Tweets and Trades:
The Information Content of Stock Microblogs

working paper

Timm O. Sprenger*, Isabell M. Welpe

Technische Universität München
TUM School of Management
Chair for Strategy and Organization
Leopoldstraße 139
80804 Munich
Germany

December 2010

Tweets and Trades:
The Information Content of Stock Microblogs

Abstract

Microblogging forums have become a vibrant online platform to exchange trading ideas and other stock-related information. Using methods from computational linguistics, we analyze roughly 250,000 stock-related microblogging messages, so-called tweets, on a daily basis. We find the sentiment (i.e., bullishness) of tweets to be associated with abnormal stock returns and message volume to predict next-day trading volume. In addition, we analyze the mechanism leading to 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 in microblogging forums.

"Just like the credibility and objectivity crisis of sell-side analysts in 2001 led to a boom in financial blogs like 'Seeking Alpha' and Barry Ritholtz's 'The Big Picture', the credibility crisis afflicting mainstream financial media today has led to a boom in investor social networks. Traders and investors alike have come to view these platforms as trusted filters that help them make more informed decisions because they can discuss and interpret the news with their peers."

BusinessWeek (2009)

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, 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[1] or through various third-party applications. Users can subscribe to (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 90 million messages per day generated by roughly twice as many registered users (TechCrunch (2010)). 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-

---

[1] www.twitter.com

based investment opportunities (Bloomberg (2010)) and by academic researchers to predict break-points in financial time-series (Vincent and Armstrong (2010)). Therefore, the investor community has come to call Twitter and related third-party applications such as StockTwits.com, which filter stock-related microblogs, "a Bloomberg for the average guy" (BusinessWeek (2009)). It is interesting to note that one of the most frequently used features on the professional Bloomberg terminals, which come at more than $2,000 per month, is the centralized chat system that allows traders to talk to each other in real-time. Twitter offers very similar features and is available at no charge. In fact, Bloomberg has even come to integrate Twitter messages into their terminals and NASDAQ has launched a mobile application that prominently incorporates content from StockTwits. News stories claim that financial microblogs capture the market conversation and suggest that these messages have a significant impact on the financial markets: "Communities of active investors and day traders who are sharing opinions and in some case sophisticated research about stocks, bonds and other financial instruments will actually have the power to move share prices […] making Twitter-based input as important as any other data to the stock" (TIME (2009)).

Stock microblogs have not yet been the subject of scholarly research. This is a puzzling oversight for at least two reasons. First, the unique characteristics of stock microblogging forums do not allow us to transfer results from previous studies of internet message boards. Second, stock microblogging forums permit researchers to observe previously unavailable aspects of information diffusion in an online investment community. Earlier studies have focused on exploring the relationship between internet stock message boards (e.g., Yahoo!Finance or Raging

Bull) and financial markets. For instance, analyzing the most frequently discussed firms on Yahoo!Finance, Wysocki (1998) illustrates that message volume forecasts next-day trading volume and abnormal returns. While this study only investigated message volume, Tumarkin and Whitelaw (2001) have taken a more nuanced approach to the information content on message boards by studying the information embedded in voluntary user ratings (from strong buy to strong sell). However, the authors found no evidence that any information with respect to subsequent returns is embedded in these recommendations. Whereas these studies are limited to rather simple, quantitative information (e.g., message volume, user ratings), Antweiler and Frank (2004), whose study is most closely related to ours, used sophisticated 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 report that message volume predicted trading volume and volatility. However, this study has some severe limitations: the sample period in the year 2000 includes the burst of the internet bubble and dot-com companies with unsustainable business models and partly unrealistic valuations represent a substantial share of the sample.

Previous research has focused specifically on internet stock message boards. As a consequence, we know very little about the information content of *stock microblogs* with respect to financial markets. Despite many parallels to these more established forums, the distinct characteristics of microblogging make the generalization of previous results from stock message boards to stock microblogs challenging for the following reasons. First, unlike Twitter's public timeline, message boards categorize postings into separate bulletin boards for each company,

which may lead to significant attention to outdated information as long as there are no more recent entries. Second, while message boards require users to actively enter the forum for a particular stock, Twitter represents a live conversation. Third, microbloggers have a strong incentive to publish valuable information in order to maintain or increase mentions, the rate of retweets (i.e., quotes by other users) and their followership. We argue that these incentives provide the Twittersphere with a mechanism to weigh information. As a result, we would expect both users and the information in stock microblogging forums to differ substantially from those on message boards.

Next to the differences to internet message boards, there is a second aspect that warrants the investigation of stock microblogs. The nature of microblogging forums makes previously unavailable aspects of information diffusion partially observable (e.g., retweets and followership relationships). However, scholarly research has not yet explored whether these mechanisms to structure information diffusion are really used effectively. Thus, it remains unclear whether, on a large scale, stock microbloggers produce valuable information or simply represent the online equivalent of uninformed noise traders.

Therefore, the purpose of our study is to explore whether and to what extent stock microblogs reflect and affect financial market developments. In particular, for comparability with related research (e.g., Antweiler and Frank (2004)), our study compares the relationship between the most important and heavily studied market features return, trading volume, and volatility with

the corresponding tweet features message sentiment (i.e., bullishness)[2], message volume, and the level of agreement among postings. In addition, we empirically explore possible mechanisms behind the efficient aggregation of information in microblogging forums. Our two overarching research questions are, first, whether and to what extent the information content of stock microblogs reflects financial market developments (RQ1) and, second, whether microblogging forums provide an efficient mechanism to weigh and aggregate information (RQ2). With respect to our first research question we explore, first, whether bullishness can predict returns, second, whether message volume is related to returns, trading volume, or volatility, and third, whether the level of disagreement among messages correlates with trading volume or volatility. With respect to our second research question, we compare the quality of investment advice with the level of mentions, the rate of retweets and the authors' followership.

We find bullishness to be associated with abnormal returns. However, new information, reflected in the tweets, is incorporated in market prices quickly and market inefficiencies are difficult to exploit with the inclusion of reasonable trading costs. An event study of buy and sell signals shows that microbloggers follow a contrarian strategy. Message volume can predict next-day trading volume. In addition, our results offer an explanation for the efficient aggregation of information in microblogging forums. Users who provide above average investment advice are retweeted (i.e., quoted) more often, have more followers and are thus given a greater share of voice in microblogging forums.

---

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

The contribution of this study is threefold. First, to the best of our knowledge, it is the first to comprehensively explore the information content of stock microblogs. Unlike much of the related literature, this study is able to go beyond the analysis of relatively simple measures of online activity (e.g., message volume or word counts), but, instead, leverages an innovative methodology from computational linguistics to evaluate the actual message content and sentiment. As a consequence, our results permit researchers and financial professionals to reliably identify tweet features, which may serve as valuable proxies for investor behavior and belief formation. Second, our study extends previous research, which has shown a correlation of online message content with financial market indicators by providing an explanation for the efficient aggregation of information in stock microblogging forums. The structure of these forums allows us to empirically explore theories of social influence concerning the diffusion and processing of information in the context of a financial community. Third, this study replicates and extends similar research in the context of internet message boards without some of the previous limitations (e.g., sample selection, timeframe). We analyze a more comprehensive set of stocks over the course of 6 months with fairly stable financial market activity. In addition, we examine the economic exploitability of trading schemes based on signals embedded in stock microblogs.

The remainder of the paper is structured as follows. First, we review related work and derive our research questions and hypotheses. Second, we describe our data set and methodology. Third, we provide results illustrating the timing of tweet features relative to market features (i.e., the contemporaneous and lagged relationships). We also explore the information diffusion in

6

stock microblogging forums. We conclude that stock microblogs contain valuable information that is not yet fully incorporated in current market indicators. Finally, we discuss the implications of our findings and provide suggestions for further research.

I. Related Work and Research Questions

A. Introduction to the research of online stock forums

In this section, we review the theoretical basis motivating studies of online stock forums. According to the Efficient Market Hypothesis (EMH) financial markets are "informationally efficient" meaning that market prices reflect all known information. The widely accepted semi-strong version of the EMH claims that prices aggregate all publicly available information and instantly reflect new public information. Therefore, according to the EMH, investors cannot earn excess profits from trading strategies based on publicly available information (Fama (1970), Fama (1991)).

However, a growing body of research suggests that financial markets do not always comply with the EMH (for a comprehensive overview see Malkiel (2003). Recent studies have suggested that particularly qualitative information[3] is not reflected fully and instantly in market prices. Tetlock, Saar-Tsechansky, and Macskassy (2008) found that firms' stock prices underreact to the textual information embedded in news stories (i.e., the fraction of negative words in firm-specific news). In addition, other studies suggest that many unofficial but nevertheless public data sources contain valuable information. Bagnoli, Beneish, and Watts (1999), for example, have

---

[3] For the purpose of this study, we define qualitative information as words. This definition is in line with related research (Tetlock, Saar-Tsechansky, and Macskassy (2008)).

illustrated that "earnings whispers" (i.e., unofficial earnings forecasts that circulate among traders) are more accurate proxies for market expectations than official First Call forecasts. They claim that whispers are increasingly becoming the true market expectation of earnings and show that trading strategies based on the relationship between whispers and First Call forecasts earn abnormal returns. Sources of qualitative data, such as those mentioned above, have been largely neglected in the financial literature, possibly because computational linguistic methods, as applied in this study, are necessary to process the information and have only recently been recognized by scholars in the financial literature.

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[4] like Yahoo!Finance, deserve special attention. Online financial communities provide a time-stamped archive of the collective interpretation of information by individual investors. Prior literature shows that the information exchange in online financial communities includes the dissemination of public information, speculation regarding private and forthcoming information, analysis of data, and personal commentary (see Lerman (2010), Felton and Kim (2002), Das, Martinez-Jerez, and Tufano, (2005), Campbell (2001)).

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.

---

[4] Some studies (e.g., Clarkson, Joyce, and Tutticci (2006) refer to these as internet discussion sites (IDS), virtual investment communities (VIC) or bulletin boards. We prefer the more common term internet message board, but will occasionally use the alternative terms in line with the cited research.

For the 50 most frequently discussed firms on Yahoo!Finance between January and August 1998, he illustrates that message volume did forecast next-day trading volume and abnormal returns. 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 confirm that these recommendations contain relevant information related to stock returns. Consistent with the 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 and confirmed that recommended stocks had a strong prior performance indicating that these traders follow a naïve momentum strategy. In addition, the author explored the reasons leading to recommendation including technical analysis, financial issues and company operations.

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 (Zhang and Swanson (2010)). 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 marker returns. Antweiler and Frank (2004), whose study is most closely related to ours, 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, these results are based on data from the year 2000 during which asset prices were highly volatile.[5] In addition, one third of the sample was taken from the Dow Jones Internet Index comprised of many companies with unsustainable business models and unrealistic valuations. 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)).

The massive amount of digital content creates specific challenges for the analysis of these data sets. Within existing research, one can broadly distinguish between two focus areas depending on the background of the academic community. On the one hand, many studies with a background in computer science put an emphasis on natural language processing and text

---

[5] 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.

classification.[6] Many of these studies lack a rigorous analysis of financial market indicators (e.g., no implementation of market models to calculate excess returns). On the other hand, studies from the finance community are mostly limited to quantitative input data (such as ratings provided by users of online communities).

Most of the above-mentioned studies of internet message boards explore the effect of these forums on the financial markets. However, Jones (2006) has pointed out that message boards "may be an observable form of a pre-existing information network or […] they may have altered the information landscape in a way which has changes pricing behavior" (p. 67). To explore this question, Jones (2006) has investigated changes in stock market behavior between the pre and post message board eras. Empirical evidence shows a significant increase in daily trading volumes, lower returns, and higher volatility after a firm's message board was established. The author concludes that message boards are not merely reflecting pre-existing information networks, but have changed market behavior.[7]

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.[8] Despite many parallels to these more established forums, the following three distinct characteristics of

---

[6] Most of these studies use subsequent stock price movements to automatically label news articles as buy or sell recommendations. Mittermayer and Knolmayer (2006) provide a comprehensive overview of developed prototypes and their performance.

[7] Engelberg and Parsons (2011) support the notion of a causal impact of media in financial markets. For local, nonoverlapping trading markets surrounding major U.S. cities and local daily newspaper of that city, the authors show that local press coverage increases the trading volume of local retail investors up to 50%.

[8] Zhang and Skiena (2010) include a limited sample of tweets in a study of four different media sources and their effect on trading volume and stock prices. However the study focuses on newspaper content and the Twitter data set does not include stock microblogs, but tweets mentioning the official company name. The authors caution that their "Twitter database […] is small and the result is less accurate". While they noted a strong correlation among all media sources, the sentiment on Twitter appeared to have a slightly stronger correlation with future returns.

microblogging do not allow us to generalize previous results from stock message boards to stock microblogs for the following reasons. First, whereas message boards categorize postings into separate bulletin boards for each company, Twitter's public timeline may more accurately capture the natural market conversation. Thus outdated information may still receive attention on stock message boards 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. 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, previous research 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 behavior, and overreaction to good or bad news.

A few recent studies suggest that the information content of microblogs may help predict macroeconomic market indicators. O'Connor et al. (2010) have found Twitter messages to be a leading indicator for the Index of Consumer Sentiment (ICS), a measure of US consumer confidence. Both Zhang, Fuehres, and Gloor (2010) and Bollen, Mao, and Zeng (2010) find that a random subsample of messages from Twitter's public timeline can be used to predict market indices such as the Dow Jones Industrial Average (DJIA) or the S&P 500. However, all of these studies are concerned with broadly defined data sets (e.g., all available messages or blog posts in the sample period, most without a specific reference to the stock market) and derive aggregate sentiment measures. While the correlation of these aggregate measures with macroeconomic indicators is encouraging, it does not allow us to draw conclusions about the information content of stock microblogs with respect to individual stocks. Das and Chen (2007) found the relationship between aggregated sentiment and index returns to be much stronger than the correlation for individual stocks. Therefore, our study focuses on the specific domain of stock microblogs and investigates their relationship with market prices of publicly traded companies. While the link between information and market developments has been examined extensively in other contexts, the mechanics of this link are largely unexplored. Therefore, unlike similar previous studies (e.g., Antweiler and Frank (2004)), we also investigate information diffusion in this social media context, which may help explain why microblogging forums can aggregate information efficiently.

B. Research questions, related research and hypotheses

In this section, we define our two research questions, review related research and derive our hypotheses. For each hypothesis, we review related empirical evidence from internet message boards. First, we derive our hypotheses for RQ1 exploring the relationship between tweet and market features (H1-H3b in sections B.1-B.3) and, second, for RQ2 investigating the information diffusion in stock microblogging forums (H4a-H4b in section B.4).

As illustrated in the previous section, the message board literature has mainly focused on three primary message features: Message sentiment (i.e., bullishness), message volume, and the level of agreement among postings. Since these features are transferrable to the microblogging domain, we adopt them for our study and structure related hypotheses around the three resulting tweet features. Our study compares these tweet features with the corresponding market features return, trading volume, and volatility. Therefore, the overarching questions are, first, whether bullishness can predict returns, second, whether message volume is related to returns, trading volume, or volatility, and third, whether the level of disagreement among messages correlates with the trading volume or volatility.

While empirical findings suggest that online information networks may change market behavior and are not just an indicator of otherwise motivated investor behavior (Jones (2006)), we will adopt this more conservative interpretation of our results and understand the relationship between the two as the reflection of information in the tweets.[9] However, if this reflection

---

[9] Many studies boldly interpret this relationship as an effect of online forums on the market (e.g., Bettman, Hallett, and Sault (2000): "we interpret the intraday results as providing consistent evidence of a strong and significant market reaction to the posting of takeover rumours in IDS", p. 44).

provides valuable information to understand market movements, it may be just as helpful to researchers and financial analysts as are stock prices to understand investors' perception of the prospects of a company.[10] As we illustrate below, the answers as to what exactly tweet features reflect are not trivial and in many cases there are competing interpretations. Increased message, for example, volume may indicate that more information is available, but depending on whether this message volume is generated by noise traders or informed investors it could lead to either lower or higher volatility.

B.1. 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 confirms that individual investors pay a significant performance penalty for active trading (Barber and Odean (2000)). On the other hand, in a direct comparison with Barber and Odean (2000), Mizrach and Weerts (2009) found that 55% of the investors of a public stock-related chat room made profits after transaction costs. 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" allowing the investor and the followers to fully capture the value of the private information (Van Bommel (2003), p. 1499). On the other hand, (Van Bommel, (2003)) acknowledges a moral hazard problem due to the

---

[10] Clarkson, Joyce, and Tutticci (2006) have pointed out that irrespective of whether market indicators lead sentiment or sentiment leads market indicators, both may feed off and reinforce each other.

opportunity to spread false rumors[11], which could lead followers to ignore rumors altogether. In addition, a prolific poster interviewed by Das, Martinez-Jerez, and Tufano (2005) stated that "I don't think there was any truly inside information…the whole group had no better idea than the next person" (p. 109). According to the EMH, stock prices should not be affected by this type of information. The research on stock message boards confirms this hypothesis. Tumarkin and Whitelaw (2001) have found no evidence that any information is embedded in voluntary disclosed user ratings (from strong buy to strong sell). Consistent with the EMH, message board activity did not predict industry-adjusted returns and postings followed the stock market. Dewally (2003) has confirmed that recommended stocks had a strong prior performance indicating that these traders follow a naïve momentum strategy. Das and Chen (2007) report only a contemporaneous relationship between message bullishness and market returns. However, Hirshleifer and Teoh (2003) show that, due to limited attention and processing power, informationally equivalent disclosures can have different effects on investor perceptions and market prices. For instance, investors primarily consider purchasing stocks that have been brought to their attention through the news (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) and Hong, Kubik, and Stein (2005)). In an internet chat room, for example, traders were likely to follow the trade direction (i.e., buy vs. sell) of their peers if there had been a recent post on the same stock (Mizrach and Weerts (2009)). DeMarzo, Vayanos, and Zwiebel (2003) have

---

[11] The term rumor is often referred to generally as "unofficial public information with an unkown quantity of truth and untruth" in the context of financial markets, which are "the perfect breeding ground for rumours because highly competitive industry participants value every piece of information in vying for a comparative advantage" (Clarkson, Joyce, and Tutticci (2006), p. 2). In this sense, most of the information in stock microblogs may be considered rumors.

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.

B.2. 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, Martinez-Jerez, and Tufano (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, Coval, and Hirshleifer (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, Sarkar, and Zhang (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. Wysocki (1998) finds a minimal explanatory power of an increase in message volume for positive next-day abnormal returns. 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 acting in concert can increase volatility. Antweiler and Frank (2004) show that, on internet message boards, message volume is a

predictive factor of volatility. Koski, Rice, and Tarhouni (2004) confirm that noise trading (proxied by 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[12], 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.

B.3. Disagreement

Das, Martinez-Jerez, and Tufano (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, Kaul, and Lipson (1994), Shalen (1993)). We derive:

---

[12] Koski, Rice, and Tarhouni (2004) suggest that the vast majority of message board participants are day traders, which can be considered noise traders.

Hypothesis 3a: Increased disagreement among stock microblogs is associated with higher volatility.[13]

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.

---

[13] Disagreement further leads to discussions, reinforcing the previous hypothesis. Das, Martinez-Jerez, and Tufano (2005) find that disagreement is positively related to the message volume.

B.4. Information diffusion

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

Gu, Konana and Chen (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, Tan, and Whinston (2005)). Microbloggers have an incentive to publish valuable information to maintain or increase mentions, the rate of retweets, and their followership – these 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)). 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)[14]). Reports in the business press support the hypothesis that microblogging forums may aggregate information more efficiently than previously studied online communities.[15] Zhang (2009) has found poster reputation on a special internet stock message board with 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 (Konana, Rajagopalan, and Chen (2007)). Following and "unfollowing" an author virtually allows users of microblogging forums to construct their own customized 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

---

[14] 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 1427.5 GAIN 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.

[15] "What I like most is this new level of transparency or accountability. [Stock microblogs] are essentially creating a trading record that's scrutinized on a daily basis […]. A trader's reputation is always on the line. You like what one trader is doing? Simply press follow. Underperformers will be ignored, and rightly so—trading is a zero-sum game and bad advice is a waste of time and money. That's precisely what validates [stock microblogs]" (BusinessWeek, 2009).

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., individual messages) are weighted more heavily and spread through retweets. In the context of financial information, Kosfeld (2005) suggests that information quantity and quality are related. Thus, we propose:

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

II. Data Set and Methodology

In this section, we describe our data set and detail the methodology used to derive the variables used in this study. These include the tweet and market features used to address RQ1 as well as measures of social influence used to answer RQ2. The text analytical methods, which are used to extract the tweet features, deserve special attention and are therefore illustrated in more detail.

A. 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. Jaiku, Tumblr, frazr) 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 90 million messages are posted on Twitter's public timeline every day. 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 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 January 1st and June 30th, 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.[16] 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.[17]

B. Naïve Bayesian text classification

In order to compare the signals from stock microblogs to market movements, we had to classify messages as either buy, hold or sell signals. Our data set contains too many messages for manual coding. Therefore, we chose to classify messages automatically using well established methods from computational linguistics. In line with Antweiler and Frank (2004), we employ the

---

[16] 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. A load balancing feature ensured that messages associated with more frequently mentioned stock symbols were downloaded more often.

[17] Specifically, we focus on those companies that have been included in the S&P 100 as of January 1, 2010.

Naïve Bayesian classification method, one of 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)).[18]

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.[19] 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 I shows a few typical examples of the tweets from both the training set including the manual coding.

[Insert Table I about here]

---

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

[19] 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 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 confirms high interrater reliability (0.78).

As Table II shows, overall in-sample classification accuracy was 81.2%. The accuracy by class further validates the use of automatically labeled 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. 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.

[Insert Table II about here]

C. Aggregation of daily tweet features

In order to compare hundreds of daily messages to the market movements on a daily basis, tweet features need to be aggregated. With respect to RQ1, 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

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

26

where M^{Buy} (M^{Sell}) represents the number of buy (sell) signals on day t.[20] 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.[21] In line with Antweiler and Frank (2004), agreement among messages is defined as

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

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).[22] 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.

---

[20] 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.

[21] The log transformation ln(1+M_t) is analogous to the transformation of the trading volume allowing us to compute elasticities and controls 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 normalized 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 normalized 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 normalization. We, therefore, use the logged version when we refer to message volume in the remainder of our paper.

[22] 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).

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.

D. Financial market data

We have downloaded financial data in daily intervals for the S&P 100 from Thompson Reuters Datastream. Returns are calculated as the log difference of total return to shareholders (TRS), which reflects both price changes and dividend payments. 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 actual return for stock i on day t 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 a OLS regressed market model ($AR^{market\,model}$) as

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

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, Philbrick, and Stephan (1984)), we use a 120-day estimation period starting 130 days prior to the relevant date to not overlap with the event-window of our event study. 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

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

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

E. 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[23] of the stock in question as

$$quality = \begin{cases} = 1 \; if \; \dfrac{s_{it}}{R_{it}} > 0 \\[2ex] \hline = 0, \; otherwise \end{cases} \tag{9}$$

where s_{it} is the sentiment of a message on day *t* associated with stock *i*. We ignore hold messages in the computation of quality scores.[24] Next to the quality of individual message, we also compute the quality of a particular user's investment advice as the average quality of all messages posted by this user. We also 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

---

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

[24] 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.

original author.[25] The first sample tweet in Table I 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.[26] 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 favorite 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.[27] Having laid out the definition of our variables, we now turn to exhibiting and interpreting the results.

---

[25] Alternative formats include "RT: @", "via @", "by @", and "retweet @".

[26] 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.

[27] 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 recognize this limitation and interpret related results with caution, we find followership too relevant a measure to be ignored altogether.

III. Results

A. Descriptive statistics

The results section is structured as follows. After a brief summary of descriptive statistics regarding our data set, sections B and C address RQ1, whereas section D covers results related to RQ2. Section B covers the overall analysis of tweet and market features. Following Antweiler and Frank (2004), we provide results illustrating the contemporaneous (pairwise correlations and contemporaneous regressions) and lagged relationships (time-sequencing regressions) of these features. Next, in section C, we provide in-depth analyses exploring the relationships that are supported by the empirical evidence. Section D tests our hypothesis explaining the efficient information aggregation by exploring whether good investment advice receives greater attention in stock microblogging forums.

We have collected 249,533 stock-related microblogging messages containing the dollar-tagged ticker symbol of an S&P 100 company.[28] 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. Figure 1 shows the

---

[28] Some messages contain more than one ticker symbol. In order to retain these tweets in the data set, we treat them as separate messages for each company that is mentioned. Our text classification method can only determine the overall message sentiment and does not distinguish between distinct references. However, since most of these messages contain the same sentiment for all stocks (e.g., "$GOOG $AMZN big boy stocks acting well") and because their share is relatively small (13.4%), this approach does not affect our results. We have confirmed the robustness of our results by repeating our analyses with the sample limited to tweets containing only one ticker symbol. The results are quite similar and we, therefore, do not report them separately.

distribution of messages throughout the day. We observe a significant spike in message volume before the markets open. The majority of tweets are posted during the trading hours between 9:30 am and 4:00 pm. This provides further evidence that stock microblogs are used by financial professionals to exchange relevant trading ideas in real time.

[Insert Figure 1 about here]

[Insert Table III about here]

Table III shows summary statistics of the market and tweet features on a per company basis. Figure 2 provides us with a first indication of the overall relationships between trading and message volume. The two measures show a strong correlation (*r = 0.468, p = 0.016*). Message volume tracks the rise in trading volume closely in January, for example, and drops at the beginning of February when trading slowed. In April, message volume picks up on a spike in trading activity. The comparison of market returns and a market-cap weighted bullishness index (see Figure 3) exhibits a slightly weaker, but nevertheless statistically significant correlation between the two indicators (*r = 0.408, p = 0.038*).

[Insert Figure 2 and Figure 3 about here]

Figure 3 shows an elevated level of bullishness during a time when the S&P 100 was rising from February to April and a sharp decline for both measures in June. Due to the volatility of the

bullishness index on a daily basis, we explored smoothed versions computed as moving averages. O'Connor et al. (2010) have found that the correlation of their Twitter-based measure of consumer confidence and the Index of Consumer Sentiment (ICS) increased up to a window of 60-day moving averages. We observe the contrary with the correlation of a moving-average bullishness index and the market losing statistical significance when we increase the window beyond 10 days. This indicates that the bullishness of stock microblogs accurately captures market movements fairly quickly. Even though our Twitter-based bullishness index shows much greater volatility than Antweiler and Frank's (2004) equivalent for internet message boards, this volatility does not appear to be mere noise. While the correlation of aggregate features is encouraging, it does not allow us to draw conclusions about the information content of stock microblogs with respect to individual stocks: Das and Chen (2007) found the relationship between aggregated sentiment and index returns to be much stronger than the correlation for individual stocks. Therefore, we devote the remainder of our paper to analyses on a company level.

B. Overall relationship of tweet and market features

B.1. Pairwise correlations

Contemporaneous pairwise correlations provide a first indication with respect to the relationships between market and tweet features (Table IV). We observe a relatively strong correlation between bullishness and returns ($r = 0.166, p = 0.0$). The sentiment in microblogs clearly picks up on absolute market movements. A more conservative test of the quality of

message sentiment is the correlation with abnormal returns, for which we notice slightly weaker correlations. But even simple, market-index-adjusted returns (*r = 0.156, p = 0.0*) and market-model-adjusted returns (*r = 0.147, p = 0.0*) are among the strongest correlations between market and tweet features. This provides support for H1.

Logged message volume and logged trading volume exhibit the strongest among all correlations (*r = 0.441, p = 0.0*), supporting H2a. Message volume correlates only weakly with returns (H2b) and volatility (H2c).

As hypothesized (H3a), trading volume decreases as agreement rises *(r = -0.113, p = 0.0)*. The pairwise correlation of agreement among investors and volatility (*r = -0.014, p = 0.13*) is not statistically significant (H3b).

Overall, it is worth noting that many correlations between tweet and market features are stronger than closely studied relationships among market features, such as the relationship between volatility and trading volume (*r = 0.046, p = 0.0*).

[Insert Table IV about here]

B.2. Contemporaneous regressions

While the pairwise correlations have suggested interesting relationships between tweet features and market features, they do not address the independence of these relationships. It remains unclear whether these relationships remain significant when all other tweet features are controlled for. Thus, in this section, we use panel regression techniques to explore the

contemporaneous[29] relationships between tweet and market features corresponding to our hypotheses in order to investigate whether tweet features can serve as proxies for market developments.

[Insert Table V about here]

Table V 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. The regression results support the strong relationship between bullishness and all three return measures (H1). Thus, increased bullishness can serve as a proxy for positive investor sentiment indicated by rising stock prices. In addition, we find support for the relationship between message volume and trading volume (H2a). This 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 more than 10% increase in trading volume ($c = 10.798, p < 0.001$). In contrast to previous research (e.g., Wysocki (1998)), we reject the hypothesis that message volume can explain returns (H2b). In line with H2c, we observe an increase in volatility as the message volume rises ($c = 1.391, p < 0.001$). While disagreement

---

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

does no longer explain volatility (H3a) the negative correlation of agreement and trading volume (H3b) prevails (c = *-4.644, p < 0.001*).

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.

B.3. Time-sequencing regressions

While contemporaneous relationships between tweet and market features are noteworthy, the litmus test for the quality of information in microblogs are time-sequencing regressions. 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 analyze all relationships in both directions. In the following, we focus on those hypotheses that have not yet been rejected by previous analyses.

Table VI shows time-sequencing regressions for tweet and market features (in line with Antweiler and Frank (2004)). We regress one and two day lags of every tweet feature on every market feature separately (and vice versa). Similar to the contemporaneous regressions, we use panel regressions with company fixed effects and the market index as a control. Because market returns have been repeatedly found to be negative on the first trading day of the week (e.g., Lakonishok and Levi (1982)), we also include a dummy variable for this day (NWK) in line with Antweiler and Frank (2004). In order to assess the relative strength of the impact of tweet and market features on each other, we report standardized next to absolute coefficients.

[Insert Table VI about here]

The most obvious question in time-series analysis of microblogs and the market is whether message sentiment can help predict returns (H1). Table VI shows that, while there is almost no effect of bullishness on next day returns, bullishness two days ago (X-2) is, contrary to our hypothesis, associated with negative returns (*c = -0.057, p < 0.05*). On the other hand, previous day returns (Y-1) have a positive effect on bullishness (c = *0.035, p < 0.01*). The standardized coefficient shows that, in addition to higher statistical significance, the effect of returns on bullishness is about four times as strong as the inverse (c = *0.091, p < 0.001* vs. c = *-0.022, p < 0.05*). Thus, bullishness in stock microblogs is affected more strongly by returns than vice versa.

[Insert Table VI about here]

Message volume one and two days ago seems to predict current day trading volume (H2a). At the same time, high trading volume triggers increased message volume over the next two days. Autocorrelation of trading volumes may explain parts of this effect, which warrants a closer analysis of the relationship in the next section. However, similar to the relationship between bullishness and returns, the standardized coefficients illustrate that the stronger effect is in the direction from trading volume to message volume. High volatility also leads to increased message volume, confirming that uncertainty causes investors to exchange information and

consult their peers. The opposite relationship, however, does not hold (H2c). It is worth noting that, in contrast to Antweiler and Frank (2004), we do not find message volume to be related to stock returns (H2b). This indicates that investors may take a more nuanced approach in processing information content of stock microblogs compared to message boards.

In line with the contemporaneous regressions, disagreement is not associated with higher volatility (H3a). However, we find some confirmatory evidence for H3b that agreement among traders does lead to lower trading volumes ($c = -2.022, p = 0.01$ for X-2).

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

C. In-depth analysis for selected market features

In this section, we provide in-depth analyses exploring the two relationships that are most intriguing and were supported by the empirical evidence in the previous sections. These are the relationships between message volume and trading volume on one hand (section C.1), and bullishness and returns on the other (section C.2).[30]

---

[30] Because tweet features did not consistently explain changes in volatility in the previous sections, we do not report in-depth analysis for this market feature. We computed ARCH and GARCH models to reflect the autoregressive nature of volatility, but did not find consistent support for our hypothesis that disagreement among messages explains volatility.

C.1. Volume

While the pairwise correlations and predictive regressions have suggested that message volume and agreement contain valuable information with respect to trading volume, it remains to be seen whether these relationships can survive the inclusion of a more inclusive set of relevant control variables derived by Chordia, Roll, and Subrahmanyam (2001). In contrast to all other analyses, where we assign messages posted after 4:00 pm to the next trading day, we define message volume and agreement as the respective tweet features for the 24 hours prior to the market open. We take this approach for the purposes of this analysis because it more closely represents the information that is available to predict trading volume. Next to message volume and agreement, we add the control variables following Antweiler and Frank (2004). These include previous changes in the stock price, the market index, and market volatility as well as the federal funds rate (FFR), the quality spread (between corporate BB bond yields and the treasury rate), and the term spread between the FFR and the 10-year treasury bill rate. To capture calendar effects, we also add day of week dummies and a dummy for days preceding or following a public holiday. Due to the autocorrelation of trading volumes discussed in the previous section, we expand the list of controls used by Antweiler and Frank (2004) to also include changes in trading volumes in the preceding days. As in the previous section, we use fixed-effect panel regressions.

[Insert Table VII about here]

Table VII reveals two aspects concerning message volume and agreement. First, while the direction of the coefficients is consistent with previous analyses for both tweet features, only message volume survives the inclusion of the above-mentioned control variables. Second, a comparison of standardized coefficients shows that message volume explains trading volume better than some of these accepted control variables. We conclude that message volume contains valuable information with respect to next-day trading volume (H2a).

C.2. Return: Event-study of buy and sell signals

Our previous analyses have indicated that bullishness may contain new information not yet reflected in market prices. However, the aggregated sentiment measure does not allow us to decompose the distinct qualities of buy and sell signals. To explore the information contained in these sentiments, we conduct an event study of returns around the day of particular strong buy and sell signals. The development of returns around the event day can inform us about traders' motivation to recommend a stock. In line with Tumarkin and Whitelaw (2001), we define an event-day as a day when the bullishness index of a particular stock exceeds the previous 5-day average by at least two standard deviations. Event days with less than 5 messages for a stock were excluded from the sample.

[Insert Figure 4 about here]

41

Figure 4 shows the development of abnormal returns around the event days. We observe that buy recommendations are preceded by an extended period of negative returns. Abnormal returns are negative from 6 to 2 days prior to the recommendation. This is in line with our time-sequencing regressions, where 2 day lags of bullishness where negatively associated with returns. One day before the event day, returns bottom out. Following this sustained decrease, we notice a reversal in abnormal returns. More importantly, the return reversal allows us to earn statistically significant abnormal returns on the day following the recommendation. Although these abnormal returns are small (0.24%), they exceed frequently assumed levels of transaction costs for online brokers in the range of 0.15%-0.2% (Clarkson, Joyce, and Tutticci (2006))[31]. In addition, abnormal returns on the day following the recommendation are a conservative measure since many traders may pick up the signal on the event day and capture some of the abnormal returns on that day (0.45%).

We find a similar pattern with respect to sell recommendations. Returns of the recommended securities have risen steadily from days -6 to -2 until they reach a peak on t-1. Microbloggers then recommend selling the stock, which, indeed, shows statistically significant negative returns on that day (-0.19%). Even though the stock continues to fall for 3 more days, abnormal returns are no longer statistically significant.

In summary, our event study shows that microbloggers follow a contrarian strategy with strong buy signals being followed by abnormal next-day returns (H1). This finding is contrary to previous research of stock message boards, where users followed a naïve momentum strategy

---

[31] Tetlock, Saar-Tsechansky, and Macskassy (2008) even use only 10 bps to assume reasonable transaction costs.

and no new information was contained in their recommendations (Dewally (2003), Mizrach and Weerts (2009), Tumarkin and Whitelaw (2001)).

D. Information diffusion in stock microblogging forums

We have seen that stock microblogs do contain valuable information with respect to financial market developments. 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 examine whether high quality pieces of information (i.e., individual messages) may be weighted more heavily and spread through retweets (H4a). Second, 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 (H4b).

Greater weight given to high quality pieces of investment advice could 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. So, 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 discard our hypothesis that higher quality pieces of information are retweeted more frequently (H4b).

Yang and Counts (2010) have shown that the properties of users are stronger predictors of information propagation than properties of the tweets. Thus, next to individual tweets, weighing of information may occur on the user level. 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 VIII 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.[32] But, as the last column indicates, the higher frequency users do not appear to be better investment advisors as the average quality does not

---

[32] 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.

vary by user group.[33] 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.

[Insert Table VIII about here]

Next, we explore whether this quality is recognized 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.[34] Zhang 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 authors 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

_____

[33] Some studies of virtual communities argue that members are motivated to post more messages when they receive feedback that their postings generate a valuable exchange of knowledge (Konana, Rajagopalan, and Chen (2007)).

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

and -1, respectively. Zhang (2009) has shown that, while accuracy with respect to same-day returns did not affect a users 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 VIII). 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 labeled stock microbloggers separately.[35] 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.

[Insert Table IX about here]

---

[35] 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 IX shows the effect of the determinants of the three indicators of user influence. Obviously and as hypothesized, 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 lead to greater reputation, appreciation of information quality in stock microblogs appears to be more short-lived. In line with our event-study, which illustrated that microbloggers follow a contrarian strategy, 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 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.

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 (H4a).

IV. Discussion

A. 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. Our results are summarized in Table X.

[Insert Table X about here]

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 compares these tweet features with the corresponding market features return, trading volume, and volatility.

We hypothesized that increased bullishness of stock microblogs is associated with higher returns. We find support for both a contemporaneous as well as a lagged relationship of bullishness and abnormal returns. Our event study of buy and sell signals shows that microbloggers follow a contrarian strategy. Buy signals are accompanied and followed by

abnormal returns, which exceed frequently assumed levels of transaction costs. Sell signals have no predictive power for returns. However, our results indicate that new information, reflected in the tweets, is incorporated in market prices quickly, and reasonable transaction costs make it difficult to exploit market inefficiencies. Of course, we cannot rule out that more sophisticated algorithms for text classification or refined trading rules would indeed be profitable.

As hypothesized, message volume is consistently significant only in explaining trading volume, but not returns or volatility. The predictive power of message volume even survives the inclusion of numerous accepted control variables.

Disagreement is primarily contemporaneous with an increase in trading volume, but neither associated nor able to predict volatility. 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 standardized 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.

B. Limitations and further research

This study, like others, 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 the first study 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, the time-sequencing regressions confirmed the robustness of key results.

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 (2004) have pointed out that "one could try to determine which classes of events have particularly large effects for stock returns". 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.

C. 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 follower 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.

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 recognizes users who consistently offer high quality investment advice, although there are no simple rules to identify valuable pieces of information. One of the most critical aspects of further research will be to better understand the mechanisms by which information is weighted and diffused in microblogging forums. Until then, picking the right tweets remains just as difficult as making the right trades.

References

Antweiler, Werner, and Frank, Murray Z., 2004, Is all that talk just noise? The information content of internet stock message boards, *Journal of Finance* 59, 1259-1294.

Bagnoli, Mark, Beneish, Messod D., and Watts, Susan G., 1999, Whisper forecasts of quarterly earnings per share, *Journal of Accounting and Economics* 28, 27–50.

Barber, Brad, and Odean, Terrance, 2000, Trading is hazardous to your wealth: The common stock investment performance of individual investors, *The Journal of Finance* 55, 773-806.

Barber, Brad, and Odean, Terrance, 2008, All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors, *Review of Financial Studies* 21, 785-818.

Bettman, Jenni, Hallett, Aiden, and Sault, Stephen, 2000, The impact of electronic message board takeover rumors on the US equity market, Working paper, Australian National University.

Black, Fischer, 1986, Noise, *Journal of Finance* 41, 529-543.

Bloomberg, 2010, Hedge Fund Will Track Twitter to Predict Stock Moves, *Bloomberg*, *Online Edition (author Jack Jordan)*, December 22.

Bollen Johan, Mao Huina, and Zeng Xiao-Jun, 2010, Twitter mood predicts the stock market, Working paper, Indiana University.

Bommel Van, Jos, 2003, Rumors, *The Journal of Finance* 58, 1499–1520.

Brown, Gregory W., 1999, Volatility, sentiment, and noise traders, *Financial Analysts Journal* 55, 82-90.

Busheee, Brian J., Core, John E., Guay, Wayne, and Hamm, Sophia J., 2010, The role of the business press as an information intermediary, *Journal of Accounting Research* 48, 1-19.

BusinessWeek, 2009, StockTwits may change how you trade, *BusinessWeek, Online Edition (author Max Zeledon)*, February 11.

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

Campbell, John A., 2001, In and out scream and shout: An internet conversation about stock price manipulation, *Proceedings of the 34th Hawaii International Conference on System Sciences,* 1-10, Maui, HI.

Cao, H. Henry, Coval, Joshua D, and Hirshleifer, David, 2002, Sidelined investors, trading-generated news, and security returns, *Review of Financial Studies* 15, 615-648.

Cha, Meeyoung, Haddadi, Hamed, Benevenuto Fabrício, and Gummadi Krishna P., 2010, Measuring user influence in Twitter: The million follower fallacy, *International Conference on Weblogs and Social Media*, 10-17, Washington, DC.

Chordia, Tarun, Roll, Richard, and Subrahmanyam, Avanidhar, 2001, Market liquidity and trading activity, *Journal of Finance* 56, 501–530.

Clarkson, Peter M., Joyce, Daniel, and Tutticci, Irene, 2006, Market reaction to takeover rumour in internet discussion sites, *Accounting and Finance* 46, 31-52.

Danthine, Jean-Pierre, and Moresi, Serge, 1993, Volatility, information, and noise trading, *European Economic Review* 37, 961-982.

Das, Sanjiv, and Chen, Mike, 2007, Yahoo! for Amazon: Sentiment extraction from small talk on the web, *Management Science* 53, 1375-1388.

Das, Sanjiv, Martinez-Jerez, Asís, and Tufano, Peter, 2005, eInformation: A clinical study of investor discussion and sentiment, *Financial Management* 34, 103–137.

De Long, J. Bradford, Shleifer, Andrei, Summers, Lawrence H. and Waldmann, Robert J., 1990, Noise trader risk in financial markets, *Journal of Political Economy* 98, 703-738.

Delort, Jean-Yves, Arunasalam, Bavani, Milosavljevic, Maria, and Leung, Henry, 2009, The impact of manipulation in internet stock message boards, *Working paper*, Macquarie University, Sydney.

DeMarzo, Peter M., Vayanos, Dimitri, and Zwiebel, Jeffrey, 2003, Persuasion bias, social influence, and unidimensional opinions, *Quarterly Journal of Economics* 118, 909-968.

Dewally, Michael, 2003, Internet investment advice: Investing with a rock of salt, *Financial Analysts Journal* 59, 65-77.

Dyckman, Thomas, Philbrick, Donna, and Stephan, Jens, 1984, A comparison of event study methodologies using daily stock returns: A simulation approach, *Journal of Accounting Research* 22, 1-30.

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

Engelberg, Joseph, and Parsons, Christopher A., 2011, The causal impact of media in financial markets, *Journal of Finance* (forthcoming).

Fama, Eugene F., 1970, Efficient capital markets: A review of theory and empirical work, *Journal of Finance* 25, 383–417.

Fama, Eugene F., 1991, Efficient capital markets: II, *Journal of Finance* 46, 1575–1617.

Fan, Ming, Tan, Young, and Whinston, Andrew B., 2005, Evaluation and design of online cooperative feedback mechanisms for reputation management, *IEEE Transactions on Knowledge and Data Engineering* 17, 244-254.

Felton, James, and Kim, Jonchai, 2002, Warnings from the Enron message board, *The Journal of Investing* 11, 29–52.

Giller, Graham L., 2009, 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*, Giller Investments, New Jersey.

Gu Bin, Konana, Prabhudev, Rajagopalan, Balaji, and Chen, Hsuan-Wei, 2007, Competition among virtual communities and user valuation: The case of investing-related communities, *Information Systems Research* 18, 68-85.

Gu, Bin, Konana, Prabhudev, and Chen, Hsuan-Wei, 2008, Melting-pot or homophily? - An empirical investigation of user interactions in virtual investment-related communities, *Working paper*, University of Texas.

Hall, Mark, Frank, Eibe, Holmes, Geoffreu, Pfahringer, Bernhard, Reutemann, Peter, and Witten Ian H., 2009, The WEKA data mining software: An update, *SIGKDD Explorations* 11, 10-18.

Harris, Milton, and Raviv, Artur, 1993, Differences of opinion make a horse race, *Review of Financial Studies 6*, 473-506.

Hirshleifer, David, and Teoh, Siew, 2003, Limited attention, information disclosure, and financial reporting, *Journal of Accounting and Economics* 36, 337–386.

Hong, Harrison, Kubik, Jeffrey, and Stein, Jeremy C. ,2005, Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers, *The Journal of Finance* 60, 2801–2824.

Jones, Anne L., 2006, Have internet message boards changed market behavior?, *info* 8, 67-76.

Jones, Charles M., Kaul, Gautam, and Lipson, Marc L., 1994, Transactions, volume, and volatility, *Review of Financial Studies* 7, 631-651.

Karpoff, Jonathan. M., 1986, A theory of trading volume, *Journal of Finance* 41, 1069-1087.

Kim, Oliver, and Verrecchia, Robert, 1991, Trading volume and price reactions to public announcements; *Journal of Accounting Research* 29, 302-321.

Kosfeld, Michael, 2005, Rumours and markets, *Journal of Mathematical Economics* 41, 646–664.

Koski, Jennifer L., Rice, Edward M., and Tarhouni, Ali, 2004, Noise trading and volatility: Evidence from day trading and message boards, *Working paper*, University of Washington.

Lakonishok, Josef, and Levi, Maurice, 1982, Weekend Effects on Stock Returns: A Note, *Journal of Finance* 37, 883-889.

Lerman, Alina, 2010, Individual investors' attention to accounting information: Message board discussions, *Working paper*, New York University.

Malkiel, Burton G., 2003, The efficient market hypothesis and its critics, *The Journal of Economic Perspectives* 17, 59-82.

Milgrom, Paul, and Stokey, Nancy, 1982, Information, trade and common knowledge, *Journal of Economic Theory* 26, 17-27.

Mittermayer, Marc-André, and Knolmayer, Gerhard F., 2006, Text mining systems for market response to news: A survey, *Working paper*, University of Bern.

Mizrach, Bruce, and Weerts, Susan, 2009, Experts online: An analysis of trading activity in a public Internet chat room. *Journal of Economic Behavior & Organization* 70, 266-281.

Newcomb, Theordore M., 1953, An approach to the study of communicative acts, *Psychological Review* 60, 393-404.

Ng, Lilian, and Wu, Fei, 2006, Peer effects in the trading decisions of individual investors, *Financial Management* 39, 807-831.

O'Connor, Brendan, Balasubramanyan, Ramnath, Routledge, Bryan R., and & Smith, Noah, 2010, From Tweets to polls: Linking text sentiment to public opinion time series, *International Conference on Weblogs and Social Media,* 122-129, Washington, DC.

Parkinson, Michael, 1980, The Extreme Value Method for Estimating the Variance of the Rate of Return, *Journal of Business* 53, 61-65.

Romero Daniel M., Galuba, Wojciech, Asur, Sitaram, and Huberman, Bernardo A., 2010, Influence and passivity in social media, *Working paper*, Cornell University.

Sabherwal, Sanjiv, Sarkar, Salil K., and Zhang, Ying, 2008, Online talk: does it matter?, *Managerial Finance* 34, 423-436.

Shalen, Catherine T., 1993, Volume, volatility, and the dispersion of beliefs, *Review of Financial Studies* 6 405-434.

TechCrunch, 2010, Twitter seeing 90 million tweets per day, 25 percent contain links, *TechCrunch Blog (author Leena Rao)*, September 14.

Tetlock, Paul C., Saar-Tsechansky, Maytal, and Macskassy, Sofus, 2008, More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance* 63, 1437-1467.

TIME, 2009, Turning Wall Street on its head, *TIME magazine, Online Edition (author Douglas McIntyre)*, May 29.

Tumarkin, Robert, and Whitelaw, Robert F., 2001, News or noise? Internet postings and stock prices, *Financial Analysts Journal* 57, 41-51.

Vincent, Arnaud, and Armstrong Margaret, 2010, Predicting break-points in trading strategies with Twitter, *Working paper*, École Nationale Supérieure des Mines de Paris..

Wysocki, Peter, 1998, Cheap talk on the web: The determinants of postings on stock message boards, *Working paper*. University of Michigan.

Yang, Jiang, and Counts, Scott, 2010, Predicting the speed, scale, and range or information diffusion in Twitter, *International Conference on Weblogs and Social Media,* 355-358, Washington, DC.

Zhang, Wenbin, and Skiena, Steven, 2010, Trading strategies to exploit blog and news sentiment, In *International AAAI Conference on Weblogs and Social Media,* 275-378, Washington, DC.

Zhang, Xue, Fuehres, Hauke, and Gloor, Peter, 2010, Predicting stock market indicators through Twitter – "I hope it is not as bad as I fear", In *Collaborative Innovations Networks Conference,* 1-8, Savannah, GA.

Zhang, Ying, 2009, Determinants of poster reputation on internet stock message boards. *American Journal of Economics and Business Administration* 1, 114-121.

Zhang, Ying, and Swanson, Peggy E., 2010, Are day traders bias free? - Evidence from internet message boards. *Journal of Economic and Finance* 34, 96-112.

TABLE I

Sample Tweets from Training and Test Set including Classification

Tweets were randomly selected and are shown in their original format (before preprocessing).

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

TABLE II
Classification Accuracy (Confusion Matrix)

This table shows the accuracy of our classification method (for example, 35.2% of the training set were manually labeled as buy signals, 28.0%. of which were correctly classified as such by the model).

Manual classification		Automatic classification		
		Full training set		
		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%

TABLE III
Summary Statistics of Market and Tweet Features

All daily statistics are reported on a per company basis. Returns were defined as the log difference in prices, 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 their sentiment (e.g. buy vs. sell). We refer to our method section for the formulas used to compute these variables. All returns (for individual stocks as well as market indices) are scaled by 100 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 (log diff. in price, in %)	-0.06	1.87	-16.04	11.30
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 IV

Pairwise Correlations for Stock and Tweet Data

This table shows the pairwise correlations for tweet and market features. P-values reported below the correlation, correlations that are significantly different from zero at the 99% confidence level are reported in bold. N = 12,443 company-trading-days.

	Return	AR (simple)	AR (market model)	Traded volume	Volatility	Bullishne ss	Message volume
AR (simple)	**0.757**						
	0.00						
AR (market model)	**0.658**	**0.917**					
	0.00	0.00					
Traded volume	**-0.044**	0.004	**0.029**				
	0.00	0.65	0.00				
Volatility	**-0.063**	-0.016	-0.012	**0.046**			
	0.00	0.08	0.17	0.00			
Bullishness	**0.166**	**0.156**	**0.147**	**0.126**	-0.012		
	0.00	0.00	0.00	0.00	0.20		
Message volume	**0.028**	0.022	**0.024**	**0.441**	0.016	**0.340**	
	0.00	0.02	0.01	0.00	0.08	0.00	
Agreement	**0.036**	**0.035**	**0.033**	**-0.113**	-0.014	**0.362**	-0.016
	0.00	0.00	0.00	0.00	0.13	0.00	0.08

TABLE V
Contemporaneous Regressions

This table shows power of tweet features for explain changes in market features. The first row shows these market features as the dependent variable in panel regressions with company fixed-effects. All tweet features are used as independent variables and the market return added as a control. N = 12,443 company-trading-days.

	Return	AR^{simple}	$AR^{market\ model}$	Trading volume	Volatility
Bullishness	0.492***	0.347***	0.318***	-1.872***	-0.593
	18.41	*17.67*	*16.84*	*-3.65*	*-1.27*
Message volume	0.002	0.014	0.012	10.798***	1.391***
	0.10	*0.98*	*0.88*	*28.68*	*4.06*
Agreement (A)	-0.170***	-0.127***	-0.106***	-4.644***	-0.839
	-4.15	*-4.21*	*-3.66*	*-5.88*	*-1.17*
Market return	0.066***	0	-0.003	-2.057***	-0.160*
	16.17	*0.06*	*-1.08*	*-26.32*	*-2.24*
R^2	0.052	0.028	0.026	0.104	0.002
F-value	168.1***	89.0***	81.1***	358.7***	5.5***

Notes: * $p<0.05$, ** $p<0.01$, *** $p<0.001$, t-statistics in italics below the coefficients.

TABLE VI
Time-Sequencing Regressions

This table shows lagged for regressions for tweets features (*X*) explaining market features (*Y*) in the columns labeled *X->Y* and the inverse relationship in the columns labeled *Y->X*. The first row for each combination of tweet and market feature reports results for regression of actual values (bold), second row reports results for regression of standardized values (italics). All regressions use company fixed effects. N = 12,443 company-trading-days.

X	Y	X->Y					Y->X				
		X-1	X-2	NWK	Market	F-value	Y-1	Y-2	NWK	Market	F-value
Bullishness	**Return**	**0.006**	**-0.057***	**0.435*****	**0.068*****	97.2***	**0.035*****	**0.003**	**0.101*****	**0.006*****	41.5***
		0.002	*-0.022***	*0.091****	*0.146****		*0.091****	*0.009*	*0.054****	*0.033****	
Bullishness	**Volume**	**0.332**	**-1.191***	**-13.182*****	**-1.828*****	200.5***	**0.000**	**0.000**	**0.082*****	**0.009*****	15.6***
		0.002	*-0.008***	*-0.046****	*-0.066****		*0.049*	*0.014*	*0.044****	*0.049****	
Bullishness	**Volatility**	**-0.047**	**-0.922***	**-1.358**	**-0.128**	3.0*	**0.000**	**0.000**	**0.086*****	**0.008*****	101.6***
		-0.001	*-0.021***	*-0.016*	*-0.016*		*0.004*	*0.008*	*0.046****	*0.045****	
Messages	**Return**	**-0.013**	**0.032**	**0.436*****	**0.067*****	96.3***	**0.006**	**0.003**	**0.348*****	**0.016*****	95.0***
		-0.010	*0.024*	*0.091****	*0.144****		*0.008*	*0.004*	*0.097****	*0.047****	
Messages	**Volume**	**5.777*****	**1.590*****	**-12.526*****	**-1.966*****	287.2***	**0.004*****	**0.001****	**0.296*****	**0.025*****	202.8***
		*0.073****	*0.020****	*-0.044****	*-0.071****		*0.312****	*0.057***	*0.082****	*0.072****	
Messages	**Volatility**	**0.064**	**0.331**	**-1.315**	**-0.143***	2.0	**0.001*****	**0.001****	**0.342*****	**0.017*****	101.6***
		0.003	*0.014*	*-0.016*	*-0.018***		*0.025****	*0.015***	*0.095****	*0.049****	
Agreement	**Return**	**0.024**	**0.047**	**0.436*****	**0.067*****	96.3***	**0.005***	**0.000**	**0.013**	**0.000**	1.5
		0.006	*0.011*	*0.091****	*0.145****		*0.019***	*0.002*	*0.011*	*0.003*	
Agreement	**Volume**	**-1.370**	**-2.022****	**-13.164*****	**-1.830*****	201.9***	**0.000**	**0.000**	**0.013**	**0.000**	1.0
		-0.006	*-0.008****	*-0.046****	*-0.066****		*-0.048*	*0.021*	*0.011*	*0.004*	
Agreement	**Volatility**	**-0.254**	**0.798**	**-1.329**	**-0.137**	2.1	**0.000**	**0.000**	**0.011**	**0.001**	1.4
		-0.004	*0.011*	*-0.016*	*-0.017*		*-0.008*	*0.015*	*0.009*	*0.006*	

Notes: * p<0.05, ** p<0.01, *** p<0.001

TABLE VII
Volume Regression

This table presents the explanatory power of message volume and agreement for trading volume. It shows that message volume can significantly predict next day trading volume. The dependent variable is the natural logarithm of the trading volume. The independent variables are defined as follows: Message volume and agreement are the tweet features laid out in the section detailing our methodology with a one day lag relative to the dependent variable. In line with Antweiler and Frank (2004), changes in trading volume, stock returns, and market returns are calculated identically, e.g., *stock up yesterday = max {0, r_{t-1} - r_{t-2})}* where r_t represents the return on day t. *Stock 5 day volatility* is the 5 day average of *PARK volatility*. Federal funds rate (FFR) is the US federal funds rate, *the quality spread = ln(1+BB)-ln(1+T10)* where *BB* is the BB corporate bond yield and *T10* represents the 10 year US government yield. The *term spread = ln(1+T10)-ln(1+FFR)*. *Monday* through *Thursday* are day of week dummies and *holiday* a dummy for days preceding or following a public holiday. N = 12,443 company-trading-days.

	Actual values	Standardized coefficients
Message volume	0.014***	0.019***
Agreement	-0.008	-0.004
Trading volume up yesterday	-0.266***	-0.059***
Trading volume down yesterday	0.181***	0.036***
Trading volume up last 5 days	-0.265***	-0.073***
Trading volume down last 5 days	0.087***	0.022***
Stock up yesterday	-1.575***	-0.017***
Stock down yesterday	0.201	0.002
Stock up last 5 days	-0.835***	-0.015***
Stock down last 5 days	-0.835***	-0.015***
Stock 5 day volatility	-0.001***	-0.010***
Market up yesterday	-9.750***	-0.064***
Market down yesterday	-2.076***	-0.017***
Market up last 5 days	2.220***	0.022***
Market down last 5 days	0.525	0.008
Market 5 day volatility	0.009***	0.047***
Federal funds rate	-0.329**	-0.010**
Quality spread	-0.497***	-0.025***
Term spread	-0.074	-0.005
Monday	-0.259***	-0.100***
Tuesday	-0.041***	-0.017***
Wednesday	-0.081***	-0.033***
Thursday	-0.081***	-0.032***
Holiday	-0.123***	-0.022***
Observations	11,736	
Adj. R^2	0.273	
F-value	188.6***	

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

TABLE VIII

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.

Tweets per user	Users		Message volume		Average quality
	Total	*Share*	*Total*	*Share*	
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%

TABLE IX

Determinants of User Influence

This table shows the power of the quality of investment advice to explain 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. Stock microbloggers are defined as users, which have posted at least 5 stock microblogs during our sample period. The sample was restricted to users with at least 20 messages, representing the two most frequently posting user groups from Table VIII. We show the results for negative binomial regressions. We report the coefficients for the robust version of the regression using a sandwich estimator of variance.

	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*
Chi²	396.7***	234.6***	34.1***	16.5**
Observations	614	614	614	604

Notes: * p<0.05, ** p<0.01, *** p<0.001.

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 been 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. We report the coefficients for the robust version of the regression using a sandwich estimator of variance. This estimator is robust to most types of misspecification as long as the observations are independent (Cameron and Trivedi (1998)). Message volume and followership are highly skewed and thus log-transformed when used as independent variables. Results are robust to other user definitions (e.g. with fewer messages) and regressions models (e.g. poisson regressions).

TABLE X
Summary of Results

Research Questions and related Hypotheses		Results	
RQ1	**Analysis of tweet and market features**		
		Contemporaneous relationship	*Lagged relationship*
	Bullishness		
H1	Increased bullishness of stock microblogs is associated with higher returns.	*Yes*	*Yes*
	Message volume		
H2a	Increased message volume in stock microblogging forums is associated with an increase in trading volume.	*Yes*	Yes
H2b	Increases in message volume in stock microblogging forums are associated with higher returns.	No	No
H2c	Increased message volume in stock microblogging forums is associated with higher volatility.	Yes	No
H2d	Increased message volume in stock microblogging forums reduces information asymmetry indicated by lower spreads.	No	No
	Agreement		
H3a	Increased disagreement among stock microblogs is associated with higher volatility.	No	No
H3b	Increased disagreement among stock microblogs is associated with an increase in trading volume.	Yes	(Yes)*
RQ2	**Analysis of information diffusion**		
H4a	Users who consistently provide high quality investment advice have more influence in the microblogging forum (through retweets, mentions, or followers).	Yes	
H4b	High quality pieces of investment advice are spread more widely than low quality pieces of advice.	No	

Notes: *Only one of the two tweet feature lags helped explain the market feature with statistical significance.

FIGURE 1
Hourly Message Volume

This figure presents the distribution of stock microblogs throughout the day. The graph shows the message volume in the 60 minutes following the indicated hours. We notice a substantial spike in message volume during trading hours indicating that investment professionals are using stock microblogs to exchange trading ideas in real-time. Results are based on our sample of 249,533 stock-related microblogging messages containing the dollar-tagged ticker symbol of an S&P 100 company.

Hourly distribution of message volume
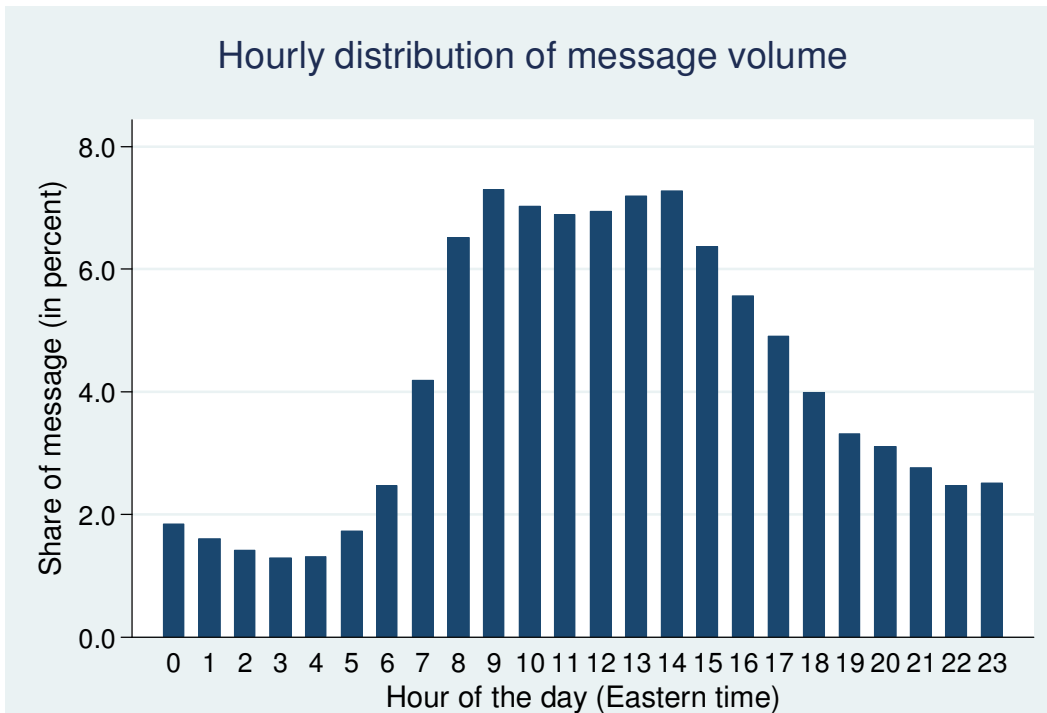
FIGURE 2
Aggregate Message Volume vs. Trading Volume

This figure presents the development of total message volume and trading volume on a weekly basis and shows some notable correlations such as the spike in both measures in the months of April and May.

FIGURE 3
Aggregate Bullishness vs. Market Return

This figure presents the development of overall bullishness and the S&P 1000 market index.
Both weekly indices are market-cap-weighted, i.e. to construct the bullishness index,
bullishness of individual stocks was weighted by the market capitalization of the company.

Development of bullishness and stock indices
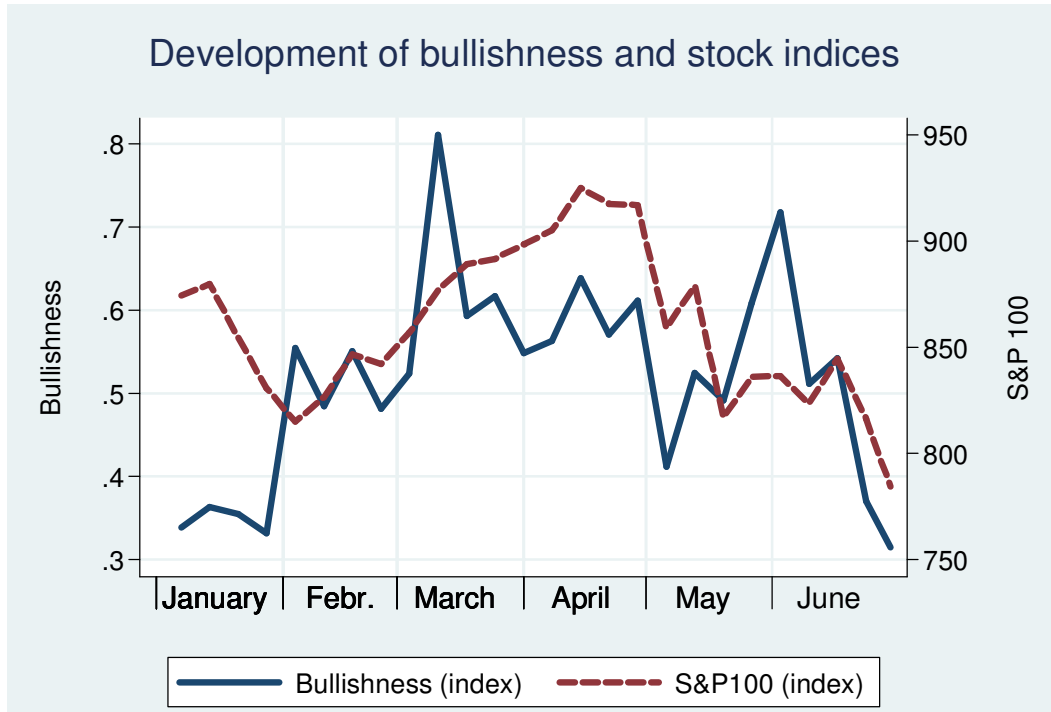
FIGURE 4
Event-Study of Buy and Sell-Signals

This figure shows the returns around event days (t=0) with a substantial increase in buy or sell signals for a particular company. A substantial increase was defined as a deviation of at least 3 standard deviations from the previous 5 day average. The figure shows that, according to these signals, stock microbloggers follow a contrarian strategy. Returns are shown in percent.

Returns around Event Day
Returns Around Day with Buy- and Sell-Signals

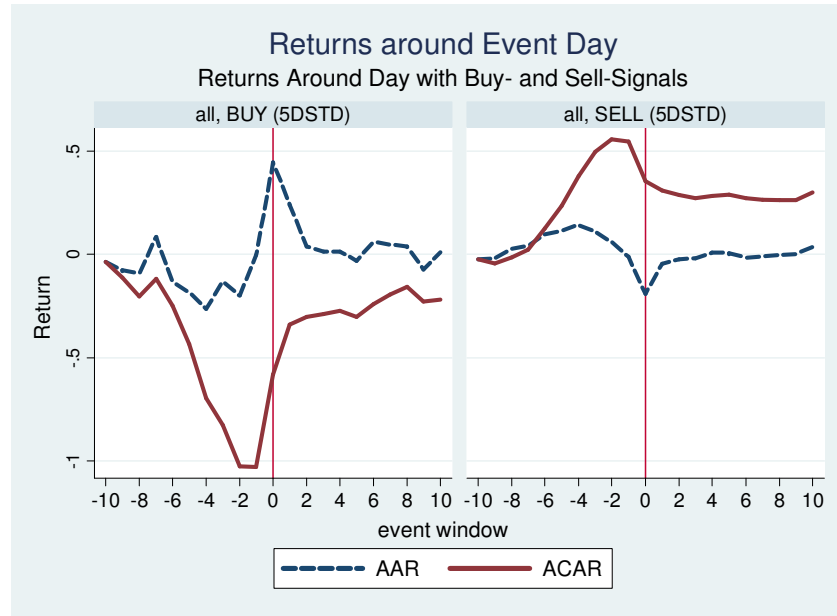| Day	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10
Stock returns around Buy Signals (394 observations)																					
AAR	-0.036	-0.075	-0.092	0.086	**-0.132**	**-0.184**	**-0.263**	**-0.130**	**-0.199**	-0.003	**0.447**	**0.242**	0.039	0.014	0.014	-0.030	0.061	0.048	0.038	-0.073	0.012
p-value (AAR)	0.53	0.31	0.19	0.22	**0.05**	**0.02**	**0.00**	**0.09**	**0.01**	0.96	**0.00**	**0.00**	0.60	0.83	0.83	0.65	0.39	0.44	0.54	0.34	0.86
ACAR	-0.036	-0.111	-0.203	-0.118	-0.250	-0.434	-0.697	-0.827	-1.026	-1.029	-0.583	-0.341	-0.302	-0.288	-0.274	-0.303	-0.242	-0.195	-0.157	-0.230	-0.218
Stock returns around Sell Signals (1216 observations)																					
AAR	-0.024	-0.019	0.027	0.042	**0.097**	**0.114**	**0.144**	**0.112**	0.060	-0.012	**-0.191**	-0.046	-0.023	-0.018	0.010	0.006	-0.016	-0.010	-0.002	0.001	0.038
p-value (AAR)	0.53	0.59	0.50	0.25	**0.01**	**0.00**	**0.00**	**0.00**	0.12	0.75	**0.00**	0.22	0.51	0.60	0.80	0.87	0.64	0.77	0.96	0.97	0.29
ACAR	-0.024	-0.044	-0.013	0.025	0.124	0.237	0.379	0.499	0.558	0.547	0.355	0.311	0.288	0.272	0.283	0.291	0.273	0.265	0.263	0.264	0.300

Supplementary Appendix

Appendix A. 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

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

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

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

where W_c is the total number of occurrences of word w in training documents of class c. We include Laplace Smoothing to minimize 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

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

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). 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, Salton, and Allan (1993)) and added words that were relevant to our particular context (e.g., company names). Third, we tokenized 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, Saar-Tsechansky, and Macskassy (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, Saar-Tsechansky, and Macskassy (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 distributions 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).

Appendix B. Classification of our data set

Table BI 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 BII shows, overall in-sample classification

accuracy was 81.2%. Even a more conservative 10-fold cross validation[36] 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, Faraj, and Teigland (2004)). The accuracy by class further validates the use of automatically labeled 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. 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.

A look at the most common words per class (see Table BIII) 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

---

[36] See the note to Table 2 for details regarding this validation approach.

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 the finding of Tetlock, Saar-Tsechansky, and Macskassy (2008). 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).

Appendix C. Market and tweet features per company

Table CI shows summary statistics by company. It illustrates that there are some stocks, especially high-tech companies and to a lesser extent financial institutions, that trigger a majority of the conversation. In line with Antweiler and Frank (2004), whose bullishness index for the Dow Jones Internet Index (XLK) was about twice as high as for the Dow Jonnes Industrial Average (DIA), bullishness in our sample period appears to be somewhat higher for technology companies such as Apple, Microsoft and Google.

Appendix D. Trading strategy

The event-study has shown that message sentiment, in particular buy signals, can inform us about future stock returns. However, the insight with respect to the exploitability and the economic impact of these signals is limited. To more thoroughly test the ability to earn abnormal profits based on message bullishness, we design a market-neutral trading strategy in line with Zhang and Skiena (2010). On every trading day, we buy (and sell) the stocks with the highest

(and lowest) level of bullishness. We distribute our investment equally across the selected stocks. As usual in this type of analysis, we initially ignore transaction costs (Zhang and Skiena (2010)). There are three key parameters, which may influence our trading result. First, the number of previous days we use to calculate bullishness (i.e., sentiment history). Second, the number of stocks we select from the top and bottom of our bullishness ranking (number of stock picks), and third, the number of days for which we will hold these stocks (holding period).[37] In order to better understand the impact of these parameters on our trading results, we conducted a sensitivity analysis backtesting the performance of our strategy. In this analysis, we alter one of the three parameters while leaving the other two constant. We chose the default values based on our results from the previous sections: Due to the negative correlation of returns and more than 1-day lags of bullishness found in our predictive regression, we choose a short 1 day default value for the sentiment history. Buy signals, as defined in our event study, were followed by statistically significant abnormal returns only one day after the signal. Assuming that new information reflected in the tweets is incorporated into prices quickly, we choose a default holding period of one day. The first and last stocks resulting from our bullishness ranking should be best and worst performers. However, due to noise in our data and potential benefits of diversification we choose a default value of 3 picks from both the top and bottom of our ranking of the S&P 100 constituents.

Figure D1 shows the results from the sensitivity analysis of our trading strategy. Next to the total return, we report profits derived from the long and short end of the strategy separately. We

---

[37] In line with our event-study we ignore observations with less than 5 messages per day of the specified sentiment history.

can see that a number of strategies earn abnormal returns of more than 15% in our 6 month sample period. Our default parameters, for instance, generate a profit of 16.2%. Generally, short sentiment signals produce higher returns. Longer sentiment signals even produce negative returns, confirming the contrarian strategy embedded in message bullishness (e.g., a few days prior to strong sell signals, stocks prices are actually rising sharply, explaining the loss from the short end of the trading strategy with 3-day-old sentiment signals). Except for some benefits of noise reduction with up to 3 stocks, increasing the number of stock picks generally decreases returns. This indicates that the ranking identified the top performers fairly accurately. In the short term, only a one-day holding period is associated with positive returns for both the long and short position. The fact that our strategy performs even better with long holding periods (of between 6 and 9 days) is in line with the drift in cumulative returns found in our event-study. However, we would be cautious to interpret these returns as the reaction to information contained in the news.

Tetlock, Saar-Tsechansky, and Macskassy (2008), who found a trading strategy based on negative words in firm specific news articles to earn abnormal annualized returns of more than 20%[38], showed that these returns evaporated with the inclusion of reasonable transaction costs. Therefore, to better understand the exploitability of our trading scheme, we determine its sensitivity with respect to trading costs. Table DI shows the profits derived from the long and short end of the trading scheme based on the default values for sentiment history, number of picks, and holding period with different transaction costs. We can see that the strategy starts

---

[38] The analysis does not distinguish between the long and short end of the trading strategy.

losing money with transaction costs of more than 7 bps, roughly in line with the threshold determined by Tetlock, Saar-Tsechansky, and Macskassy (2008). In addition, our analysis shows a slightly better performance of the short end of the trading scheme. However, even this side supports transaction costs only up to 8 bps.

In conclusion, we find that a strategy based on bullishness signals can earn substantial abnormal returns. Strategies based on short-term signals and short holding periods produce higher returns, indicating that new information, reflected in the tweets, is incorporated in market prices quickly. However, market inefficiencies are difficult to exploit with the inclusion of reasonable trading costs. Of course, we cannot rule out that more refined trading rules would be profitable.

References Appendix

Antweiler, Werner, and Frank, Murray Z., 2004, Is all that talk just noise? The information content of internet stock message boards, *Journal of Finance* 59, 1259-1294.

Buckley, Chris, Salton, Gerard, and Allan, James, 1993, Automatic retrieval with locality information using SMART, *First Text Retrieval Conference,* 59-72, Gaitherburg, MD.

Koppel, Moshe, and Shtrimberg, Itai, 2006, Good news or bad news? Let the market decide, *Computing Attitude and Affect in Text: Theory and Applications* 20, 297-301.

Porter, Martin F., 1980, An algorithm for suffix stripping, *Program* 14, 130–137.

Rennie, Jason D., Shih, Lawrence, Teevan, Jamie, and Karger, David R., 2003, Tackling the poor assumptions of Naive Bayes text classifiers, *Proceedings of the Twentieth International Conference on Machine Learning*, 616-623, Washington DC.

Tetlock, Paul C., Saar-Tsechansky, Maytal, and Macskassy, Sofus, 2008, More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance* 63, 1437-1467.

Wasko, Molly, Faraj, Samer, and Teigland, Robin, 2004, Collective action and knowledge contribution in electronic networks of practice, *Journal of the Association for Information Systems* 5, 493-513.

Zhang, Wenbin, and Skiena, Steven, 2010, Trading strategies to exploit blog and news sentiment, *International AAAI Conference on Weblogs and Social Media,* 275-378, Washington, DC.

TABLE BI

Sample Tweets from Training and Test Set including Classification

Tweets were randomly selected and are 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

Sample tweets* (main data set)	Automatic classification		
	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%

Classification Accuracy (Confusion Matrix)

This table shows the accuracy of our classifier (for example, 35.2% of the training set were manually labeled as buy signals, 28.0%. of which were correctly classified as such by the model). Classification accuracy for the full training set shows the accuracy of the model applied to the whole training set. Since the training set provides the input for the model, this performance measure bears the risk of crediting overfitting. Therefore, 10-fold cross validation provides a more conservative measure of accuracy. In this case, the training set is split in 10 parts of equal size, each of which are classified based on a model trained on the remaining 9/10 of the dataset.

Manual classification		Automatic classification		
		Full training set		
		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%
		10-fold cross validation		
		Buy	*Hold*	*Sell*
Buy	35.2%	22.1%	10.6%	2.5%
Hold	49.6%	9.4%	34.0%	6.1%
Sell	15.2%	2.2%	5.0%	8.0%
Training set		33.8%	49.6%	16.6%
All messages		23.0%	67.0%	10.0%

Classification Accuracy (Accuracy by Class)

This table presents the classification accuracy per class. It shows 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 labeled as such in the training set. False positives are message 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 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).

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%

TABLE BIII
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

TABLE CI
Summary Statistics by Company

This table presents summary statistics for the market and tweet features for all companies in our sample.

Ticker	Company name	Mes-sages	Bullish-ness	Agree-ment	Return	Volume	Vola-tility
AAPL	Apple Inc.	53,185	0.960	0.128	17.7	3,043,133	4.50
GOOG	Google Inc.	28,727	0.813	0.113	-33.2	439,630	2.29
GS	Goldman Sachs Group	18,676	0.272	0.058	-24.7	1,891,813	4.17
MSFT	Microsoft Corp.	14,930	1.165	0.214	-27.2	8,256,065	2.27
T	AT&T Inc.	13,427	0.774	0.152	-11.6	3,961,095	1.16
C	Citigroup Inc.	11,770	0.567	0.097	12.7	82,918,294	5.42
AMZN	Amazon Corp.	8,465	0.437	0.075	-20.8	1,046,927	4.57
BAC	Bank of America Corp.	7,541	0.486	0.084	-4.6	24,403,827	4.44
F	Ford Motor	6,690	0.572	0.150	0.8	14,399,336	6.56
JPM	JPMorgan Chase & Co.	4,785	0.572	0.136	-12.7	5,722,338	3.34
VZ	Verizon Communications	3,356	0.792	0.312	-13.7	2,679,188	1.18
CSCO	Cisco Systems	2,945	0.756	0.349	-11.6	6,806,863	2.51
INTC	Intel Corp.	2,877	0.850	0.367	-3.3	8,672,286	2.69
MS	Morgan Stanley	2,861	0.399	0.270	-24.0	2,518,266	4.42
HPQ	Hewlett-Packard	2,840	0.894	0.430	-17.1	1,969,842	3.01
GE	General Electric	2,610	0.844	0.348	-3.6	10,533,259	3.77
ORCL	Oracle Corp.	2,446	0.705	0.367	-13.0	3,754,551	2.34
FCX	Freeport-McMoran Cp & Gld	2,413	0.382	0.254	-30.2	2,038,341	5.96
WAG	Walgreen Co.	2,307	0.225	0.316	-31.1	1,098,243	2.29
AA	Alcoa Inc	2,202	0.283	0.350	-46.7	4,739,617	5.00
WMT	Wal-Mart Stores	1,950	0.306	0.181	-9.5	1,912,608	0.85
BA	Boeing	1,818	1.031	0.447	16.0	815,745	3.30
XOM	Exxon Mobil Corp.	1,776	0.552	0.334	-16.5	3,822,851	1.49
MA	Mastercard Inc.	1,774	0.422	0.372	-24.8	243,691	3.98
IBM	International Bus. Machines	1,732	0.725	0.413	-4.9	906,381	1.28
QCOM	QUALCOMM Inc.	1,698	0.366	0.397	-33.3	3,249,124	2.24
DELL	Dell Inc.	1,630	0.408	0.362	-17.5	3,584,715	4.15
DIS	Walt Disney Co.	1,613	0.718	0.493	-2.4	1,677,664	2.46
WFC	Wells Fargo	1,611	0.459	0.376	-5.0	5,671,621	3.88
PFE	Pfizer Inc.	1,603	0.528	0.375	-22.3	8,068,317	1.84
CAT	Caterpillar Inc.	1,521	0.507	0.378	6.6	1,228,493	4.26
CVX	Chevron Corp.	1,255	0.409	0.404	-10.7	1,453,107	1.63
MON	Monsanto Co.	1,190	0.114	0.314	-56.3	1,019,700	2.64
MCD	McDonald's Corp.	1,183	0.604	0.480	7.0	933,283	0.93
DOW	Dow Chemical	1,152	0.184	0.331	-14.2	1,482,144	5.85
JNJ	Johnson & Johnson	1,130	0.424	0.457	-7.0	1,659,560	0.88
HAL	Halliburton Co.	1,092	0.225	0.447	-19.7	2,353,208	7.49
MRK	Merck & Co.	1,078	0.420	0.380	-2.3	2,187,360	2.60
ALL	Allstate Corp.	1,031	0.346	0.446	-3.2	612,893	2.15
PG	Procter & Gamble	951	0.386	0.522	0.4	1,617,361	7.00
KO	Coca Cola Co.	900	0.555	0.544	-11.2	1,418,457	0.94
TGT	Target Corp.	831	0.213	0.363	2.3	820,907	1.82
NKE	NIKE Inc.	806	0.347	0.414	3.0	375,354	2.00
CMCSA	Comcast Corp.	805	0.559	0.523	4.0	2,846,721	3.36
AXP	American Express	791	0.192	0.433	-0.7	1,388,546	3.81
LOW	Lowe's Cos.	786	0.316	0.412	-12.9	1,805,030	2.59
TWX	Time Warner Inc.	775	0.433	0.474	0.7	1,150,284	2.42
FDX	FedEx Corporation	743	0.289	0.464	-17.1	421,743	2.90
HD	Home Depot	726	0.661	0.568	-1.6	2,314,273	2.01
AMGN	Amgen	691	0.255	0.439	-7.3	926,956	1.73
COST	Costco Co.	685	0.237	0.501	-7.0	497,737	1.09

Ticker	Company name	Mes-sages	Bullish-ness	Agree-ment	Return	Volume	Vola-tility
GILD	Gilead Sciences	640	0.162	0.338	-23.3	1,565,378	2.33
MO	Altria Group Inc.	638	0.350	0.539	5.5	2,070,532	1.23
SLB	Schlumberger Ltd.	625	0.381	0.417	-15.5	1,694,161	4.04
CL	Colgate-Palmolive	625	0.299	0.473	-3.0	323,919	0.93
MMM	3M	578	0.308	0.499	-3.3	561,743	3.45
BMY	Bristol-Myers Squibb	575	0.331	0.484	1.2	1,822,071	1.72
KFT	Kraft Foods Inc-A	555	0.408	0.457	4.9	2,153,003	1.46
COP	ConocoPhillips	546	0.323	0.490	-1.9	1,458,102	2.35
UPS	United Parcel Service	477	0.366	0.427	0.7	697,816	1.91
PEP	PepsiCo Inc.	465	0.375	0.476	1.7	978,175	0.93
EMC	EMC Corp.	447	0.358	0.435	4.6	3,056,842	2.57
UNH	United Health Group Inc.	446	0.201	0.398	-6.6	1,485,055	3.53
DD	Du Pont (E.I.)	444	0.271	0.475	5.0	993,291	2.78
COF	Capital One Financial	441	0.282	0.442	5.2	886,405	5.31
LMT	Lockheed Martin Corp.	416	0.526	0.547	0.5	316,952	1.62
DVN	Devon Energy Corp.	412	0.327	0.451	-18.3	587,302	3.44
SO	Southern Co.	408	0.244	0.456	2.5	587,372	0.86
TXN	Texas Instruments	386	0.385	0.481	-10.3	1,995,061	3.07
NYX	NYSE Euronext	381	0.298	0.444	10.9	430,898	3.62
CVS	CVS Caremark Corp.	379	0.295	0.449	-8.9	1,503,006	2.08
BHI	Baker Hughes	370	0.167	0.516	3.3	842,072	5.98
OXY	Occidental Petroleum	368	0.205	0.391	-4.4	744,503	3.04
RF	Regions Financial Corp.	365	0.189	0.391	22.1	3,459,752	8.83
PM	Philip Morris International	346	0.387	0.507	-2.6	1,085,600	9.89
BAX	Baxter International Inc.	344	0.143	0.504	-35.5	760,735	1.57
MET	MetLife Inc.	327	0.327	0.461	6.6	871,863	4.44
AVP	Avon Products	284	0.193	0.550	-15.8	615,081	3.39
NWSA	News Corporation	283	0.325	0.399	-12.8	2,734,286	5.50
MDT	Medtronic Inc.	274	0.272	0.342	-18.4	775,111	1.64
USB	U.S. Bancorp	258	0.097	0.398	-0.3	1,863,897	3.19
BK	Bank of New York Mellon	256	0.240	0.411	-11.9	1,143,712	2.61
HON	Honeywell Int'l Inc.	240	0.237	0.369	1.0	728,071	2.76
CPB	Campbell Soup	221	0.192	0.458	6.6	299,602	0.87
AEP	American Electric Power	220	0.233	0.602	-4.9	458,002	1.77
XRX	Xerox Corp.	219	0.197	0.258	-4.2	2,110,388	4.84
EXC	Exelon Corp.	218	0.209	0.415	-22.8	632,020	27.49
GD	General Dynamics	211	0.349	0.452	-13.4	270,957	2.45
NOV	National Oilwell Varco Inc.	199	0.216	0.408	-28.2	780,903	5.11
HNZ	Heinz (H.J.)	196	0.179	0.385	3.0	333,837	1.42
SLE	Sara Lee Corp.	187	0.275	0.346	16.2	1,089,132	1.57
RTN	Raytheon Co.	186	0.317	0.409	-5.6	338,520	1.51
UTX	United Technologies	163	0.210	0.327	-5.5	676,545	1.83
BNI	Burlington Northern Santa Fe	155	0.444	0.359	1.9	482,426	0.02
COV	Covidien	145	0.187	0.256	-16.8	444,585	1.99
WMB	Williams Cos.	132	0.067	0.188	-13.1	887,820	3.73
WY	Weyerhaeuser Corp.	125	0.152	0.277	-20.1	257,024	3.76
NSC	Norfolk Southern Corp.	122	0.167	0.267	2.5	346,295	2.85
ETR	Entergy Corp.	103	0.176	0.326	-11.3	184,008	1.61

TABLE DI
Sensitivity of Trading Strategy to Transaction Costs

This table shows the sensitivity of our trading strategy based on tweet signals relative to various levels of trading costs. It shows that our strategy is profitable only up to trading costs of about 7 bps. The trading strategy is based on the default values for sentiment history (1 day), number of picks (top and bottom 3 companies), and holding period (1 day). Market model abnormal returns are shown for the 6 month backtesting period. Trading costs are round-trip transaction costs in basis points.

Trading costs (bps)	AR (total), in %	AR (long)	AR (short)
0	16.2	6.6	9.6
1	13.9	5.5	8.4
2	11.6	4.3	7.3
3	9.3	3.2	6.2
4	7.1	2.0	5.0
5	4.8	0.9	3.9
6	2.5	-0.2	2.7
7	0.2	-1.4	1.6
8	-2.1	-2.5	0.5
9	-4.3	-3.7	-0.7
10	-6.6	-4.8	-1.8
11	-8.9	-5.9	-3.0
12	-11.2	-7.1	-4.1
13	-13.4	-8.2	-5.2
14	-15.7	-9.4	-6.4
15	-18.0	-10.5	-7.5

Trading Strategy based on Tweet Signals (Backtesting Results)

This figure presents the returns of a trading strategy based on tweet signals. The strategy is based on using the sentiment signals of the past x days, buying (selling) the most (least) bullish stocks (number of picks), and holding these stocks for a certain number of days (holding period). One of these parameters is altered in turn with the other two left constant. The default values are 1 (sentiment signal), 3 (number of picks), and 1 (holding period). The figure shows that there are a number of trading strategies that lead to excess returns of more than 10% with short sentiment signals and holding periods generally resulting in higher returns. Excess returns are market model abnormal returns as defined in our methodology section. Using simple abnormal returns as the performance measure produces a similar pattern, so we do not report these results separately.