European Financial Management, Vol. 20, No. 5, 2014, 926–957
doi: 10.1111/j.1468-036X.2013.12007.x

Tweets and Trades: the Information Content of Stock Microblogs

Timm O. Sprenger, Andranik Tumasjan, Philipp G. Sandner and Isabell M. Welpe
Technische Universität München, TUM School of Management, Germany
E-mail: timm.sprenger@gmail.com

Abstract

Microblogging forums (e.g., Twitter) have become a vibrant online platform for exchanging stock-related information. Using methods from computational linguistics, we analyse roughly 250,000 stock-related messages (so-called tweets) on a daily basis. We find an association between tweet sentiment and stock returns, message volume and trading volume, as well as disagreement and volatility. In contrast to previous related research, we also analyse the mechanism leading to an efficient aggregation of information in microblogging forums. Our results demonstrate that users providing above average investment advice are retweeted (i.e., quoted) more often and have more followers, which amplifies their share of voice.

Keywords: *Twitter*, *microblogging*, *stock market*, *investor sentiment*, *text classification*, *computational linguistics*

JEL classification: *G*12, *G*14

1. Introduction

Scholars and practitioners alike increasingly call attention to the popularity of online investment forums among investors and other financial professionals (Antweiler and Frank, 2004; BusinessWeek, 2009). Stock microblogging, mostly based on the social networking service Twitter (www.twitter.com), has recently been at the forefront of this development. Some commentators have even described the conversations on this platform as ‘the modern version of traders shouting in the pits’ (BusinessWeek, 2009). Twitter is a microblogging service allowing users to publish short messages with up to 140 characters, so-called ‘tweets’. These tweets are visible on a public message board of the website or through various third-party applications. Users can subscribe to

We thank John Doukas (the editor of this Journal) and an anonymous referee for insightful comments and suggestions that helped to improve this article. Correspondence: Timm O. Sprenger.

(i.e., 'follow') a selection of favorite authors or search for messages containing a specific key word (e.g., a stock symbol). The public timeline has turned into an extensive real-time information stream of currently more than 340 million messages per day generated by roughly 140 million active users (ZDNet, 2012). Many of these messages are dedicated to the discussion of public companies and trading ideas. As a result, there are investors who attribute their trading success to the information they find on social media websites and Twitter-based trading systems have been developed by financial professionals to alert users of sentiment-based investment opportunities (Jordan, 2010), and by academic researchers to predict break-points in financial time series (Vincent and Armstrong, 2010).

As yet, there is only little research examining whether and how microblogging messages are related to financial indicators. For instance, Sprenger *et al.* (2012) investigate how different company-specific news events published on Twitter (e.g., corporate governance or legal issues) are related to S&P 500 stock prices. Their comprehensive study shows that the content of Twitter messages provides valuable information with regard to the effects of different stock-related news types on company financial indicators. In a related vein, Sprenger and Welpe (2011) demonstrate that joint company mentions in Twitter messages predict their stocks' comovement. Moreover, their study indicates that this measure of company relatedness can also be used to delineate homogenous industry groups.

A few recent studies have made a first step in exploring whether the content of Twitter may help predict macroeconomic market indicators. For instance, Bollen *et al.* (2011) investigated whether collective mood states on Twitter are related to the value of the Dow Jones Industrial Average (DJIA). They find that certain mood states are indeed predictive of the DJIA closing values. In a similar vein, Zhang *et al.* (2010) explored the relationship between hope and fear on the one hand and the Dow Jones, NASDAQ, and S&P 500 on the other hand. Their results indicate that the level of tweet emotionality was significantly related to all three aggregated indicators. However, while these studies offer a first indication of the relationship between tweets and aggregated financial indicators, they are also limited in a number of ways. First, both studies use randomised subsamples of all available tweets of the Twitter message stream. Since the majority of these messages may arguably not be related to stock market topics, we cannot infer whether the stock-specific information contained in tweets is indeed associated with these indicators. Second, both studies only explored the relationship between aggregate sentiment measures and aggregate stock market indices, which does not allow us to draw conclusions on whether the information contained in stock-related messages is related to the performance of individual stocks. Since Das and Chen (2007) found the relationship between aggregated sentiment and index returns to be much stronger than the correlation for individual stocks, a more conservative approach focusing on the specific domain of stock-related messages and their relationship with the market prices of publicly traded companies (rather than relying on aggregated indices) is needed. Third, and most importantly, there is as yet no research investigating the mechanism underlying the link between social media message sentiment (e.g., tweet sentiment) and market prices. As a result, we do not know how information diffuses in social media in general, and Twitter microblogging in particular, leading to efficient information processing.

Our study addresses these limitations of prior related work and makes the following three main contributions to the literature. First, unlike previous related research (e.g., Bollen *et al.*, 2010; Zhang *et al.*, 2010), we analyse only microblogging messages with a direct reference to the stock market (i.e., we analyse only explicit stock microblogging messages rather than all available Twitter messages as in prior studies) which allows us to determine the predictive validity of stock microblogs without 'noise' that is unrelated to the stock market, as has been the case in prior studies (e.g., Bollen *et al*., 2010; Zhang *et al*., 2010).

Second, to the best of our knowledge, our study is the first to comprehensively explore the information content of stock microblogs with respect to individual stocks rather than aggregate stock market indices. In contrast to related previous studies, this study is able to go beyond the analysis of relatively simple measures of online activity (e.g., message volume or word counts), instead leveraging an innovative methodology from computational linguistics to evaluate the actual message content and sentiment. Moreover, this study replicates and extends similar research in the context of internet message boards without some of the previous methodological limitations (e.g., sample selection, timeframe; Antweiler and Frank, 2004). We analyse a more comprehensive set of stocks over the course of 6 months with fairly stable financial market activity.

Third, in contrast to the existing research that has merely shown a correlation of online message content with financial market indicators (e.g., Antweiler and Frank, 2004; Bollen *et al*., 2010; Zhang *et al*., 2010), our study goes beyond this correlational evidence by providing an explanation for the underlying mechanism that leads to the efficient aggregation of information in stock microblogging forums. Unlike these previous studies, we explicitly exploit the structure of the microblogging forum Twitter,which allows us to empirically explore theories of social influence concerning the diffusion and processing of information in the context of a financial community, which is not possible in traditional message boards (e.g., Antweiler and Frank, 2004).

In particular, the present study takes the following steps in extending the existing literature on the relationship between social media content and stock performance using the microblog forum Twitter. First, for comparability with prior related research (e.g., Antweiler and Frank, 2004), our study examines the relationship between the most important and heavily studied market features return, trading volume, and volatility and the corresponding tweet features message sentiment (i.e., bullishness),[1] message volume, and the level of agreement among postings. Second, we empirically explore the mechanism behind the efficient aggregation of information in microblogging forums. Specifically, we investigate the association between the quality of investment advice and the level of mentions, the rate of retweets, and the authors' followership.

2. Related Work and Research Questions

2.1 *Introduction to the research of online stock forums*

One of the most intriguing sources of unofficial and qualitative information is the vast amount of user-generated content online. In the context of the stock market, internet forums dedicated to financial topics, such as internet stock message boards like Yahoo! Finance, deserve special attention. A number of previous studies have investigated the relationship between stock message boards and financial markets. Wysocki (1998) was the first to investigate internet stock message boards. For the 50 most frequently discussed

[1] We use the terms sentiment and bullishness interchangeably.

firms on Yahoo!Finance between January and August 1998, he demonstrated that message volume did forecast next-day trading volume. Whereas this study only investigated message volume, others have taken a more differentiated approach to the information content on message boards. For a limited sample of internet service sector stocks, Tumarkin and Whitelaw (2001) have explored the information embedded in voluntary user ratings (from strong buy to strong sell), but were unable to show that these recommendations contain relevant information related to stock returns. Consistent with the Efficient Market Hypothesis (EMH), message board activity did not predict industry-adjusted returns and postings followed the stock market. Dewally (2003) has replicated this study in up and down markets demonstrating that recommended stocks had a strong prior performance which indicates that these traders follow a naïve momentum strategy.

All of these studies focused on readily available quantitative information (e.g., message volume, user ratings). However, this approach ignores much of the sample, because, for instance, only less than a quarter of all messages come with a user rating (Tumarkin and Whitelaw, 2001). In addition, this information does not capture the information content and sentiment of the actual messages. Moreover, evidence from stock message boards has shown that self-disclosed ratings are often biased. 'Hold' sentiments, for example, are systematically optimistic and significantly differ from neutral sentiments (Zhang and Swanson, 2010). In contrast to these relatively simple measures, automated classifiers can provide an unbiased interpretation of a message based on its content. Das and Chen (2007) have illustrated the use of natural language processing algorithms to classify stock messages based on input from human coders. In an explorative sample of 24 stocks they found only contemporaneous but no predictive relationships between message bullishness and market returns.

The study that is most closely related to ours, Antweiler and Frank (2004), used text classification methods to study the information content on both the Yahoo!Finance and Raging Bull message boards for the 45 companies of the Dow Jones Industrial Average and Dow Jones Internet Index. They demonstrate that message volume predicts trading volume and volatility. Its effect on stock returns was negative and, although statistically significant, economically small. However, this study is limited with regard to its sample period in the year 2000, which includes the burst of the internet bubble and dot-com companies with unsustainable business models representing a third of their sample[2]. Methodologically, the study focuses on real returns and does not examine potential differences between buy and sell signals. However, buy and sell signals may carry very different information with respect to subsequent stock returns and the true information value of online messages becomes apparent only when measured against market-adjusted abnormal returns (Tumarkin and Whitelaw, 2001). We address these limitations by differentiating between buy and sell signals. Moreover, since Antweiler and Frank's (2004) study used message boards, their approach did not permit them to investigate how information diffuses in the online investment community. Our approach, using stock microblogs, allows us to observe previously unavailable aspects of information diffusion among investors connected in social media. We thus extend the extant literature by exploiting the nature of microblogging forums (e.g., retweets and followership

[2] The burst of the internet bubble falls right into the middle of this sample with the Dow Jones Internet Index gaining almost 20% in the first quarter and losing half of its value in the last 4 months of the year.

relationships) which offers the unique opportunity to investigate whether and how stock (micro-)bloggers produce valuable information that may predict companies' market prices.

Whereas all of these studies have investigated internet stock message boards, the information content of stock microblogs with respect to financial markets is largely unexplored. The following three distinct characteristics of microblogging do not allow us to generalise previous results from stock message boards to stock microblogs for the following reasons. First, whereas message boards categorise postings into separate bulletin boards for each company, Twitter's public timeline may more accurately capture the natural market conversation and reflect up to date developments. Conversely, on stock message boards, outdated information may still receive attention as long as there are no more recent entries. Second, whereas message boards have an archival nature that requires users to actively enter the forum for a particular stock, Twitter reflects a more ticker-like live conversation. Message board users who do not actively enter the forum for a particular stock may not become aware of breaking news for that particular company, whereas stock microbloggers are usually exposed to the most recent information for all stocks. Third, unlike other financial bloggers who attract a readership by writing commentary and opinion pieces or message board users who can be indifferent to their reputation in the forum, microbloggers have a strong incentive to publish valuable information to maintain or increase mentions, the rate of retweets, and their followership. These factors may represent the Twittersphere's 'currency' and provide it with a mechanism to weigh information (Tumasjan *et al*., 2011). In addition to the differences to message boards, there is another characteristic of stock microblogs that deserves attention. Microblogging forums make previously unavailable aspects of information diffusion observable (e.g., retweets and followership relationships). However, prior research (e.g., Bollen *et al*., 2010; Zhang *et al*., 2010) has not yet explored whether these mechanisms that inevitably structure information diffusion are really used effectively to produce valuable information or whether stock microbloggers simply represent the online equivalent of uninformed noise traders with poor timing, herding behaviour, and overreaction to good or bad news. We address this limitation of prior related studies by explicitly investigating the underlying mechanisms that may account for this correlation.

2.2 *Research questions, related research and hypotheses*

In this section, we review related research, develop theory, and derive our hypotheses.

Bullishness. In much of the financial literature individual investors are considered the least informed market participants (e.g., Easley and O'Hara, 1987, Hirshleifer and Teoh, 2003) and empirical evidence shows that individual investors pay a significant performance penalty for active trading (Barber and Odean, 2000). Theoretical models suggest that 'informed investors with limited investment capacity [cannot fully exploit their advantage by trading, have private information left after trading and are] motivated to spread informative, but imprecise stock tips [because] followers trade on the advice and move prices' (van Bommel 2003, p. 1499) allowing the investor and the followers to fully capture the value of the private information. Tumarkin and Whitelaw (2001) have found no evidence that any information is embedded in voluntary disclosed user ratings (from strong buy to strong sell). Das and Chen (2007) report only a contemporaneous association between message bullishness and market returns. However, Hirshleifer and

Teoh (2003) show that, due to limited attention and processing power, disclosures with similar information can have different effects on investor perceptions and market prices (see also Barber and Odean, 2008). In particular, empirical studies have shown that investors are often influenced by word of mouth (e.g., Ng and Wu, 2006; Mizrach and Weerts, 2009; Hong *et al*., 2005). DeMarzo *et al*. (2003) have proposed a model of bounded rationality in which individuals are subject to persuasion bias and fail to account for repetition in the information they receive. As a result of this persuasion bias, 'influence on group opinions depends not only on accuracy, but also on how well-connected one is in the social network that determines communication' (p. 909). Given that stock microblogs reflect the theoretical properties of this model with the size of the followership indicating social influence, we propose:

Hypothesis 1: *Increased bullishness of stock microblogs is associated with higher returns.*

Note that we state our hypotheses as contemporaneous relationships between tweet and market features. However, in all cases we are also, in addition to that, interested in lagged relationships and examine the predictive information quality of tweet signals.

Message volume. Obviously, people may have a desire to post messages concerning the stocks in which they trade (van Bommel, 2003). In line with this argument, both Wysocki (1998) as well as Antweiler and Frank (2004) find that message volume can forecast next-day trading volume. On the other hand, online forums reflect primarily the activity of day traders, but not large volume institutional investors (Das et al., 2005). However, beyond the direct link between posting and trading, there are reasons to believe that an increase in message volume may even lead 'lurkers' to trade. Cao *et al*. (2002) have suggested that conversation among market participants induces trading from all kinds of so-called 'sidelined investors' who decide to trade as they learn that other traders share a similar signal. Since the message volume of stock microblogs should reflect this conversation, we expect:

Hypothesis 2a: *Increased message volume in stock microblogging forums is associated with an increase in trading volume.*

Increases in message volume indicate arrival of new information in the market. The vast majority of messages on internet message boards represent buy signals (Dewally, 2003). As a result, increases in message volume should be associated with increases in bullishness. While Antweiler and Frank (2004) report that the effect of message volume on stock returns was negative and, although statistically significant, economically small, there is empirical evidence from message boards supporting this notion. For the 50 most frequently discussed firms on Yahoo!Finance, Sabherwal *et al*. (2008) report that, in the case of internet message boards of thinly traded micro-cap stocks, the most talked about stocks were associated with high contemporaneous abnormal returns and statistically significant positive returns on the next day. Therefore:

Hypothesis 2b: *Increases in message volume in stock microblogging forums are associated with higher returns.*

Danthine and Moresi (1993) suggest that more information reduces volatility because it increases the chances of rational agents to counteract the actions of noise traders. Brown

(1999), however, provides empirical evidence that noise traders who act in concert can increase volatility. Antweiler and Frank (2004) show that, on internet message boards, message volume is a predictive factor of volatility. Koski *et al*. (2004) demonstrate that noise trading (used as a proxy for message volume) induces volatility, but note that the reverse causation is even stronger. In contrast to the EMH, theoretical models support the notion that trading of biased noise traders can be correlated on either the sell or the buy side of a particular stock and lead to an increase in volatility because the unpredictability of noise traders' beliefs creates a risk that deters arbitrageurs from correcting market prices (e.g., Black, 1986; De Long *et al*., 1990). Given that a large share of participants in stock microblogging forums consists of day traders, who document an increase in their trading activity through message volume, we derive:

Hypothesis 2c: *Increased message volume in stock microblogging forums is associated with higher volatility.*

Disagreement. Das *et al*. (2005) suggest that disagreement about market information leads to extensive debate and the release of more information. In line with Danthine and Moresi (1993) more information should reduce volatility. However, intuition suggests that disagreement and volatility should be positively correlated. Both theory and empirical evidence support the notion that volatility reflects the dispersion of beliefs among investors (e.g., Jones *et al*., 1994; Shalen, 1993). We derive:

Hypothesis 3a: *Increased disagreement among stock microblogs is associated with higher volatility.*

In line with the psychology literature, which suggests that uncertainty leads to an increase in communication activity (Newcomb, 1953), the traditional hypothesis in financial theory is that disagreement causes trading volume to rise because trading occurs when two market participants assign different values to an asset (Harris and Raviv, 1993; Karpoff, 1986; Kim and Verrecchia, 1991). Research on stock message boards is in line with this hypothesis as disagreement among online messages has been associated with increased trading volume (Antweiler and Frank, 2004). However, Milgrom and Stokey (1982) have developed the 'no-trade-theorem' suggesting that disagreement can reduce trading as the risk-averse participants of a trade are aware that the other party would only enter the trade to their advantage and any attempt to speculate on new, private information will impound this information in market prices. However, this theory is based on the assumption that 'new information is never small' and would instantly move market prices. Given that this is a rather strict assumption, which pertains even less to the large number of small day traders participating in stock microblogs, we would expect that:

Hypothesis 3b: *Increased disagreement among stock microblogs is associated with an increase in trading volume.*

Information diffusion. The hypotheses suggested above concern the much studied link between information and market developments. However, the mechanics underlying this link are largely unexplored. Microblogging forums make information processing partially observable. Thus, next to the investigation of tweet and market features, we analyse information diffusion among stock microblogs to explore whether microblogging forums weigh information effectively. In order to establish a link between information and returns, we investigate the relationship between the quality of investment advice and the level of mentions, the rate of retweets, and the author's followership.

Gu *et al*. (2008) have suggested that the interactions in message boards may create information aggregation and potentially lead to higher social welfare. While message boards and blogs have been questioned for their lack of objectivity and vulnerability to stock touting in classic 'pump and dump' trading strategies (Campbell, 2001; Delort *et al*., 2009), there are reasons to believe that microblogging forums produce higher quality information. Theoretical models have shown that online feedback mechanisms can serve as a sustained incentive for users to behave honestly (Fan *et al*., 2005). Microbloggers have an incentive to publish valuable information to maintain or increase mentions, the rate of retweets, and their followership – these features affect information diffusion in microblogging forums and provide the readers with a mechanism to weigh information. Studies have shown that user influence in terms of retweets and mentions is not simply driven by popularity in terms of followership (Cha *et al*., 2010). Boyd *et al*. (2010) have suggested that users often retweet messages to relay valuable content in order to validate and endorse a particular user or posting. In addition, there is empirical evidence that, despite the abundance of available information and considerable noise, Twitter users follow the accounts to which they subscribe closely and are highly attentive to their content. A working paper studying a single Twitter account making directional forecasts of the stock market indicates that the number of followers may be correlated with the accuracy of the published information (i.e., the forecasts of the stock market; Giller, 2009).[3] Moreover, reports in the business press suggest that microblogging forums may aggregate information more efficiently than previously studied online communities. Zhang (2009) has found poster reputation on a special internet stock message board with an explicit feedback mechanism to be determined, among other things, by information quality, not quantity. Due to increased information processing costs and potential information overload associated with more postings, internet stock message boards with less noise and more high quality postings attract more users (Gu et al., 2008). Following and 'unfollowing' an author virtually allows users of microblogging forums to construct their own customised message boards. We thus propose:

Hypothesis 4a: *Users who consistently provide high quality investment advice have more influence in the microblogging forum (indicated by retweets, mentions, or followers).*

Yang and Counts (2010) illustrate that, next to the properties of Twitter users, some properties of their messages (such as the inclusion of a hyperlink to another website), can predict greater information propagation. On the other hand, Romero *et al*. (2010) claim that the majority of users act as passive information consumers and do not forward the content to the network. Both studies were conducted with large, randomly sampled data sets and do not capture a specific domain such as stock microblogging. Therefore, next to high quality advisors, we examine whether high quality pieces of investment information (i.e.,

[3] However, the study is limited to one single account, which posted very explicit messages recording specific trading transactions and results (e.g., '16:14:47 BOT 9 $NQU 1427.5GAIN 15.58', p. 4). Our database is different in that we consider the vast majority of general messages containing mostly qualitative information including opinions and news items.

individual messages) are weighted more heavily and spread more widelythrough retweets. Thus, we propose:

Hypothesis 4b: *High quality pieces of microblogging investment advice are spread more widely than low quality pieces of advice (through retweets).*

3. Data Set and Methodology

3.1 *Data set and sample selection of stock microblogs*

We chose the microblogging platform Twitter as our data source for stock microblogs as opposed to other potential microblogging contexts (e.g.,Tumblr, Jaiku) because, as illustrated in the introduction, it has the widest acceptance in the financial community and all messages are accessible via the website's application programming interface (API). Currently, more than 340 million messages are posted on Twitter's public timeline every day (ZDNet, 2012). While there are few restrictions with respect to the format of these messages (except for the 140-character-limit), users have developed a number of syntax elements to structure the information flow. One of the most commonly used elements is the so-called hashtag (e.g., '#earnings'), which is a keyword included in many messages to associate (i.e., 'tag') them with a relevant topic or category and allows them to be found more easily. Similarly, traders have adopted the convention of tagging stock-related messages by a dollar sign followed by the relevant ticker symbol (e.g., '$AAPL'). Our study focuses on this type of explicit market conversation. This focus allows us to investigate the most relevant subset of stock microblogs and avoid 'noise'. We study the 6 month period between 1 January and 30 June 2010, to deal with stable developments on the US financial markets and to avoid potentially distorting repercussions of the subprime mortgage crisis in 2009. During this period, we have collected 249,533 English-language, stock-related microblogging messages containing the dollar-tagged ticker symbol of an S&P 100 company.[4] We focus on the S&P 100 to adequately reflect the entire spectrum of US equities, including a wide range of industries, while limiting our study to well-known companies that trigger a substantial number of stock microblogs.[5]

3.2 *Naïve Bayesian text classification*

In order to examine the relationship between signals from stock microblogs and market movements, we had to classify messages as either buy, hold or sell signals. As our data set contains too many messages for manual coding we have chosen to classify messages automatically using well established methods from computational linguistics. In line with Antweiler and Frank (2004), we employ the Naïve Bayesian classification method, one of

[4] Twitter provides only a limited history of data at any point in time. We, therefore, developed a webcrawler, which made requests to and downloaded data from the Twitter API 24 hours a day for a period of six months. A load balancing feature ensured that messages associated with more frequently mentioned stock symbols were downloaded more often.
[5] Specifically, we focus on those companies that have been included in the S&P 100 as of 1 January 2010.

the most widely used algorithms for supervised text classification. In short, the probability of a message belonging to a particular class depends on the conditional probability of its words occurring in a document of this class. These conditional probabilities are estimated based on a training set of manually coded documents. Compared to more advanced methods in computational linguistics, this method is relatively simple (e.g., high replicability and few arbitrary fine-tuning parameters), but has consistently shown robust results while providing a high degree of transparency into the underlying data structure. We use the multinomial Naïve Bayesian implementation of the Weka machine learning package (Hall *et al*., 2009).[6]

Input for our Naïve Bayesian model comes from a training set of 2,500 tweets, which we manually classified as either buy, hold, or sell signals.[7] Roughly half of these messages were considered to be hold signals (49.6%). Among the remainder, buy signals were more than twice as likely (35.2%) as sell signals (15.2%). This indicates that stock microblogs appear to be more balanced in terms of bullishness than internet message boards where the ratio of buy vs. sell signals ranges from 7:1 (Dewally, 2003) to 5:1 (Antweiler and Frank, 2004). Table 1 shows a few typical examples of the tweets from the training set including the manual coding and the automatic classification (for details on automatic classification, see Appendix).

As Table 2 shows, overall in-sample classification accuracy was 81.2%. The accuracy by class further validates the use of automatically labelled messages. For our purposes, falsely labelling a buy or sell signal as hold is more acceptable than falsely interpreting messages as buy or sell signals. The confusion matrix shows that the worst misclassification (of buy signals as sell signals and vice versa) occurs only rarely. In addition, the more balanced distribution of buy and sell signals compared to previous studies of internet message boards provides us with a greater share of sell signals in the main data set (10.0% compared to only 1.3% in the study of Antweiler and Frank, 2004). This permits us to explore the information content of buy and sell signals separately.

3.3 *Aggregation of daily tweet features*

In order to examine the relationship between hundreds of daily messages and the market movements on a daily basis, tweet features need to be aggregated to firm-specific variables. The focus of our study is on the market features return, trading volume, and volatility and the corresponding tweet features bullishness, message volume, and agreement.We follow Antweiler and Frank (2004) by defining bullishness as

$$B_t = \ln \frac{(1 + M_t^{Buy})}{(1 + M_t^{Sell})}, \tag{1}$$

[6] See our appendix for a detailed description of our classification method.

[7] In line with most text classification methods using a manual training set (e.g., Antweiler and Frank, 2004) we use one primary judge. The manual classification was reviewed by a second judge and critical cases were revisited and discussed to reach a consensus regarding their classification. For a subset of the training set, the second judge classified all messages independently. We observed a correlation of 0.92 with the first judge illustrating the robustness of the manual coding. Cohen's Kappa supports high interrater reliability (0.78).

Table 1

Sample tweets from training and test set including classification.

This table presents randomly selected sample tweets shown in their original format (before preprocessing). In the case of automatic classification, tweets are assigned to the class with the highest probability.

Sample tweets (training set)	Manual classification
RT @bampairtrading $TGT Target Q4 Profits Surge http://bit.ly/ciQFjY	Buy
Great place to short $X. Stop loss at 54.25. I am still short via puts from Friday HOD.	Sell
Big banks up or down with Bernanke's re-nomination? $C $BAC	Hold
$DELL (Dell Inc) $13.87 crossed its 1st Pivot Point Resistance #emppv #stocks http://empirasign.com/s/42f	Buy
Heinz Q3 EPS of 83c beats by 6c. Revenue of $2.6B meets. $HNZ #earnings http://bit.ly/avlHFH	Buy
Microsoft Corporation $MSFT Not Moving. Docuware Integration In Microsoft Outlook: http://bit.ly/db66Ox	Hold
$AXP looking strong here	Buy
$BA Boeing Sees Sales Drop, Maintains 737 Output http://bit.ly/9kmvUa	Sell
Trader Bots has recently calculated a Neutral Overall Stock Prediction on $TGT http://bit.ly/7k5H	Hold
I think if $AMZN closes above 116 today! You could go long tomorrow.	Buy

	Automatic classification		
Sample tweets (main data set)	***Buy***	***Hold***	***Sell***
Trader Bots has recently calculated a Bullish Overall Stock Prediction on $AA http://bit.ly/92SIuf	98.0%	2.0%	0.0%
$PFE raised quarterly div by 13% to 18 cents and said more annual increases are likely barring significant unforeseen events	100.0%	0.0%	0.0%
$COF, very strong the last few days but i'm sticking to my 2 week short. Here's my pretty chart doodle for download. http://bit.ly/6DDhFW	55.7%	5.2%	39.1%
$XOM ratings stand strong with $XTO acquisition http://twurl.nl/0c8vbm $$	97.7%	2.2%	0.1%
I just bought 12000 shares of General Electric Co ($GE) on @WeSeed http://tinyurl.com/dcevoo	100.0%	0.0%	0.0%
Merck CMO announcement strikes me as big deal and positive for $MRK. New, senior executive with proven drug development record from Merck.	100.0%	0.0%	0.0%
$CSCO - in depth, instant analysis for ANY stock - http://bit.ly/39XZdG	0.0%	80.7%	19.3%
New 52 wk high for $hpq	97.2%	2.7%	0.1%
sold $30% of my $AMD position at 9.29…	0.7%	20.7%	78.6%
Anyone ready to short $NVDA? Looks to be getting ahead of itself a bit here. Thoughts?	0.5%	8.3%	91.3%

where M^{Buy} (M^{Sell}) represents the number of buy (sell) signals on day t.[8] This measure reflects both the share of buy signals as well as the total number of messages giving greater weight to a more robust larger number of messages expressing a particular sentiment.

Message volume is defined as the natural logarithm of the total number of tweets per day.[9] In line with Antweiler and Frank (2004), agreement among messages is defined as

$$A_t = 1 - \sqrt{1 - \left(\frac{M_{ct}^{Buy} - M_{ct}^{Sell}}{M_{ct}^{Buy} + M_{ct}^{Sell}}\right)^2}. \tag{2}$$

If all messages are either bullish or bearish, agreement equals 1.

Even after the aggregation of individual messages to daily indicators, there are days for some stocks without any tweets. In the absence of messages, we define all three tweet features for these silent periods as zero following Antweiler and Frank (2004).[10] However, since our data set contains a full set of both tweet and market features for more than 80% of all company-day-combinations, the influence of silent periods on our results is limited.

Finally, because we use financial data from the NASDAQ and NYSE, we align messages with US trading hours (9:30 am to 4:00 pm) by assigning messages posted after 4:00 pm to the next trading day, in line with Antweiler and Frank (2004). Thus, messages posted after the markets close are included together with pre-market messages in the calculation of tweet features for the following day because these messages can only have an effect on the market indicators of that day or be affected by other factors that are not apparent in the market indicators until the next day.

[8] We conducted all our analyses also with two alternative measures of bullishness, the simple share of buy vs. sell messages and the surplus of buy messages. While both of these measures lead to very similar findings, the logged bullishness measure outperforms these two, so we only report these results.

[9] The log transformation $\ln(1 + M_t)$ is analogous to the transformation of the trading volume allowing us to compute elasticities and control for scaling. There are two concerns with this volume measure. First, given the growth of microblogging forums such as Twitter, the total volume may not be a stable indicator over time. Second, the message volume may vary slightly due to crawling efficacy. Therefore, for each company, we also computed a normalised message volume relative to the total number of daily messages. While this indicator provides a comparable measure of the relative share of postings for each company, it does not reflect the absolute volume. This normalised relative message volume shows a much weaker correlation with the trading volume. Despite possible shortcomings of the absolute volume measure, this indicator still contains more information with respect to changes in the trading volume, which we are giving up in the case of normalisation. We, therefore, use the logged version when we refer to message volume in the remainder of our paper.

[10] We have explored two alternatives by either maintaining missing values as such or filling silent periods with medians of the respective tweet features. All results are very similar, so we only report the treatment of silent periods as zeroes, in line with the results reported by Antweiler and Frank (2004).

Table 2

Classification accuracy (confusion matrix).

This table shows the accuracy of our classification method (for example, 35.2% of the training set were manually labelled as buy signals, 28.0%. of which were correctly classified as such by the model). In the case of classification accuracy by class, it can be seen that the vast majority of messages in the training set were classified correctly by our model. True positives (or precision) represent, for example, the share of messages classified as Buy, which were labelled as such in the training set. False positives are messages classified incorrectly as Buy. Recall represents the share of all messages of a particular class, which were classified correctly. The F-measure combines precision and recall and is calculated as F = (2*recall*precision)/(recall + precision). The ROC (receiver operating characteristic) area measures the quality of the trade-off between true and false positives (i.e., the area under the curve plot of true and false positives).

		Automatic classification		
		Full training set		
Manual classification		Buy	Hold	Sell
Buy	35.2%	**28.0%**	5.9%	1.3%
Hold	49.6%	5.2%	**40.7%**	3.7%
Sell	15.2%	0.6%	2.1%	**12.5%**
Training set		33.7%	48.7%	17.6%
All messages		23.0%	67.0%	10.0%

Classification accuracy (accuracy by class)

Class	True positives	False positives	Recall	F-measure	ROC area
Full training set					
Buy	79.5%	8.8%	83.1%	81.2%	93.5%
Hold	82.1%	15.8%	83.6%	82.8%	92.0%
Sell	82.5%	6.0%	71.2%	76.4%	96.4%
weighted average	81.2%	11.9%	81.5%	81.3%	93.2%
10-fold cross validation					
Buy	62.8%	18.0%	65.6%	64.2%	79.0%
Hold	68.7%	30.9%	68.6%	68.6%	74.9%
Sell	52.7%	10.1%	48.3%	50.4%	80.2%
weighted average	64.2%	23.2%	64.4%	64.3%	77.2%

3.4 *Financial market data*

We have downloaded financial data in daily intervals for the S&P 100 from Thompson Reuters Datastream. Although Antweiler and Frank (2004) applied log difference of total return to shareholders we decided to use standard returns to allow for an easier interpretation. We are primarily interested not in absolute returns, but excess returns. Therefore we compute abnormal returns defined as

$$AR_{it} = R_{it} - E(R_{it}), \tag{3}$$

where R_{it} is the excess return for stock i on day t (i.e., subtracting the risk free rate on day t from the raw return; specifically, we have obtained Daily Treasury Bill Rates and deducted the coupon equivalent rate from raw returns) and $E(R_{it})$ is the expected return of the stock. In a simple version the expected return is the return of the relevant market index, so that

$$AR_{it}^{simple} = R_{it} - (R_t^{market}) \tag{4}$$

with the S&P 100 index serving as our benchmark for the market return. This simple abnormal return calculation does not reflect a stock's distinct market risk. Therefore, we also estimate the expected return based on an OLS regressed market model ($AR^{market\ model}$) as

$$E(R_{it}) = \alpha_i + \beta_i(R_{mt}) + \mu_{it}\, for\ t = 1, 2, ..., T, \tag{5}$$

where α_i is the intercept term, β_i is the association between stock and market returns, μ_{it} is the standard error term and T is the number of periods in the estimation period. In line with common practice (e.g., Dyckman *et al.*, 1984), we use a 120-day estimation period starting 130 days prior to the relevant date. Cumulative abnormal returns are calculated as

$$CAR_{it} = \sum AR_{it} \tag{6}$$

and average cumulative abnormal returns for N companies are calculated as

$$ACAR_t = \frac{\sum_{i=1}^{N} CAR_{it}}{N}. \tag{7}$$

Average abnormal returns (AAR) are computed identically with abnormal returns taking the place of cumulative abnormal returns.

Trading volume is the logged number of traded shares. We estimate daily volatility based on intraday highs and lows using the well-established PARK volatility measure (Parkinson 1980), defined as

$$VOL^{PARK} = \frac{(\ln(H_t - \ln(L_t)))^2}{4\ln(2)}, \tag{8}$$

where H_t and L_t represent the daily high and low of a stock price.

3.5 *Information aggregation in microblogging forums*

In order to explore whether high quality investment advice is attributed greater weight in stock microblogs, we define one measure of quality and three measures of influence.

Every tweet in our sample is classified as a recommendation to buy, hold, or sell a stock. We code this sentiment s of buy, hold, and sell signals as 1, 0 and -1, respectively. In line with Zhang (2009), who studied the determinants of poster reputation on online message boards, we define the quality of a tweet as the accuracy of this recommendation relative to same-day returns[11] of the stock in question as

$$quality = \begin{cases} = 1 \ if \ \frac{S_{it}}{R_{it}} > 0 \\ \hline = 0, otherwise \end{cases}, \tag{9}$$

where s_{it} is the sentiment of a message on day t associated with stock i. We only take into account messages published during trading hours and ignore hold messages in the computation of quality scores.[12] Next to the quality of individual messages, we also compute the quality of a particular user's investment advice as the average quality of all messages posted by this user. In addition, we compute the average sentiment of a user's messages.

In the context of microblogging, Cha *et al.* (2010) have defined three different measures of user influence: retweets, mentions, and followership. The first measure (i.e., the fact whether a message was retweeted) can also serve as a proxy for the weight given to an individual tweet. Microblogging users frequently forward (i.e., 'retweet') messages which they find noteworthy to their followers. The retweets usually contain the abbreviation 'RT' followed by the name of the original author.[13] The first sample tweet in Table 1 provides an example of such a retweet. Because Twitter does not provide information regarding the relationship of individual tweets, we identified retweets in our data set by filtering all retweets and matching the 40 characters following the retweet token and the name of the original author with all other tweets in the data set.[14] This allows us to separate retweets from non-retweets and identify the originals alongside the frequency with which they were retweeted. Second, next to retweets, users can be credited by mentioning their name (e.g., 'I think @peter is right on $AAPL'). Mentions increase the user's exposure on the public timeline. For every username in our sample we, therefore, extract the number of mentions. Regarding the third measure, users of microblogging forums subscribe to (i.e., 'follow') a selection of favourite authors whose messages appear in reverse chronological order on their home screen. Thus, the number of followers is a good indicator of a user's regular readership. We measure the number of followers for all users in our sample at the end of our sample period.[15] Having laid out the definition of our variables, we now turn to exhibiting and interpreting the results.

---

[11] Mizrach and Weerts (2009) make the same assumption and close positions announced in a public internet chat room at the end of the day.

[12] People searching for investment advice online are arguably interested primarily in buy or sell recommendations. In addition, daily returns are rarely zero and any other range of returns, defined to justify a hold recommendation to be correct, would be arbitrary.

[13] Alternative formats include 'RT: @', 'via @', 'by @', and 'retweet @'.

[14] We limit this match to 40 characters for two reasons. First, users often append their own commentary altering parts of the original tweet. Second, contrary to common practice, in some cases the retweet token is not placed directly at the beginning of the tweet, leaving fewer characters for the match.

[15] We understand that the total number of followers at any point in time needs to be interpreted with caution. Followership is not necessarily a direct measure of the quality of content. But even relative measures, such as the growth in followership, can be misleading because base rates can vary substantially depending on when a user joined Twitter. Even though we recognise this limitation and interpret related results with caution, we find followership too relevant a measure to be ignored altogether.

4. Results

4.1 *Descriptive statistics*

We have collected 249,533 stock-related microblogging messages containing the dollar-tagged ticker symbol of an S&P 100 company. Ranging from 342 to 4051 daily postings, this represents an average of 2,012 tweets per trading day with a standard deviation of 718 messages. An average of more than 20 tweets per day and company indicates that our data set comprises a dense information stream. Three quarters of the companies in our sample receive an average of at least 3 (and only one company less than one) mentions per trading day. The summary descriptive statistics of market and tweet features are displayed in Table 3.

4.2 *Overall relationship of tweet and market features*

Pairwise correlations are displayed in Table 4. Trading and message volume shows a fairly strong correlation (r $= 0.441, p < 0.001$). Message volume thus tracks the rise in trading volume closely. The association of market returns and a market-cap weighted bullishness index exhibits a weaker but nevertheless statistically significant correlation between the two indicators (r $= 0.165, p < 0.001$).

Contemporaneous regressions. While the pairwise correlations suggest interesting relationships between tweet features and market features, they do not address the interdependence of these relationships. It remains unclear whether these relationships

Table 3

Summary statistics of market and tweet features.

This table presents the summary statistics of market and tweet features. All daily statistics are reported on a per company basis. Returns were defined as the standard returns (in contrast to the log differences in prices as used by Antweiler and Frank, 2004), traded volume represents the number of shares traded, and we use PARK volatilities derived from daily highs and lows. Bullishness is the sentiment of a particular message, message volume the number of daily messages per company, and agreement the concurrence of messages for a particular company with respect to their sentiment (e.g., buy vs. sell). All returns (for individual stocks as well as market indices) are scaled by 100 (i.e., shown in percent) and PARK volatility is scaled by 10,000 for easier readability. N = 10,123 company-days with tweet features and 12,443 company-trading-days.

Variable	Mean	Std. Dev.	Minimum	Maximum
Market features				
Return	−0.004	0.019	−0.152	0.116
Traded volume	24,807	79,272	497	1,864,159
Volatility	3.33	32.01	0	3,249.05
Tweets features				
Bullishness	0.39	0.73	−3.35	3.69
Message volume	20.05	63.00	0	1,543
Agreement	0.38	0.45	0	1

Table 4

Pairwise correlations for stock and tweet data.

This table shows the pairwise correlations for tweet and market features. P-values are reported below the correlation, correlations that are significantly different from zero at the 99% confidence level are reported in bold. Abnormal returns are calculated in a simple version (AR_{simple}) adjusted by market development and in a version based on market model reflecting the stock's distinct market risk ($AR_{market\ model}$). Traded volume represents the number of shares traded,volatility refers to PARK volatilities derived from daily highs and lows, bullishness is the sentiment of a particular message, message volume is the number of daily messages per company, agreement is the concurrence of messages for a particular company with respect to their sentiment (e.g., buy vs. sell), and return is defined as the standard returns. N = 12,443 company-trading-days.

	Return	AR (simple)	AR (market model)	Traded volume	Volatility	Bullish-ness	Message volume
AR (simple)	**0.755**						
	0.000						
AR (market model)	**0.659**	**0.918**					
	0.000	0.000					
Traded volume	−**0.038**	0.012	**0.033**				
	0.000	0.197	0.000				
Volatility	−**0.060**	−0.013	−0.010	**0.046**			
	0.000	0.154	0.258	0.000			
Bullishness	**0.165**	**0.154**	**0.144**	**0.126**	−0.012		
	0.000	0.000	0.000	0.000	0.195		
Message volume	**0.027**	0.020	0.020	**0.441**	0.016	**0.340**	
	0.003	0.024	0.026	0.000	0.077	0.000	
Agreement	**0.035**	**0.032**	**0.030**	−**0.113**	−0.014	**0.362**	−0.016
	0.000	0.000	0.001	0.000	0.125	0.000	0.083

remain significant when all other tweet features are controlled for. Thus, in this section, we investigate the contemporaneous[16] relationships between tweet and market features corresponding to our hypotheses in order to investigate whether tweet features can serve as proxies for market developments.Table 5 shows contemporaneous fixed-effects panel regressions of the market features as the dependent variable and the three tweet features as independent variables.

The market index is used as a control variable. Due to significant cross-sectional differences in message volume, we use fixed-effects for each stock. In line with H1, the regression results support the strong relationship between bullishness and all three return measures. Thus, increased bullishness is associated with rising stock prices. In addition, supporting H2a, we find a relationship between message volume and trading volume. This result strengthens our hypothesis that users post messages concerning stocks that are traded more heavily. Since both volume measures were log transformed, we can interpret the coefficients as elasticities. A 1% increase in the message volume is associated with a

[16] Contemporaneous refers to the contemporaneity of tweet and market features.

Table 5

Contemporaneous regressions.

This table shows the power of tweet features for explaining changes in market features. The first row shows these market features as the dependent variables in panel regressions with company fixed-effects. All tweet features are used as independent variables and the market return is added as a control. Bullishness is the sentiment of a particular message, message volume is the number of daily messages per company, agreement is the concurrence of messages for a particular company with respect to their sentiment (e.g., buy vs. sell), and market return is the market return based on S&P 100. Return represents standard returns, AR_{simple}refers to the simple calculation of abnormal returns adjusted by market development, $AR_{market\ model}$ represents abnormal returns based on market model reflecting the stock's distinct market risk, trading volume represents the number of shares traded, volatility refers to PARK volatilities derived from daily highs and lows, and Amihud illiquidity represents Amihud's (2002) illiquidity measure defined as daily returns divided by daily trading volume across a five days moving average. N = 12,443 company-trading-days.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, t-statistics in italics below the coefficients.

	Dependent variable					
	Return	AR^{simple}	$AR^{market\ model}$	Trading volume	Volatility	Amihud illiquidity
Bullishness	0.005***	0.003***	0.003***	−1.872***	−0.593	0.155***
	18.46	*17.72*	*16.96*	*−3.65*	*−1.27*	*8.63*
Message volume	0	0	0	10.798***	1.391***	0.063***
	0.08	*0.90*	*0.51*	*28.68*	*4.06*	*4.82*
Agreement	−0.002***	−0.001***	−0.001***	−4.644***	−0.839	−0.05
	−4.31	*−4.45*	*−3.99*	*−5.88*	*−1.17*	*−1.81*
Market return	0.001***	−0.000***	−0.000***	−2.057***	−0.160*	0.003
	13.53	*−3.32*	*−4.94*	*−26.32*	*−2.24*	*1.15*
R^2	0.046	0.028	0.027	0.104	0.002	0.011
F-value	147.1	89.9	84.6	358.7	5.5	34.3

more than 10% increase in trading volume (*c* = *10.798, p* < *0.001*). Contrary to H2b, we find no relationship between message volume and returns, which is in contrast to previous research (e.g., Wysocki, 1998). In line with H2c, we observe an increase in volatility as the message volume rises (c = *1.391, p* < *0.001*). Whereas we find no support for H3a, i.e., disagreement does not explain volatility, H3b, i.e., a negative correlation between agreement and trading volume is supported (c = *-4.644, p* < *0.001*). As disagreement might be associated with illiquidity in the market, we also regressed the Amihud illiquidity measure (Amihud, 2002) on the tweet features. As this metric is computed for a specific time window, we use a five day moving average to compute the Amihud illiquidity measure. Here, disagreement does not appear to be associated with liquidity in the market, yet we find an interesting effect with regard to the tweet features bullishness and message volume: both tweet features are associated with reduced liquidity in the market.

Overall, we conclude that the contemporaneous relationships between bullishness and returns, message volume and trading volume, as well as agreement and trading volume appear to be the most robust.

Fama-MacBeth cross-sectional regressions. While contemporaneous relationships between tweet and market features are noteworthy, the litmus test for the quality of information in microblogs is analysing the temporal structure of both market and tweet features. If microblogs contain new information not yet reflected in market prices, tweet features should anticipate changes in market features. Therefore, in this section, we explore the lagged relationships between tweet and market features corresponding to our hypotheses. In order to evaluate the direction of the effect, we analyse all relationships in both directions in line with Antweiler and Frank (2004). In the following, we focus on those hypotheses that have not yet been rejected by previous analyses.

Antweiler and Frank (2004) relied on time-sequencing regressions to analyse the temporal structure of their variables. Yet, the following main three disadvantages arise from time-series regressions. First, an omitted return factor from the benchmark for abnormal returns could also explain the results. Second, an unusual time trend or other time-fixed effect could explain the results – unless one explicitly controls for this possibility. Third, one needs to use clustered standard errors to account for the lack of independence in firms' stock returns. Therefore, to circumvent these disadvantages, we employ Fama-MacBeth cross-sectional regressions. In order to proxy companies' size we also include the logarithm of market capitalisation. We regress one and two day lags of every tweet feature on every market feature separately (and vice versa) and report standardised coefficients.

An obvious question in the analysis of microblogs and the market is whether message sentiment can help predict returns (H1). The results (displayed in Table 6) show that bullishness cannot be used to predict returns (see Table 6, column X-> Y). Yet, the standardised coefficient shows that the effect of returns on bullishness is positive and significant (see Table 6, column Y-> X). Not surprisingly, the coefficient of the one day-lagged variable is higher than the effect of the two days-lagged variable ($c = 0.132, p < 0.001$ vs. $c = 0.038, p < 0.05$). Thus, returns affect bullishness in stock microblogs, but not vice versa.

In line with H2a, message volume one and two days ago predict current day trading volume. At the same time, high trading volume triggers increased message volume over the following days. The standardised coefficients illustrate that the stronger effect is in the direction from trading volume to message volume – but only for a one day-lag. Contrary to H2b (and prior research, i.e., Antweiler and Frank, 2004), we do not find message volume to be related to stock returns. This result indicates that investors may take a more nuanced approach in processing information content of stock microblogs compared to message boards. In line with H2c, high volatility also leads to increased message volume, supporting the notion that uncertainty causes investors to exchange information and consult their peers. However, the opposite relationship does not hold.

As we now add temporal structure in comparison to the previous section, higher volatility contributes to disagreement, supporting H3a, but not vice versa. However, we find supporting evidence for H3b, i.e., that disagreement among traders leads to higher trading volumes.

In summary, we conclude that while some few tweet features appear to contain predictive information with respect to market features (especially disagreement for trading volume), the standardised coefficients show a much stronger effect of market features on tweet features.

Table 6

Fama-MacBeth cross-sectional regressions.

This table shows lagged regressions for tweet features (X) explaining market features (Y) in the columns labeled X-> Y and the inverse relationship in the columns labelled Y-> X. The first row for each combination of tweet and market feature reports the regression coefficients for the standardised values (bold), the second row reports the standard errors (italics). Method: Fama-MacBeth cross-sectional regressions. Bullishness is the sentiment of a particular message, message volume is the number of daily messages per company, agreement is the concurrence of messages for a particular company with respect to their sentiment (e.g., buy vs. sell), and return represents standard returns. Market cap. = Market capitalisation. N = 12,443 company-trading-days.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

		X->Y				Y->X			
X	Y	X-1	X-2	Market Cap.	F-value	Y-1	Y-2	Market Cap.	F-value
Bullishness	Return	**0.005**	**−0.013**	**0.001**	1.0	**0.132*****	**0.038***	**0.238*****	167.6***
		0.561	*−1.693*	*0.053*		7.772	2.417	20.805	
Bullishness	Volume	**0.038*****	**0.024***	**0.529*****	8279.1***	**0.123*****	**−0.079***	**0.216*****	157.8***
		4.108	*2.47*	*129.515*		3.41	−2.289	19.55	
Bullishness	Volatility	**0.002**	**−0.013**	**−0.018*****	94.2***	**0.32**	**0.05**	**0.243*****	144.3***
		0.717	*−1.035*	*−5.910*		1.878	0.353	20.312	
Messages	Return	**−0.002**	**−0.011**	**0.008**	0.3	**0.032**	**0.018**	**0.587*****	809.0***
		−0.172	−0.684	0.491		1.66	1.112	49.104	
Messages	Volume	**0.189*****	**0.120*****	**0.362*****	5895.1***	**0.312*****	**−0.061**	**0.448*****	939.8***
		12.879	8.325	59.492		8.571	−1.633	40.595	
Messages	Volatility	**−0.007**	**0.015***	**−0.024***	85.9***	**2.157*****	**1.484*****	**0.636*****	856.7***
		−0.375	2.555	−2.580		10.515	7.339	50.314	
Agreement	Return	**0.017**	**0.002**	**−0.001**	1.3	**0.038***	**0.018**	**−0.065*****	14.9***
		1.968	0.260	−0.114		2.422	1.322	−5.964	
Agreement	Volume	**−0.076*****	**−0.076*****	**0.529*****	7000.7***	**−0.079***	**−0.042**	**0.006**	41.4***
		−7.943	−8.178	139.348		−2.197	−1.177	0.631	
Agreement	Volatility	**−0.009*****	**0.009**	**−0.021*****	89.1***	**−0.855*****	**−0.818*****	**−0.089*****	26.8***
		−4.321	0.626	−13.488		−5.098	−5.102	−7.476	

Table 7

Frequency distribution of users and messages by user group.

This table shows the message volume and average quality for various user groups by posting frequency. Message volume is defined as the total number of messages posted during our sample period by a user group. Average quality is the percentage of correct stock predictions interpreting all messages classified as buy or sell signals as end of day prediction for the stock that is mentioned.

	Users		Message volume		
Tweets per user	Total	Share	Total	Share	Average quality
1	10,604	67.3%	27,601	11.1%	52.7%
2	1,790	11.4%	9,318	3.7%	54.0%
3-4	1,222	7.8%	10,615	4.3%	53.3%
5-9	947	6.0%	15,935	6.4%	52.1%
10-19	513	3.3%	18,337	7.3%	54.6%
20-49	415	2.6%	33,664	13.5%	55.6%
50+	235	1.5%	134,063	53.7%	55.0%

4.3 *Information diffusion in stock microblogging forums*

We have seen that stock microblogs do contain valuable information with respect to stock performance. However, with thousands of average day traders participating in these forums, the question is how the information stream as a whole becomes informative. One possible answer is that information is weighted effectively by online users. In this section, we explore whether good investment advice receives greater attention. We investigate two aspects: First, we investigate whether we can identify above average advisors (i.e., market mavens or investment gurus) and whether these users receive greater attention in the online community through higher levels of retweets, mentions or followers (H4a). Second, we examine whether high quality pieces of information (i.e., individual messages) may be weighted more heavily and spread through retweets (H4b).

Regarding H4a, Yang and Counts (2010) have shown that the properties of users may significantly predict information propagation in Twitter. Users have an interest to subscribe to the content of high quality investment advisors. While they may not constantly identify high quality pieces of information in the message stream, they may notice and pay more attention to market mavens who consistently provide good investment advice. If these investment gurus had more followers, their contributions would find a larger audience. Table 7 shows the distribution of users and messages across various user groups according to the frequency with which a user posts messages.

In line with previous research, participation is highly skewed. While two thirds of all users have only posted one stock-related message in our sample period, the 1.5% heavy users are responsible for more than 50% of all contributions.[17] But, as the last column

[17] We can only observe user names and, for simplicity, refer to these as users. While a person may maintain multiple accounts, we have no reason to believe that this practice is common enough to affect our findings.

indicates, the higher frequency users do not appear to be better investment advisors as the average quality does not vary by user group. However, even among users with hundreds of messages, we can identify some that seem to consistently provide higher quality investment advice than others with more than three quarters of messages containing correct predictions.

Next, we explore whether this quality is recognised in the microblogging forum in the form of retweets, mentions, or followership. We use these three variables as dependent variables in regressions with user quality and include all control variables used by Zhang (2009), which are relevant to our context, or their microblogging equivalents.[18] Zhang (2009) found the number of watch lists to which the poster had been added to explain poster reputation. Watch lists are lists of favorite authors and represent an indicator of popularity. In a sense, watch lists (i.e., followership relationships) are the very fabric of microblogging forums. Therefore, we add the number of followers as a control variable. The followers are the most immediate recipients of an author's tweets and a larger audience should increase the chances of a message being retweeted. Next to the followership, the total message volume provides exposure to a user's messages. Thus, we include it as a control. In addition, Zhang (2009) reports that the average sentiment affected a user's reputation, with more bullish users gaining higher reputation scores. We therefore compute the average sentiment for a user's messages coding buy, hold, and sell signals as 1, 0 and -1, respectively. Zhang (2009) has shown that, while accuracy with respect to same-day returns did not affect a user's reputation, one-day follow opinions has a positive effect (i.e., buy recommendations for stocks that had risen the day before). We add this 'lagged' accuracy to our model. Obviously, retweets are correlated highly with user mentions ($r = 0.854,\ p < 0.001$) and follower count ($r = 0.44,\ p < 0.001$). Therefore we run separate regressions for the three indicators. Most users only dedicate a small fraction of their messages to stock-related issues. Hence, we follow Cha *et al.* (2010) in limiting the analysis to 'active users' by restricting our sample to 'serious' stock microbloggers with at least 20 messages in our sample period (the two highest frequency groups shown in Table 7). On the other hand, the followership includes many users that are not necessarily subscribing to the stock-related content of a particular user. Therefore, next to the total number of followers, we have also downloaded the entire network structure of all users in our sample consisting of more than 8.8 million follower relationships and labelled stock microbloggers separately.[19] On average, the users in our sample have more than 1,500 followers. Interestingly, the share of peers among the followership of serious stock microbloggers is less than 2% (1.8%). Thus the community of stock microbloggers appears to be not particularly tight knit. Table 8 shows the effect of the determinants of the three indicators of user influence.

---

[18] We do not use average message length because the 140-character limit of microblog srenders it useless as a mark of distinction between tweets.

[19] The tweets in our sample were created by roughly 15,700 different users. We were able to download user information for about 14,200 and the network of followers for about 13,600 users, because some users delete their account and others activate a privacy protection option that prevents public access to their data. In addition, some accounts are suspended by Twitter itself.

Table 8

Determinants of user influence.

This table shows the power of the quality of investment advice for explaining user influence (measured by the total number of retweets, mentions, or followers). The table shows that higher quality investment advice is associated with a larger number of retweets and followers. Quality is the percentage of correct stock predictions interpreting all messages classified as buy or sell signals as end of day prediction for the stock that is mentioned. The sample was restricted to users with at least 20 messages during our sample period, representing the two most frequently posting user groups from Table 7. The results of negative binomial regressions are shown. We report the coefficients for the robust version of the regression using a sandwich estimator of variance.

Regarding the estimation methods: The dependent variables are all highly skewed, non-negative count data. We find overdispersion with significant alphas for all dependent variables. Since poisson regressions, which have been used in similar studies of Twitter accounts (Giller, 2009), assume equal conditional means and variance, this suggests the use of either zero-inflated poisson regressions or negative binomial regressions. The count data for retweets and mentions contain a substantial share of zeroes. This may justify the use of zero-inflated models for these two dependent variables. However, there is no reasonable data generating process to explain excess zeroes. In addition, the predictors for excess zeroes are not significant in most cases where we attempted to fit this type of model and the Vuong test does not suggest the use of zero-inflated models. In fact, we would argue that there are no "excess" zeroes because all zeroes among users with a substantial number of tweets truly indicate that their messages are simply not being retweeted (i.e., quoted). This suggests the use of negative binomial regressions. They come with the additional advantage of being more parsimonious and allowing for the use of one consistent regression model across all four independent variables. Therefore, we show the results for negative binomial regressions and report the coefficients for the robust version using a sandwich estimator of variance that is robust to most types of misspecification as long as the observations are independent (Cameron and Trivedi, 1998). Because message volume and followership are highly skewed they were log-transformed when used as independent variables. The results are robust to other user definitions (e.g., with fewer messages) and regressions models (e.g., poisson regressions).

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

	Retweets	Mentions	Followers (total)	Follower (stock m.)
Message volume (log)	0.890^{***}	0.838^{***}	0.760^{**}	0.266^{***}
	9.551	*7.59*	*2.792*	*3.408*
Followers (log)	*0.461^{***}*	*0.445^{***}*		
	8.148	*9.408*		
Quality investment advice (same day)	0.777^{*}	0.215	2.514^{***}	0.458
	2.021	*0.53*	*3.588*	*1.272*
Quality investment advice (yesterday)	-1.035^{**}	-0.946^{*}	-2.386^{*}	−0.141
	−2.63	*−2.45*	*−2.079*	*−0.361*
User sentiment	0.599	−0.147	−1.964	−0.615
	1.147	*−0.279*	*−1.836*	*−1.739*
Chi^2	396.7^{***}	234.6^{***}	34.1^{***}	16.5^{**}
Observations	614	614	614	604

Obviously, and as hypothesised, message volume and followership are positively related to retweets and mentions. However, more importantly and beyond the natural volume effect, users who provide higher quality investment advice are retweeted more frequently ($c = 0.777$, $p < 0.05$). This relationship only holds for accurate advice relative to same day returns. In contrast to message boards, where a one-day follow up leads to greater reputation (Zhang, 2009), appreciation of information quality in stock microblogs appears to be more short-lived. Investment advice in line with a simple momentum strategy is actually associated negatively with influence ($c = -1.035$, $p < 0.01$). Users appear to be immune to advisors with consistently bullish sentiment. While the number of retweets can be explained by the quality of the investment advice, the coefficients are not consistently significant in the case of user mentions (albeit the direction of the coefficients indicates a similar relationship). The total followership behaves similarly. Followership increases with higher quality same day investment advice ($c = 2.514$, $p < 0.01$). Again, a one-day follow-up representing a momentum strategy hurts followership (c = -2.386, $p < 0.05$). The determinants of followership among serious stock microbloggers are not significant. This may have to do with the difficulty to clearly define this group, as indicated above.

Therefore, we conclude that users who provide above average investment advice are given credit and receive greater attention in microblogging forums through higher levels of retweets as well as a larger followership, supporting H4a.

Regarding H4b, greater weight given to high quality pieces of investment advice could, next to the properties of users, also explain efficient information aggregation in stock microblogging forums. A retweet indicates that a user found an original tweet noteworthy enough to forward it to his or her followers and thus award it with greater weight in the information stream. Accordingly, we compare the quality of retweets with non-retweeted tweets. While the average quality across all messages is 55.8%, the difference between retweets (55.9%) and non-retweets (55.1%) is miniscule and statistically not significant. There are a number of reasons why retweets may not be of higher quality than other tweets. First, many authors only forward parts of the message and add their own commentary. This may change the bullishness of the message and no longer correspond to the original market movement. Second, if a message is retweeted a day after the original message, the signal may no longer correspond with same-day returns. Therefore, we also compared the quality of the original (retweeted) messages with the rest of the sample. However, there is no difference in quality between the two. We therefore reject our hypothesis H4b that higher quality pieces of information are retweeted more frequently.

5. Discussion

5.1 *Summary of results*

Stock microblogs have become a vibrant online forum to exchange trading ideas and other stock-related information. This study set out to investigate the relationship between stock microblogs and financial market activity and offer an explanation for the efficient aggregation of information in microblogging forums. We find, first, that stock microblogs contain valuable information that is not yet fully incorporated in current market indicators and, second, that retweets and followership relationships provide microblogging forums with an efficient mechanism to aggregate information.

We have used methods from computational linguistics to determine the sentiment (i.e., bullishness), message volume, and level of agreement of nearly 250,000 stock-related microblogging messages on a daily basis. Our study examines the relationship between these tweet features and the corresponding market features return, trading volume, and volatility.

We hypothesised that increased bullishness of stock microblogs is associated with higher returns. While we find a significant association between bullishness and returns, we cannot find a lagged relationship of bullishness with abnormal returns. Yet the opposite is true as Fama-MacBeth cross-sectional regressions showed: abnormal returns lead to an increased bullishness among microbloggers.

Disagreement shows a significant association with an increase in trading volume. Overall, it is worth noting that many correlations between tweet and market features are stronger than relationships among market features, which are studied intensively in financial market research. However, we note that while some tweet features appear to contain predictive power with respect to market features, the standardised coefficients show a much stronger effect of market features on tweet features.

Our analysis of information diffusion in the form of retweets, mentions, and followership shows that users who provide above average investment advice are given credit and a greater share of voice in microblogging forums through higher levels of retweets and followers. However, the analysis of individual messages shows that higher quality pieces of information are not retweeted more frequently.

5.2 *Limitations and further research*

This study does not come without limitations. First, we use only daily granularity of analysis. The real-time nature of microblogs warrants an intraday analysis. However, as one of the first studies of stock microblogs, our focus was on the comprehensive coverage of stocks. This restriction limited us to daily data because there are only a handful of stocks that attract sufficient message volume for daily analysis. One caveat of our daily granularity is that it may give tweets a slight advantage in contemporaneous analyses because tweet features like sentiment are based on messages posted throughout the day and aggregated at the end of day. As a result, bullish messages toward the afternoon may merely reflect market developments over the course of the trading day, which are unlikely to reverse. However, on the other hand, the alignment of tweets with US trading hours by assigning messages posted after 4:00 pm to the next trading day leads to the inclusion of those tweets in the calculation of tweet features for the following day. These older tweets were published long before they have had a chance to reflect market developments of that day. This would even suggest our contemporaneous results to include a predictive component. In addition, Fama-MacBeth cross-sectional regressions were employed to assess the temporal relationships between tweet and market features. These results specified our key results and supported their robustness.

Second, in line with previous research, we have considered microbloggers to be day traders. Most financial indicators that we consider, such as prices and volatility, are only available as aggregate market measures. However, in the case of trading volume, one could replace volume with the number of trades of different size categories to distinguish between small and institutional investors. We favored a more simple approach because the insights to be gained from this distinction have been fully captured in previous research of internet message boards (such as Antweiler and Frank, 2004).

Third, as is the case in most large-scale studies of financial market data, many conclusions need to be interpreted with caution. The large number of observations often leads to statistically significant results despite high variance among financial measures such as returns. Even though the aggregate conclusions are correct, we cannot expect significant relationships, such as the one between message and trading volume, to hold for each and every individual stock.

Fourth, we have explored the information content of stock microblogs in terms of sentiment (i.e., bullishness), arguably the most critical piece of information value contained in these postings. However, the definition of information could be expanded to include other dimensions such as the topic or type of news that is discussed. In a working draft of their manuscript, Antweiler and Frank (2002) have pointed out that 'one could try to determine which classes of events have particularly large effects for stock returns' (p. 30). Thus, future work should distinguish the market reaction to different types of company-specific news events.

Finally, we study the reflection of market developments in stock microblogs. While we find some notable relationships, our results do not allow us to determine whether these forums are merely reflecting investor behavior or have changed market behavior. It is probably too early to explore this question, so we leave this type of analysis of pre and post microblogging eras for future research.

5.3 *Conclusion*

It appears that online investors have matured since the introduction of messages boards more than 10 years ago. We observe a more balanced ratio of buy and sell signals and traders no longer follow a naïve momentum strategy, but seem to recommend contrarian trading positions. Quality and content appear to be more important than quantity, since bullishness is related to returns more strongly than message volume. Investigating the mechanism of information weighting and diffusion, we find that users providing high quality investment advice receive greater attention through higher levels of retweets and also gain a larger followership.

In conclusion, stock microblogs do contain valuable information that is not yet fully incorporated in current market indicators. Our results permit researchers and financial professionals to use tweet features as valuable proxies for investor behavior and belief formation. Increased bullishness can serve as a proxy for positive investor sentiment indicated by rising stock prices. Users primarily post messages concerning stocks that are traded more heavily. Our results suggest that stock microblogs can claim to capture key aspects of the market conversation.

We provide early indications with respect to the information aggregation in stock microblogging forums. According to our results, the microblogging community recognises users who consistently offer high quality investment advice, although there are no simple rules to identify valuable pieces of information. Based on our initial findings, we encourage future research to further investigate the role of information weighting and diffusion in microblogging forums. Until then, picking the right tweets remains just as difficult as making the right trades.

Appendix

A.1. Naïve Bayesian text classification

In this section we describe in detail the method underlying our Naïve Bayesian text classification. The probability of a document d belonging to class c is computed as

$$P(c|d) = \ln P(c) \sum_{1 \leq i \leq n_d} \ln P(w_i|c), \tag{A1}$$

where $P(w_i|c)$ is the conditional probability of word w_i occurring in a document of class c. $P(c)$ is the prior probability of a document belonging to class c. The algorithm assigns the document to the class with the highest probability. The parameters $P(c)$ and $P(w_i|c)$ are estimated based on a training set of manually coded documents, so that the prior probability

$$\hat{P}\ (c) = \frac{N_c}{N}, \tag{A2}$$

where N_c is the number of documents in class c and N is the total number of documents. The conditional probability $P(w_i/c)$ is estimated as

$$\hat{P}\ (w|c) = \frac{W_c}{\sum_{c \in C} W_c}, \tag{A3}$$

where W_c is the total number of occurrences of word w in training documents of class c. We include Laplace Smoothing to minimise the effect of cases where $P(w_i|c) = 0$. This conditional probability illustrates the algorithm's 'naïve' assumption that all words, or features, are independent of each other.

In most applications, the dictionary is limited to improve the classification performance by avoiding overfitting the model to the training set. The dictionary can be pruned by choosing the most representative set of words in terms of the information gain criterion (IG). IG measures the entropy difference between the unconditioned class variable and the class variable conditioned on the presence or absence of the word. It is equivalent to the mutual information between a class and a word and calculated as

$$IG(w_i, c) = H(c) - H(c|w_i) = \sum_{c \in C} \sum_{w_i \in \{0,1\}} p(c, w_i) \ln \frac{p(c|w_i)}{p(c)}, \tag{A4}$$

where p(c, w_i) is the joint probability for the occurrence of word w_i and class c. Due to the use of multiple classes, a sum weighted by the probability of the respective classes c is calculated to each word. In line with Antweiler and Frank (2004) we chose the 1,000 words with the highest information gain to compose our dictionary.

Our classification method uses individual words as input variables (a so-called 'bag of words' approach). An automated algorithm will, therefore, treat any distinct sequence of characters separately (by default, even 'buy' and 'Buy' would be two different features).

Table A1

Classification results – most common words/features per class.

This table presents the most common words associated with a class. It indicates that the Information Gain model derived a plausible dictionary from our training set.

Buy	Hold	Sell
[number]	[number]	[negemo]
[dollarvalue]	[retweet]	short
Cross	[mention]	[number]
Day	[url]	[dollarvalue]
[url]	[negemo]	support
Movingaverage	[posemo]	down
[posemo]	stock	point
[retweet]	active	[url]
Buy	[dollarvalue]	drop
Resistance	market	bearish
Up	up	[retweet]
[percentagefigure]	earn	[percentagefigure]
High	watch	cross
Long	trade	[posemo]
Today	ipad	sell
[mention]	today	[mention]
[negemo]	detail	put
Trade	my	trade
Stock	twitter	low
Call	look	sold
New	bank	loss
Earn	[percentagefigure]	call
Share	iphone	bear
Market	new	sale
Above	move	weak
Bullish	app	fall
Bought	#stockpick	lost
Strong	call	lower
Beat	day	below
Acquire	report	decline

We performed seven preprocessing steps to improve the quality of the input data and reduce the feature space. First, all messages were lowercased and punctuation removed. Second, we compiled a custom stopword list to remove noise words (such as ‘a’, ‘the’, or ‘and’). We built on commonly used collections (e.g., the SMART stopword list; see Buckley *et al*., 1993) and added words that were relevant to our particular context (e.g., company names). Third, we tokenised a number of repeating elements: most importantly, we replaced all stock tickers with the token ‘[ticker]’ because a specific company references should not be counted as a signal with respect to the bullishness of the message. Next we replaced all hyperlinks, dollar values, and percentages figures with a token, respectively. Fourth, we aggregated a selected number of words with different spellings to a common format (e.g., the characters '$$s' and '$$$' are commonly used as abbreviations of the term 'money'). Fifth, building on the finding of Tetlock *et al*. (2008) that the fraction of emotional words in firm-specific news, can predict stock returns, we tag more than 4,000 emotional words as either positive or negative. Following Tetlock *et al*.(2008) we use the General Inquirer's Harvard-IV-4 classification dictionary and add each occurrence of an emotional word to the bag of words for that message. Thus we combine text mining approaches based on pre-defined dictionaries and statistical methods. Sixth, we apply the widely used Porter stemmer in order to remove the morphological endings from words (e.g., 'buys' and 'buying' are reduced to 'buy'; Porter, 1980). Finally, following established preprocessing procedures (see Rennie *et al*., 2003), word counts are transformed to a power-law distribution that comes closer to empirical text distributions than most training sets (term frequency [TF] transformation) and words occurring in many messages are discounted (inverse document frequency [IDF] transformation).

A.2. Classification of our data set

Table 1 shows a few typical examples of the tweets from both the training set and the sample data used in our study including the manual coding (for the training set) and the results of the automatic classification (for the main data set). As these examples illustrate, the Naïve Bayesian algorithm can classify messages quite well. As Table 2 shows, overall in-sample classification accuracy was 81.2%. Even a more conservative 10-fold cross validation of the model within the training set correctly classifies 64.2% of all messages. Our classification is in line with similar studies that have applied Naïve Bayesian learning algorithms to financial text samples (Koppel and Shtrimberg, 2006; Wasko *et al*., 2004). The accuracy by class further validates the use of automatically labelled messages. False positives are less likely among buy and sell signals than among hold messages. For our purposes, falsely labeling a buy or sell signal as hold is more acceptable than falsely interpreting messages as buy or sell signals. The confusion matrix shows that the worst misclassification (of buy signals as sell signals and vice versa) occurs only rarely.

A look at the most common words per class (see Table A1) indicates that the Information Gain model derived a plausible dictionary from our training set. Obviously, some features occur frequently in all classes (e.g., numbers and hyperlinks). However, beyond these universal features, the most common words reasonably reflect the linguistic bullishness of the three classes. Positive emotions, for example, are much more likely among buy signals. In addition, buy signals often contain bullish words with an origin in technical analysis (e.g., 'moving average', 'resistance', 'up', or 'high'), operations (e.g., 'acquire'), financials (e.g., 'beat', 'earn'), or trading (e.g., 'buy', 'long', 'call'). Sell signals contain many corresponding bearish words in the areas of technical analysis (e.g., 'support' and 'cross'), financials (e.g., 'loss') or trading (e.g., 'short' and 'put'). As a results of the frequent occurrence of negative adjectives (e.g., 'weak', 'low') and verbs (e.g., 'decline', 'fall'), negative emotions are among the most common features in sell signals supporting Tetlock *et al*.'s (2008) findings. Positive and negative emotions are much more equally balanced in hold messages, which also contain more neutral words such as product names (e.g., 'ipad', 'iphone') and make fewer references to specific price targets (i.e., dollar values).

References

Amihud, Y., 'Illiquidity and stock returns: cross-section and time-series effects', *Journal of Financial Markets*, Vol. 5, 2002, pp. 31–56.

Antweiler, W. and Frank, M.Z., 'Is all that talk just noise? The information content of internet stock message boards', *Journal of Finance*, Vol. 59, 2004, pp. 1259–94.

Antweiler, W. and Frank, M.Z., 'Is all that talk just noise? The information content of internet stock message boards', *Working Paper* (2002), Available at http://strategy.sauder.ubc.ca/antweiler/public/noise-a.pdf

Barber, B. and Odean, T., 'Trading is hazardous to your wealth: The common stock investment performance of individual investors', *Journal of Finance*, Vol. 55, 2000, pp. 773–806.

Barber, B. and Odean, T., 'All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors', *Review of Financial Studies*, Vol. 21, 2008, pp. 785–818.

Black, F., 'Noise', *Journal of Finance*, Vol. 41, 1986, pp. 529–43.

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

Boyd, D., Golder, S. and Lotan, G., 'Tweet, tweet, retweet: conversational aspects of retweeting on Twitter', *Proceedings of the 43rd Hawaii International Conference on System Sciences*, 2010, (pp. 1–10), Kauai, HI: HICSS.

Brown, G.W., 'Volatility, sentiment, and noise traders', *Financial Analysts Journal*, Vol. 55, 1999, pp. 82–90.

Buckley, C., Salton, G. and Allan, J., 'Automatic retrieval with locality information using SMART', *First Text Retrieval Conference*, 1993, (pp. 59–72), Gaitherburg, MD: NIST.

BusinessWeek, 'Stock Twits may change how you trade', *BusinessWeek online edition* (author Max Zeledon), 2009, Available athttp://www.businessweek.com/technology/content/feb2009/tc20090210_875439.htm

Cameron, A.C. and Trivedi, P.K., *Regression Analysis of Count Data* (Cambridge University Press Cambridge 1998).

Campbell, J.A., 'In and out scream and shout: An internet conversation about stock price manipulation', *Proceedings of the 34th Hawaii International Conference on System Sciences*, 2001, (pp. 1–10), Maui, HI: HICSS.

Cao, H.H., Coval, J.D. and Hirshleifer, D., 'Sidelined investors, trading-generated news, and security returns', *Review of Financial Studies*, Vol. 15, 2002, pp. 615–48.

Cha, M., Haddadi, H., Benevenuto F. and Gummadi K.P., 'Measuring user influence in Twitter: The million follower fallacy', *Proceedings of the International Conference on Weblogs and Social Media*, 2010, (pp. 10–17), Washington, DC: AAAI.

Danthine, J-P. and Moresi, S., 'Volatility, information, and noise trading', *European Economic Review*, Vol. 37, 1993, pp. 961–82.

Das, S. and Chen, M., 'Yahoo! for Amazon: Sentiment extraction from small talk on the web', *Management Science*, Vol. 53, 2007, pp. 1375–88.

Das, S., Martinez-Jerez, A. and Tufano, P., 'eInformation: a clinical study of investor discussion and sentiment', *Financial Management*, Vol. 34, 2005, pp. 103–37.

De Long, J.B., Shleifer, A., Summers, L.H. and Waldmann, R.J., 'Noise trader risk in financial markets', *Journal of Political Economy*, Vol. 98, 1990, pp. 703–38.

Delort, J-Y., Arunasalam, B., Milosavljevic, M. and Leung, H., 'The impact of manipulation in internet stock message boards', *Working Paper* (2009), Available at http://ssrn.com/abstract=1497883

DeMarzo, P.M., Vayanos, D. and Zwiebel, J., 'Persuasion bias, social influence, and unidimensional opinions', *Quarterly Journal of Economics*, Vol. 118, 2003, pp. 909–68.

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

Dyckman, T., Philbrick, D. and Stephan, J., 'A comparison of event study methodologies using daily stock returns: a simulation approach', *Journal of Accounting Research*, Vol. 22, 1984, pp. 1–30.

Easley, D. and O'Hara, M., 'Price, trade size, and information in securities markets', *Journal of Financial Economics*, Vol. 19, 1987, pp. 69–90.

Fan, M., Tan, Y. and Whinston, A.B., 'Evaluation and design of online cooperative feedback mechanisms for reputation management', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17, 2005, pp. 244–54.

Giller, G.L., 'Maximum likelihood estimation of a poissonian count rate function for the followers of a twitter account making directional forecasts of the stock market', *Working Paper* (2009), available at http://ssrn.com/abstract=1423628

Gu, B., Konana, P. and Chen, H-W., 'Melting-pot or homophily? An empirical investigation of user interactions in virtual investment-related communities', *Working Paper* (2008), available at http://ssrn.com/abstract=1259224

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten I.H., 'The WEKA data mining software: An update', *SIGKDD Explorations*, Vol. 11, 2009, pp. 10–18.

Harris, M. and Raviv, A., 'Differences of opinion make a horse race', *Review of Financial Studies*, Vol. 6, 1993, pp. 473–506.

Hirshleifer, D. and Teoh, S., 'Limited attention, information disclosure, and financial reporting', *Journal of Accounting and Economics*, Vol. 36, 2003, pp. 337–86.

Hong, H., Kubik, J. and Stein, J.C., 'Thy neighbor's portfolio: word-of-mouth effects in the holdings and trades of money managers', *Journal of Finance*, Vol. 60, 2005, pp. 2801–24.

Jones, C.M., Kaul, G. and Lipson, M.L., 'Transactions, volume, and volatility', *Review of Financial Studies*, Vol. 7, 1994, pp. 631–51.

Jordan, J., 'Hedge fund will track Twitter to predict stock moves', *Bloomberg (online edition)*, 2010, available at http://www.bloomberg.com/news/2010-12-22/hedge-fund-will-track-twitter-to-predict-stockmarket-movements.html

Karpoff, J.M., 'A theory of trading volume', *Journal of Finance*, Vol. 41, 1986, pp. 1069–87.

Kim, O. and Verrecchia, R., 'Trading volume and price reactions to public announcements', *Journal of Accounting Research*, Vol. 29, 1991, pp. 302–21.

Koppel, M. and Shtrimberg, I., 'Good news or bad news? Let the market decide', *Computing Attitude and Affect in Text: Theory and Applications*, Vol. 20, 2006, pp. 297–301.

Koski, J.L., Rice, E.M. and Tarhouni, A., 'Noise trading and volatility: Evidence from day trading and message boards', *Working Paper* (2004), available athttp://ssrn.com/abstract=533943

Milgrom, P. and Stokey, N., 'Information, trade and common knowledge', *Journal of Economic Theory*, Vol. 26, 1982, pp. 17–27.

Mizrach, B. and Weerts, S., 'Experts online: an analysis of trading activity in a public Internet chat room', *Journal of Economic Behavior & Organization*, Vol. 70, 2009, pp. 266–81.

Newcomb, T.M., 'An approach to the study of communicative acts', *Psychological Review*, Vol. 60, 1953, pp. 393–404.

Ng, L. and Wu, F., 'Peer effects in the trading decisions of individual investors', *Financial Management*, Vol. 39, 2006, pp. 807–31.

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

Porter, M.F., 'An algorithm for suffix stripping', *Program*, Vol. 14, 1980, pp. 130–37.

Rennie, J.D., Shih, L., Teevan, J. and Karger, D.R., 'Tackling the poor assumptions of Naive Bayes text classifiers', *Proceedings of the Twentieth International Conference on Machine Learning*, 2003, (pp. 616–23), Washington DC: AAAI.

Romero D.M., Galuba, W., Asur, S. and Huberman, B.A., 'Influence and passivity in social media', *Working Paper* (2010), available at http://arxiv.org/PS_cache/arxiv/pdf/1008/1008.1253v1.pdf

Sabherwal, S., Sarkar, S.K. and Zhang, Y., 'Online talk: does it matter?', *Managerial Finance*, Vol. 34, 2008, pp. 423–36.

Shalen, C.T., 'Volume, volatility, and the dispersion of beliefs', *Review of Financial Studies*, Vol. 6, 1993, pp. 405–34.

Sprenger, T.O., Sandner, P.G., Tumasjan, A. and Welpe, I.M., 'News or noise? The stock market reaction to different types of company-specific news events.', *Working Paper* (TechnischeUniversitätMünchen, TUM School of Management, 2012).

Sprenger, T.O. and Welpe, I.M., 'Tweets and peers: Defining industry groups and strategic peers based on investor perceptions of stocks on Twitter.', *Algorithmic Finance*, Vol. 1, 2011, 57–76

Tetlock, P.C., Saar-Tsechansky, M. and Macskassy, S., 'More than words: quantifying language to measure firms' fundamentals', *Journal of Finance*, Vol. 63, 2008, pp. 1437–67.

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

Tumasjan, A., Sprenger, T.O., Sandner, P.G. and Welpe, I.M., 'Election forecasts with Twitter: how 140 characters reflect the political landscape', *Social Science Computer Review*, Vol. 29, pp. 402–18.

van Bommel, J., 'Rumors', *Journal of Finance.* Vol. 58, 2003, pp. 1499–1520.

Vincent, A. and Armstrong M., 'Predicting break-points in trading strategies with Twitter', *Working Paper* (2010), available at http://ssrn.com/abstract=1685150

Wasko, M., Faraj, S. and Teigland, R., 'Collective action and knowledge contribution in electronic networks of practice', *Journal of the Association for Information Systems*, Vol. 5, 2004, pp. 493–513.

Wysocki, P., 'Cheap talk on the web: The determinants of postings on stock message boards', *Working Paper* (1998), Available at http://ssrn.com/abstract=160170

Yang, J. and Counts, S., 'Predicting the speed, scale, and range or information diffusion in Twitter', *Proceedings of the International Conference on Weblogs and Social Media*, 2010, (pp. 355–358), Washington, DC: AAAI.

ZDNet, 'Twitter turns six: 140 million users, 340 million tweets daily', *ZDNet blog* (author Zack Whittaker), 2012, Available at http://www.zdnet.com/blog/btl/twitter-turns-six-140-million-users-340-million-tweets-daily/72123

Zhang, X., Fuehres, H. and Gloor, P., 'Predicting stock market indicators through Twitter – "I hope it is not as bad as I fear"', *Collaborative Innovations Networks Conference*, 2010, (pp. 1–8), Savannah, GA: COIN.

Zhang, Y., 'Determinants of poster reputation on internet stock message boards', *American Journal of Economics and Business Administration*, Vol. 1, 2009, pp. 114–21.

Zhang, Y. and Swanson, P.E., 'Are day traders bias free? Evidence from internet message boards', *Journal of Economic and Finance*, Vol. 34, 2010, pp. 96–112.