of social media tweets should have a direct effect on stock market returns. This is similar to the effects that professional news media have on same day and future returns, . Thus negative emotional valence should be associated with negative same day abnormal returns and negative future abnormal returns. Similarly, positive emotional valence should be associated with positive same day abnormal returns and should predict positive future abnormal returns.

Our primary focus is on future returns, rather than the returns on the same day as the tweet. The tweet on the same day could precede or follow the price change that produces the positive or negative return. We are interested in both short term future returns (i.e., the next day) and longer term future returns. We have chosen to use 10-day future returns for longer term returns; the choice between 10-day and some other period is somewhat arbitrary (e.g., 9-day could be argued to be equally appropriate) but is consistent with prior research. Therefore:

H1a. The emotional valence of Twitter tweets about a specific firm is positively correlated with the individual stock return of the same trading day.

H1b. The emotional valence of Twitter tweets about a specific firm is positively correlated with the individual stock return of the next trading day.

H1c. The emotional valence of Twitter tweets about a specific firm is positively correlated with the individual stock return of the next 10 trading days.

We argue that the effects of emotional valence spread in the same manner in which information spreads through the investing public. Thus the speed of information diffusion is important. If a tweet about a specific firm is sent by a user who has many followers, the emotional valence it contains will spread faster than the emotional valence sent by a user who has few followers. Thus the number of follower is a proxy for the speed of information diffusion. Information that is spread more quickly will be incorporated into prices faster, so that it will have a larger effect on the same day returns and less of an effect on future returns [26, 41]. Less visible information that is spread slower will have a smaller same day effect and a larger future effect [26].

Therefore, because the valence of tweets sent by users with many followers is likely to reach a broader audience more quickly, its impact will be more highly associated with same day returns than the valence of tweets sent by users with few followers. Likewise, tweets from users with fewer followers should be less associated with the firm's same day returns and have higher future return predictability.

H2a. The relationship between the emotional valence of tweets about a specific firm and the individual stock return on the same trading day is positively moderated by number of followers of the tweet sender, such that tweets by individuals with many followers will have a greater impact than those by individual with fewer followers.

H2b. The relationship between the emotional valence of tweets about a specific firm and the individual stock return on future trading days is negatively moderated by number of followers of the tweet sender, such that tweets by individuals with many followers will have a lesser impact than those by individual with fewer followers.

3. Data

3.1. Financial Data

To ensure sufficient reliability of Twitter data, we focused only on firms that are part of the S&P500. Financial data for the closing price of each stock in the S&P 500 were obtained from *Compustat*, *Center for Research in Security Prices (CRSP)*, and *Kenneth French*'s website [16]. The sample period is from March 2011 to February 2012.

3.2. Twitter Data

This study used data collected from *Twitter*. The focus of this paper is on whether the emotional content of 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&P500 stock ticker from Twitter using a developer account. We found 3,475,428 Tweets. We excluded any tweets that contained more than one ticker symbol because we could not be sure 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. A random inspection of 500 tweets found none from the firm itself.

3.3. The Emotional Valence of Tweets

There are many approaches to sentiment analysis [20]. We used the word analysis strategy to quantify tweets about specific firms into an emotional valence variable. Each word in a tweet was matched to a dictionary of terms to determine its emotional valence. The Harvard-IV dictionary [14] is a commonly used source for word classification in financial analysis, used, for example, by Tetlock [40, 41] and Da, et al. [11]. We counted all words in the tweets that had the 'NEG' tag in the Harvard-IV dictionary as words that conveyed a negative emotional valence. We counted 'POS' tagged words as words conveying a positive valence.

Since 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 4pm. Any tweet that was posted after 4pm was treated as day *t+1*. Following Tetlock et al. [41], we used three variables to measure emotional valence. The emotional valence 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.

$$Valence = \begin{cases} neg1 \equiv \dfrac{N}{T} \\ pos1 \equiv \dfrac{P-N}{P+N} \\ pos2 \equiv \log(\dfrac{1+P}{1+N}) \end{cases}$$

Conceptually, neg1 is the ratio of the amount of negative sentiment to the total communication (positive, negative, and neutral). Pos1 is a normalized ratio (on a -1 to =1 scale) of the overall positive or negative valance of the sentiment expressed (omitting neutral 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 reliability,

3.4. Control Variables

We used two Cumulated Abnormal Return (CAR) variables as control variables [9, 41]. The abnormal returns are computed as the raw returns (from CRSP) minus the size and book-to-market matched characteristic portfolio's return, which are the six portfolios based on the 30th and 70th NYSE book to-market ratio percentiles and on the median NYSE market equity from Kenneth French's website.

4. Method

To answer the question of whether social media have emotional information that can predict future returns, we focus on two research questions. First, can the emotional valence of individual investors' tweets about a specific firm explain contemporaneous returns and predict the firm's future returns? Second, if there is return predictability, does the speed of information dissemination (reflected by the number of followers) influence future returns?

To answer the first research question, we test whether social media has information that could explain contemporaneous returns and predict future returns. We test the following equations,

$$CAR_t^i = \alpha + \beta_0 valence_t^i + \gamma CV + \epsilon, \qquad (1)$$

$$CAR_{t+n}^i = \alpha + \beta_0 valence_t^i + \gamma CV + \epsilon, \qquad (2)$$

where CAR_t^i would be the cumulative abnormal return about firm i on day t, and $valence_t^i$ would be the daily measure of the emotional valence of tweets about firm i on day t. For the control variables CV we include past returns, cumulative abnormal return from the [-30,-2] trading window ($\mathrm{CAR}^{\mathrm{i}}_{t-30,t-2}$) (i.e., from 30 to 2 days prior to the day of interest) and the abnormal return on the prior day, i.e., day -1, ($\mathrm{CAR}^{\mathrm{i}}_{t-1,t-1}$). Thus equation (1) examines current abnormal returns (i.e., same day) and equation (2) future abnormal returns for days 1 to n.

Our first hypotheses argue for a direct effect of emotional valence on current and future stock return. We hypothesize that $\beta_0 > 0$ when valence is pos1 or pos2 and $\beta_0 < 0$ when valence is neg1.

To test the second hypotheses, we classify the tweets by the number of followers. An interesting feature of Twitter is that we are able to collect the number of followers for each user. We split the tweets into two groups based on the number of followers, those with many followers and those with few followers. The questions is, what is “many” and “few”? In our test, we use three break points 177 (the median number of followers in our sample), 1,000, and 100,000 as the threshold for assigning tweets into groups with few followers and many followers. We test the return predictability of the two valence variables, $valence\_o_t^i$ (the valence of tweets from users over the threshold) and $valence\_u_t^i$ (from users at or under the threshold), in a single regression.

$$CAR_{t+n}^i = \alpha + \beta_0 valence_{u_t}^{\ i} + \beta_1 valence_{o_t}^{\ i} + \gamma CV + \epsilon, \qquad (3)$$

For future returns, we hypothesize that $\beta_0 > \beta_1$ when valence is pos1 or pos2 and $\beta_0 < \beta_1$ when valence is neg1 since tweets from those with fewer followers should contain private information that is not quickly incorporated into current prices. For association between valence and same-day return, CAR_t^i, we expect to find opposite results.

5. Results

5.1. Return Predictability

The first tests examine H1, the overall return predictability. We ended up with 111,228 data points (remember that we are examining daily returns for the S&P 500). Table 1 shows the impact of the emotional valence of the tweets in explaining returns using the three measures of valence over the three time periods. The R^2 for these analyses are comparable to those in other studies examining the impact of information on current and future returns [5, 9, 10, 40]. For H1a (same day returns), all three measures of emotional valence have a significant and direct effect on returns. For H1b (next day returns), none of the three measures have effects on returns. For H1c (10-day returns), all three measures of emotional valence have a significant and direct effect on returns. We conclude H1a and H2c supported, but H1b is not.

Since the scales are different for the three dependent variables (pos1, pos2, neg1), it makes the most sense to compare the betas over the three time windows (same day, next day, 10 day). This shows that same day and 10-day effects are similar for all three variables, although effects may be stronger for neg1 at the 10-day mark than for same day.

Table 1. Statistical results for returns using the emotional valence of Tweets

	Betas (and Std)		
	Same Day Return	**NextDay Return**	**10 Day Return**
Pos1 Valence	0.912*** (0.082)	0.043 (0.074)	1.071 *** (0.231)
Control 1	0.001* (0.001)	0 (0.001)	0.002 (0.002)
Control 2	-0.003 (0.003)	0.001 (0.003)	0.010 (0.008)
Intercept	0*** (0)	0*** (0)	-0.002*** (0)
adj. R^2	0.001	0	0
Pos2 Valence	0.995*** (0.067)	0.047 (0.061)	1.083*** (0.189)
Control 1	0.001* (0.001)	0 (0.001)	0.002 (0.002)
Control 2	-0.003 (0.003)	0.001 (0.003)	0.009 (0.008)
Intercept	0*** (0)	0*** (0)	-0.002*** (0)
adj. R^2	0.002	0	0
Neg1Valence	-12.741*** (1.340)	-1.421 (1.216)	-21.613*** (3.798)
Control 1	0.001* (0.001)	0 (0.001)	0.002 (0.002)
Contro 2	-0.002 (0.003)	0.001 (0.003)	0.010 (0.008)
Intercept	0.001*** (0)	0** (0)	0*** (0)
adj. R^2	0.001	0	0

* $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$ Note: The coefficients are adjusted by multiplying variables by 1000; Control 1: past returns, cumulative abnormal return from the [-30,-2] trading window; Control 2: the abnormal return on the prior trading day

5.2. Impact of the Number of Followers

In Table 2, we take a more nuanced look at the impact of emotional valence by splitting the sample into two groups based on the number of followers of the tweet sender using 177 (median split), 1000, and 100,000 as the thresholds for the groups. The R^2 are similar to other studies [5, 9, 10, 40, 41].

We begin with H2a, which argued that same day returns would be more strongly influenced by tweets from users with many followers. With the median split (over and under 177 followers), we see that effects that were hypothesized: the beta coefficient on the valence of tweets from users with over 177 followers is higher than the beta coefficient on those under 177 followers for all three measures of valence (pos1, pos2 and neg1). The pattern using the 1,000 follower threshold is not as clear because the beta coefficients are very similar. The pattern using the 100,000 follower threshold is not what we hypothesized: tweets from those *under* the threshold have greater same day impact.

H2b argued that future returns would be more strongly influenced by tweets from users with few followers. With the median split (177 followers), we generally see the hypothesized effects: for next day returns (CAR1) the beta coefficients for pos1 and pos2 are not significant for users over this threshold, but are significant for users under this threshold. Negative valence is not significant for either number of users. For 10-day returns (CAR10), the beta coefficients on all three measures are greater for users under the threshold than for those over. The patterns for the 1,000 and 100,000 groups are similar.

We conclude H2a is partially supported – only when "many" followers means those over the median. We conclude that H2b is supported.

6. Discussion

Our study provides evidence that of a significant relationship between emotion in Twitter tweets and the future returns of individual stocks. The combined emotional valence of tweets tagged with a company's stock ticker is positively correlated with stock return on the same day as the tweets. Perhaps more importantly, the combined valence of the tweets can be used to predict the firm's stock return ten days after the tweets were posted.

We also found that the number of followers of the users sending the tweets moderated the relationship between the tweets' emotional valence and stock returns. In general, the valence of tweets from users with more followers than the median in our sample (177) had a stronger immediate same-day impact on stock returns compared to tweets from users with few followers. Thus Twitter users with many followers have a market impact similar to traditional news media; the impact of the emotional content in their tweets disseminates rapidly and is quickly incorporated into stock price.

Table 2. Statistical results for returns by speed of information diffusion

	Split using 177 followers			Split using 1,000 followers			Split using 100,000 Followers		
	Same Day Return	NextDay Return	10 Day Return	SameDay Return	NextDay Return	10 Day Return	SameDay Return	Next Day Return	10 Day Return
$Pos1^s$ _o	1.710***	-0.161	0.970**	1.139***	-0.200	0.527	0.761**	-0.210	-0.318
	(0.182)	(0.144)	(0.436)	(0.170)	(0.134)	(0.406)	(0.318)	(0.228)	(0.682)
$Pos1^s$ _u	0.500***	0.359***	1.456***	1.342***	0.308**	1.017**	2.523***	0.152	3.368***
	(0.165)	(0.131)	(0.397)	(0.190)	(0.150)	(0.454)	(0.485)	(0.348)	(1.041)
Control 1	0.001	0.001	0.013***	0.001	0.002***	0.016***	0.004*	0.002	0.038***
	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.003)	(0.002)	(0.002)	(0.005)
Control 2	0.004	0.001	0.018	0.006	0.003	0.022*	0.004	0.012*	0.040*
	(0.005)	(0.004)	(0.012)	(0.005)	(0.004)	(0.012)	(0.010)	(0.007)	(0.022)
Intercept	0**	0***	-0.003***	0	0***	-0.003***	0*	0***	-0.004***
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0.001)
adj. R^2	0.003	0	0.001	0.003	0	0.001	0.004	0	0.006
N	46354	46354	46354	45683	45683	45683	12646	12646	12646

	Split using 177 followers			Split using 1,000 followers			Split using 100,000 Followers		
	Same Day Return	NextDay Return	10 Day Return	Same Day Return	NextDay Return	10 Day Return	Same Day Return	NextDay Return	10 Day Return
$Pos2^s$ _o	1.710***	-0.125	0.608*	1.397***	-0.182	0.201	0.943***	-0.162	-0.251
	(0.182)	(0.113)	(0)	(0.147)	(0.116)	(0.352)	(0.307)	(0.220)	(0.660)
$Pos2^s$ _u	0.500***	0.326***	1.373***	1.085***	0.254**	0.974***	1.850***	-0.014	1.634**
	(0.165)	(0.108)	(0.327)	(0.146)	(0.115)	(0.348)	(0.311)	(0.223)	(0.668)
Control 1	0	0.001	0.013***	0.001	0.002***	0.016***	0.004	0.002	0.038***
	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.003)	(0.002)	(0.002)	(0.005)
Control 1	0.004	0.001	0.018	0.005	0.003	0.021*	0.003	0.012*	0.040*
	(0.005)	(0.004)	(0.012)	(0.005)	(0.004)	(0.012)	(0.010)	(0.007)	(0.022)
Intercept	0**	0***	-0.003***	0***	0***	-0.003***	-0.001***	0**	-0.003***
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0.001)
adj. R^2	0.003	0	0.001	0.005	0	0.001	0.006	0	0.005
N	46354	46354	46354	45683	45683	45683	12646	12646	12646

	Split using 177 followers			Split using 1,000 followers			Split using 100,000 Followers		
	Same Day Return	NextDay Return	10 Day Return	Same Day Return	NextDay Return	10 Day Return	Same Day Return	Next Day Return	10 Day Return
$neg1^s$ _o	-23.803***	2.478	-13.549*	-16.880***	2.404	-6.493	-9.533*	1.119***	2.672
	(2.969)	(2.346)	(7.122)	(2.848)	(2.246)	(6.802)	(5.316)	(0)	(11.399)
$neg1^s$ _u	-7.625***	-2.957	-22.374***	-21.452***	-2.666	-17.844**	-39.716***	7.600	-47.516***
	(2.682)	(2.119)	(6.435)	(3.100)	(2.444)	(7.404)	(5.316)	(6.061)	(18.133)
Control 1	0.001	0.001	0.013***	0.001	0.002***	0.016***	0.004*	0.002	0.038***
	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.003)	(0.002)	(0.002)	(0.005)
Control 2	0.005	0.001	0.018	0.007	0.003	0.022*	0.005	0.012*	0.040*
	(0.005)	(0.004)	(0.012)	(0.005)	(0.004)	(0.012)	(0.010)	(0.007)	(0.022)
Intercept	0.002***	0**	-0.001**	0.002***	0**	-0.002***	0.002***	-0.001***	-0.001
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0.001)
adj. R^2	0.002	0	0.001	0.002	0	0.001	0.003	0	0.005
N	46354	46354	46354	45683	45683	45683	12646	12646	12646

* $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$ Note: The coefficients are adjusted by multiplying by 1000; Control 1: past returns, cumulative abnormal return from the [-30,-2] trading window ; Control 2: the abnormal return on the prior day

In contrast, the emotional content in tweets from users with few followers takes much longer to spread through the investing public and thus it takes longer for stock prices to incorporate it. The valence of tweets from users with few followers had a stronger impact on stock returns over 10-day window compared to tweets from users with many followers.

We believe there are at least two possible theoretical explanations for these effects. The first is that the emotional valence of tweets “causes” changes in stock prices. Individuals post tweets when they believe have useful information about an individual stock. This information many have facts, as well as an underlying emotional content. Emotion is highly contagious [24, 39], and it influences how investors make buy/sell decisions on the stock as the emotion spreads through the public. A cumulative positive emotional valence triggers positive thoughts about the company and leads to a purchase decision, raising the stock price. A cumulative negative emotional valence induces negative thoughts and thus leads to sell decisions, decreasing the stock price.

A second possible explanation is that tweets "reflect" the underlying information that influences individual stock returns. In this case, it is not the emotional valence 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 emotions about the stock and communicate this emotion in their tweets. Likewise, investors planning to sell a stock communicate negative emotions in their tweets.

The economic magnitude of the relationships in our study are moderate to high [41]. For example, consider the magnitude of the negative valence (neg1) in equation (2). Using the untabulated standard deviation of negative valence of 0.000041768 and the regression coefficient of -21.613 for next 10 day abnormal return, we estimate the abnormal returns are 5.32 basis points lower after each one-standard deviation increase in neg1 valence. Likewise, using the untabulated standard deviation of pos1 valence of 0.000690443 and the regression coefficient of 0.912 for the same day abnormal return, we estimate the abnormal returns are 6.30 basis points higher after each one-standard deviation increase in pos1 valence. Both are greater than the -3.20 basis points found in Tetlock [41]. Thus the economic significance for pos1 and neg1 valence are quite similar.

Thus we conclude that emotion communicated via Twitter plays a significant and meaningful role in understanding current and future stock returns. The findings provide evidence that human behavior in the stock market can be understood by emotion. Yet, the lens of emotion has not been often studied by IS researchers. We believe that these results have important implications for research and practice.

One limitation is that there are 2,503,385 tweets in our sample. The large sample size may lead to significant results even though the relationships between are weak. A second limitation is that for our same day return analysis we did not determine whether the tweets occurred before or after the stock price changes; this limitation does not apply to our next day or 10-day analyses. Thus, it is difficult to say whether the positive and negative postings on Twitter caused the same day stock return or the same day stock market return causes people to post emotional content. We encourage readers to focus more on next day and 10-day returns for which this issue of temporal precedence does not exist.

6.1. Implications for Research

We believe this study opens a new door in predicting stock market returns. Recent research has shown that the calmness of emotion in tweets in general can be useful in predicting the future performance of a broad portion of the market (i.e., the DJIA) [5]. Our study shows that the emotional valence of tweets pertaining to specific stocks can be used to predict their current and future returns. We believe that this calls for more research into how emotion in social media can be used to better predict stock returns. Calmness and valence are only two aspects of emotion. There are many other aspects that bear investigation. The impact of emotional arousal may also have predictability power in stock return.

Prior studies that examined how emotional content predicts stock returns mainly focuses on the algorithm design in measuring the emotional content. The theoretical foundation behind the observed relationships has not been well established. We offered two possible explanations for the theoretical mechanism that links the emotional valence in tweets to future stock returns. We need more research to better understand the underlying theoretical mechanism that links emotion to stock returns.

Twitter users with many followers (over 177) have a stronger impact on same day stock return predictability and a weaker impact on future days return predictability than users with few followers (177). But this pattern was not found when we used a 100,000 follower split. It may be that the impact of "famous" users with over 100,000 followers is more similar to that of traditional media, (e.g., newspaper, TV). We need more research to understand why tweets from different users have different effects.

In this study, we used Twitter as the social media platform to predict stock returns. There are many other social media platforms that may also provide us insights in 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, such as SeekingAlpha?

We examined the impact of emotional valence on stock returns at a daily level. Future research could use market microstructural data to examine how emotions impact markets in real time.

6.2. Implications for Practice

Our results show that the future returns of a specific S&P 500 stock are related to the emotional valence of tweets about that stock. We believe that his study has three practical implications.

The first is providing guidance to individual investors. The emotional valence of tweets about specific stocks from users with many followers is directly related to their same-day return. Therefore, acting the same day on the emotions in those tweets may lead to higher returns, although there has been considerable debate about profits after incurring trading costs [31, 35].

The second implication is that the cumulative emotional valence of tweets from users with few followers is related to stock returns over a longer period (e.g., 10-day returns). Tweets from users with few followers are available publically and can be retrieved using Twitter development accounts. A cumulative analysis of the emotional valence of these may give important actionable insight about the future returns of that firm. Combining this with a focus on firms that have little coverage from the traditional media may also increase returns [18, 32].

A third implication is that firms need to 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 emotional valence of tweets can significantly influence the stock prices, firms need to monitor the emotions of their tweets in addition to the "rational" information they contain.

7. Conclusion

We found that the valence of emotion in social media postings is associated with same day abnormal returns and also has future abnormal return predictability. Interesting, postings from users with many followers have a greater impact on same-day returns, while postings from users with few followers have a greater impact on future returns. The findings are consistent with our hypothesis that private information that is diffused faster will be more quickly incorporated into prices, and will have higher association with same-day returns and lower future return predictability, while information that is diffused more slowly takes longer to be incorporated into prices and thus leads to greater future return predictability.

8. References

[1] E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne. "Finding high-quality content in social media". in *Proceedings of the international conference on Web search and web data mining*. 2008. ACM.

[2] G.A. Bakamitsos, "A cue alone or a probe to think? The dual role of affect in product evaluations", *Journal of Consumer Research*, 33, 3, 2006, pp. 403-412.

[3] Y. Bar-Anan, N. Liberman, and Y. Trope, "The association between psychological distance and construal level: Evidence from an implicit association test", *Journal of Experimental Psychology-General*, 135, 4, 2006, pp. 609-622.

[4] A. Bechara and A.R. Damasio, "The somatic marker hypothesis: A neural theory of economic decision", *Games and Economic Behavior*, 52, 2, 2005, pp. 336-372.

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

[6] J. Bollen, A. Pepe, and H. Mao. "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena". in *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. 2011.

[7] M.M. Bradley and P.J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential", *Journal of behavior therapy and experimental psychiatry*, 25, 1, 1994, pp. 49-59.

[8] J.T. Cacioppo, R.E. Petty, M.E. Losch, and H.S. Kim, "Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions", *Journal of Personality and Social Psychology*, 50, 2, 1986, pp. 260-268.

[9] H. Chen, P. De, Y.J. Hu, and B.-H. Hwang, "customers as advisors: the role of social media in financial markets", *Available at SSRN 1807265*, 2012.

[10] T. Chordia, R. Roll, and A. Subrahmanyam, "Order imbalance, liquidity, and market returns", *Journal of Financial economics*, 65, 1, 2002, pp. 111-130.

[11] Z. Da, J. Engelberg, and P. Gao, "In search of attention", *The Journal of Finance*, 66, 5, 2011, pp. 1461-1499.

[12] J. Daley, S. Jencks, D. Draper, G. Lenhart, N. Thomas, and J. Walker, "Predicting hospital-associated mortality for Medicare patients", *JAMA: the journal of the American Medical Association*, 260, 24, 1988, pp. 3617-3624.

[13] A. Dhir, K. Buragga, and A.A. Boreqqah, "Tweeters on Campus: Twitter a Learning Tool in Classroom?", *Journal of Universal Computer Science*, 19, 5, 2013, pp. 672-691.

[14] E. Dockery and M.G. Kavussanos, "Testing the efficient market hypothesis using panel data, with application to the Athens stock market", *Applied Economics Letters*, 3, 2, 1996, pp. 121-123.

[15] N.B. Ellison, "Social network sites: Definition, history, and scholarship", *Journal of Computer-Mediated Communication*, 13, 1, 2007, pp. 210-230.

[16] G.D. Eppen and E.F. Fama, "Cash balance and simple dynamic portfolio problems with proportional

costs", *International Economic Review*, 10, 2, 1969, pp. 119-133.

[17] T. Eyal, N. Liberman, Y. Trope, and E. Walther, "The pros and cons of temporally near and distant action", *Journal of Personality and Social Psychology*, 2004, pp. 781-795.

[18] L. Fang and J. Peress, "Media Coverage and the Cross-section of Stock Returns", *The Journal of Finance*, 64, 5, 2009, pp. 2023-2052.

[19] A. Fedorikhin and V.M. Patrick, "Positive mood and resistance to temptation: The interfering influence of elevated arousal", *Journal of Consumer Research*, 37, 4, 2010, pp. 698-711.

[20] R. Feldman, "Techniques and Applications for Sentiment Analysis", *Communications of the ACM*, 56, 4, 2013, pp. 82-89.

[21] N.H. Frijda, "Varieties of affect: Emotions and episodes, moods, and sentiments", *The nature of emotions: Fundamental questions*, 1994, pp. 197-202.

[22] K. Fujita, Y. Trope, N. Liberman, and M. Levin-Sagi, "Construal levels and self-control", *Journal of personality and social psychology*, 90, 3, 2006, pp. 351.

[23] T. Gordon, W.P. Castelli, M.C. Hjortland, W.B. Kannel, and T.R. Dawber, "Predicting coronary heart disease in middle-aged and older persons", *JAMA: the journal of the American Medical Association*, 238, 6, 1977, pp. 497-499.

[24] E. Hatfield, J.T. Cacioppo, and R.L. Rapson, "Emotional contagion", *Current Directions in Psychological Science*, 2, 3, 1993, pp. 96-99.

[25] C. Honey and S.C. Herring. "Beyond microblogging: Conversation and collaboration via Twitter". in *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on*. 2009. IEEE.

[26] H. Hong and J.C. Stein, "A unified theory of underreaction, momentum trading, and overreaction in asset markets", *The Journal of Finance*, 54, 6, 1999, pp. 2143-2184.

[27] H. Hong, T. Lim, and J.C. Stein, "Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies", *The Journal of Finance*, 55, 1, 2000, pp. 265-295.

[28] A.A. Labroo and V.M. Patrick, "Psychological distancing: why happiness helps you see the big picture", *Journal of Consumer Research*, 35, 5, 2009, pp. 800-809.

[29] A.O. Larsson, "Tweeting the Viewer—Use of Twitter in a Talk Show Context", *Journal of Broadcasting & Electronic Media*, 57, 2, 2013, pp. 135-152.

[30] N. Liberman and Y. Trope, "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, 1998, pp. 5.

[31] B.G. Malkiel and E.F. Fama, "EFFICIENT CAPITAL MARKETS: A REVIEW OF THEORY AND EMPIRICAL WORK*", *The Journal of Finance*, 25, 2, 1970, pp. 383-417.

[32] R.C. Merton, "A simple model of capital market equilibrium with incomplete information", *The Journal of Finance*, 42, 3, 1987, pp. 483-510.

[33] O. Oh, M. Agrawal, and H.R. Rao, "Community Intelligence and Social Media Services: A Rumor Theoretic Analysis of Tweets during Social Crises".

[34] D.V. Parboteeah, J.S. Valacich, and J.D. Wells, "The influence of website characteristics on a consumer's urge to buy impulsively", *Information Systems Research*, 20, 1, 2009, pp. 60-78.

[35] B. Qian and K. Rasheed, "Stock market prediction with multiple classifiers", *Applied Intelligence*, 26, 1, 2007, pp. 25-33.

[36] C. Qiu and C.W. Yeung, "Mood and comparative judgment: Does mood influence everything and finally nothing?", *Journal of Consumer Research*, 34, 5, 2008, pp. 657-669.

[37] J.A. Russell, "A circumplex model of affect", *Journal of Personality and Social Psychology*, 39, 6, 1980, pp. 1161-1178.

[38] T. Sakaki, M. Okazaki, and Y. Matsuo. "Earthquake shakes Twitter users: real-time event detection by social sensors". in *Proceedings of the 19th international conference on World wide web*. 2010. ACM.

[39] G. Schoenewolf, "Emotional contagion: Behavioral induction in individuals and groups", *Modern Psychoanalysis*, 15, 1, 1990, pp. 49-61.

[40] P.C. Tetlock, "Giving content to investor sentiment: The role of media in the stock market", *The Journal of Finance*, 62, 3, 2007, pp. 1139-1168.

[41] P.C. Tetlock, M. Saar-Tsechanksky, and S. Macskassy, "More than words: Quantifying language to measure firms' fundamentals", *The Journal of Finance*, 63, 3, 2008, pp. 1437-1467.

[42] A. Tumasjan, T.O. Sprenger, P.G. Sandner, and I.M. Welpe. "Predicting elections with twitter: What 140 characters reveal about political sentiment". in *Proceedings of the fourth international AAAI conference on weblogs and social media*. 2010.