TECHNISCHE UNIVERSITÄT MÜNCHEN
Lehrstuhl für Betriebswirtschaftslehre – Strategie und Organisation
Univ.-Prof. Dr. Isabell Welpe

Essays on the Information Content of Microblogs and their Use as an Indicator of Real-World Events

Timm O. Sprenger

Vollständiger Abdruck der von der Fakultät für Wirtschaftswissenschaften der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Wirtschaftswissenschaften (Dr. rer. pol.) genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr. Joachim Henkel
Prüfer der Dissertation:	1. Univ.-Prof. Dr. Isabell Welpe
	2. Univ.-Prof. Dr. Christoph Kaserer

Die Dissertation wurde am 09.06.2011 bei der Technischen Universität München eingereicht und durch die Fakultät für Wirtschaftswissenschaften am 20.07.2011 angenommen.

"What we have to do is deliver to people the best and freshest most relevant information possible. We think of Twitter not as a social network, but an information network."

Twitter co-founder Evan Williams

Table of Contents - Overview

Table of Contents - Details

List of Figures

I. Introduction

1. Predicting Elections with Twitter

2. Tweets and Trades

4. Followers and Foes

5. TweetTrader.net

List of Tables

1. Predicting Elections with Twitter

2. Tweets and Trades

3. News or Noise?

4. Followers and Foes

List of Abbreviations

AAR	Average Abnormal Return
AR	Abnormal Return
API	Application Programming Interface
CAR	Cumulative Abnormal Return
CDU	Christlich Demokratische Union (Christian Democratic Party)
CEO	Chief Executive Officer
CSU	Christlich Soziale Union (Christian Social Union)
CMP	Comparative Manifesto Project
DJIA	Dow Jones Industrial Average
e.g.	exempli gratia
EMB	Enterprise Microblogging
EMH	Efficient Market Hypothesis
et al.	et alii
FDP	Freie Demokratische Partei (Liberal Party)
FFR	Federal Funds Rate
GICS	Global Industry Classifications Standard
HSX	Hollywood Stock Exchange
i.e.	id est
ICS	Index of Consumer Sentiment
IDS	Internet Discussion Site
IEM	Iowa Electronic Market
LIWC	Linguistic Inquiry and Word Count
M&A	Mergers & Acquisitions
MAPE	Mean Absolute Percentage Error
MAE	Mean Absolute Error
NAICS	North American Industry Classification System
NASDAQ	National Association of Securities Dealers Automated Quotation
NLP	Natural Language Processing
NYSE	New York Stock Exchange
OLS	Ordinary Least Square
p.	page

QAP	Quadratic Assignment Procedure
RT	Retweet
RQ	Research Question
S&P	Standard & Poor's
SIC	Standard Industrial Classification
SNA	Social Network Analysis
SPD	Sozialdemokratische Partei Deutschlands (Social Democratic Party)
U.S.	United States of America
VIC	Virtual Investment Community
vs.	versus

I. Introduction

1 Motivation

This section will provide a brief introduction to the microblogging forum Twitter[1], illustrate practical applications and major research areas and derive the overarching research questions of this dissertation. Twitter[2] is a microblogging service launched in 2006 with now more than 145 million registered users (TechCrunch, 2010). Through the Twitter platform, every user can publish short messages with up to 140 characters, so-called "tweets", which are visible on a public message board of the website. However, only 25% of the traffic enters directly through the Twitter website with the rest published through nearly 300,000 third-party applications, which can freely connect to Twitter's digital ecosystem (TechCrunch, 2010). The public timeline is an extensive real-time information stream of currently more than 155 million messages per day (TechCrunch, 2011). Since it is impossible to follow this all-encompassing message board, users tend to subscribe to a selection of microblogs that interests them and only see messages posted by the authors they "follow". Unlike other social networking sites, such as Facebook[3] or MySpace[4], followership relationships do not require reciprocation. The original idea behind microblogging was to provide personal status updates. Even though updates related to their personal lives (72% of Twitter users) and work (62%) still represent the most frequent activities, the high share of references to news stories (55%) distinguishes Twitter from other social networks such as Facebook (Pew Research Center, 2010). Supporting this view, Kwak, Lee, Park, and Moon (2010) have come to the conclusion that Twitter is a news media rather than a social network. In line with this notion, postings cover every imaginable topic, ranging from political news to investment advice in a variety of formats, e.g., short sentences, links to websites, and "retweets" (i.e., quotations) of other users. According to Edison Research (2010), awareness of Twitter has exploded from 5% of Americans in 2008 to 87% in 2010 (compared to awareness of Facebook around 88%). The Pew Research Center (2010) reports that 8% of online Americans use Twitter including one

[1] Although there are a number of competing microblogging forums (e.g., Jaiku, Tumblr, frazr) this dissertation and the majority of existing research on microblogging focuses on Twitter because it has by far the widest acceptance and has, in fact, become synonymous with microblogging. While the terms Twitter and microblogging are used interchangeably throughout much of this dissertation, many of the findings have broader implications related to social media content in general (including user-generated information on social networks such as Facebook and MySpace).
[2] http://www.twitter.com
[3] http://www.facebook.com
[4] http://www.myspace.com

quarter doing so several times a day. As these figures show, Twitter has become a mainstream media channel, which is firmly established in the social media landscape. However, it has not been explored exhaustively in academic research.

The exponential growth of Twitter has started to draw the attention of researchers from various disciplines. There are three major streams of research. The first, largely rooted in computer and information science, focuses on understanding microblogging usage and community structures (e.g., Boyd, Golder, & Lotan, 2010; Dann, 2010; Java, Song, Finin, & Tseng, 2007). Areas of research include the exploration of user intentions[5] (e.g., Java et al., 2007) and usage patterns. With respect to the latter, Boyd et al. (2010) have illustrated retweeting practices and pointed out that users often retweet messages and thus relay valuable content in order to validate and endorse a particular user or posting. In an overview of content classification schemes, Dann (2010) suggests five major content categories of Twitter communication (conversational tweets, status updates, endorsements of content, news, and spam). Overall, this strand of literature primarily considers the online world and refrains from assessing its roots in the offline world.

A second stream of research concentrates on the exploration of best practices of microblogging as a communication tool. In the area of corporate applications, often referred to as Enterprise Microblogging (EMB), examples include the company-internal use for informal communication (e.g., Zhao & Rosson, 2009) or project management (e.g., Böhringer & Richter, 2009). Others have investigated the use of microblogging in education (Grosseck & Holotescu, 2008) and as a communication tool for conferences (Reinhardt, Ebner, Beham, & Costa, 2009).

Finally, a third stream of research investigates microblogging content as a source of public opinion and explores the relationship between tweets and related constructs in the real world. This intriguing area of research is rooted in the notion that social media content represents an aggregation of the information, opinions and beliefs of thousands of individuals. Leveraging this crowd wisdom (Surowiecki, 2004) and tapping into collective intelligence holds exciting promises. Social media content may represent one of the most direct measures of people's thoughts and feelings. The information extracted from social media may serve as an indicator of the current "state of the union" and may even be used as a leading indicator with predictive

[5] Java et al. (2007) distinguish three user categories depending on whether users primarily share information, seek information or maintain friendship relationships.

qualities. Examples include the analysis of Twitter messages as electronic word of mouth in the area of product marketing (e.g., Jansen, Zhang, Sobel, & Chowdury, 2009) or as a leading indicator for box office revenues (Asur & Huberman, 2010) and consumer confidence (O'Connor, Balasubramanyan, Routledge, & Smith, 2010).

This dissertation as a whole is most closely related to this third stream of research and contributes to the literature by exploring the relationship between microblogging content and real-world events. In addition, it adds to the first stream by offering an explanation for the effective aggregation of information in microblogging forums. Next to these overarching contributions, the individual essays make additional contributions to the fields of political science and capital market research, which are detailed in the respective essays.

The essays of this dissertation address two overarching research questions. The focus is on the comparison of the information extracted from social media with real-word events, i.e., the question *whether the insights extracted from microblogging forums can serve as an indicator of real-world events*.

This dissertation leverages methods from computational linguistics in order to extract information from microblogs.[6] In order to evaluate the information content, one needs to compare it to an objective benchmark. This dissertation focuses on the political (i.e., elections) and financial domains (i.e., the stock market) for two reasons: practical relevance and the availability of real-world benchmarks. First, assessing people's opinions is a particularly relevant task in both fields. Sentiment analysis (i.e., the detection of positive and negative emotions in online content; for an overview, see Pang & Lee, 2008) of textual information has become increasingly important to the political and the financial domains. With respect to the political domain, the New York Times has identified sentiment analysis as a major trend of the 2010 U.S. mid-term election (Brustein, 2010).[7] National TV channels such as CNN relied on Twitter content for election coverage and campaigns were supported by media consultants[8] dedicated to collecting and mining text-based online data to provide political campaigns with intelligence on the electorate. With respect to the financial domain, the first hedge fund based on a Twitter sentiment trading strategy was launched recently (Jordan, 2010). Furthermore, investment banks are dedicating a substantial amount of

[6] With respect to the analysis of textual data, one can broadly distinguish between classification techniques based on a pre-defined, externally validated dictionary (e.g., LIWC used in Essay 1) and statistical classifiers that use input-data, such as a manual coding, to generate a model (e.g., Naïve Bayesian classifier used in Essays 2 and 3).
[7] For an interactive example, see http://www.nytimes.com/interactive/us/politics/2010-twitter-candidates.html (last accessed, May 15, 2011).
[8] E.g., http://globalpointresearch.com/

research to the analysis of news sentiment as a signal for quantitative investors and financial data provider Thomson Reuters has developed the NewsScope Sentiment Engine and related Event Indices to power news-based trading algorithms.[9] Even though NewsScope data is being used in scientific research (e.g., Groß-Klußmann & Hautsch, 2009), illustrating the close relationship between academic and practical applications[10], the existing academic work on social media content as a source of public opinion is insufficient. Previous results (e.g., Asur & Huberman, 2010; O'Connor et al., 2010) are not necessarily transferrable to the political or financial domains. Next to practical relevance, the political and financial domains provide objective real-world benchmarks (e.g., election results and stock returns) for the accuracy of Twitter content.

Some scholars provide theoretical arguments that question the ability of blogs to aggregate dispersed bits of information (e.g., Sunstein, 2008). Thus, to ensure that the correlations between microblogging content and real-world events are not spurious, this dissertation offers both theoretical as well as empirical evidence supporting the idea that microblogging forums can function as information markets by answering a second overarching research question: *What mechanism can explain the efficient aggregation of information in microblogging forums?*

Microblogging forums have evolved rapidly through user innovation with many features being introduced by popular consensus and community behavior (Dann, 2010), including retweets (i.e., quotes of other users preceded by the acronym "RT") and hashtags (i.e., keywords preceded by "#" included in many messages to associate them with a relevant topic). These features and the followership relationships make previously unavailable aspects of information diffusion partially observable and allow us to systematically investigate whether they are used to effectively weigh and efficiently aggregate information.

2 Structure of this dissertation, key findings and contributions

Next to the brief introduction in the present chapter (I), this dissertation consists of five distinct essays (chapters II.1 to II.5), each of which represents a scholarly contribution of its own and addresses one or both of the overarching research questions outlined above. Given

[9] For more details on Thomas Reuters NewsScope, see http://online.thomsonreuters.com/newsscopereports/ (last accessed, May 15, 2011).

[10] In addition, Thomas Reuters has organized a "News Research Roundtable" bringing together many of the academics cited in Essays 2 and 3.

that all of the essays are self-contained academic contributions, each manuscript offers its own introduction, literature review, and methodology section. While the first two essays address both overarching research questions, Essays 3 and 4 focus on the correlation of microblogging content and real-word events and provide examples of the use of this innovative data source in social science research. Essay 5 presents an online application, which builds on many research results of the preceding studies.

Essay 1 addresses the overriding question whether microblogging content contains any valuable information at all. It develops theoretical arguments building on theories of crowd wisdom and collective intelligence that support the hypothesis that microblogging content can accurately reflect public opinion. Like financial markets, many social media networks, including Twitter, largely comply with the conditions for a crowd to be wise such as diversity, independence, and decentralization (Surowiecki, 2004). The information structure among social media users is not very different from that of traders in financial markets, which are widely accepted as information aggregation mechanisms (Hayek, 1945). Whereas financial markets aggregate information through the price system, messages in microblogging forums can be weighed by the followership of their authors and the rate of retweets. These represent the Twittersphere's “currency” and provide it with its own kind of a “price system”.

In the context of the 2009 German federal election, the exploratory study of over 100,000 messages containing a reference to either a political party or a politician shows that Twitter is used extensively for political deliberation and that related microblogs reflect the political preferences of the general population. Political microblogs may even serve to predict election results with a mean error of 1.65% that comes close to traditional election polls. The tweets' sentiment (e.g., positive and negative emotions associated with a politician or a party) corresponds closely to voters' political preferences.

This essay contributes to the literature by establishing microblogging forums as an information market turning them into a valuable source of political data and, more generally, a meaningful resource for social science research. However, election results provide only a limited number of real-world benchmarks and are not suitable for large sample statistical tests. While the study explores theoretical reasons for microblogging forums to reflect the combined opinions of the general population, it defines the thorough empirical investigation of the mechanism by which information is weighted and distributed in microblogging forums as one of the most critical aspects of further research.

Essay 2 addresses this research gap in the context of stock microblogging. It explores, first, whether and to what extent the information content of stock-related microblogs reflects financial market developments and, second, whether microblogging forums provide an efficient mechanism to weigh and aggregate information. Thus, the study is not limited to the correlation of online message content with financial market indicators, but offers an explanation for the efficient aggregation of information in stock microblogging forums. It leverages the nature of these forums to explore empirically theories of social influence concerning the diffusion and processing of information in the context of a financial community.

The essay shows that the sentiment (i.e., bullishness) of tweets is associated with abnormal stock returns with a number of strategies earning abnormal returns of more than 15% in the 6 month sample period. In addition, message volume can predict next-day trading volume. With respect to the mechanism leading to the efficient aggregation of information in microblogging forums, the 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.

The essay offers two primary contributions to the existing literature. For one, it is the first to comprehensively explore the information content of microblogs relative to the returns of individual stocks. The results permit researchers and financial professionals to reliably identify microblogging content, which may serve as a valuable proxy for investor behavior and belief formation (e.g., otherwise hard-to-measure constructs such as investor sentiment). Second, the study provides an explanation for the efficient aggregation of information in stock microblogging forums. While *Essay 2* explores the information content of stock microblogs in terms of sentiment (i.e., bullishness), the definition of information can be expanded to include other dimensions such as the topic or type of news that is discussed.

Essay 3 pursues this research gap by controlling for different types of company-specific news events. It investigates the market impact of different types of company-specific news events (e.g., news related to corporate governance, operations, and legal issues) on S&P 500 stock prices in order to discern genuine news that moves the market from insignificant noise without market reaction. Distinguishing between good and bad news, it controls for the sentiment (i.e., the positive vs. negative tone) of different news stories.

The results show that the information published in a stock microblogging forum can be used to detect, which types of stock-related news affect a company on a particular day. The absolute value of cumulative returns prior to a news event are more pronounced for positive news than negative news, suggesting more widespread information leakage before good news. The results show that the market reaction differs substantially across multiple types of news events, supporting the notion that there are certain event types to which investors attribute greater importance (e.g., news related to M&A or earnings announcements) and others, which rarely contain new information that moves the market (e.g., issues related to joint ventures). In addition, a cross-industry comparison indicates that industry classification may partially explain the market reaction to the same event type.

The main contributions of this study are as follows. First, the study establishes online stock forums as an alternative to traditional media sources to identify company-specific news events. Second, it provides the first comprehensive comparison of the market impact of various types of firm-specific news from an investor perspective. Third, among a handful of event studies that cover multiple event types, it offers the first systematic distinction between good news and bad news and thus illustrates that controlling for news sentiment is important in the context of an event study. Finally, the study is the first to assess the market impact of various event types across different industry groups.

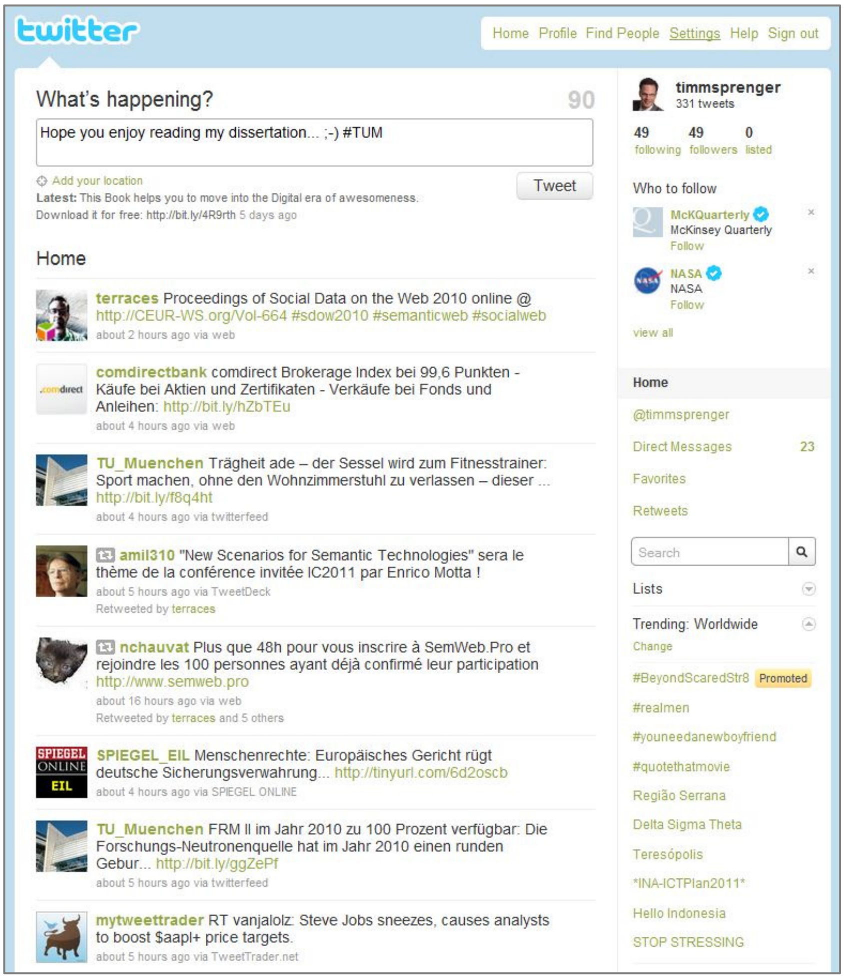
Given the finding of *Essay 3* that the industry classification may partially explain the market reaction to company-specific news and the fact that recent studies have called into question the accuracy of popular methods for industry classification (e.g., Bhojraj, Lee, & Oler, 2003), *Essay 4* proposes an alternative approach to defining industry groups based on investor perceptions of the relatedness of stocks. The study explores whether the degree to which pairs of companies are associated with each other in stock microblogs is related to the comovement of their stocks and investigates whether these relationships can be used to define homogenous subgroups of companies.

The results show that that the degree to which companies are mentioned jointly in an internet stock forum can explain the comovement of their stocks. The proposed measure of relatedness can help identify a firm’s strategic peers and delineate industry groups which explain stock returns as well as established classification methods, but offers a number of promising advantages (e.g., availability, timeliness).

Essay 5 illustrates how the insights from the research presented above were embedded into a fully functional Web 2.0 online application. It describes the main features of the implementation of TweetTrader.net[11], a stock microblogging forum that leverages crowd wisdom to aggregate the information contained in stock-related tweets. The application integrates inputs from text classification, user votings and a proprietary stock game in order to extract the sentiment (i.e., the bullishness) of online investors with respect to all publicly traded companies of the S&P 500. TweetTrader.net is designed to serve as an information aggregator for stock-related social media content and help investors see through the data and extract meaningful insights and opinions from millions of messages.

Following the five essays, a final conclusion (chapter III) summarizes results, especially those related to the overarching research questions, and discusses overall implications.

Figure 1: Screenshot of the Twitter website (example of a user's homescreen)



[11] http://TweetTrader.net

II.1 Essay 1

Predicting Elections with Twitter:
How 140 Characters Reflect the Political Landscape

Abstract

This study investigates whether microblogging messages on Twitter validly mirror the political landscape offline and can be used to predict election results. In the context of the 2009 German federal election, we conducted a sentiment analysis of over 100,000 messages containing a reference to either a political party or a politician. Our results show that Twitter is used extensively for political deliberation and that the mere number of party mentions accurately reflects the election result. The tweets' sentiment (e.g., positive and negative emotions associated with a politician) corresponds closely to voters' political preferences. In addition, party sentiment profiles reflect the similarity of political positions between parties. We discuss the use of microblogging services to aggregate dispersed information and derive suggestions for further research.

JEL Classification: C80; J11

Keywords: Twitter; microblogging; information market; prediction markets, election forecasts, politics; elections; sentiment analysis

Current status: Accepted for presentation at the ICWSM 2010 (May 2010) and published as a full paper in the *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media* (pp. 178-185), Washington, DC.
Accepted for publication at the *Social Science Computer Review*.

Acknowledgements: This paper contains elements of joint work with Prof. Dr. Isabell M. Welpe, Dr. Philipp Sandner, and Dipl.-Psych. Andranik Tumasjan.

1 Introduction

> *"Campaigns and the news media are becoming convinced that the internet can be mined systematically for useful data about public opinion. Sentiment analysis will soon be a part of every campaign […] because it helps determine quickly which messages are resonating with potential voters."*
>
> Joshua Brustein (2010)

The successful use of social media in the U.S. presidential campaign of Barack Obama has established Twitter, Facebook, MySpace, and other social media as integral parts of the political campaign toolbox. Some analysts attribute Obama's victory to a large extent to his online strategy. Obama's social-networking website helped him set records in terms of donations and grassroot mobilization (Williams & Gulati, 2008). Shortly after his victory, Obama used Twitter to let the web community know how he felt: "This is history." As this example demonstrates, after the rise of candidate websites in 1996, e-mail in 1998 (the Jesse Ventura campaign), online fund-raising in 2000 (the John McCain campaign), and blogs in 2004 (the Howard Dean campaign; Gueorguieva, 2007), the microblogging platform Twitter has become a legitimate and frequently used communication channel in the political arena as a result of the 2008 campaign.[12] While some political analysts are already turning to the "Twittersphere" as an indicator of political opinion, others have suggested that the majority of the messages are "pointless babble" (Pearanalytics, 2009). Therefore, the purpose of our study is to answer the question whether microblogging messages can actually inform us about public opinion and the political landscape in the offline world.

In particular, our study explores four aspects of this research question in the context of the 2009 federal election of the national parliament in Germany. First, we examine whether Twitter is used as a vehicle for political deliberation[13] by looking at how people use microblogging to exchange information about political issues. Second, we evaluate whether Twitter messages reflect the political preferences and the political landscape offline in a meaningful way. Third, we investigate whether individual accounts show evidence of a party bias. Finally, we analyze whether the content of Twitter messages can be used to forecast the election result.

---

[12] The political discourse on Twitter has lead to the establishment of numerous dedicated websites, both in the United States (e.g., http://tweetcongress.org, http://www.congressional140.com, and http://govtwit.com) and Germany (e.g., http://parteigefluester.de, http://wahlgetwitter.de).

[13] In line with Delli Carpini, Cook, and Jacobs (2007), we use the words deliberation, debate, and discussion interchangeably.

2 Background

2.1 The German election

In our study, we examine more than 100,000 tweets published in the weeks leading up to the federal election in Germany, which took place on September 27, 2009. After 4 years in a grand coalition with the social democrats (SPD), Chancellor Angela Merkel – member of the conservatives (CDU) – was running for reelection, but favoring a coalition with the liberals (FDP). Many commentators have called the parties' campaigns uninspiring due to the unwillingness of the main candidates to attack their then-coalition partners. The left side of the political spectrum was fragmented by the rise of the socialist party (Die Linke). The SPD publicly rejected Die Linke as a possible coalition partner, thus limiting its options to build a governing coalition. The potential coalition of CDU and FDP was leading by a slight majority in most polls and was ultimately able to form a center-right government after the election.

2.2 Related work and research questions

Recently, the exponential growth of Twitter has started to draw the attention of researchers from various disciplines. There are several streams of research investigating the role of Twitter in social media, product marketing, and project management. One stream of research concentrates on understanding microblogging usage and community structures (e.g., Honeycutt & Herring, 2009). In sum, this research demonstrates that the intensity of Twitter usage varies considerably. Market researchers have reported that in June 2009 (only a couple of weeks before the German federal election) 71% of all 1.8 million German users had visited Twitter only once and 15% of them at least 3 times (Nielsen Media Research, 2009). Honeycutt and Herring (2009) showed that Twitter is used not only for one-way communication but often serves as a means of conversation. In their study exploring conversation via Twitter, they find that 31% of a random sample of tweets contain an "@"-sign and that the vast majority (91%) of those were used to direct a tweet to a specific addressee. While these findings have provided us with a general understanding of why and how people use microblogging services, they have not explored the use of this new communication device in specific contexts such as, for instance, corporate public relations or the political debate online. This strand of literature only considers the online world and refrains from assessing its roots in the offline world. Another stream of research focuses on corporate applications of microblogging such as the company-internal use for project management or the analysis of Twitter as electronic word of mouth in the area of product marketing (e.g., Jansen, Zhang, Sobel, & Chowdury, 2009). In their study, Jansen et al. (2009) have found that 19% of a random sample of tweets contained mentions of a brand or product and that an automated classification was able to extract statistically significant differences of customer sentiment (i.e., the attitude of a writer toward a brand). While this study provides reason to believe that sentiment may also be embedded in tweets covering other topics besides branding, Twitter sentiment analysis has not yet been applied to research regarding the political debate online. While several scholars have debated the potential of weblogs as a forum for democratic debate, "empirical research on deliberative democracy has lagged significantly behind theory" (Delli Carpini, Cook, & Jacobs, 2007, p. 316). A few researchers have empirically examined internet discussion boards as a vehicle for political deliberation (e.g., Jansen & Koop, 2005). Koop and Jansen (2009) have defined the exchange of substantive issues as an indicator of deliberation and the equality of participation as a measure of the deliberative quality of blog-based discussion. While they have found discussion boards and blogs to be dominated by a relatively small number of users, it is unclear whether these findings also apply to the political debate on Twitter. Recent scholarly work on political blogs has focused on their effect on real-world politics, such as complementing the watchdog function of the mainstream media and mobilizing supporters, but largely ignored the reflection of offline politics in the digitally enhanced public sphere. However, there are a few studies exploring the reflection of the political landscape in "traditional" weblogs and social media sites. For instance, Williams and Gulati (2008) have found that the number of Facebook supporters can be considered a valid indicator of electoral success. Sunstein (2008) is more pessimistic and questions the ability of blogs to aggregate dispersed bits of information. Next to social media, in which the reflection of the political landscape may be only a by-product, prediction markets deserve mention as a special electronic platform that is designed to aggregate information on political elections. Prediction markets are similar to financial markets and allow trading in virtual securities tied to the outcome of a particular event (e.g., a candidate winning the election). Market prices can be interpreted as predictions (e.g., the share of the vote). The most cited and best known example of a prediction market is the Iowa Electronic Market (IEM), which was established in 1988 by the University of Iowa. It runs markets in federal and state elections. When compared directly to the corresponding large-scale polls, IEM prices were more accurate 76% of the time (Berg & Rietz, 2006).

However, individual users of both social media and prediction markets show a significant bias in their party orientation. Adamic and Glance (2005) provide evidence of the fragmentation or polarization of the political blogosphere. They found that linkage patterns among bloggers reflect the blogosphere along party lines with 91% of all links directed to like-minded websites. Studies of the IEM found traders to be biased by their party preference, which was reflected in both their trading activity and their portfolio holdings (Forsythe, Rietz, & Ross, 1999). Despite previous research providing evidence that "traditional" social media content can be used to validly predict political outcomes despite individual party biases, we know little about the predictive power of Twitter for political debates and outcomes. Previous scholarly examinations of social media may not be easily transferable to Twitter for the following reasons: First, tweets are much shorter and contain much less content than, for instance, news articles and traditional blogs. Hence, their informational value is less clear-cut. One marketing consultancy has even suggested that up to 40% of all Twitter messages are "pointless babble" (Pearanalytics, 2009). Second, only part of the information conveyed is found in the words themselves because 19% of all messages contain links to other websites (Zarrella, 2009). Thus, a basic question is whether 140-character messages can contain differentiated information regarding the electorate's political preferences. Preliminary results from two recent reports suggest that microblogging content may be a good predictor of election results. A conference paper analyzing the correlation between candidate mentions on Twitter and the results of the Japanese national election reports that in more than 80% of all constituencies the most mentioned candidate won the election (Suenami & Yutaka, 2010). A similar survey of candidate mentions on Twitter during the 2010 U.K. election, presented by a website that aggregates political tweets, finds the predictions of the national share of vote to be better than most opinion polls with an average error of only 1.75 percentage points (Tweetminster, 2010). While these studies indicate that political microblogs may hold intriguing information to describe the political landscape, both are largely limited to the evaluation of the frequency of candidate mentions. We extend these findings by examining not just the number of mentions but also the information content of the actual messages through linguistic sentiment analysis. We examine the nature of the political debate on Twitter and explore party biases of individual users. Next to the election results, we also investigate the relationship of sentiment profiles to election programs and likely coalitions.

Although the reference to tweets in many political commentaries shows that even analysts are already using Twitter as an indicator of political opinion, to the best of our knowledge, there are no scientific studies systematically investigating the sentiment in political microblogs. Therefore, the present study aims at addressing this general question in the following four ways: First, we examine whether Twitter provides a platform for political deliberation online. Second, we evaluate how accurately Twitter can inform us about the electorate's political preferences and the political landscape offline. Third, we investigate whether individual accounts show evidence of party preference. Fourth, we explore whether Twitter can serve as a predictor of the election result with respect to both the share of vote ands likely coalitions.

2.3 Microblogging forums as information markets

According to the American Association of Public Opinion Researchers (AAPOR), about $2 billion were spent on online research in 2009, 85% of which replaces traditional survey methods. However, the AAPOR concludes that "researchers should avoid nonprobability online panels when one of the research objectives is to accurately estimate population values [and] claims of representativeness should be avoided when using these sample sources" (p. 5). The AAPOR focuses on online surveys and largely ignores user-generated content. We agree that many online samples, including the data used in this study, are not representative[14] and that representative results can only come from a survey of a representative sample. However, user-generated content is not the result of a survey process, but the collection of a broad information exchange, in which users not only post their own opinions, but reflect on and discuss the comments of others and external sources (e.g., the press). We argue that microblogging forums allow users to weigh information and can thus produce accurate predictions, even if the users are not representative of the general population. Similar to information produced by other "unrepresentative" sources of "accurate" opinion, such as think tanks or financial markets, results can be accurate if information is aggregated efficiently. Our study investigates whether we can turn to microblogging content as an acceptable alternative to traditional surveys.

[14] As far as our data source is concerned, an online survey of 1,707 German Twitter users conducted in November 2009, only 2 months after the German election, has shown that these users are predominantly male (64%), young (31 years), and have a university degree (67%; Web Evangelisten, 2009). On the other hand, over the last decades the German population eligible to vote has aged continuously and voter turnout has been higher for older age group. Even though we do not have demographic data for the specific users in our sample, these facts indicate that the demographics of political microbloggers and actual voters are probably not skewed in the same, but - to the contrary - in opposite directions relative to the overall population.

So what warrants an investigation of microblogging forums as a viable source of public opinion next to the growth of online opinion mining and advantages relative to traditional survey methods? The focus of our study is on the comparison of the information extracted from social media with accepted benchmarks such as election results. However, next to this empirical focus, there are also theoretical arguments that support the hypothesis that social media content can produce accurate predictions. We can approach this phenomenon with theories on crowd wisdom and collective intelligence. In a summary of related studies, Surowiecki (2004) has suggested a number of conditions for the crowd to be wise, that is a large group of people to come to an accurate judgment: diversity (e.g., Hong & Page, 2001), independence (e.g., Bikhchandani, Hirshleifer, & Welch, 1998), and decentralization combined with a mechanism to aggregate dispersed bits of information (e.g., Hayek, 1945). Like financial markets, many social media networks, including Twitter, largely comply with these properties. The information structure among social media users is not very different from financial markets, which are widely accepted as information aggregation mechanisms. Whereas financial markets aggregate information through the price system, microblogging forums lack a natural aggregation mechanism. Sunstein (2008, p. 88) suggests that "the blogosphere cannot operate in Hayekian fashion, because it lacks […] the price system (or any reasonable analogue to it)." He argues that "participants in the blogosphere [are not encouraged to produce reliable and unbiased information because they] lack an economic incentive" (Sunstein, 2008, p. 90). However, a study of prediction markets has shown that both play- and real-money markets predicted outcomes equally well (Servan-Schreiber, Wolfers, Pennock, & Galebach, 2004) which indicates that nonmonetary incentives such as the position in a publicized ranking may encourage earnest participation. Most microblogging forums have their own implicit rankings, which may have similar effects: So even if Twitter does not have an explicit mechanism for aggregating information, the size of the followership and the rate of retweets may represent the Twittersphere's "currency" and provide it with its own kind of a "price system." Studies have shown 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 study of a Twitter account making directional forecasts of the stock market has shown the number of followers to be correlated with the accuracy of the published information (i.e., the forecasts of the stock market; Giller, 2009). The fact that users not only notice these subtleties in the dense information stream they are exposed to but also act on them by maintaining or terminating their subscription is only one powerful example of how the quality and accuracy of content determines the number of followers. In addition, new and valuable pieces of information are retweeted more often, providing Twitter with a mechanism to weigh the importance of information. Even if this mechanism does not provide us with a signal as clear as a market price, the display of word clouds or the ranking of trending topics illustrate successful attempts to improve the information aggregation. These mechanisms can certainly inform us about what people find important because, even with hundreds of thousands of tweets being sent every day, time and again, newsworthy messages from private individuals with no more than a few dozen followers (such as the one making the first report of the plane crash in the Hudson River in 2009) bubble to the surface and get spread to a wider audience.

These considerations encourage us to believe that the information stream on Twitter can be aggregated in a meaningful fashion in order to make accurate, albeit not necessarily representative, predictions and that we can leverage Twitter as an information market.

3 Data set and methodology

We examined 104,003 political tweets, which were published on Twitter's public message board between August 13 and September 19, 2009, prior to the German national election. We systematically collected all tweets that contained the names of either the six parties represented in the German parliament (CDU/CSU, SPD, FDP, Die Grünen, and Die Linke) or those politicians of these parties who are regularly included in a weekly survey on the popularity of politicians conducted by the research institute "Forschungsgruppe Wahlen". CDU and CSU, often referred to as the "Union", are sister parties that form one faction in the German parliament. Our query resulted in roughly 70,000 tweets mentioning one of the six major parties and 35.000 tweets referring to their politicians.

Given the samples size, we decided to use sentiment analysis, an automated mechanism to quantify the information contained in these messages. In the domain of natural language processing, the term sentiment analysis is "used in reference to the automatic analysis of evaluative text and tracking of the predictive judgments therein" (Pang & Lee, 2008, p. 10). This analysis includes the extraction of the polarity (either positivity or negativity) but, more generally, refers to the computational extraction of information from a given text sample. To extract the sentiment of the tweets objectively and systematically, we used LIWC2007 (Linguistic Inquiry and Word Count; Tausczik & Pennebaker, 2010), a text analysis software developed to assess emotional, cognitive, and structural components of text samples using a psychometrically validated dictionary. This software calculates the degree to which a text sample contains words belonging to empirically defined psychological and structural categories. Specifically, it determines the rate at which certain cognitions and emotions (e.g., future orientation, positive or negative emotions) are present in the text. For each psychological dimension, the software calculates the relative frequency with which words related to that dimension occur in a given text sample (e.g., the words “maybe”, “perhaps”, or “guess” are counted as representatives of the construct “tentativeness”). LIWC has been used widely in psychology and linguistics but also for topics related to political science. Examples include studies of the sentiment levels in U.S. Senatorial speeches (Yu, Kaufmann, & Diermeier, 2008), the linguistic differences between positive and negative political ads and television interviews of presidential candidates. LIWC-based analyses have also been used to examine shorter text samples such as instant message conversations, which are similar in length to tweets, and the Twitter accounts of gubernatorial candidates in various U.S. state elections (for a comprehensive overview of related studies, see Tausczik & Pennebaker, 2010).

We use the following 12 LIWC dimensions in order to profile sentiment in political tweets: Future orientation, past orientation, positive emotions, negative emotions, sadness, anxiety, anger, tentativeness, certainty, work, achievement, and money. These categories have either been successfully used in prior studies of political text samples or seemed best suited to profile messages in the political domain by covering both emotions (e.g., anger, anxiety) and content dimensions (e.g., work, money). Following the methodology used by Yu et al. (2008), we concatenated all tweets published over the relevant time frame into one text sample to be evaluated by LIWC. Our sample was restricted to German language tweets, which were translated into English and then processed by the LIWC English dictionary.

It is important to distinguish online sentiment, as described above, and offline sentiment. The political sentiment offline may become apparent in the form of party preference (i.e., election results) or positions on individual policy issues or people (i.e., left- vs. right-wing positions or positive/negative emotions with respect to a candidate). To avoid confusion with online sentiment, we refer to political sentiment offline as political preferences.

4 Results

4.1 Twitter as a platform for political deliberation

In this section, we will evaluate our sample along two widely accepted indicators of blog-based deliberation, the exchange of substantive issues and the equality of participation (Koop & Jansen, 2009).

Table 1 shows the number of mentions and a randomly selected tweet for all parties in our sample. These messages are only supposed to provide a glance at the underlying data. While this is only a small selection of the information stream in our sample, these messages illustrate that tweets can contain a lot of relevant information. So, despite their brevity, substantive issues can be expressed in 140 characters or less.

Table 1: Sample tweets by party

Party	Number of tweets	Examples
CDU	30,886	CDU wants strict rules for internet
CSU	5,748	CSU continues attacks on partner of choice FPD
SPD	27,356	Only a matter of time until the SPD dissolves
FDP	17,737	Whoever wants civil rights must choose FDP!
Die Linke	12,689	Society for Humans Rights recommends: No government participation for Die Linke
Grüne	8,250	After the crisis only Green can help HTTP:[…] Grüne+

Notes: Examples were randomly selected from the tweets mentioning each party. Messages were shortened for citation (e.g., omission of hyperlinks).

Next, we analyze the level of addressivity in the messages as an indication of the exchange of ideas on Twitter. About one third of all tweets in our sample (30.8%) contain an "@" sign which is in line with previous research that has also suggested that the vast majority of "@"signs are used to direct a tweet to a specific addressee (Honeycutt & Herring, 2009). A more conservative measure of direct communication are direct messages to another user starting with an "@" sign. Roughly 10% of the messages in our sample are direct messages, indicating that people are not just using Twitter to post their opinions but also engage in interactive discussions.[15] Many users on Twitter forward messages to their followership. These so-called retweets often contain information that the sender finds noteworthy such as links to other websites. Consequently, the rate at which messages are retweeted indicates whether information is considered being interesting. According to Zarrella (2009), only 1.44% of all tweets are retweets. In our sample, however, that share of both retweets and

[15] See the appendix for a specific example of political deliberation on Twitter.

messages containing a hyperlink are significantly higher: 19.1% of all messages were retweets and 54.2% contain a link to a website. Summarizing, our results indicate that people are finding interesting political information on Twitter, which they share with their network of followers.

We now turn to the analysis of the equality of participation. While we find evidence of a lively political debate on Twitter, it is unclear whether this deliberation is led by a few "political junkies" rather than the wider general public. Jansen and Koop (2005) found less than 3% of all users on the political message board BC Votes to be responsible for almost a third of all posted messages. Table 2 shows the share of users and the share of messages across various user groups for our sample according to the frequency with which a user posts messages. We adopted the categorization from Jansen and Koop (2005). While the distribution of users across user groups is almost identical with the one found by Jansen and Koop (2005), we find even less equality of participation for the political debate on Twitter. There is a high concentration of messages in the groups of heavy (23.1%) and very heavy users (21.2%). These make up 3.3% and 0.6% of the users, respectively. So roughly 4% of all users accounted for more than 40% of the messages. In sum, Twitter is used as a forum for political deliberation, and this forum is dominated by a small number of heavy users.

4.2 Twitter sentiment as a reflection of the political landscape offline

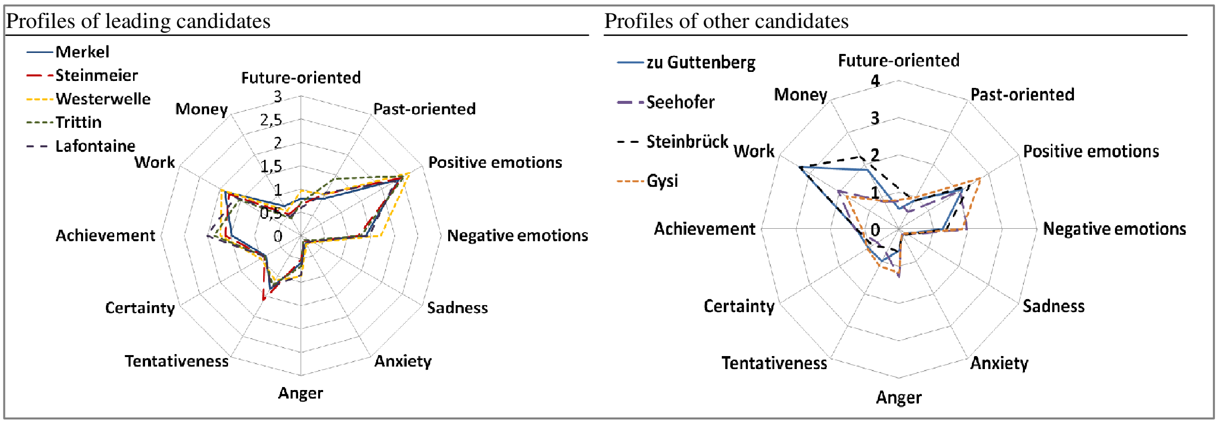
The fact that users are discussing political issues online does not mean that we can necessarily extract meaningful information from this debate. To explore this question, we aggregated the information stream about politicians and parties and compared the resulting profiles with evidence from the press and election programs. We investigate the reflection of two aspects of the political landscape in the messages: First, whether the messages reflect the voters' political preferences (e.g., positive or negative emotions with respect to the main candidates based on press reports and anecdotal evidence from the campaign) and, second, whether the tweets mirror the ideological proximity of political parties (e.g., similarity of the parties' political agendas based on their election programs).

In order to analyze the sentiment of the political tweets, we generated multidimensional profiles of the politicians in our sample using the relative frequencies of LIWC category word counts (i.e., the percentage of words in all tweets about a particular candidate, which are related to the 12 chosen dimensions according to the LIWC dictionary). It is important to note that these sentiment profiles are not policy or ideological measures. They merely represent an aggregated linguistic profile of the messages associated with a particular party or politicians. The left panel of Figure 1 shows these profiles for the leading candidates of the 5 main parties. Overall, positive emotions clearly outweigh negative emotions. This is in line with Yu et al. (2008) who find that positive emotions outweigh negative emotions by more than 2 to 1 in an LIWC-based analysis of 18 years of congressional debates. Only liberal party leader Guido Westerwelle and socialist party leader Oskar Lafontaine show more distinctive deviations from this profile on some dimensions. The dimension of perceived anger, for example, is most prominent in the case of these two politicians who, as free-market advocate and socialist leader, represent two contrasting political programs in the political spectrum. Messages regarding Frank-Walter Steinmeier, who at the time of our recording was sending mixed signals regarding potential coalition partners for his party after the election, reflect more tentativeness than those of other politicians. The higher share of tentative messages also corresponds to findings indicating that tentativeness correlates to lower status and rank (Tausczik & Pennebaker, 2010). Compared to acting chancellor Merkel, this profile is in line with Steinmeier's role as vice chancellor and clear runner-up in the polls.

The right panel of Figure 1 shows the profiles of other prominent politicians: Their profiles show some distinct differences from those of the leading candidates. Again, positive outweigh negative emotions – with the exception of Seehofer (CSU) who in addition is most frequently associated with anger. This might reflect the fact that Seehofer irritated many voters and party members by attacking the coalition partner desired by sister party CDU for much of the election campaign. Especially for Steinbrück (SPD) and zu Guttenberg (CSU), the issues money and work are probably reflecting their roles as finance and economics ministers. As can be seen in Figure 1, while small in absolute terms, the sentiment embedded in tweets does reflect nuanced differences between the politicians in our sample. To conclude, one can say that Twitter messages can be considered a plausible reflection of voters' sentiment.[16] Next we explore whether the tweets mirror the ideological proximity of political parties. Since it is not easy to visually analyze the profiles using the radar charts, we computed a measure of similarity for the LIWC profiles of all combinations of two parties (Table 3).

[16] An analysis of tweets surrounding the TV debate between the two candidates for chancellor showed that political tweets can even track changes in sentiment on a daily basis (see appendix for details).

Figure 1: Profiles of politicians



Notes: The profiles show the percentage of words in all tweets about a particular candidate, which are related to the indicated dimension according to the LIWC dictionary. The radar charts connect these percentages for every candidate across all 12 dimensions.

Hinich and Munger (1997) suggest the use of Euclidian distances along multiple policy dimensions to measure the political distance between parties. The distance of party k and party l can be calculated as

$$(1) \qquad d_{kl} = \frac{1}{S}\sqrt{\sum (x_{kj} - x_{lj})^2} \quad ,$$

where x_{kj} represents the position of party k on the policy dimension j and S is the number of parties that were included in the calculation. In our case, we use LIWC sentiment dimensions instead of policy dimensions. The measure d_{LIWC} represents the root of the sum of all squared differences between the 12 LIWC dimensions across a sample of parties. The higher the value of d_{LIWC}, the higher the heterogeneity of the LIWC profiles of the included parties. In other words, the lower the value of d_{LIWC}, the more similar the LIWC profiles. In order to evaluate whether the LIWC profiles reflect the political landscape offline, we compare them to an objective measure of political similarity. As indicated, the distance measure has been used widely in political science, even in the context of German politics (see, for example, Bräuninger & Debus, 2009, who refer to it as ideological heterogeneity). To provide an objective benchmark for the similarity of linguistic profiles we have constructed the same distance measure based on data from the Comparative Manifesto Project (CMP). The CMP conducts quantitative content analyses of election programs of parties from more than 50 countries covering all free democratic elections since 1945 (Klingemann, Volkens, Bara, Budge, & McDonald, 2006). The purpose is to measure political preferences of parties. The CMP database provides frequency tables that indicate how many sentences in its election program a party dedicates to each of 56 categories. Examples of these categories include human rights, free enterprise, economic goals, welfare state expansion/limitation, education expansion/limitation, and traditional morality. CMP provides data for the election programs of the Union faction (CDU/CSU), SPD, FDP, Die Grünen and Die Linke for the German federal election 2009. Thus, we were able to calculate the ideological heterogeneity of the parties in our sample according to the CMP (d_{CMP}) using the measure of heterogeneity outlined above. For simplicity, we have limited this analysis to all combinations of two parties for which the CMP provides data.

As can be seen in Table 3, the heterogeneity of LIWC profiles of the major parties corresponds closely to their political proximity. Most notably, the distance measure confirms the tight fit between the Union faction of sister parties CDU and CSU ($d_{LIWC} = 0.34$). Both measures indicate a fair degree of correspondence between the CDU and the other main parties, except for the socialist party Die Linke ($d_{LIWC} = 1.48$, $d_{CMP} = 1.38$). The LIWC profile of the Green party is most similar to their former and desired coalition partner SPD. Both measures show Die Linke to be farthest from the center-right parties CDU/CSU and FDP and closer to the left-of-center parties SPD and Green party. Overall, the similarity of LIWC profiles is a plausible reflection of the political proximity of the parties' election programs in the weeks before the federal election.

Table 2: Equality of participation and format of communications

User group	Users		Messages		Format of communication			
	Total	Share	Total	Share	Mention	DM	RT	URL
One-time (1)	7,064	50.3%	7,064	10.2%	29.2%	11.7%	16.7%	45.0%
Light (2-5)	4,625	32.9%	13,353	19.3%	29.4%	9.7%	18.6%	48.9%
Medium (6-20)	1,820	12.9%	18,191	26.2%	30.7%	10.2%	19.0%	49.4%
Heavy (21-79)	463	3.3%	15,990	23.1%	32.7%	10.8%	20.1%	55.7%
Very heavy (80+)	84	0.6%	14,710	21.2%	31.1%	10.2%	19.9%	67.6%
Total	**14,056**	**100.0%**	**69,318**	**100.0%**	**30.8%**	**10.4%**	**19.1%**	**54.2%**

Notes: Mentions are message containing "@" (includes direct messages), direct messages (DM) start with "@", retweets contain "RT", "via", or "by", URL are messages containing hyperlink (e.g., "http:").

Table 3: Heterogeneity of linguistic and policy profiles

Linguistic heterogeneity (LIWC) d_{LIWC}					**Ideological heterogeneity (CMP)** d_{CMP}				
	CDU/CSU	**SPD**	**FDP**	**Grüne**		**CDU/CSU**	**SPD**	**FDP**	**Grüne**
CDU/CSU	0.34*				**CDU/CSU**	-			
SPD	0.86	-			**SPD**	0.88	-		
FDP	0.90	0.62	-		**FDP**	0.81	1.12	-	
Grüne	0.87	0.65	0.83	-	**Grüne**	1.03	0.68	1.07	-
Die Linke	1.48	1.35	1.46	1.00	**Die Linke**	1.38	0.73	1.54	0.76

Notes: Due to different base rates, we made the two distance measures (d_{LIWC} and d_{CMP}) more easily comparable by dividing the distances in each table by the average of all values in that table.

** Heterogeneity of Union parties CDU and CSU, the Comparative Manifestos Project (CMP) only provides data for the Union faction as a whole.*

4.3 Party bias of individual users

In this section, we will examine whether individual users show evidence of a party bias in the volume or the sentiment of their postings. Studies of prediction markets have linked biases in trading activity and portfolio holdings to traders' party preference (Forsythe et al., 1999). Many politicians and regional party branches maintain Twitter accounts, which contain the party acronym in their account name (e.g., "SPDBerlin"). By focusing on the subset of accounts, which contain the name of the two largest parties (CDU and SPD), we were able to identify a group of 82 SPD- and 57 CDU-labeled accounts that are likely associated with one of the two parties. We filtered the messages posted through these accounts by the parties or candidates mentioned (Table 4). As can be seen, the party accounts dedicate roughly 80% of their mentions to their own party. Also, they make about three times as many references to their party's candidate compared to the opponent. With respect to the sentiment contained in these messages, we find that negative emotions are clearly correlated with party affiliation. Positive emotions, on the other hand, do not follow the same pattern. These results are consistent with a large body of literature that shows negative information to be processed more thoroughly and to have more impact than positive information ("negativity bias"; e.g., Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001). In summary, we can conclude that individual users show a party bias with respect to the volume[17] as well as the negative sentiment of the messages they post. The analysis of party-affiliated accounts suggests that this bias is linked to party preference.

Table 4: Share of tweets and sentiment of party-affiliated accounts

Account		Content (message contains reference to)			
		CDU	SPD	Merkel (CDU)	Steinmeier (SPD)
CDU	Share of messages	83.9%	12.9%	5.9%	1.6%
	Positive emotions	1.61	1.42	1.35	1.82
	Negative emotions	0.68	0.87	0.79	1.09
SPD	Share of messages	25.5%	81.9%	6.2%	15.7%
	Positive emotions	2.01	1.79	2.79	2.61
	Negative emotions	1.4	0.53	1.06	0.61

Notes: The content split addresses all messages containing a particular search term. Since some messages contain multiple party or candidate mentions the sum can be greater than 100%.

[17] Additional analyses confirm this volume effect for the whole dataset. The distribution of user attention (i.e., the share of mentions that a user dedicates to the various parties) shows a significant bias. Users put a clear emphasis on the discussion of one particular party (see appendix for details).

4.4 Twitter as a predictor of the election result

In order to understand whether the activity on Twitter can serve as a predictor of the election outcome, we examine two aspects. First, we compare the share of attention the political parties receive on Twitter with the result of the 2009 German federal election. Second, we analyze whether tweets can inform us about the ideological ties between parties and potential political coalitions after the election. It is important to distinguish this analysis based on party mentions from the sentiment profiles in the previous section. The number of mentions is not a sentiment measure but merely measures the overall attention a party garners, that is, the "buzz" it generates.

Table 5 shows the number of tweets mentioning a particular party. As can be seen, the ranking by tweet volume (i.e., the number of tweets) and the ranking by share of vote in the election results are identical. In fact, the relative volume of tweets mirrors the results of the federal election closely. If we consider the number of tweets to be a predictor of the election result, the mean absolute error (MAE) of this prediction is 1.65%. The MAE is a measure of forecast accuracy and has been widely applied to compare the accuracy of political information markets relative to election polls (see Berg et al., 2008).

To understand how the above-mentioned prediction based on message volume compares with traditional methods to collect this data, we compare Twitter with a number of election polls and the IEM prediction market. The MAE of six research institutes, which published election polls in our sample period, ranges from 1.1% to 1.7%.[18] Thus, Twitter comes close to these accepted benchmarks. The predictive accuracy is even more impressive when compared to the track record of the IEM, a prediction market set up with the explicit purpose to predict election results. The IEM produced a MAE of 1.37% in U.S. presidential elections and 2.12% in non-U.S. elections based on election eve market prices (Berg, Forsythe, Nelson, & Rietz, 2008). In conclusion, the mere number of tweets mentioning a political party can be considered a plausible reflection of the vote share and its predictive power even comes close to traditional election polls.

[18] See appendix for details.

Table 5: Share of tweets and election results

Party	All mentions		Election	
	Number of tweets	Share of Twitter traffic	Election result	Prediction error
CDU	30,886	30.1%	29.0%	1.0%
CSU	5,748	5.6%	6.9%	1.3%
SPD	27,356	26.6%	24.5%	2.2%
FDP	17,737	17.3%	15.5%	1.7%
Die Linke	12,689	12.4%	12.7%	0.3%
Die Grünen	8,250	8.0%	11.4%	3.3%
			MAE:	1.65%

Notes: Election result was adjusted to reflect only the 6 main parties in our sample. Tweets represents the total number of messages mentioning the party. If a message mentioned several parties, it was counted for each one of those parties. MAE = mean absolute error.

While the predictive qualities of Twitter mentions confirm reports of similar findings (Suenami & Yutaka, 2010; Tweetminster, 2010), other studies have not explored the information content of microblogs with respect to likely coalitions after the election. Therefore, after extracting the absolute strength of each party from the message volume, we now turn to the relationships between the parties. This is all the more relevant, as all parties were far from an absolute majority in the weeks preceding the federal election and a coalition government was on the horizon. In order to assess the predictive accuracy of the information content of political microblogs with respect to likely coalitions, we need a benchmark. We can look at two sources to provide this benchmark. Based on preelection polls, government formation on state and federal level since 1990, and political heterogeneity of political programs, Bräuninger and Debus (2009) have developed econometric models to predict the likelihood of various coalitions after the federal election 2009. They come to the conclusion that, if CDU/CSU and FDP can gain a majority in the election, a coalition of the two is almost certain. A remake of the grand coalition is the most likely scenario, if the center-right fails to gain a majority. According to Bräuninger and Debus (2009), a left-of-center coalition is very unlikely whereas the so-called Jamaica-coalition consisting of CDU/CSU, FDP, and the Green party is only slightly less likely than a center-right coalition. In addition to this academic study, we also provide the results of a survey conducted by research institute TNS Emnid in 2009, which offered the most accurate preelection poll among the six research institutes cited above. In a survey of 1,000 people, 43% of the respondents selected the center-right coalition of CDU/CSU and FDP, 20% the grand coalition (CDU and SPD), 6% a coalition of CDU and Green party, and 5% a “Jamaica”-coalition (CDU/CSU, FDP, Green party) as the most likely coalition to be formed after the election. Although both studies confirm conventional wisdom among political analysts in Germany, they provide an objective benchmark for our results, especially for readers not familiar with the German political system.

While some tweets mention only one particular party (sole mentions), many messages refer to several parties (joint mentions). We investigate whether joint mentions reflect prevailing or even upcoming political ties. To make the comparison easier and the interpretation more straightforward, we focus on tweets mentioning only two parties. Based on the overall probability that any one party is mentioned in these tweets, a conditional probability that two parties are mentioned together can be computed. If all combinations were equally likely, this conditional probability should equal the observed share of tweets mentioning these two parties. Due to different base rates, we divide the observed share of joint mentions by the conditional probability to derive a comparative measure. If share(CDU, CSU) represents the share of observed joint mentions of these two parties, the relative frequency *f*, is calculated as follows:

$$(2) \qquad f = \frac{share(CDU, CSU)}{\left[P(CDU|CSU) + P(CSU|CDU)\right]/2}$$

The relative frequency illustrates how often two parties are mentioned together relative to the random probability based on the overall "share of voice" of the individual parties. If *f* equals 1.5, the share of observed joint mentions is 50% higher than pure chance would suggest. Table 6 shows the relative frequency for all combinations of two parties based on all tweets mentioning more than on party (*n = 61,700*). Not surprisingly, the combined mentioning of sister parties CDU and CSU was the most frequent (*f = 1.25*), whereas CSU and the left-of-center parties (SPD, Grüne, and Die Linke) were mentioned together the least. While the governing coalition of CDU and SPD are naturally mentioned jointly quite frequently, the Union parties (CDU and CSU) are associated most closely with its desired coalition partner at that time, the FDP. The parties of the left side of the political spectrum are associated with each other more often than with the right-of-center parties (CDU, CSU, and FDP). In sum, the joint mentions of political parties accurately reflect the political ties between the parties. We conclude that, despite the fact that the Twittersphere is no representative sample of the German electorate, the activity prior to the 2009 German election seems to validly reflect the election outcome.

Table 6: Relative frequency of joint mentions

	CDU	CSU	SPD	FDP	Die Linke
CDU	-				
CSU	1.25*	-			
SPD	1.23*	0.71*	-		
FDP	1.04*	1.01	0.90*	-	
Die Linke	0.81*	0.79*	1.04*	0.97	-
Die Grünen	0.84*	0.79*	0.98	1.06*	1.18*

*Notes: * $p < .05$*

5 Conclusion

5.1 Discussion of results

We analyzed over 100,000 Twitter messages mentioning parties or politicians prior to the German federal election 2009. Overall, we found that Twitter is indeed used as a platform for political deliberation. The mere number of tweets reflects voters' preferences and comes close to traditional election polls, while the sentiment of political Twitter messages closely corresponds to the electorate's sentiment and evidence from the media coverage of the campaign trail. With respect to our first research question, we found more than one third of all messages to be part of a conversation indicating that Twitter is not just used to spread political opinions but also to discuss these opinions with other users. While we find evidence of a lively political debate on Twitter, this discussion is dominated by a small number of users: Only 4% of all users accounted for more than 40% of the messages. With respect to our second research question, we found the multidimensional sentiment profiles of politicians and parties to plausibly reflect many nuances of the election campaign. Overall, the similarity of profiles between parties matches the similarity of their political agendas. With respect to our third research question, we have found a party bias in individual user accounts with respect to the volume as well as the sentiment of their political communication on Twitter. Our results suggest that this bias is linked to party affiliation. With respect to our fourth research question, we found that the mere number of messages reflects the election result and that this rather simple metric, with a MAE of 1.65%, even comes close to traditional election polls. This finding is in contrast to previous studies of political deliberation online. In a study on internet message boards by Jansen and Koop (2005), even the positions of the two largest parties were reversed and the party winning an absolute majority was only associated with 27.2% of the party mentions. The authors attributed this phenomenon to the dominance of a few users who "determined the overall ideological 'feel' of the discussion board" (Jansen &

Koop, 2005, p. 624). Given that there was even less equality of participation in our sample and a bias in party orientation among individual users, it is all the more surprising that heavy users were unable to impose their views on the discussion and affect the accuracy of aggregate results. However, we strongly believe that our results did not come about by chance of this particular data set and elaborate on the reasons in our discussion of microblogging forums as information market. Our results provide evidence supporting our theory that microblogging forums provide a mechanism for weighing information and that, despite individual biases, errors can cancel each other out. The predictive accuracy is even more impressive when compared to the track record of the IEM, a prediction market set up with the explicit purpose to predict election results. Our results clearly suggest that Twitter may complement traditional methods of political forecasting (e.g., polls or surveys). There are multiple advantages of extracting public opinion from microblogging content (see O'Connor, Balasubramanyan, Routledge, & Smith, 2010, for more detail). These include cost (because most of the content is freely available and easily accessible), speed (traditional polls often take days to plan and conduct), more recent information (with permanent online access including mobile connections users often share new opinions instantly when they occur to them), frequency (the density of the information stream on Twitter allows us to draw samples at almost arbitrary intervals at almost no additional costs), unedited expression (natural responses not constrained to predefined topics and by standardized response formats), and a greater variety of topics (microblogging forums covers almost every imaginable topic). Overall, our results demonstrate that Twitter can be considered a valid indicator of the political landscape offline.

5.2 Limitations and further research

This study, like others, does not come without caveats. First, the demographics of Twitter users may raise concerns that our sample may not have been representative of the German electorate. While we have explored theoretical reasons for this sample to produce predictions that are accurate, but not representative of the general population, the majority of our article has been dedicated to the empirical comparison of the microblogging content and the political landscape offline. One of the most critical aspects of further research will be to better understand and investigate empirically the mechanism by which information is weighted and distributed in microblogging forums. Our finding that counting party mentions is different from a survey process provides a starting point for this research. Second, our investigation was based on one particular text analysis software and used an existing dictionary not specifically tailored to classify short political tweets. There are many specifics of communication through microblogging services, including the use of a special syntax and conventions (e.g., the use of emoticons) which are not reflected in our default LIWC dictionary. In addition, categories such as patriotism, xenophobia, parochialism, empathy or humanitarian, and philanthropic instincts seem particularly relevant in the realm of political analyses. Since we translated the German language messages into English, some nuances in meaning may have been lost in the translation. However, we believe this effect to be negligible since LIWC is based on word count only and therefore should not be affected by grammatical errors. Third, due to the requirement of a minimum document length for LIWC to produce meaningful, comparable results, we treated all messages published in a given time frame as one document and were unable, for instance, to distinguish between positive and negative comments. Further research should refine the text analysis to the political discussion and investigate the sentiment of the messages tweet by tweet because Asur and Huberman have shown that "sentiments extracted from Twitter can be further utilized to improve the forecasting power of social media [mentions]" (2010, p. 1). In addition, similar to the stratified sampling approach and weighted designs of traditional polling methods, one could leverage user demographics embedded in the profiles of social media users, to calibrate results derived from online content (O'Connor et al., 2010). Fourth, while we have examined overall sentiment of political tweets, voters' attitudes and opinions may vary depending on specific political issues. Future sentiment analysis could address this issue by conducting a more detailed classification of content. This may allow us to produce results similar to detailed opinion surveys on various political issues. Finally, our study was limited to the federal election in Germany. Although other studies that correlated party or candidate mentions on Twitter with electoral success indicate that similar results can be found in other countries with different electoral systems (even on the level of individual constituencies), we cannot generalize that these results will hold everywhere (e.g., Suenami & Yutaka, 2010; Tweetminster, 2010). To summarize, our results demonstrate that Twitter can be seen as a valid real-time indicator of voters' political preferences. Little research has yet been conducted in this area, leaving many questions unresolved. Further research should test whether more advanced text analysis procedures can produce even more meaningful results. Researchers should also try to capture the context of a particular statement in a more comprehensive manner including threads of conversation and links to information beyond the tweets. Analyzing conversations may be a rich source of data for understanding how users interact, political ideas evolve, and arguments are exchanged in online discussions. Including information from external sources, such as news articles linked to in the tweets, may be a starting point to distinguish between the content generated by microbloggers themselves and the content that is merely reflected in the microblogging forum. It would help us understand whether and to what extent the content in microblogging forums provides us with truly new or unique insights or whether Twitter is simply mirroring other sources.

Next to the immediate conclusions from our empirical results, our study also contributes to the understanding of the ability of the blogosphere to aggregate information. Even though we do not yet fully understand how these mechanisms works, our results indicate that information on Twitter actually can be aggregated in a meaningful way. The fact, that even the fairly simple methodology used in our study was able to generate plausible results is encouraging and points to additional possibilities to leverage Twitter as an information market.

6 Appendix

6.1 Political deliberation on Twitter

Table 7 provides one example of political deliberation on Twitter's public timeline. Of course, examples like these can only provide exploratory evidence and we may have missed some replies belonging to a discussion thread because respondents do not necessarily repeat the party names in every message. In their study of political discussion boards, Jansen and Koop (2005) have found that only 60% of all messages mentioned a political party by name. However, since Twitter users are aware of the unstructured nature of microblogging communication and therefore include searchable keywords, so-called hashtags, in many messages (e.g., “#CDU”), we believe the share of relevant replies to be small.

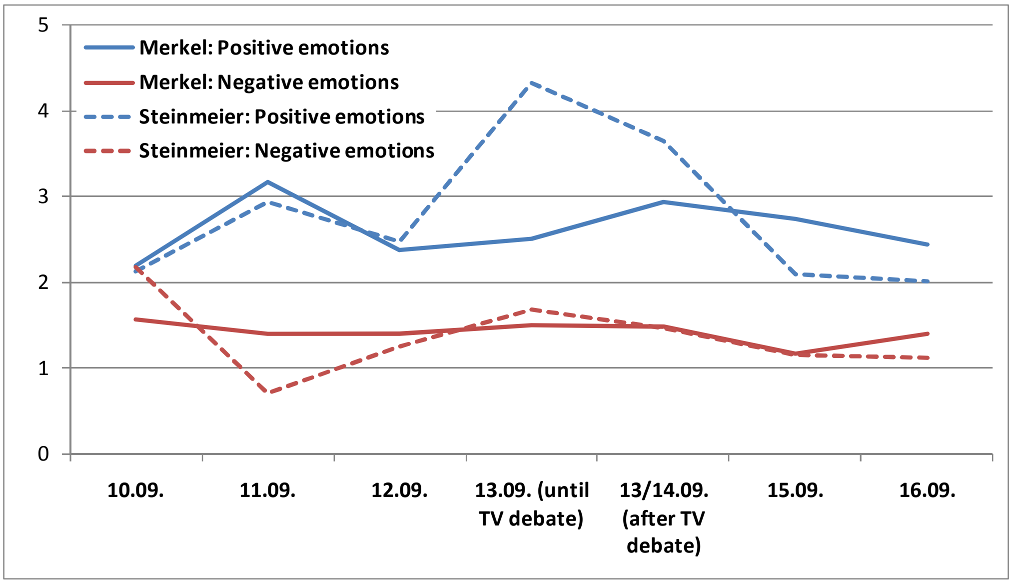
Table 7: Sample of a political discussion on Twitter

Date and time	Message
Tue, 25 Aug 2009 11:48:58	@HOLGI And that's why it's important that the small coalition partner is as strong as possible. Pro Citizen Rights, Pro FDP+
Tue, 25 Aug 2009 11:54:23	@TMOEHLE The FDP is the enemy of a free society and will remain this enemy in any coalition. Don't fool yourself.
Tue, 25 Aug 2009 12:00:26	@HOLGI Is this your old gut feeling or do you have concrete evidence based on the current FDP federal political agenda
Tue, 25 Aug 2009 12:34:19	@TMOEHLI Politics for the wealthy is politics against freedom, FDP-politics is for the wealthy, thus: FDP politics is against freedom
Tue, 25 Aug 2009 13:23:06	@HOLGI FDP does not make politics for the wealthy, quite the contrary. The citizen money proposal will help the weak in the society!
Tue, 25 Aug 2009 13:52:30	@TMOEHLE Citizen money? A means of coercion against the poverty that was brought about by your neoliberal ideology in the first place? What a farce! FDP-
Tue, 25 Aug 2009 13:56:55	@TMOEHLE If the FDP really wants to do something for the weak, it would not try to make the stronger even stronger.
Tue, 25 Aug 2009 15:19:13	@TMOEHLE This is ludicrous. For the past couple of year, no social welfare cuts have been enough for the FDP - and now you are becoming meek and mild.
Tue, 25 Aug 2009 15:25:24	@HOLGI What are you talking about? Cuts in social welfare were made for 11 years while the SPD was governing. Are you confusing us with them?
Tue, 25 Aug 2009 15:45:07	@TMOEHLE The FDP was in the opposition and supported all of that. You don't get it - and I don't argue with fundamentalists. EOT [end of thread]

6.2 Changes of sentiment over time

While much of the analysis of Twitter traffic focuses on the reflection of sentiment in the entire sample text corpus, one benefit of Twitter is its timeliness. Consequently, in order to examine whether political tweets can reflect sentiment over time, we conduct an analysis of tweets surrounding the TV debate between the two candidates for chancellor, Angela Merkel and Frank-Walter Steinmeier (see Figure 2).

Figure 2: LIWC sentiment of political tweets in the days surrounding the TV debate



Notes: This figure shows the share (in percent) of negative and positive emotion words according to the LIWC dictionary associated with the two candidates for chancellor.

Again, positive emotions outweigh negative emotions in the messages about both candidates. However, while the two were head-to-head with respect to positive emotions in the 3 days prior to the debate, Steinmeier takes the lead on the day of the debate and the following day. Even though there was no clear winner, most commentators and polls saw Steinmeier slightly ahead in the debate, especially relative to low prior expectations. But already two days after the debate, positions reverse, indicating only a temporary swing in emotions. Interestingly, negative emotions seem to be less volatile than positive emotions, which could be a reflection of the fact that the debate of the then-coalition-partners was fairly harmonious. To conclude, one can say that even on a daily basis, Twitter messages can be considered a plausible reflection of the changes in the voters' sentiment.

6.3 Distribution of user attention

Users usually subscribe to a selection of microblogs and only see messages posted by the authors they “follow”. While there are no dedicated message boards on Twitter, users obtain their desired contents and build their own discussion forums by subscribing to specific individuals. Still, the theory that the political discussion online is fragmented along ideological lines may still hold. We can analyze the messages on these “personalized message boards” to evaluate whether they represent ideological pockets. While it is not feasible to collect all messages subscribed by a particular user, we can take his or her own comments as a proxy for the party orientation and determine whether there is an emphasis on the discussion of one particular party. We can use the distribution of attention as a measure of this emphasis. The distribution of attention represents the share of mentions that a user dedicates to the various parties. Table 8 shows the average distribution of attention for the various user groups introduced in the first part of the results section. Obviously, due to the limited number of postings, there is a bias towards one party in the case of one-time and light users. But even heavy users dedicate almost half of all mentions to one particular party. So we can conclude that individual users do not spread their attention equally, but put a clear emphasis on the discussion of one particular party. On the other hand, these results confirm that counting party mentions is not to be confused with a survey process, as all users dedicate a significant share of voice to more than one party.

Table 8: Distribution of user attention

User group	Distribution of attention (share of mentions)					
	Most mentioned party	2nd party	3rd party	4th party	5th party	6th party
One-time (1)	86.6%	8.5%	2.3%	1.4%	1.1%	0.1%
Light (2-5)	62.0%	25.3%	8.2%	3.0%	1.4%	0.3%
Medium (6-20)	51.2%	24.4%	12.8%	7.0%	3.6%	1.0%
Heavy (21-79)	47.6%	23.5%	13.2%	8.5%	5.1%	2.0%
Very heavy (80+)	47.3%	21.2%	13.5%	9.0%	5.8%	3.2%
Total	72.4%	16.7%	6.0%	2.9%	1.7%	0.3%

Notes: Users can mention several parties in one tweet.

6.4 Forecast accuracy of traditional methods

Table 9: Forecast accuracy of various election polls and prediction markets

Source	Sample size	MAE
Twitter	*[1]*	*1.65%*
Election polls		
Forsa	7	1.37%
Allensbach	5	1.27%
Emnid	6	1.08%
Forschungsgruppe Wahlen	5	1.33%
GMS	2	1.70%
Infratest/dimap	6	1.23%
IEM prediction markets		
U.S. presidential elections	5	1.37%
Other U.S. elections	14	3.43%
Non-U.S. elections	30	2.12%
German elections	5	1.10%

Notes: Sample size represents the number of polls in our sample period for election polls and the number of prediction markets for the IEM. IEM results are based on Berg, Forsythe, Nelson, & Rietz (2008).

7 References

AAPOR (2010). *AAPOR report on online panels*. Retrieved May 15, 2011 from http://www.aapor.org/AM/Template.cfm?Section=AAPOR_Committee_and_Task_Force_Reports&Template=/CM/ContentDisplay.cfm&ContentID=2223

Adamic, L. A., & Glance, N. (2005). The political blogosphere and the 2004 US election: Divided they blog. In *Proceedings of the 3rd International Workshop on Link Discovery* (pp. 36-43). Chicago, IL: ACM.

Asur, S., & Huberman, B. (2010). Predicting the future with social media. *Working paper.* Retrieved May 15, 2011 from http://arxiv.org/pdf/1003.5699

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5, 323-370.

Berg, J., Forsythe, R., Nelson, F., & Rietz, T. (2008). Results from a dozen years of election futures markets research. In C. Plott & V. Smith, *Handbook of Experimental Economics Results*, 742-751, Amsterdam, The Netherlands: Elsevier.

Berg, J., & Rietz, T. (2006). The Iowa Election Markets: Stylized facts and open issues. In R. Hahn & P. Tetlock (Eds.), *Information Markets - A New Way of Making Decisions* (pp. 142-169), Washington, DC: The AEI Press.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1998). Learning from the behavior of others: Conformity, fads, and informational cascades. *The Journal of Economic Perspectives*, 12, 151-170.

Bräuninger, T., & Debus, M. (2009). Die Regierungsbildung nach der Bundestagswahl 2009: Wie wahrscheinlich ist eine Neuauflage der großen Koalition? [in German]. In *DVPW Workshop zur Bundestagswahl 2009* (pp. 1-35), Kiel: DVPW.

Delli Carpini, M. X., Cook, F. L., & Jacobs, L. R. (2004). Public deliberation, discursive participation, and citizen engagement: A review of the empirical literature. *Annual Review of Political Science*, 7, 315-344.

Forsythe, R., Rietz, T., & Ross, T. (1999). Wishes, expectations and actions: a survey on price formation in election stock markets. *Journal of Economic Behavior & Organization*, 39, 83-110.

Giller, G. 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.* Retrieved May 15, 2011 from http://ssrn.com/abstract=1423628

Gueorguieva, V. (2007). Voters, MySpace, and YouTube: The impact of alternative communication channels on the 2006 election cycle and beyond. *Social Science Computer Review*, 26, 288-300.

Hayek, F. V. (1945). The use of knowledge in society. *American Economic Review*, 35, 519-530.

Hinich, M., & Munger, M. (1997). *Analytical politics*. Cambridge, England: Cambridge University Press.

Hong, L., & Page, S. (2001). Problem solving by heterogeneous agents. *Journal of Economic Theory*, 97, 123-163.

Honeycutt, C., & Herring, S. C. (2009). Beyond microblogging: Conversation and collaboration via Twitter. In *42nd Hawaii International Conference on System Sciences* (pp. 1-10), Kauai, HI: HICSS.

Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, 60, 1-20.

Jansen, H. J., & Koop. R. (2005). Pundits, ideologues, and ranters: The British Columbia election online. *Canadian Journal of Communication*, 30, 613-632.

Klingemann, H.-D., Volkens, A., Bara, J., & McDonald M. (2006). *Mapping policy preferences II: Estimates for parties, electors, and governments in Eastern Europe, European Union and OECD 1990-2003.* Oxford, England: Oxford University Press.

Koop, R., & Jansen, H. J. (2009). Political blogs and blogrolls in Canada: Forums for democratic deliberation? *Social Science Computer Review*, 27, 155-173.

Nielsen Media Research (2009). *Das Phänomen Twitter in Deutschland* [in German]. Retrieved May 15, 2011 from http://de.nielsen.com/news/NielsenPressemeldung04.08.2009-Twitter.shtml

O'Connor, B., Balasubramanyan, R., Routledge, B., & Smith, N. (2010). From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of International AAAI Conference on Weblogs and Social Media* (pp. 122-129). Washington, DC: AAAI.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2, 1–135.

Pearanalytics (2009). *Twitter study*. Retrieved May 15, 2011 from http://www.pearanalytics.com/blog/wp-content/uploads/2010/05/Twitter-Study-August-2009.pdf

Servan-Schreiber, E., Wolfers, J., Pennock, D. M., & Galebach, B. (2004). Prediction markets: Does money matter? *Electronic Markets*, 14, 243-251

Suenami, A., & Yutaka, M. (2010). Voter behavior analysis and election campaign strategies using information on the web [in Japanese]. In *Proceedings of the SIG-KBS, No. 87* (pp. 1-6). Tokyo: JSAI.

Sunstein, C. (2007). Neither Hayek nor Habermas. *Public Choice*, 134, 87-95.

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. New York, NY: Random House, Inc.

Tausczik, Y. R., & Pennebaker, J. W. (2009). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29, 24-54.

TNS Emnid (2009). *Wahrscheinliche Regierungskoalition nach der Bundestagswahl* [in German]. Retrieved May 15, 2011 from http://de.statista.com/statistik/daten/studie/29701/umfrage/wahrscheinliche-regierungskoalition-nach-der-bundestagswahl/#info

Tweetminister (2010). *Is word-of-mouth correlated to general election results? The results are in*. Retrieved May 15, 2011 from http://www.scribd.com/doc/31208748/Tweetminster-Predicts-Findings

Web Evangelisten (2009). *Twitterumfrage 2* [in German]. Retrieved May 15, 2011 from http://webevangelisten.de/twitterumfrage/

Williams, C., & Gulati, G. (2008). What is a social network worth? Facebook and vote share in the 2008 presidential primaries. In *Annual Meeting of the American Political Science Association* (pp. 1-17). Boston, MA: APSA.

Yu, B., Kaufmann, S., & Diermeier, D. (2008). Exploring the characteristics of opinion expressions for political opinion classification. In *Proceedings of the 2008 International Conference on Digital Government Research* (pp. 82-91). Montreal: ACM.

Zarrella, D. (2009). *State of the Twittersphere*. Retrieved May 15, 2011 from http://blog.hubspot.com/Portals/249/sotwitter09.pdf

II.2 Essay 2

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.

JEL Classification: G12; G14

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

<u>Current status</u>: Submitted to and reviewed by the *Journal of Finance*.

Acknowledgements: This paper contains elements of joint work with Prof. Dr. Isabell M. Welpe, Dr. Philipp Sandner, Dipl.-Psych. Andranik Tumasjan, Philip Heinemann, and Sebastian Peters.

1 Introduction

> "*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 & 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[19] 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 155 million messages per day generated by roughly twice as many registered users (TechCrunch, 2011). Many of these messages are dedicated to the discussion of public companies and trading ideas. As a result, there are investors who attribute their trading success to the information they find on social media websites and Twitter-based trading systems have been developed by financial professionals to alert users of sentiment-based investment opportunities (Jordan, 2010) and by academic researchers to predict break-points in financial time series (Vincent & 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 $1,500 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

[19] http://www.twitter.com

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. 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 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 & 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)[20], 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

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

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.

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 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.

2 Related work and research questions

2.1 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[21] 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 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

[21] For the purpose of this study, we define qualitative information as words. This definition is in line with related research (Tetlock et al., 2008).

dedicated to financial topics, such as internet stock message boards[22] 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 Campbell, 2001; Das, Martinez-Jerez, and Tufano, 2005; Felton & Kim, 2002; Lerman, 2010).

A number of previous studies have investigated the relationship between stock message boards and financial markets. Wysocki (1998) was the first to investigate internet stock message boards. For the 50 most frequently discussed firms on Yahoo!Finance between January and August 1998, he illustrates that message volume did forecast next-day trading volume. Whereas this study only investigated message volume, others have taken a more differentiated approach to the information content on message boards. For a limited sample of internet service sector stocks, Tumarkin and Whitelaw (2001) have explored the information embedded in voluntary user ratings (from strong buy to strong sell), but were unable to 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 recommendations 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 & 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 & Swanson, 2010). Automated classifiers can provide an unbiased interpretation of a message based on its content. Das and

[22] Some studies (e.g., Clarkson, Joyce, & 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.

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.[23] 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 & 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 classification.[24] 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). Methodologically our study attempts to close the gap between these two 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

[23] 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.

[24] 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.

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.[25]

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.[26] The following three distinct characteristics of 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

[25] 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%.

[26] 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" (p. 377). While they noted a strong correlation among all media sources, the sentiment on Twitter appeared to have a slightly stronger correlation with future returns.

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, Balasubramanyan, Routledge, and Smith (2010) have found Twitter messages to be a leading indicator for the Index of Consumer Sentiment (ICS), a measure of U.S. consumer confidence. Both Zhang, Fuehres, and Gloor (2010) and Bollen, Mao, and Zeng (2011) 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 & Frank, 2004), we also investigate information diffusion in this social media context, which may help explain why microblogging forums can aggregate information efficiently.

2.2 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 2.2.1-2.2.3) and, second, for RQ2 investigating the information diffusion in stock microblogging forums (H4a-H4b in section 2.2.4).

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.[27] However, if this reflection 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.[28] 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 volume, for example, 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.

2.2.1 Bullishness

In much of the financial literature individual investors are considered the least informed market participants (e.g., Easley & O'Hara, 1987; Hirshleifer & Teoh, 2003) and empirical evidence confirms that individual investors pay a significant performance penalty for active trading (Barber & 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

[27] Many studies boldly interpret this relationship as an effect of online forums on the market (e.g., Bettman, Hallett, & 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).

[28] Clarkson et al. (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.

(van Bommel, 2003, p. 1499). On the other hand, van Bommel (2003) acknowledges a moral hazard problem due to the opportunity to spread false rumors[29], which could lead followers to ignore rumors altogether. In addition, a prolific poster interviewed by Das et al. (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 & Odean, 2008). In particular, empirical studies have shown that investors are often influenced by word of mouth (e.g., Ng & Wu, 2006; Hong, Kubik, & 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 & Weerts, 2009). DeMarzo, Vayanos, and Zwiebel (2003) have proposed a model of bounded rationality in which individuals are subject to persuasion bias and fail to account for repetition in the information they receive. As a result of this persuasion bias, "influence on group opinions depends not only on accuracy, but also on how well-connected one is in the social network that determines communication" (p. 909). Given that stock microblogs reflect the theoretical properties of this model with the size of the followership indicating social influence, we propose:

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

[29] 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 et al., 2006, p. 2). In this sense, most of the information in stock microblogs may be considered rumors.

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.

2.2.2 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 et al., 2005). However, beyond the direct link between posting and trading, there are reasons to believe that an increase in message volume may even lead "lurkers" to trade. Cao, 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, Shleifer, Summers, & Waldmann, 1990). Given that a large share of participants in stock microblogging forums consists of day traders[30], 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.

2.2.3 Disagreement

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

Hypothesis 3a: Increased disagreement among stock microblogs is associated with higher volatility.[31]

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 & Raviv, 1993; Karpoff, 1986; Kim & 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 & Frank, 2004). However, Milgrom and Stokey (1982) have developed the "no-

---

[30] Koski et al. (2004) suggest that the vast majority of message board participants are day traders, which can be considered noise traders.

[31] Disagreement further leads to discussions, reinforcing the previous hypothesis. Das et al. (2005) find that disagreement is positively related to the message volume.

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.

2.2.4 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, Arunasalam, Milosavljevic, & Leung, 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, & 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, Haddadi, Benevenuto, & Gummadi, 2010). Boyd, Golder, and Lotan (2010) have suggested that users often retweet messages in order to relay valuable content in order to validate and endorse a particular user or posting. In addition, there is empirical evidence that, despite the abundance of available information and considerable noise, Twitter users follow the accounts to which they subscribe closely and are highly attentive to their content. A working paper studying a single Twitter account making directional forecasts of the stock market indicates that the number of followers may be correlated with the accuracy of the published information (i.e., the forecasts of the stock market; Giller; 2009[32]). Reports in the business press support the hypothesis that microblogging forums may aggregate information more efficiently than previously studied online communities.[33] 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, & 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, Galuba, Asur, and Huberman (2010) claim that the majority of users act as passive information consumers and do not forward the content to the network. Both studies were conducted with large, randomly sampled data sets and do not capture a specific domain such as stock microblogging. Therefore, next to high quality advisors, we examine whether high quality pieces of investment information (i.e., individual messages) are weighted more heavily and spread through retweets. Thus, we propose:

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

---

[32] 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.

[33] "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).

3 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.

3.1 Data set and sample selection of stock microblogs

We chose the microblogging platform Twitter as our data source for stock microblogs as opposed to other potential microblogging contexts (e.g., 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 155 million messages are posted on Twitter's public timeline every day (TechCrunch, 2011). 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 U.S. 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.[34] We focus on the S&P 100 to adequately reflect the entire spectrum of U.S. equities, including a wide range of industries, while limiting our study to well-known companies that trigger a substantial number of stock microblogs.[35]

[34] 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.
[35] Specifically, we focus on those companies that have been included in the S&P 100 as of January 1, 2010.

3.2 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 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).[36]

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.[37] 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 & Frank, 2004). Table 1 shows a few typical examples of the tweets from the training set including the manual coding.

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

[37] In line with most text classification methods using a manual training set (e.g., Antweiler & 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).

Table 1: Sample tweets from training and test set including classification

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

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

As Table 2 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.

Table 2: Classification accuracy (confusion matrix)

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%

Notes: 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).

3.3 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 to firm-specific variables. 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

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

where M^{Buy} (M^{Sell}) represents the number of buy (sell) signals on day t.[38] 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.[39] In line with Antweiler and Frank (2004), agreement among messages is defined as

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

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).[40] 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.

---

[38] 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.

[39] 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.

[40] 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 U.S. 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.

3.4 Financial market data

We have downloaded financial data in daily intervals for the S&P 100 from Thompson Reuters Datastream. Following Antweiler and Frank (2004), 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

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

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

(4) $AR_{it}^{simple} = R_{it} - R_{t}^{market}$

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

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

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, & 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

(6) $CAR_{it} = \sum AR_{it}$

and average cumulative abnormal returns for *N* companies are calculated as

(7) $ACAR_t = \dfrac{\sum_{i=1}^{N} CAR_{it}}{N}$.

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

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

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

3.5 Information aggregation in microblogging forums

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

Every tweet in our sample is classified as a recommendation to buy, hold, or sell a stock. We code this sentiment *s* of buy, hold, and sell signals as 1, 0 and -1, respectively. In line with Zhang (2009), who studied the determinants of poster reputation on online message boards, we define the quality of a tweet as the accuracy of this recommendation relative to same-day returns[41] of the stock in question as

$$(9) \quad quality = \begin{cases} = 1 \; if \; \dfrac{s_{it}}{R_{it}} > 0 \\ \hline = 0, \; otherwise \end{cases},$$

where s_{it} is the sentiment of a message on day *t* associated with stock *i*. We only take into account messages published during trading hours and ignore hold messages in the computation of quality scores.[42] Next to the quality of individual messages, we also compute the quality of a particular user's investment advice as the average quality of all messages posted by this user. In addition, we compute the average sentiment of a user's messages.

In the context of microblogging, Cha et al. (2010) have defined three different measures of user influence: retweets, mentions, and followership. The first measure (i.e., the fact whether a message was retweeted) can also serve as a proxy for the weight given to an individual

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

[42] 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.

tweet. Microblogging users frequently forward (i.e., “retweet”) messages which they find noteworthy to their followers. The retweets usually contain the abbreviation “RT” followed by the name of the original author.[43] The first sample tweet in Table 1 provides an example of such a retweet. Because Twitter does not provide information regarding the relationship of individual tweets, we identified retweets in our data set by filtering all retweets and matching the 40 characters following the retweet token and the name of the original author with all other tweets in the data set.[44] 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.[45] Having laid out the definition of our variables, we now turn to exhibiting and interpreting the results.

4 Results

4.1 Descriptive statistics

The results section is structured as follows. After a brief summary of descriptive statistics regarding our data set, sections 4.2 and 4.3 address RQ1, whereas section 4.4 covers results related to RQ2. Section 4.2 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 4.3, we provide in-depth analyses exploring the relationships that are supported by the empirical evidence. Section 4.4 tests our

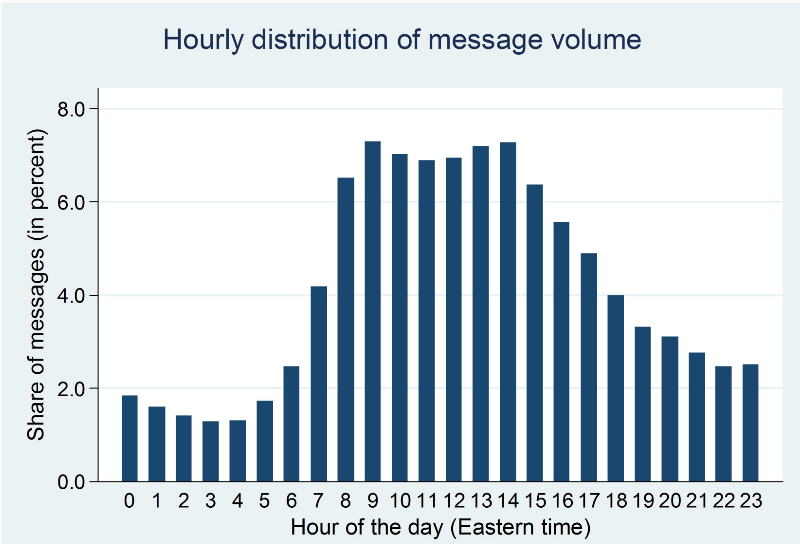
[43] Alternative formats include “RT: @”, “via @”, “by @”, and “retweet @”.

[44] 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.

[45] 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.

hypothesis explaining the efficient information aggregation by exploring whether good investment advice receives greater attention in stock microblogging forums.

Figure 1. Hourly message volume



Notes: 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.

We have collected 249,533 stock-related microblogging messages containing the dollar-tagged ticker symbol of an S&P 100 company.[46] 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 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

[46] 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.

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. Table 3 shows summary statistics of the market and tweet features on a per company basis.

Table 3: Summary statistics of market and tweet features

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

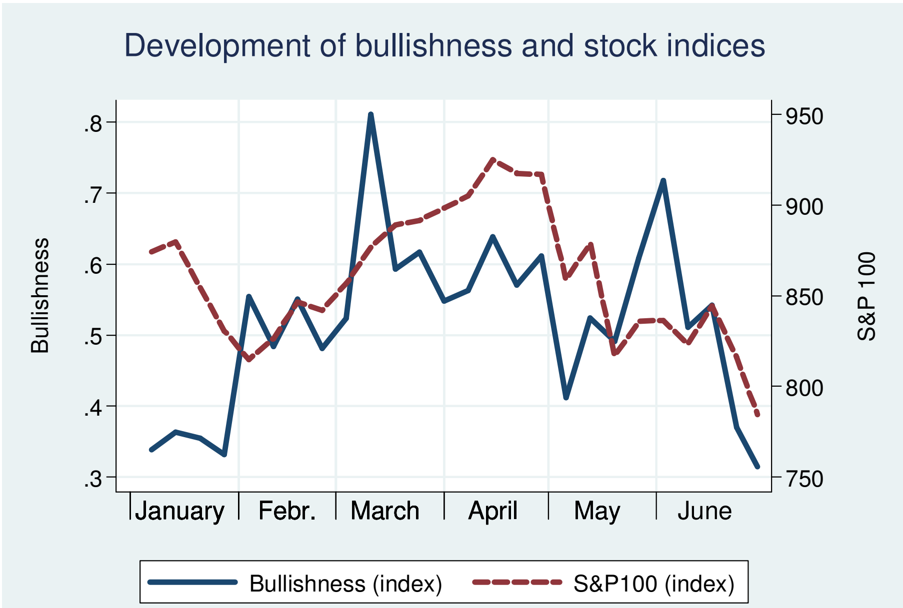
Notes: All daily statistics are reported on a per company basis. Returns were defined as the log difference in prices, traded volume represents the number of shares traded, and we use PARK volatilities derived from daily highs and lows. Bullishness is the sentiment of a particular message, message volume 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 (i.e., shown in percent) and PARK volatility is scaled by 10,000 for easier readability. N = 10,123 company-days with tweet features and 12,443 company-trading-days.

Figure 2 provides us with a first indication of the overall relationships between trading and message volume. The two measures show a strong correlation ($r = 0.468$, $p = 0.02$). 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 ($r = 0.408$, $p = 0.04$).

Figure 2: Aggregate message volume vs. trading volume

Notes: 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



Notes: This figure presents the development of overall bullishness and the S&P 100 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.

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.

4.2 Overall relationship of tweet and market features

4.2.1 Pairwise correlations

Contemporaneous pairwise correlations provide a first indication with respect to the relationships between market and tweet features (Table 4). 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 ($r = 0.046$, $p = 0.0$).

Table 4: Pairwise correlations for stock and tweet data

	Return	AR (simple)	AR (market model)	Traded volume	Volatility	Bullishness	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

Notes: 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.

4.2.2 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[47] relationships between tweet and market features corresponding to our hypotheses in order to investigate whether tweet features can serve as proxies for market developments.

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

Table 5: Contemporaneous regressions

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

Table 5 shows contemporaneous fixed-effects panel regressions of the market features as the dependent variable and the three tweet features as independent variables. The market index is used as a control variable. Due to significant cross-sectional differences in message volume, we use fixed-effects for each stock. 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 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.

4.2.3 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 6 shows time-sequencing regressions for tweet and market features (in line with Antweiler & 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 & 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.

The most obvious question in time series analysis of microblogs and the market is whether message sentiment can help predict returns (H1). Table 6 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.

Table 6: Time-sequencing regressions

		X->Y					Y->X				
X	**Y**	**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: 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 the regression of actual values (bold), the second row reports results for the regression of standardized values (italics). All regressions use company fixed effects. N = 12,443 company-trading-days.

** p<0.05, ** p<0.01, *** p<0.001*

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.

4.3 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 4.3.1), and bullishness and returns on the other (section 4.3.2).[48]

4.3.1 Trading volume

While the pairwise correlations and predictive regressions have suggested that message volume and agreement contain valuable information with respect to trading volume, it

[48] 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.

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.

Table 7 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).

Table 7: Volume regression

	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: 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 U.S. 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 U.S. 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.
** p<0.05, ** p<0.01, *** p<0.001.*

4.3.2 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 days 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.

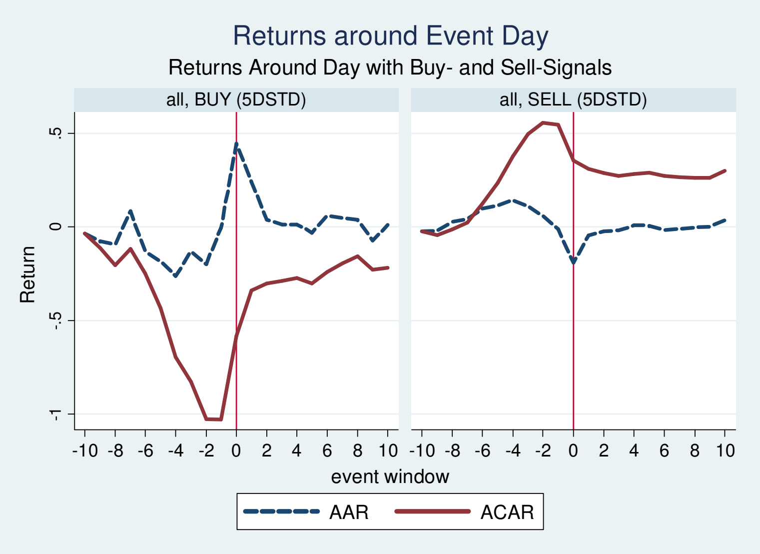
Figure 4 shows the development of abnormal returns around the event days. We observe that buy recommendations are preceded by an extended period of negative returns. Abnormal returns are negative from 6 to 2 days prior to the recommendation. This is in line with our time-sequencing regressions, where 2 day lags of bullishness were 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, & Tutticci, 2006).[49] 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 and no new information was contained in their recommendations (Dewally, 2003; Mizrach & Weerts, 2009; Tumarkin & Whitelaw, 2001).

[49] Tetlock et al. (2008) even use only 10 bps to assume reasonable transaction costs.

Figure 4: Event-study of 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 (1,216 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

Notes: 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.

4.4 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 8 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.[50] But, as the last column indicates, the higher frequency users do not appear to be better investment advisors as the average quality does not vary by user group.[51] 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.

Table 8: Frequency distribution of users and messages by user group

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%

Notes: 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.

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.[52] 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

[50] 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.

[51] 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 et al., 2007).

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

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 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 8). 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.[53] 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.

[53] 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 9: Determinants of user influence

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

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

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 8. 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.

Regarding the estimation methods: The dependent variables are all highly skewed, non-negative count data. We find overdispersion with significant alphas for all dependent variables. Since poisson regressions, which have been used in similar studies of Twitter accounts (Giller, 2009), assume equal conditional means and variance, this suggests the use of either zero-inflated poisson regressions or negative binomial regressions. The count data for retweets and mentions contain a substantial share of zeroes. This may justify the use of zero-inflated models for these two dependent variables. However, there is no reasonable data generating process to explain excess zeroes. In addition, the predictors for excess zeroes are not significant in most cases where we attempted to fit this type of model and the Vuong test does not suggest the use of zero-inflated models. In fact, we would argue that there are no "excess" zeroes because all zeroes among users with a substantial number of tweets truly indicate that their messages are simply not being retweeted (i.e., quoted). This suggests the use of negative binomial regressions. They come with the additional advantage of being more parsimonious and allowing for the use of one consistent regression model across all four independent variables. Therefore, we show the results for negative binomial regressions. 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 & 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 9 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 (Zhang, 2009), 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).

5 Discussion

5.1 Summary of results

Stock microblogs have become a vibrant online forum to exchange trading ideas and other stock-related information. This study set out to investigate the relationship between stock microblogs and financial market activity and offer an explanation for the efficient aggregation of information in microblogging forums. We find, first, that stock microblogs contain valuable information that is not yet fully incorporated in current market indicators and, second, that retweets and followership relationships provide microblogging forums with an efficient mechanism to aggregate information. Our results are summarized in Table 10.

Table 10: 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.*

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.

5.2 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 U.S. 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 & 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.

5.3 Conclusion

It appears that online investors have matured since the introduction of messages boards more than 10 years ago. We observe a more balanced ratio of buy and sell signals and traders no longer 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.

6 Appendix

6.1 Naïve Bayesian text classification

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

$$\text{(A1)}\quad P(c \mid d) = \ln P(c) \sum_{1 \le i \le n_d} \ln P(w_i \mid c)\,,$$

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

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

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

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

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

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

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 was 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, & 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 et al. (2008) that the fraction of emotional words in firm-specific news, can predict stock returns, we tag more than 4,000 emotional words as either positive or negative. Following Tetlock et al. (2008) we use the General Inquirer’s Harvard-IV-4 classification dictionary and add each occurrence of an emotional word to the bag of words for that message. Thus we combine text mining approaches based on pre-defined dictionaries and statistical methods. Sixth, we apply the widely used Porter stemmer in order to remove the morphological endings from words (e.g., “buys” and “buying” are reduced to “buy”; Porter, 1980). Finally, following established preprocessing procedures (see Rennie, Shih, Teevan, & Karger, 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).

6.2 Classification of our data set

Table 11 shows a few typical examples of the tweets from both the training set[54] 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

[54] These examples for the training set are identical to Table 1 of the manuscript and only repeated here for better readiblity.

Naïve Bayesian algorithm can classify messages quite well. As Table 12 shows, overall in-sample classification accuracy was 81.2%. Even a more conservative 10-fold cross validation[55] 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 & Shtrimberg, 2006; Wasko, Faraj, & 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.

A look at the most common words per class (see Table 13) indicates that the Information Gain model derived a plausible dictionary from our training set. Obviously, some features occur frequently in all classes (e.g., numbers and hyperlinks). However, beyond these universal features, the most common words reasonably reflect the linguistic bullishness of the three classes. Positive emotions, for example, are much more likely among buy signals. In addition, buy signals often contain bullish words with an origin in technical analysis (e.g., “moving average”, “resistance”, “up”, or “high”), operations (e.g., “acquire”), financials (e.g., “beat”, “earn”), or trading (e.g., “buy”, “long”, “call”). Sell signals contain many corresponding bearish words in the areas of technical analysis (e.g., “support” and “cross”), financials (e.g., “loss”) or trading (e.g., “short” and “put”). As a result 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 et al. (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).

[55] See the note to Table 12 for details regarding this validation approach.

Table 11: Sample tweets from training and test set including classification

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%

Notes: 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.

Table 12: Classification accuracy

Classification accuracy (confusion matrix)

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%

Notes: 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.

Classification accuracy (accuracy by class)

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

*Notes: 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).*

Table 13: Classification results – most common words/features per class

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

Notes: 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.

6.3 Market and tweet features per company

Table 14 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 Jones Industrial Average (DJIA), bullishness in our sample period appears to be somewhat higher for technology companies such as Apple, Microsoft, and Google.

Table 14: Summary statistics by company (1/2)

Ticker	Company name	Messages	Bullishness	Agreement	Return	Volume	Volatility
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
…							

Table 14: Summary statistics by company (2/2)

Ticker	Company name	Messages	Bullishness	Agreement	Return	Volume	Volatility
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

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

6.4 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 & 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).[56] 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 1 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 5 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 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%.

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

Figure 5: Trading strategy based on tweet signals (backtesting results)

Notes: 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.

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 et al. (2008), who found a trading strategy based on negative words in firm specific news articles to earn abnormal annualized returns of more than 20%[57], 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 15 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 losing money with transaction costs of more than 7 bps, roughly in line with the threshold determined by Tetlock et al. (2008). In addition, our analysis shows a slightly better performance of the short end of the trading scheme. But even this side supports transaction costs only up to 8 bps.

Table 15: Sensitivity of trading strategy to transaction costs

Trading costs (bps)	AR (total), in %	AR (long), in %	AR (short), in %
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

Notes: 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.

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

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.

7 References

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *Journal of Finance,* 59, 1259-1294.

Bagnoli, M., Beneish, M. D., & Watts, S. G. (1999). Whisper forecasts of quarterly earnings per share. *Journal of Accounting and Economics*, 28, 27–50.

Barber, B., & Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The Journal of Finance*, 55, 773-806.

Barber, B., & Odean, T. (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, J., Hallett, A., & Sault, S. (2000). The impact of electronic message board takeover rumors on the US equity market, *Working paper.* Retrieved May 15, 2011 from http://scheule.com/Research/seminars/2008/Sault%202008.pdf

Black, F. (1986). Noise. *Journal of Finance*, 41, 529-543.

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

Bommel van, J. (2003). Rumors. *The Journal of Finance*. 58, 1499–1520.

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

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

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

Busheee, B. J., Core, J. E., Guay, W., & Hamm, S. 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).* Retrieved May 15, 2011 from http://www.businessweek.com/technology/content/feb2009/tc20090210_875439.htm

Cameron, A. C., & Trivedi, P. K. (1998). *Regression analysis of count data*. Cambridge, England: Cambridge University Press.

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

Cao, H. H., Coval, J. D, & Hirshleifer, D. (2002). Sidelined investors, trading-generated news, and security returns. *Review of Financial Studies*, 15, 615-648.

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

Chordia, T., Roll, R., & Subrahmanyam, A. (2001). Market liquidity and trading activity. *Journal of Finance*, 56, 501–530.

Clarkson, P. M., Joyce, D., & Tutticci, I. (2006). Market reaction to takeover rumour in internet discussion sites, *Accounting and Finance*. 46, 31-52.

Danthine, J.-P., & Moresi, S. (1993). Volatility, information, and noise trading. *European Economic Review*, 37, 961-982.

Das, S., & Chen, M. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science*, 53, 1375-1388.

Das, S., Martinez-Jerez, A., & Tufano, P. (2005). eInformation: A clinical study of investor discussion and sentiment. *Financial Management*, 34, 103–137.

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, 98, 703-738.

Delort, J.-Y., Arunasalam, B., Milosavljevic, M., & Leung, H. (2009). The impact of manipulation in internet stock message boards. *Working paper.* Retrieved May 15, 2011 from http://ssrn.com/abstract=1497883

DeMarzo, P. M., Vayanos, D., & Zwiebel, J. (2003). Persuasion bias, social influence, and unidimensional opinions. *Quarterly Journal of Economics*, 118, 909-968.

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

Dyckman, T., Philbrick, D., & Stephan, J. (1984). A comparison of event study methodologies using daily stock returns: A simulation approach. *Journal of Accounting Research*, 22, 1-30.

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

Engelberg, J., & Parsons, C. A. (2011) The causal impact of media in financial markets. *Journal of Finance*, 66, 67-97.

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

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

Fan, M., Tan, Y., & Whinston, A. B. (2005). Evaluation and design of online cooperative feedback mechanisms for reputation management. *IEEE Transactions on Knowledge and Data Engineering*, 17, 244-254.

Felton, J., & Kim, J. (2002). Warnings from the Enron message board. *The Journal of Investing*, 11, 29–52.

Giller, G. 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.* Retrieved May 15, 2011 from http://ssrn.com/abstract=1423628

Gu Bin, K. P., Rajagopalan, B., & Chen, H.-W. (2007). Competition among virtual communities and user valuation: The case of investing-related communities. *Information Systems Research,* 18, 68-85.

Gu, Bin, K. P., & Chen, H.-W. (2008). Melting-pot or homophily? - An empirical investigation of user interactions in virtual investment-related communities. *Working paper.* Retrieved May 15, 2011 from http://ssrn.com/abstract=1259224

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten I. H. (2009). The WEKA data mining software: An update. *SIGKDD Explorations*, 11, 10-18.

Harris, M., & Raviv, A. (1993). Differences of opinion make a horse race. *Review of Financial Studies*, 6, 473-506.

Hirshleifer, D., & Teoh, S. (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics*, 36, 337–386.

Hong, H., Kubik, J., & Stein, J. 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, A. L. (2006). Have internet message boards changed market behavior?. *Info,* 8, 67-76.

Jones, C. M., Kaul, G., & Lipson, M. L. (1994). Transactions, volume, and volatility. *Review of Financial Studies*, 7, 631-651.

Jordan, J. (2010). Hedge fund will track Twitter to predict stock moves, *Bloomberg (online edition)*. Retrieved May 15, 2011 from http://www.bloomberg.com/news/2010-12-22/hedge-fund-will-track-twitter-to-predict-stockmarket-movements.html

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

Kim, O., & Verrecchia, R. (1991). Trading volume and price reactions to public announcements. *Journal of Accounting Research,* 29, 302-321.

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

Koski, J. L., Rice, E. M., & Tarhouni, A. (2004). Noise trading and volatility: Evidence from day trading and message boards. *Working paper.* Retrieved May 15, 2011 from http://ssrn.com/abstract=533943

Lakonishok, J., & Levi, M. (1982). Weekend Effects on Stock Returns: A Note. *Journal of Finance*, 37, 883-889.

Lerman, A. (2010). Individual investors' attention to accounting information: Message board discussions. *Working paper.* Retrieved May 15, 2011 from http://ssrn.com/abstract=1540689

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

Milgrom, P., & Stokey, N. (1982). Information, trade and common knowledge. *Journal of Economic Theory*, 26, 17-27.

Mittermayer, M.-A., & Knolmayer, G. F. (2006). Text mining systems for market response to news: A survey. *Working paper. Retrieved* May 15, 2011 from http://www2.ie.iwi.unibe.ch/publikationen/berichte/resource/WP-184.pdf

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

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

Ng, L., & Wu, F. (2006) Peer effects in the trading decisions of individual investors. *Financial Management*, 39, 807-831.

O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. (2010). From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of the International Conference on Weblogs and Social Media* (pp. 122-129), Washington, DC: AAAI.

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

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

Rennie, J. D., Shih, L., Teevan, J., & Karger, D. R. (2003). Tackling the poor assumptions of Naive Bayes text classifiers. In *Proceedings of the Twentieth International Conference on Machine Learning* (pp. 616-623), Washington DC: AAAI.

Romero D. M., Galuba, W., Asur, S., & Huberman, B. A. (2010). Influence and passivity in social media. *Working paper. Retrieved* May 15, 2011 from arxiv.org/PS_cache/arxiv/pdf/1008/1008.1253v1.pdf

Sabherwal, S., Sarkar, S. K., & Zhang, Y. (2008). Online talk: does it matter?. *Managerial Finance*, 34, 423-436.

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

TechCrunch (2011). Twitter Tweets Some Big Q1 Stats: 155 Million Tweets a Day Now, *TechCrunch Blog*, Retrieved May 15, 2011 from http://techcrunch.com/2011/04/06/twitter-q1-stats/

Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (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)*, Retrieved May 15, 2011 from http://www.time.com/time/specials/packages/article/0,28804,1901188_1901207_1901198,00.html

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

Vincent, A., & Armstrong M. (2010). Predicting break-points in trading strategies with Twitter. *Working paper. Retrieved* May 15, 2011 from http://ssrn.com/abstract=1685150

Wasko, M., Faraj, S., & Teigland, R. (2004). Collective action and knowledge contribution in electronic networks of practice. *Journal of the Association for Information Systems*, 5, 493-513.

Wysocki, P. (1998). Cheap talk on the web: The determinants of postings on stock message boards. *Working paper.* Retrieved May 15, 2011 from http://ssrn.com/abstract=160170

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

Zhang, W., & Skiena, S. (2010). Trading strategies to exploit blog and news sentiment. In *Proceedings of the International Conference on Weblogs and Social Media* (pp. 275-378), Washington, DC: AAAI.

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

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

Zhang, Y., & Swanson, P. E. (2010). Are day traders bias free? - Evidence from internet message boards. *Journal of Economic and Finance*, 34, 96-112.

II.3 Essay 3

News or Noise?
The Stock Market Reaction to Different Types of Company-Specific News Events

Abstract

While most event studies are limited to one single type of event (e.g., earnings announcements, acquisitions, or initial public offerings), we examine the market impact of a comprehensive set of company-specific news events (e.g., news related to corporate governance, operations, and legal issues) on S&P 500 stock prices in order to discern genuine news that moves the market from insignificant noise without market reaction. We control for the sentiment (i.e., the positive vs. negative tone) of different news stories by leveraging computational linguistics methods, which allows us to distinguish between good and bad news. Our results show that the absolute value of cumulative returns prior to a news event are more pronounced for positive news than they are for negative news, suggesting more widespread information leakage before good news. Moreover, we find that the market reaction differs substantially across various types of news events. In addition, a cross-industry comparison indicates that industry classification may partially explain the market reaction to the same event type.

JEL Classification: G14

Keywords: Twitter; microblogging; text classification; stock market; news; event study; news sentiment

Current status: Submitted and currently under review at the *Review of Financial Studies*.

Acknowledgements: This paper contains elements of joint work with Prof. Dr. Isabell M. Welpe, Dr. Philipp Sandner, Dipl.-Psych. Andranik Tumasjan, Philip Heinemann, and Sebastian Peters.

1 Introduction

"People everywhere confuse what they read in the newspaper with news."

A. J. Liebling

While most event studies are limited to *one particular* type of event (e.g., earnings announcements, acquisitions, or initial public offerings), we investigate the market impact of *a comprehensive set* of company-specific news events in order to discern genuine "news" that moves the market from insignificant "noise" without market reaction. According to the Efficient Market Hypothesis (EMH), stock prices are supposed to reflect all available information and react quickly to company-specific news (Fama, 1970; Fama, 1991). However, a growing body of research – especially event studies – suggests that financial markets do not always comply with the EMH (Malkiel, 2003). Event studies, which investigate the development of stock prices before, on, or after the day defined as the event day (usually defined as the official release of new information), have produced partly contradictory results identifying both underreaction (e.g., Ikenberry & Ramnath, 2002; Michaely, Thaler, & Womack, 1995) and overreaction (e.g., Agrawal, Jaffe, & Mandelker, 1992; De Bondt & Thaler, 1985; Ritter, 1991) of the market to news. In addition, further empirical findings include market reactions prior to an event day, usually referred to as information leakage, as well as sustained market reactions after the event day, known as momentum or drift (for an overview, see Kothari & Warner, 2007). As these studies illustrate, previous empirical research on the market reaction to company-specific news events is inconsistent.

Investigating reasons for this inconsistency, we find that previous event studies are subject to one ore more of the following limitations. First, most of them use the business press to identify stock-related events among company-specific news, which (a) makes timing of events difficult, (b) may not represent what individual investors find important, and (c) limits the analysis largely to sporadic, extra-ordinary events (e.g., Ryan & Taffler, 2004; Antweiler & Frank, 2006). For example, events reported on in a newspaper article may have occurred on the same day or before and are published because a professional journalist deemed the story to be newsworthy. These limitations of previous research do not allow us to accurately investigate the link between news in the eyes of market participants who ultimately determine market prices and the majority of ordinary, daily news-events. Second, most event studies do not distinguish systematically between good and bad news sentiment despite the fact that the market reaction to these very different news signals may vary substantially and not controlling for news sentiment may confound results.[58] Third, almost all event studies are limited to the analysis of one particular type of event, such as stock splits (Ikenberry & Ramnath, 2002), dividend initiations and omissions (Michaely et al., 1995), acquisitions (Agrawal et al., 1992) or initial public offerings (Ritter, 1991) preventing us from comparing their relative market impact. As a result and in the context of event studies, Ikenberry and Ramnath (2002) have pointed out that researchers "become errantly excited over spurious results" (p. 490) because only events which are more likely to produce significant results, such as earnings announcements or acquisitions, are explored and published more often. In addition, previous studies do not permit a comparison of the relative market reaction of various types of news events due to different data sources and sample definitions. In line with other researchers (e.g., Ryan & Taffler, 2004; Antweiler & Frank, 2006), Thompson, Olsen, and Dietrich (1987), who have compiled an exhaustive enumeration of the types of firm-specific news reported in the Wall Street Journal (e.g., news items related to earnings announcements, management, and labor issues), come to the conclusion that "additional research is needed to determine which news release subcategories […] are associated with the largest contemporaneous security price changes" (p. 268). Finally, we are not aware of any event study that controls for differences across industries. Thus, we currently do not know whether the same event causes different reactions across industries, which would challenge results that do not control for industry affiliation.

We address these research gaps and limitations by exploring data from a real-time online stock forum. We identify news events from more than 400,000 stock-related messages published over the 6 month period between January 1st and June 30th, 2010, on the microblogging platform Twitter and investigate the market impact of different types of company-specific news events[59] (related to *Corporate Governance, Financial Issues, Operations, Restructuring Issues, Legal Issues, and Technical Trading*) on S&P 500 stock prices. We leverage computational linguistics methods in order to distinguish between good

[58] Good (bad) news refers to news items that positively (negatively) affect a company's prospects. We use the term bullish (bearish) interchangeably.
[59] The event space follows the categories used by previous research (Morse, 1982; Dewally, 2003; Ryan & Taffler, 2004; Antweiler & Frank, 2006). For details, see the chapter dedicated to "Event types used in this study and classification of our data set" in our methodology section.

and bad news and thus control for the sentiment (i.e., the positive vs. negative tone) of different news stories.[60]

Our study will answer the following four research questions. First, we explore whether the information published in an online stock forum can be used to detect which types of stock-related news affect a company on a particular day. Second, we systematically investigate whether and to what extent the distinction between good and bad news (i.e., its sentiment or bullishness) matters in the context of an event study with respect to the absolute value of returns. Third, we examine to what extent the market reaction in terms of returns differs between various types of news events. Finally, we explore whether the news coverage in terms of the relative share of different event types varies by industry group. Significant differences would indicate that the importance that investors attribute to an event type depends on the industry and leads us to explore whether the market reaction to various event types differs across industry groups.

Our results show that the information in an online stock forum can be used to detect which types of stock-related news affect a company on a particular day (e.g., accurately identify earnings announcement dates). We find that event studies should control for sentiment and distinguish positive and negative news items. The absolute value of cumulative returns prior to a news event are more pronounced for positive news than negative news, suggesting more widespread information leakage before good news. Our results show that the market reaction differs substantially across various types of news events, suggesting that there are certain event types to which investors attribute greater importance (“news”) and other events, which rarely contain new information that moves the market (“noise”). In addition, a cross-industry comparison indicates that industry classification may partially explain the market reaction to the same event type. Our finding that stock market reactions differ across industries suggests that future event studies should control for these effects. The relative assessment of various news types across industry groups hold intriguing implications for practitioners. For example, investor relations departments of publicly traded companies can adjust their communication strategy accordingly and fund manager may adjust industry weightings in their portfolios with respect to current news events.

[60] Obviously good (bad) news may not always contain positive (negative) sentiment. However, in line with our above-mentioned definition of good (bad) news as positively (negatively) affecting a company’s prospects (i.e., its stock price), Sprenger and Welpe (2010) show that this assumption generally holds in large datasets with a sufficient number of news items.

The remainder of the paper is structured as follows. First, we review related work and derive our research questions. Second, we describe our data set and methodology. Third, we provide results illustrating the market reaction to a comprehensive set of news events. Finally, we conclude that the market reaction differs across multiple event types and that sentiment is a crucial factor in evaluating news events. We discuss the implications of our findings and provide suggestions for further research.

2 Related work and research questions

2.1 News as a source to identify event days

The majority of the event study literature is concerned with price reactions to stock-related news (e.g., Mitchell & Mulherin, 1994; Huberman & Regev, 2001; Chan, 2003; Barber & Odean, 2008). News is regarded as the release of new information to the market and, in most studies, defined literally as the publication of a news story regarding a particular stock (Schmitz, 2007). It is important to note that this definition includes so-called attention events (Barber & Odean, 2008), i.e., events that do not necessarily carry new information, but may simply draw attention to a particular stock (e.g., stock splits). In addition, there are some events for which it is unclear ex ante, whether they will reveal new information or just create attention (e.g., earnings announcements) because current prices are based on prevalent expectations (e.g., analyst recommendations) and only the arrival of information that leads to a revision of those expectations constitutes news (e.g., unexpected earnings figures). However, Barber and Odean (2008) have shown that individual investors are more likely to buy attention-grabbing stocks (i.e., "stock in the news", p. 785) irrespective of the arrival of new information. Engelberg and Parsons (2011) support the notion of a causal impact of media in financial markets. In another drastic example, Huberman and Regev (2001) found a strong market impact to the republication of information in popular newspapers that had been published in a public journal read by specialists a full 5 months before. Tetlock (2011) extends these findings by illustrating that especially individual investors overreact to stale news. These examples illustrate that the publication of news stories can represent news and thus serve as a legitimate source of event days (i.e., news events).

Within the event study literature, there are only a few studies that use comprehensive data sets to explore the market impact of stock-related news stories. For example, in order to explore the effect of news on drift, Chan (2003) used news stories archived in the Dow Jones

Interactive Publication Library and focused on the differences between returns after major news stories about a company on the one hand and returns after large price movements in the absence of news on the other. The author finds that stocks with news, in particular bad news, exhibit momentum or drift of up to 12 months. However, event days were defined based on the presence or absence of one or more news headlines in major publications, which provides no indication with respect to the intensity and salience of news coverage (Barber & Odean, 2008). Addressing this limitation, Mitchell and Mulherin (1994) have compared the number of news announcements (i.e., the news volume) reported by Dow Jones & Company to the market activity and found a weak positive relationship between the news volume and both the trading volume and the absolute value of firm-specific returns. Fang and Peress (2009) support the notion that the breadth of information dissemination affects stock returns, but find empirical evidence in contrast to Mitchell and Mulherin (1994) suggesting that stocks with no media-coverage earn higher returns. Mitchell and Mulherin (1994) refer to the news volume as a "measure of information" (p. 923). We argue that this is a rather optimistic definition because the mere number of news stories fails to capture many nuances of the information content such as the sentiment or importance of any particular news story.[61] In sum, event studies that comprehensively explore company-specific media coverage are largely limited to the number of news items and do not take into account news sentiment or topics.

2.2 Limitations of the business press as a source of news

Most event studies (e.g., Mitchell & Mulherin, 1994; Chan, 2003; Antweiler & Frank, 2006) use professionally edited news content such as the Dow Jones News Service or the Wall Street Journal. We argue that there are three concerns with respect to these data sources. First, professional news agencies do not necessarily reflect what investors find important (Antweiler & Frank, 2006).[62] However, ultimately market participants and their perception of what constitutes news determine market prices. Second, determining the information release and thus the event day from a newspaper publication is difficult. Most studies make a somewhat

[61] To address this deficiency, Mitchell and Mulherin (1994) have experimented with a proxy for news importance considering a news day to be important when it coincides with the announcement of one of 17 monthly macroeconomic indicators (e.g., employment, new home sales). However, similar to the classification of news according to the market reaction, this is potentially prone to bias through endogeneity.

[62] Much of the information compiled by the Dow Jones New Service originates from publicly traded companies because the major stock exchanges require their members to provide all material information to Dow Jones (for details see Thompson et al., 1987). Until recently, the Wall Street Journal was owned by Dow Jones and 44.3% of all stories were transmitted across the Broadtape and then reported in the Wall Street Journal (Mitchell & Mulherin, 1994), illustrating the large overlap of these outlets.

arbitrary decision by assigning event days either to the day before or the day of the publication (e.g., Antweiler & Frank, 2006; Morse, 1982). Irrespective of the design, some news items will thus almost certainly be assigned to the wrong day. Third, most companies are mentioned in newspapers only sporadically. However, we argue that news of all kinds and quality are generated continuously and multiple, often contrasting, signals have to be processed by investors on any given day and for any given company. Any newspaper represents only a partial reflection of this news stream. In line with this reasoning, Roll (1988) concludes that stories from the financial press alone have little effect on returns. We suggest the use of Twitter as a more comprehensive empirical database of real-time news, which provides a constant stream of news stories for a given company and requires the reader to detect trends and prioritize signals. These signals do not only include major news stories studied by the existing literature, but also many minor news items (e.g., product refinements, new marketing campaigns, public appearances of executives, and technical trading signals). Stock prices, or more precisely investors, are not only reacting to major news stories, but respond to the arrival of new information every single day. The existing literature has so far neglected this granular level of analysis. It focuses on major news events and does, therefore, not inform us about the market reaction to minor everyday news items, which nonetheless affect investor decisions. This paper addresses these concerns by using data from an online stock forum, which reflects an investor perspective rather than a professional news perspective and provides a dense, time-stamped information stream of news signals with varying significance. Content from online stock forums has been used successfully in a few event studies. While the focus of related studies has been on the market reaction to overall message volume and sentiment (e.g., Tumarkin & Whitelaw, 2001; Dewally, 2003; Antweiler & Frank, 2004), there is also one example of an event study that focuses on a particular event type, namely the market reaction to takeover rumors on internet discussion sites (Clarkson, Joyce, & Tutticci, 2006). Whereas all of these studies have explored internet stock message boards, which became popular at the turn of the century, these forums have a serious limitation. Message boards categorize postings into separate bulletin boards for each company and have an archival nature. Message board users who do not actively enter the forum for a particular stock may not become aware of breaking news for that company. Thus, message boards may not accurately capture the natural market conversation. To avoid this limitation, we chose the microblogging platform Twitter as our data source. Stock microbloggers are usually exposed to and frequently comment on the most recent information for all stocks, which allows us to more accurately capture the investor perception of what constitutes news. The fact that stock microblogs are largely unexplored in the financial literature as a source of event days motivates our first research questions where we explore whether the investor discussion in an online stock forum meaningfully reflects real-world news events.

2.3 News sentiment

Most event studies treat news as neutral informational input and do not differentiate between buy and sell signals (i.e., the bullishness or sentiment of the news items), which is arguably the most crucial feature of its information content with respect to a particular stock (Ryan & Taffler, 2004; Storkenmaier, Wagener, & Weinhardt, 2010[63]). In some studies this information is fairly obvious or implicitly taken into consideration (e.g., when positive and negative earnings surprises are distinguished), but in many cases the distinction is not as clear (e.g., the market assessment of takeover announcements as positive or negative depends on the takeover price and strategy). The distinction of sentiment should affect most types of news and is thus relevant to many event studies, even those focused on one particular type of event. Surprisingly, the existing event study literature rarely makes this distinction. Empirical evidence in the context of macroeconomic news (e.g., unemployment or inflation rates) suggests that responses to positive and negative information are asymmetric and that negative information has a much greater impact on individuals' attitudes than does positive information (for an overview, see Soroka, 2006). Some studies of company-specific news use stock price reactions to consider news either as good or bad (e.g., Pritamani & Singal, 2001). However, this is an endogenous measure that does not assess the genuine information content of the actual news item. There are only very few studies that distinguish the ex-ante sentiment of news stories (e.g., Schmitz, 2007; Storkenmaier et al., 2010), which is problematic since bad news and good news may cancel each other out suggesting very little market impact in aggregate (Ryan & Taffler, 2004), for example when positive and negative earnings announcement are not considered separately. Empirical evidence suggests that the news sentiment of traditional media sources, such as the Wall Street Journal, has an effect on market reactions (Tetlock, 2007). Riordan, Storkenmaier, and Wagener (2010) and Storkenmaier et al. (2010) illustrate that market liquidity decreases around negative newswire messages, but they do not analyze the effect on returns. In a study of roughly 300,000 news

[63] As Walter Wriston, long-time CEO of Citicorp, pointed out: "Markets function only through the transmission of information – both good and bad." (p. 2)

items coded manually by 240 analysts of a research institute, Schmitz (2007) found only a short-term post-event price drift of a few days after good news, but a drift of several days after bad news. While these studies take news sentiment into account, they are based on the business press and do not investigate whether and to what extent the distinction between good and bad news matters in the context of an event study based on the investor perception of news.

2.4 Different types of news events

The vast majority of the existing event study literature is concerned with the market impact of one specific type of event (e.g., earnings announcements) and addresses the debate over how fast information is incorporated into prices. Most event studies fail to evaluate the relative impact of different event types simultaneously (e.g., news related to corporate governance, operations, or legal issues), which prevents us from comparing the relative market reaction to various types of news events. This comparison would allow us to determine the relative importance that investors attribute to various news events. Only a few studies exist, which are dedicated to the simultaneous analysis of multiple events.[64] Among these, one can distinguish between exploratory studies investigating the effect of major "world events" on market indices (e.g., Cutler, Poterba, & Summers, 1998; Niederhoffer, 1971) and studies focusing on firm-specific news (e.g., Morse, 1982; Antweiler & Frank, 2006). As far as world events (e.g., "increase in hot tension", "Soviet discovery", "peace meeting", "election", or "change of foreign leader") are concerned, Cutler et al. (1998) cast doubt on the view that "qualitative news" (p. 4) can explain company returns. In a related study of 432 world events, Niederhoffer (1971) supports this finding, but points out that limiting the analysis of market reactions to stock market indices "may conceal pronounced and possibly divergent effects on particular companies or industries" (p. 204). As far as firm-specific news are concerned, there are very few studies covering multiple types of news events. Ryan and Taffler (2004) have filtered major stock price and trading volume movements for roughly 250 UK stocks and matched those movements with 32 news categories (e.g., analyst recommendations, director share dealings, or financing issues) through an exploration of significant information events

[64] In addition to the studies detailed in this section, there are a few studies that select their sample of news stories from a specified set of events, but limit the analysis to the mere news volume due a limited sample size (e.g., Brookfield & Morris, 1992).

covered in the financial press.[65] They are able to match two-thirds of all major price changes with a news story and conclude that “65% of significant price changes […] can be explained by readily available public domain information” (p. 51), in particular by sell-side analyst reports (17.4%), preliminary results (8.6%), and director share dealings (8.5%). Yet, there are two major limitations of this study. First, the endogenous selection of event days which “allows [the authors] to focus on economically significant events” (p. 50) is one of the more drastic examples of a selection bias in line with the event study criticism of Fama (1998). Second, the fact that some news story was matched ex post to a trading day with major stock returns establishes a questionable causal link at best.[66]

The two studies most closely related to ours are Morse (1982) and Antweiler and Frank (2006). For 50 publicly traded companies, Morse (1982) examined price behavior and trading volume movements over a 10-day event-window for a selection of 9 types of company announcements.[67] In the 3-year sample period, quarterly earnings reports appeared to have the most significant and sustained effect on market prices, next to dividend increases and product sales. While the announcements of acquisitions (0.38%) and product sales (0.19%) only resulted in slight adjustments in prices, labor strikes had no discernible effect on the stock market. However, the study of Morse (1982) is limited to reports of official company announcements. Thus, except for more than 600 quarterly earnings reports, none of the event different types generate more than 75 observations resulting in a sample size that is too small to derive robust conclusions.[68] This leaves us with one single study, which explores the market reaction to a comprehensive set of firm-specific news events: Using computational linguistic methods Antweiler and Frank (2006) classified over 250,000 Wall Street Journal corporate news stories from 1973 to 2001 according to topic and ran an event study for the 48 event types with a large number of observations (e.g., earnings forecasts, stock splits, lawsuits, and new product releases). The results challenge the notion that stock prices reflect news immediately. Consistent with the existing event study literature, the authors find short term momentum for many days after the publication of a news story and a longer-term reversal for many events. Even though this study provides the first comprehensive analysis of

[65] They use the London Stock Exchange Regulatory News Service, The Financial Times, and McCarthy Information.

[66] The authors’ use of quotation marks in their summary is very telling with respect to this issue (“price movements and trading volumes are ‘explained’ by publicly available information”, p. 55).

[67] These were an increase in dividends, a sale of a product, an unfavorable (favorable) earnings forecast by a company official, an acquisition, a construction or building project, a stock split, a labor strike, and quarterly earnings.

[68] Morse (1982) notes that some effects “may be attributable to the small sample size” (p. 75).

multiple event types, it also has some limitations.[69] First, the authors do not systematically distinguish between good and bad news.[70] However, as stated above, we argue that sentiment (or the related concept of bullishness) is the most crucial piece of information that financial news boil down to. For example, Antweiler and Frank's (2006) category "Lawsuit End" confounds very different events with respect to a particular stock depending on the outcome (i.e., the sentiment the article carries). Second, Antweiler and Frank (2006) use various event windows between 5 and 40 days after the event day to compute cumulative returns. There are likely confounding effects affecting these long event windows.[71] Third, comparable to related studies based on fairly comprehensive databases (e.g., Mitchell & Mulherin, 1994; Chan, 2003) Antweiler and Frank (2006) use professionally edited news content, which not necessarily reflects what investors find important. Antweiler and Frank (2006) themselves note that "the Wall Street Journal presumably chooses what to report in an effort to make money" (p. 10). Fourth, with results for more than 40 different events, Antweiler and Frank (2006) limit their analysis to overall patterns (such as overreaction), but did not systematically explore differences between event types. Therefore, we examine to what extent the market reaction in terms of returns differs between various types of news events.

2.5 Industry effects

There are two aspects to the systematic analysis of market reactions to company-specific news: the categorization of market reactions (into different types of news) on one hand and of sample companies on the other. In the previous section we have discussed the categorization of news events by type or topic. In addition, one can also structure the sample companies. While some event studies use endogenous measures, such as return deciles, to distinguish firms in the sample (e.g., Chan, 2003), we are not aware of any event study that distinguishes between one of the most obvious categories for publicly traded companies, the industry classification. Surprisingly, a comprehensive overview of the use of industry classifications in financial research finds that only very few studies "use industrial classification to determine

[69] Next to sample size, these limitations also affect the study conducted by Morse (1982).

[70] Interestingly, the authors note that "obviously some kinds of news are good news, while other types of news are bad news. There is no apparent reason that [there should be] an equal number of good news stories and bad news stories" (p. 10). However, as (Schmitz, 2007) has pointed out, Antweiler and Frank (2006) do not address this issue systematically. Only a few of the specified event types include an implicit evaluation of sentiment (e.g., "Earnings Forecast Down" and "Earnings Forecast Up").

[71] Although the authors provide some results controlling for overlapping news events, these events are limited to the original dataset of Wall Street Journal articles. Obviously, many other events may have affected the stock prices, which are not reported in the newspaper.

the extent to which industrial structure explains the cross-sectional dispersion of financial characteristics" (Kahle & Walkling, 1996, p. 311). While the corporate finance literature illustrates the importance of industries in explaining IPO valuations, M&A activity, and leverage (e.g., Bradley, Jarrell, & Kim, 1984) suggesting that industry classification may help explain financial characteristics, the empirical evidence with respect to returns is sparse. Previous literature has found relatively little impact of industries on stock prices (for an overview, see Moskowitz and Grinblatt (1999), who represent a notable exception in finding that industries may explain long-term momentum[72]). However, none of these previous studies accounts for firm-specific news affecting the stock price. We argue that controlling for this vital piece of information may add significantly to the explanatory power of industry categorizations for returns. Empirical evidence of intra-industry information transfer (e.g., a market reaction to news affecting related firms as illustrated by Baginski (1987) and Ramnath (2002)) supports the idea that investors take an industry perspective when assessing the impact of individual news items. This leads us to explore whether the market reaction to various event types differs across industry groups.

3 Data set and methodology

In this section, we describe our data set and detail the methodology used to derive the variables for this study. This includes the automated content analysis of our news messages, which leverages computational linguistics methods, as well as the definition of financial variables and choice of parameters for our event study. The text analytical methods, which are used to classify news messages, deserve special attention and are illustrated in more detail.

3.1 Data set and sample selection

We use stock-related messages from the microblogging platform Twitter as our data source because it provides us with a dense information stream of stock-related news items published by individual investors. Twitter allows users to post short messages with up to 140 characters[73], so-called "tweets".[74] These tweets appear on a public message board of the website or on third-party applications. Users can subscribe to (i.e., "follow") a selection of

[72] In contrast to our short-horizon event study, Moskowitz and Grinblatt (1999) study long-term return developments with a focus on a 12 month time frame.
[73] Note, that the brevity forces users to write succinct messages and makes the underlying text input comparable to that of existing studies which focus on news headlines because they are so condensed.
[74] We will refer to these tweets throughout the paper as messages or news items.

favorite authors or search for messages containing a specific key word (e.g., a stock symbol). The public message board has become an extensive information stream of currently more than 155 million messages per day (TechCrunch, 2011). Many of these messages are dedicated to the discussion of public companies, trading ideas and current news stories. Some commentators have even called this platform "the modern version of traders shouting in the pits" (BusinessWeek, 2009) and news stories claim that financial microblogs capture the market conversation and suggest that "Twitter-based input [is] as important as any other data to the stock" (TIME, 2009). The investor community has come to call Twitter and related third-party applications, which filter stock-related microblogs, "a Bloomberg for the average guy" (BusinessWeek, 2009).

While there are few restrictions with respect to the format of messages (e.g., posts are confined to a maximum of 140 characters), 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-related messages. Messages are accessible via the website's application programming interface (API). We study the 6 month period between January 1st and June 30th, 2010, to deal with stable developments on the U.S. financial markets and to avoid potentially distorting repercussions of the subprime mortgage crisis in 2009. During this period, we have collected 439,960 English-language, stock-related microblogging messages[75] containing the dollar-tagged ticker symbol of an S&P 500 company.[76] We focus on the S&P 500 to adequately reflect a wide spectrum of U.S. equities, which permits a cross-industry analysis, while limiting our study to well-known companies that trigger a substantial number of tweets.[77] Ranging from 845 to 7,729 daily postings, this represents an average of 3,548 tweets per

[75] (Mitchell & Mulherin, 1994) illustrate that their "primary contribution […] is that we employ a distinctive proxy for information – the number of accouncements released daily by Dow Jones & Company [which] is more comprehensive than most measures used in prior studies." (p. 923). Their study is based on roughly 750,000 story headlines over the course of 7 years. For our 6 month sample period, our data set is of approximately the same order illustrating the density of the information stream we are investigating.

[76] 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.

[77] Specifically, we focus on those companies that have been included in the S&P 500 as of January 1, 2010.

trading day with a standard deviation of 1,300 messages. We observe an average of more than 8 tweets per day and company with a maximum of 1,543[78]. Our message data contains news for 64% of all company-day-combinations. Apple (59,158 messages), Google (30,945), and Goldman Sachs (19,785) were the companies mentioned most frequently. As far as the distribution of messages throughout the day is concerned, we observe a significant spike in message volume before the markets open and the majority of tweets are posted during the trading hours between 9:30 am and 4:00 pm.

3.2 Analysis of news

3.2.1 Naïve Bayesian text classification

For the purpose of our study, we have to extract both the sentiment (i.e., bullish or bearish) and the event type or news category (e.g., news related to corporate governance, operations, or legal issues) from the text messages. We chose to classify messages automatically using well established methods from computational linguistics.[79] In line with Antweiler and Frank (2006), we use the 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 is calculated with the conditional probability of its words occurring in a document of this class. These conditional probabilities are estimated based on a manually coded training set of 2,500 tweets, which we classified according to sentiment on one hand and event type on the other.[80] Compared to more advanced methods in computational linguistics, this method is relatively simple (e.g., easily replicable and subject to few arbitrary fine-tuning parameters), but has consistently shown robust results. We use the multinomial Naïve Bayesian implementation of the Weka machine learning package (Hall et al., 2009).[81]

[78] This spike of messages related to Apple Inc. occurred on April 5th, when the company announced that it sold more than 300,000 iPads on the first day.

[79] In the context of manually coding messages as either good or bad, Niederhoffer (1971) already noted “it would have been possible to perform a completely objective coding by programming definitions and procedures for a computer” (p. 199). We have taken this more objective approach. Note also that most newspaper-based event studies use key words (e.g., “merger”) to search through the news archive and filter news items before the manually identification/classification of events. In a sense, their methodology follows a very crude semi-automated text analytical approach.

[80] In line with most text classification methods using a manual training set (e.g., Antweiler & 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).

[81] See our supplementary appendix for a detailed description of our classification method and results.

3.2.2 Event types used in this study and classification of our data set

In this section, we introduce the set of event types extracted from our data set and illustrate the classification results. Related studies have defined between 9 (Morse, 1982) and 67 different event types (Antweiler & Frank, 2006). We have followed the general approach used in Antweiler and Frank (2006) and used the manual classification of the training set to define an event space that is both appropriate for our dataset and in line with the categories used by previous research (Morse, 1982; Dewally, 2003; Ryan & Taffler, 2004; Antweiler & Frank, 2006). There always remains a subjective component in event definition and other categories may be equally appropriate in many cases (Niederhoffer, 1971). However, in contrast to related studies, we have deliberately defined a limited number of aggregated event types. There are four reasons for this focus. First, any meaningful comparison between events requires a limited set of combinations – otherwise, the analysis becomes limited to overall patterns (as in Antweiler & Frank, 2006). Second, for automated text classification it is important that different categories be semantically different (i.e., contain typical, category-specific words). If one attempts to distinguish too many similar categories, classification performance suffers.[82] Third, a comparison of categories used in previous studies shows that, even though events may differ on the most detailed level, there is high agreement on the level of broader categories.[83] This aggregated level is the focus of our analysis. Fourth, our distinct classification of tweets according to sentiment and event type allows us to combine these two features to derive event types, which were defined separately in other studies (e.g., Earnings Forecast Up and Earnings Forecast Down). Following this logic, the event categories we use in this study are related to news about *Corporate Governance, Financial Issues, Operations, Restructuring Issues, Legal Issues, and Technical Trading*. In addition, to allow for more granular analysis, we have defined subcategories for which we found a substantial number of messages among our data set (e.g., the two subcategories stock-related and market-related

[82] This is all the more true for relatively simple classifiers such as the Naïve Bayesian. It is interesting to note that Antweiler and Frank (2006) do not report classification results for the accuracy of their classifier – neither in the paper nor in a 40-page appendix detailing robustness checks supporting their financial analysis. Given the use of 67 different event types with nuanced differences we would surmise a relatively a high error rate (e.g., “Product New” and “Product New Possible”, 5 different merger-related events such as “Merger Announce” and “Merger Complete”, or 4 different lawsuit-related events such as “Lawsuit End” and “Lawsuit Ongoing”).

[83] Dewally (2003), for example, structures 28 reasons for posting recommendations on internet message boards along the following categories: Change in Corporate Structure, Operations, Financials, Market (i.e., Technical Trading Signals), and Other. Due to a small sample size, the author did not investigate the market reaction to these categories. Antweiler and Frank (2004) assign their 67 events to: Corporate Governance, Earnings Reports, Financial Issues, General Issues, Legal Issues, Operational Issues, and Restructuring Issues.

technical trading signals as event details for the event category *Technical Trading*[84]). We refer to these subcategories as event details and use the term event type more generally for both event categories and event details. The classification of our data set took place at the level of event details.[85] Messages were then automatically assigned to the broader event category. Table 1 provides an overview of the events and also includes sample messages from our training set assigned to the respective class. Technical trading signals are the most frequently mentioned event category with references in roughly one third of all messages (34.0%), among which stock-related signals make up the vast majority (28.2%). Next are comments regarding company *Operations* (20.3%), especially *Prroduct Development* (9.4%), *Operational Performance* (4.2%) and *Marketing* (4.2%). *Financial Issues* come third as an event category (13.3%) with a majority of messages dedicated to the discussion of earnings results (6.9%). *Restructuring Issues* (6.1%), *Legal Issues* (3.5%) and topics related to *Corporate Governance* (3.3%) are mentioned less frequently. The distribution of topics is roughly in line with findings for internet message boards (Dewally, 2003), but shows some interesting differences to professionally edited newspapers. Among the 48 different event types reported by Antweiler and Frank (2006), none captures technical trading signals. Articles related to product development make up less than 1% in the Wall Street Journal. Whereas contracts are discussed less frequently by stock microbloggers, they represent more than 10% of all newspaper articles. These results support our view that the perception of individual investors with respect to what matters to a stock is different than that of a professionally edited, for-profit newspaper. It provides evidence supporting our motivation to study this largely unexplored database as a source of news events.

[84] Technical trading signals are slightly different than other news types. As pointed out earlier, news is usually regarded as the release of new information and, in most studies, defined literally as the publication of a news story. Huberman and Regev (2001) have shown that the publication of a news story can represent news in itself. Technical trading signals are sometimes discussed in newspapers, just like other news types, but more often traders are aware of these news from other sources (Barber & Odean, 2008). The distinct analysis of technical trading signals is also motivated by theories suggesting that investors underreact to qualitative news and overreact to pure price movements (i.e., technical trading signals; Hong & Stein, 1999).

[85] In addition to the 16 event types, there is one more class for messages that could not be assigned to any one of these ("not classified"). Similar to related studies (Antweiler & Frank, 2006) this class makes up roughly 20% of the training set.

Table 1: Event categorization and sample messages (training set)

Event category	Share	Event detail	Share	Sample message
Corporate Governance	**3.3%**	CEO	1.4%	$C Citigroup CEO Pandit earns $128,000 in 2009 pay (AP) http://url4.eu/1UPxD
		Other Executive	1.9%	$GME CFO leaving to go to $WMT. Shares drop
Financial Issues	**13.3%**	Earnings	6.9%	Heinz Q3 EPS of 83c beats by 6c. Revenue of $2.6B meets. $HNZ #earnings http://bit.ly/avlHFH
		Analyst Rating	2.7%	BofA/Merrill Lynch upgraded Dell $DELL from Neutral to Buy and raised their price target from $16.50 to $18
		Financial Other	3.7%	$C Kevin Goldstein-Jackson: Picking up the private-equity pieces http://url4.eu/1RNL0
Operations	**20.3%**	Labor Issues	1.3%	Verizon to cut 13,000 more jobs this year $VZ http://tinyurl.com/yk7kkys
		Product Development	9.4%	Crazy Google now building super-high-speed fiber Internet network to scare Comcast and AT&T http://bit.ly/dvWSzL $GOOG
		Operational Performance	4.2%	Reading: Smartphone Sales Up 24% in 2009, iPhone Share Nearly Doubles http://seekingalpha.com/a/42q7 $AAPL
		Marketing	4.2%	$GS can definitely do better at PR and giving green journos some quality info. this is preposterous http://ow.ly/1bVw2
		Contract	1.3%	http://bit.ly/9LWjhS $XOM Praxair Awarded ExxonMobil Contract for Enhanced Oil Recovery Project
Restructuring Issues	**6.1%**	Joint Venture	1.0%	Skype and Verizon Wireless partnering up $VZ - http://bit.ly/cVfhUX
		M&A	5.1%	Schlumberger $SLB will acquire rival Smith International $SII in an all-stock deal valued at $11 billion. this was rumored on Friday.
Legal Issues	**3.5%**	Jurisdiction	1.0%	Apple Sues Phone Maker of Google Phone Over Patents - http://nyti.ms/aZ5IRO $goog $aapl
		Government Authorities	2.5%	$BAC Bankers at Davos told more regulation on the way http://www.financial24.org/story/1169384/
Technical Trading	**34.0%**	Stock-Related	28.2%	$AEP (American Electric Power) $34.20 crossed its 1st Pivot Point Resistance #emppv #stocks http://empirasign.com/s/40a
		Market-Related	2.3%	Financials Find Support! Financial sector jumped to a 0.7% gain, better than any other sector-$JPM primary leader... http://bit.ly/226smu
Not classified			22.9%	Three ways Apple shares could skyrocket $AAPL http://cot.ag/9mVuLq

Notes: Randomly selected tweets are shown in their original format (before preprocessing). Share refers to the share of messages assigned to a particular event type in the training set.

Overall in-sample classification accuracy with respect to event types was 80.8%. Even a more conservative 10-fold cross validation of the model within the training set correctly classifies 61.1% of all messages. These results are similar to related studies that applied Naïve Bayesian learning algorithms to financial text samples in order to separate buy, hold and sell signals (e.g., Koppel & Shtrimberg, 2006). However, the fact that our algorithm had to distinguish 16 different event details, illustrates that this classifier shows a very high accuracy. The results demonstrate that the event classes we chose are semantically distinct. The accuracy by class further validates the use of automatically labeled messages. True positives are higher than about 40% for all classes, except for the identification of message related to *Joint Ventures* (18.2%). More importantly, false positives are below 6% for all classes, except “not classified”. For our purposes, falsely assigning a news item to the class “not classified” is more acceptable than falsely interpreting messages as the wrong event type. This worst type of misclassification occurs rarely.

A look at the most common words per class (Table 2) indicates that the information gain model derived a plausible dictionary from our training set. The most common words reasonably reflect the linguistic profile of the 16 classes. *Corporate Governance* news, for instance, contain words referring to job titles (e.g., “CEO”, “executive”), payment (e.g., “million”, “bonus”), and specific people (such as “Steve” Jobs and Jamie “Dimon”). Next to the obvious financial speak, *Earnings* related messages contain a large share of references to specific price targets (i.e., dollar values or cent values). Messages about *Product Development* are frequently identified by specific products (e.g., “ipad”, “windows”). *Marketing*-related news is associated with positive emotions, whereas firm-specific legal issues (i.e., *Jurisdiction*) are associated primarily with negative emotions. The most typical words for *Government*-related *Legal Issues* include many federal agencies (e.g., “FED [federal reserve, central bank]”, “FDA” [federal department of agriculture], “EPA” [environmental protection agency]) and are related to “rate” changes or product “approval”. *Technical Trading* terms plausibly reflect news in this field (e.g., “cross”, “pivot”, “moving average”, “support”). In sum, the dictionary derived from our training set accurately distinguishes the semantic profile of the specified event types.

Table 2: Classification results – most common words/features per class

Corporate Governance		Financial Issues			Operations		
CEO	***Other Executive***	***Earnings***	***Analyst Rating***	***Financial Other***	***Labor Issues***	***Product Dev.***	***Ops Performance***
ceo	bonus	earnings	analyst	dividend	cut	nvidia	sales
steve	execute	report	upgrade	cash	worker	ipad	[percentage]
dimon	executive	beat	cramer	buffett	labor	launch	share
bonus	award	[number]	target	share	employee	graphic	retail
screw	boss	eps	rate	stake	walmart	iphone	market
[mention]	million	#earn	conviction	fund	[number]	app	growth
sun	board	[dollarvalue]	market	hedge	depot	phone	chart
million	live	[centvalue]	downgrade	money	create	buzz	lose
[posemo]	retire	quarter	research	investor	consolidate	technology	demand
money	ex	[percentage]	cut	asset	hire	windows	profit

Operations (cont'd)		Restructuring Issue		Legal Issue		Technical Trading	
Marketing	***Contract***	***Joint Venture***	***M&A***	***Jurisdiction***	***Gov't Authorities***	***Stock-Related***	***Market-Related***
[posemo]	deal	partnership	buy	patent	rate	[dollarvalue]	stock
ad	annual	team	acquire	lawsuit	bank	cross	sector
bowl	bid	mobile	[dollarvalue]	against	approve	[number]	weak
super	million	group	billion	[negemo]	fed	pivot	financial
advertise	#deal	extend	bid	sue	fda	point	bank
user	receive	[posemo]	deal	decision	bernanke	moving average	market
store	agreement	agreement	rumor	fight	reform	resist	tech
customer	contract	force	million	approve	drug	high	airline
free	sign	exclusive	acquisition	action	congress	support	solar
commercial	approve	supplier	buyout	settlement	epa	volume	gain

Notes: This table shows 10 of the 20 most common words per class (words/features common to all classes, such as [URL] or [ticker] were not included). Words in brackets are tokens generated by text preprocessing.

The method for the classification of message sentiment (i.e., buy, hold, sell) followed that of event types. In the training set, roughly half of all 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%).[86] The distribution of sentiment provides further evidence that the news perception of individual investors differs from that of professionally edited newspapers, for which studies find a significant bias to report more on negative than on positive events in line with the saying “bad news is good news” (Soroka, 2006; Riordan et al., 2010). Overall in-sample classification accuracy with respect to message sentiment was 81.2%. The more conservative 10-fold cross validation of the model correctly classifies 64.2% of all messages. We refer to the related study of the information content of stock microblogs for details (Sprenger & Welpe, 2010).

There are days for some stocks without any tweets. However, our data set contains a full set of both tweet and market features for more than 64% of roughly 61,000 company-day-combinations illustrating the density of the news stream. Finally, because we use financial data from the NASDAQ and NYSE, we align messages with U.S. trading hours (9:30 am to 4:00 pm) by assigning messages posted after 4:00 pm to the next trading day. Thus, messages posted after the markets close are included in the calculation of tweet features for the following day because these news items cannot have an effect on the market until the next day.

3.2.3 Detection of news event dates

In this section, we will describe the aggregation of messages to the level of daily information and the identification of the relevant event type for a particular day. The number of news items is too high to treat each message as a separate event. In addition, with hundreds of daily messages for some companies and daily market data, we would not be able to distinguish the effects for individual messages. Related studies of news volume have identified news events as days with unusually high media coverage (e.g., Schmitz, 2007). However, we want to identify the most relevant event type from our message stream. Therefore, the identification of events follows two steps. First, for every company and day we compute the share of messages for each event type. This provides us with a daily profile of what type of news investors are

[86] 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.

talking about for a particular company. Second, we assign a day to the event type that generates an unusually high share of traffic. We follow Tumarkin and Whitelaw (2001), who have defined an unusually high number of messages relative to the 5-day standard deviation of the share of those messages on a stock message board. Whereas Tumarkin and Whitelaw (2001) want to identify the most bullish days only, we are interested in the most relevant event type for every single day, which we take to be the extraordinary news type that moves the stock price. Thus, we assign a particular company-day combination to that event type with the highest positive 5-day standard deviation of the share of messages.

A particular news event is classified as either bullish or bearish depending on the share of bullish messages for that particular day. We follow Antweiler and Frank (2004) and consider only messages that were classified as either buy or sell signals. Due to the excess amount of bullish messages, we use the median share of bullish messages (0.66) as the cut-off point between bullish and bearish days.

3.3 Financial market data

We have obtained financial data in daily intervals for the S&P 500 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

$$(8)\qquad AR_{it} = R_{it} - E(R_{it})\,,$$

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

$$(9)\qquad AR_{it}^{simple} = R_{it} - R_{t}^{market}$$

with the S&P 500 index serving as our 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

$$(10)\qquad E(R_{it}) = \alpha_i + \beta_i(R_{mt}) + \mu_{it} \quad for\ t = 1, 2, ..., T\,,$$

where α_i is the intercept term, β_i is the slope of the coefficient, 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, & 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

(11) $CAR_{it} = \sum AR_{it}$

and average cumulative abnormal returns for *N* companies are calculated as

(12) $ACAR_t = \frac{\sum_{i=1}^{N} CAR_{it}}{N}$.

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, following Antweiler and Frank (2006). In line with Morse (1982) and Antweiler and Frank (2006), average abnormal volumes (AAV) are calculated using the same market model described for returns, but using logarithms of stock and market trading volumes instead of returns.[87]

3.3.1 Event study methodology

Every event study needs to define the relevant event window. We focus on a short event window of three to five days before and after the events for three reasons. First and foremost, short-horizon methods are quite reliable, whereas long-horizon methods have serious limitations (Kothari & Warner, 2007). Antweiler and Frank (2006), for example, have shown that the choice of the event window (i.e., the range of dates around the event day included in the study) can have a considerable effect on the results and even lead to contrary conclusions. As a result, short-horizon event windows are typical in related studies (e.g., Morse, 1982; Ryan & Taffler, 2004). Second, given our high frequency analysis of daily data, a short event window limits the overlap and thus the distortion of effects by confounding events. Third, the effect of the large number of minor news events we study is arguably relatively small and short term.

Most event studies, even those that distinguish sentiment, find that the main price reaction occurs on the day of the arrival of the new information (Schmitz, 2007). Thus, while we look at the effect before and after the event day to investigate both leakage and drift, the focus of our study will be on the clearest signal of the market reaction, the event day itself.

---

[87] Alternatively, one can compute abnormal volume as the difference between the actual trading volume and the average daily trading volume in the 4 weeks preceding the event window (Schmitz, 2007). We have calculated all of our results with this measure. The results do not differ substantially from those we report.

3.3.2 Selection of industry groups

There are various industry classification schemes for the analysis of market reactions across industry groups (for a comprehensive overview, see Kahle & Walkling, 1996). The Standard Industrial Classification (SIC) is the predominant classification system in capital market research with more than 90% of relevant studies making use of this classification scheme (Bhojraj, Lee, & Oler, 2003).[88] There are varying levels of granularity in the SIC scheme. For the purpose of our study, we use the most common 2-digit level of analysis (e.g., Moskowitz & Grinblatt, 1999).

Our study is designed to answer the question whether differences exist in market reactions for various industry groups. We do not intend to provide a comprehensive return profile across the entire spectrum of all industry groups. For the purpose of financial research, Bhojraj et al. (2003) define an industry as functional if it contains at least 5 companies while others use an average of more than 30 companies per group (Bradley et al, 1984). We limit the analysis to industry groups with a minimum number of 5 high-volume companies with at least 600 messages in our sample (i.e., 5 per day) and a total of 20 companies. Among the 54 SIC groups at the 2-digit level, this leaves us with the following industries: Business Services, Depository Institutions, and Industrial Machinery and Equipment.

4 Results

The results section is structured as follows: First, we propose and conduct an innovative test of out-of-sample classification accuracy to determine whether the information published in an online stock forum can be used to detect which types of stock-related news affect a company on a particular day. Second we provide results illustrating the aggregate impact of new information and show why the distinction of positive and negative news is important in an online stock forum. Third, a more granular analysis explores the market reaction to all event types covered in this study. Finally, we show that the news coverage of different industries varies in terms of the relative share of several event types and investigate whether the market reaction to these event types differs across the selected industries.

---

[88] Some studies suggest that other industry classification schemes, such as the Global Industry Classifications Standard (GICS), are significantly better at explaining stock return comovements (Bhojraj et al., 2003). However, GICS classifications are a commercial product and not widely available. Nevertheless, we have repeated our analysis using the GICS and the results do not differ substantially from those reported.

4.1 Identification of news events

Before we study the market reaction to the events detected in an online stock forum, the most fundamental question is whether the information published there can be used to identify relevant news events. In this section, we will explore whether the messages published in an online stock forum can be used to detect which types of stock-related news affect a company on a particular day. To test the effectiveness of stock microblogs as indicators of real-world events, we use external information as a benchmark.[89] Among the event types used in our study, earnings announcements are the most widely and objectively available event dates that are commonly considered important sources of new information. Therefore we compare the earnings announcement dates of sample companies, first, with the message volume and, second, the event type identified by our classifier for that particular day.

According to Bloomberg, 672 earnings announcements were made by sample companies in the timeframe covered in this study. When we define a news spike as a one standard deviation increase in the message volume over the previous 5 days, we detect a substantial increase of messages on 224 of those days. Adding news spikes which occurred on the day before (101) and the day after the earnings announcements[90] (109) illustrates that an increase of investor generated message volume indicates the arrival of new information.

Table 3 shows the classification of events for 190 days on which sample firms made an earnings announcements and which our classifier identified as event days (i.e., company-days that were not labeled as "not classified"). Even though *Financial Issues* represent less than 15% of all news items (see Table 1), almost three quarter of the earnings announcement days were identified as related to *Financial Issues* (73.6%). The vast majority was accurately considered to be *Earnings*-related (69.5%). This is not to say that all of the remaining classifications are incorrect. The information released by an earnings announcement may well trigger investor concern over other issues that these days were associated with more frequently, such as *Operational Performance* (3.6%) and *Stock-Related Technical Trading* signals (7.1%).

---

[89] Note that this approach differs from the evaluation of classification accuracy illustrated in our methodology section. Traditional methods used in computational linguistics (e.g., 10-fold cross validation) are limited to the accuracy of the classifier within the training set. As pointed out in our methodology section, Antweiler and Frank (2006) do not provide evidence of the accuracy of their classification mechanism relative to the training set or relative to alternative methods to identify events.

[90] We would include these two days to the analysis for two reasons. First, many earnings announcement are made after the market closes, which may trigger a discussion among investors primarily on the following trading day. Second, the short moving average period (5 days) may lead to the identification of a news spike due to an increase of messages in anticipation of the announcement (e.g., information leakage).

We conclude that online stock forums can be used to reliably identify real-world news events that are on investors' minds. This external validation of computational linguistic methods for capital market research has not been conducted before. It validates the use of this innovative data source and may establish it for further research.

Table 3: Classification of earnings announcement dates

Event category	Share	Event detail	Share
Corporate Governance	**3.0%**	CEO	0.5%
		Other Executive	2.5%
Financial Issues	**73.6%**	**Earnings**	**69.5%**
		Analyst Rating	2.5%
		Financial Other	1.5%
Operations	**4.6%**	Labor Issues	0.5%
		Product Development	0.0%
		Operational Performance	3.6%
		Marketing	0.5%
		Contract	0.0%
Restructuring Issues	**5.6%**	Joint Venture	0.5%
		M&A	5.1%
Legal Issues	**2.5%**	Jurisdiction	0.5%
		Government Authorities	2.0%
Technical Trading	**13.2%**	Stock-Related	7.1%
		Market-Related	3.6%

Notes: This table shows the classification of events for 190 days on which sample firms made an earnings announcements and which our classifier identified as event days (i.e., company-days with the minimum number of observations that were not labeled as "not classified").

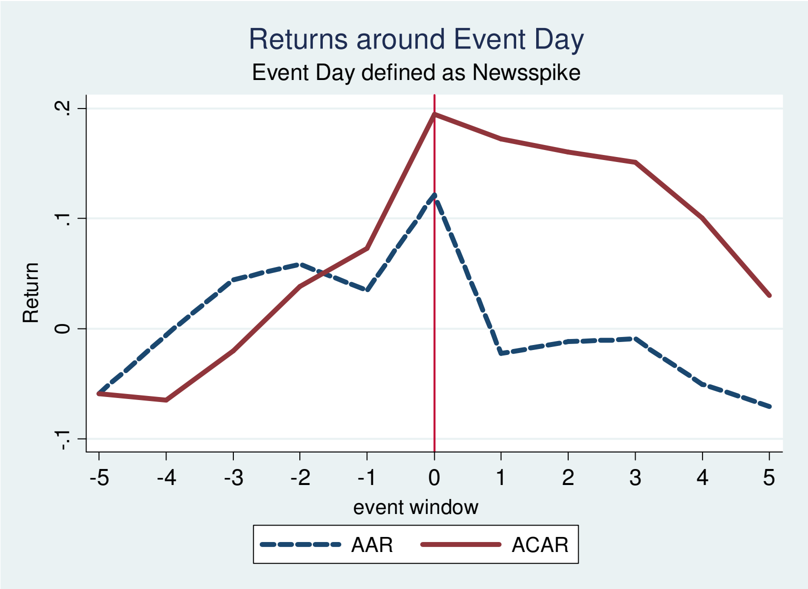
4.2 Overall impact of news spikes and distinction of news sentiment

In this section we explore the market reaction to an increase in messages as a generic sign of news arrival and investors processing this information (i.e., discussion). Table 4 shows the average abnormal returns on the 5 days surrounding a news spike as defined in the previous section. We would expect that the market reaction to the arrival of new information occurs primarily on the event day itself. The results confirm this hypothesis and show the strongest market reaction ($AAR = 0.1215$, $p < 0.01$) on the event day, which also is the only day with returns that are significant at the 1% level. Returns are positive for two of three days in a row before the news release and negative on all days after the event day. Cumulative returns confirm this pattern with a return reversal following the positive returns on the event day.

Trading volume picks up one day before the news spike, either in anticipation of new information or trading based on insider and leaked information, and continues to be high in the days after the news arrival. In line with the reversal pattern of returns, this indicates that further trading is necessary for all new information to be impounded fully in the stock price and that – in contrast to the EMH – these do not immediately reflect all new information.

Given the positive return on the event day, these results may suggest that news is generally good news. However, we know that investors tend to share more bullish than bearish messages in online forums. As a result, this excess of bullish news items may be reflected in the aggregate news spike. To separate these effects, Table 5 distinguishes between bullish and bearish news spikes, i.e., bullish and bearish days with a substantial increase in messages. We see that the news sentiment greatly influences the market reaction. Especially bullish news spikes show much more pronounced returns on the event day in terms of their absolute value ($AAR = 0.3169$, $p < 0.01$). In addition, cumulative returns around the event day show clear signs of overreaction with positive returns in the 5 days leading up to the event ($ACAR = 0.2283$, $p < 0.01$) and negative returns following the return spike on the event day ($ACAR = -0.2936$, $p < 0.01$). These findings suggest that positive news events are subject to information leakage. In addition, the market seems to overreact to the official news release on the event day resulting in negative returns in almost exactly the same magnitude in the following 5 days. Negative news events, on the other hand, experience a significant market reaction only on the event day itself, indicating that related information was not leaked prior to the event day. These effects would not have become apparent in the aggregate analysis of overall news volume (e.g., Mitchell & Mulherin, 1994) highlighting the importance of controlling for news sentiment.

Table 4: Market reaction to news spike



	Newsspike					
	Return			Volume		
Day	**AAR**	*t-value*	**ACAR**	*t-value*	**AAV**	*t-value*
-5	-0.0589 **	2.24			-0.36	0.52
-4	-0.0058	*0.23*			-0.67	*0.99*
-3	0.0443 *	*1.65*	0.0730	1.17	-0.85	*1.24*
-2	0.0585 **	*2.06*			0.57	*0.82*
-1	0.0348	*1.16*			3.18 ***	*4.44*
0	**0.1215 *****	***3.30***			**7.84 *****	***9.99***
1	-0.0224	*0.79*			4.51 ***	*5.99*
2	-0.0117	*0.42*			2.28 ***	*3.25*
3	-0.0094	*0.34*	-0.1645 ***	2.60	2.10 ***	*3.03*
4	-0.0505 *	*1.77*			2.10 ***	*2.96*
5	-0.0704 **	*2.36*			2.30 ***	*3.19*

Notes: This table shows the stock return before, on, and after an event day. An event day is defined as a day with a substantial increase in message volume (one standard deviation above the previous 5 day average). Average abnormal returns (AAR) and average cumulative returns (ACAR) are scaled by 100 (i.e., shown in percent) for easier readability. Average abnormal volume (AAV) is measures as the logged number of traded shares. Number of observations: 3,413.
**** (**, *) indicates significance at the 1% (5%, 10%) level.*

Table 5: Market reaction to news spike by sentiment

	Bullish Newsspike						Bearish Newsspike					
Obs	1,886						1,527					
Day	**AAR**	*t-value*	**ACAR**	*t-value*	**AAV**	*t-value*	**AAR**	*t-value*	**ACAR**	*t-value*	**AAV**	*t-value*
-5	-0.0227	0.64			0.21	0.23	-0.1037 ***	2.66			-1.06	1.02
-4	0.0241	*0.72*			-0.61	*0.65*	-0.0427	*1.14*			-0.75	*0.76*
-3	0.0433	*1.19*	0.2283 ***	2.72	-0.86	*0.92*	0.0457	*1.15*	-0.1189	1.28	-0.84	*0.84*
-2	0.0899 **	*2.35*			0.18	*0.20*	0.0197	*0.46*			1.06	*0.99*
-1	0.0936 **	*2.39*			2.53 ***	*2.63*	-0.0379	*0.82*			3.98 ***	*3.71*
0	**0.3169 *****	***6.39***			**7.24 *****	***6.89***	**-0.1198 ****	***2.21***			**8.58 *****	***7.27***
1	-0.0115	*0.29*			4.46 ***	*4.31*	-0.0359	*0.88*			4.58 ***	*4.18*
2	-0.0566	*1.54*			1.81 *	*1.86*	0.0436	*1.04*			2.86 ***	*2.83*
3	-0.0426	*1.19*	-0.2936 ***	3.46	1.70 *	*1.81*	0.0315	*0.72*	-0.0050	0.05	2.59 **	*2.53*
4	-0.0708 *	*1.78*			1.20	*1.21*	-0.0254	*0.62*			3.22 ***	*3.18*
5	-0.1122 ***	*2.97*			1.34	*1.34*	-0.0189	*0.40*			3.48 ***	*3.39*

*Notes: *** (**, *) indicates significance at the 1% (5%, 10%) level.*

Table 6: Market reaction by event category

	Corporate Governance				Financial Issues				Operations			
Obs	1,262				2,692				1,848			
Day	**AAR**	*t-value*	**ACAR**	*t-value*	**AAR**	*t-value*	**ACAR**	*t-value*	**AAR**	*t-value*	**ACAR**	*t-value*
-3	0.1187 **	*2.18*			-0.0176	*0.55*			0.0225	*0.53*		
-2	-0.0328	*0.69*	0.0272	0.75	0.0598 *	*1.89*	0.0646	0.26	0.0281	*0.61*	0.0999	0.16
-1	-0.0587	*1.22*			0.0225	*0.68*			0.0493	*1.22*		
0	**-0.0131**	***0.30***			**0.0538**	***1.53***			**0.0154**	***0.40***		
1	-0.0490	*1.07*			-0.0321	*0.95*			0.0357	*0.94*		
2	-0.0712 *	*1.71*	-0.1478 *	1.95	-0.0038	*0.11*	-0.0224	0.38	-0.0379	*0.96*	-0.0191	0.29
3	-0.0276	*0.63*			0.0135	*0.44*			-0.0169	*0.41*		

	Restructuring				Legal Issues				Technical Trading			
Obs	1,479				820				4,339			
Day	**AAR**	*t-value*	**ACAR**	*t-value*	**AAR**	*t-value*	**ACAR**	*t-value*	**AAR**	*t-value*	**ACAR**	*t-value*
-3	0.0250	*0.54*			-0.0220	*0.31*			-0.0062	*0.24*		
-2	-0.0418	*0.87*	0.0089	0.92	-0.0598	*0.92*	-0.0699	0.56	-0.0075	*0.31*	0.0167	0.71
-1	0.0257	*0.50*			0.0119	*0.18*			0.0304	*1.20*		
0	**0.0471**	***0.93***			**-0.0899**	***1.46***			**-0.0236**	***0.97***		
1	-0.0631	*1.44*			0.0007	*0.01*			0.0462 *	*1.86*		
2	-0.0704	*1.62*	-0.1260 *	1.67	0.0444	*0.73*	-0.0132	0.14	0.0546 **	*2.17*	0.0832 *	1.94
3	0.0074	*0.17*			-0.0583	*0.97*			-0.0175	*0.69*		

*Notes: *** (**, *) indicates significance at the 1% (5%, 10%) level.*

Table 7: Market reaction by event category and sentiment (1/2)

Bullish Events

	Corporate Governance (Bullish)				Financial Issue (Bullish)				Operations (Bullish)			
Obs	570				1,349				952			
Day	**AAR**	*t-value*	**ACAR**	*t-value*	**AAR**	*t-value*	**ACAR**	*t-value*	**AAR**	*t-value*	**ACAR**	*t-value*
-3	0.1354 **	*2.05*			-0.0083	*0.19*			0.0139	*0.24*		
-2	-0.0387	*0.54*	0.2455 **	0.03	0.1257 ***	*2.86*	0.2112 ***	0.01	0.1039 *	*1.79*	0.2925 ***	0.00
-1	0.1488 **	*2.17*			0.0937 *	*1.95*			0.1747 ***	*3.07*		
0	**0.1655 *** **	***2.69***			**0.2478 *** **	***5.17***			**0.1583 *** **	***3.03***		
1	0.0167	*0.27*			-0.0406	*0.91*			0.0511	*0.99*		
2	-0.1475 ***	*2.68*	-0.1735 *	1.79	-0.0185	*0.40*	-0.0026	0.03	-0.1140 **	*2.15*	-0.0880	0.96
3	-0.0427	*0.78*			0.0565	*1.32*			-0.0250	*0.47*		

	Restructuring Issue (Bullish)				Legal Issue (Bullish)				Technical Trading (Bullish)			
Obs	786				383				2,295			
Day	**AAR**	*t-value*	**ACAR**	*t-value*	**AAR**	*t-value*	**ACAR**	*t-value*	**AAR**	*t-value*	**ACAR**	*t-value*
-3	0.0126	*0.84*			0.0937	*0.37*			0.0177	*0.59*		
-2	0.0047	*0.94*	0.1102	0.34	0.0648	*0.45*	0.3268 **	0.03	0.0348	*0.25*	0.2076 ***	0.00
-1	0.0929	*0.16*			0.1682 **	*0.05*			0.1550 ***	*0.00*		
0	**0.2705 *** **	***0.00***			**0.0078**	***0.92***			**0.2219 *** **	***0.00***		
1	0.0196	*0.72*			0.0595	*0.44*			0.0500	*0.11*		
2	-0.0919	*0.10*	-0.1513	0.10	-0.0148	*0.86*	-0.0600	0.65	0.0387	*0.22*	0.0790	0.15
3	-0.0790	*0.15*			-0.1047	*0.20*			-0.0098	*0.76*		

*Notes: *** (**, *) indicates significance at the 1% (5%, 10%) level.*

Table 7: Market reaction by event category and sentiment (2/2)

Bearish Events

	Corporate Governance (Bearish)				**Financial Issue (Bearish)**				**Operations (Bearish)**			
Obs	692				1,343				896			
Day	**AAR**	*t-value*	**ACAR**	*t-value*	**AAR**	*t-value*	**ACAR**	*t-value*	**AAR**	*t-value*	**ACAR**	*t-value*
-3	0.1050	*1.26*			-0.0270	*0.57*			0.0316	*0.52*		
-2	-0.0280	*0.43*	-0.1526	0.23	-0.0065	*0.14*	-0.0826	0.31	-0.0525	*0.73*	-0.1048	0.32
-1	-0.2296 ***	*3.45*			-0.0490	*1.07*			-0.0839	*1.46*		
0	**-0.1603 ** **	***2.55***			**-0.1411 *****	***2.76***			**-0.1365 ** **	***2.44***		
1	-0.1032	*1.56*			-0.0237	*0.47*			0.0194	*0.35*		
2	-0.0085	*0.14*	-0.1267	1.13	0.0110	*0.22*	-0.0423	0.48	0.0390	*0.67*	0.0516	0.54
3	-0.0151	*0.23*			-0.0296	*0.66*			-0.0068	*0.11*		

	Restructuring Issue (Bearish)				**Legal Issue (Bearish)**				**Technical Trading (Bearish)**			
Obs	693				437				2,044			
Day	**AAR**	*t-value*	**ACAR**	*t-value*	**AAR**	*t-value*	**ACAR**	*t-value*	**AAR**	*t-value*	**ACAR**	*t-value*
-3	0.0391	*0.58*			-0.1235	*0.22*			-0.0330	*0.43*		
-2	-0.0945	*0.18*	-0.1059	0.43	-0.1690 *	*0.08*	-0.4175 **	0.02	-0.0551	*0.15*	-0.1976 ***	0.01
-1	-0.0506	*0.52*			-0.1251	*0.22*			-0.1094 ***	*0.01*		
0	**-0.2064 *****	***0.01***			**-0.1754 ***	***0.05***			**-0.2992 *****	***0.00***		
1	-0.1569 **	*0.02*			-0.0508	*0.57*			0.0419	*0.29*		
2	-0.0460	*0.50*	-0.0974	0.42	0.0964	*0.28*	0.0279	0.84	0.0724 *	*0.07*	0.0872	0.19
3	0.1054	*0.13*			-0.0177	*0.84*			-0.0271	*0.49*		

*Notes: *** (**, *) indicates significance at the 1% (5%, 10%) level.*

Trading volumes increase prior to the event day and continue to be high for a couple of days in both cases. This pattern is largely consistent with most of the event types that we have studied and in line with related findings (Schmitz, 2007). Thus, in the remainder of the paper we do not report these results for every event type separately and focus instead on market returns.

Our results suggest that event studies that focus on the amount of news as an indicator of information release should control for sentiment and distinguish positive and negative news items. We will take this approach in the remainder of this paper. Aggregate results suggest that information leakage is more pronounced for positive news than negative news. More generally, theses findings confirm that event studies based on news stories must take into account the nature of their content. Sentiment is certainly the most basic and in many cases arguably the most important aspect of this context, but the subject matter, i.e., the event type should also be considered, as we do in the next section.

4.3 Distinction of news types

In this section we evaluate the market reaction to the specified types of news events. We start with the 6 event categories specified in our methodology section and then distinguish, first, between bullish and bearish messages and, second, between the event details within an event category.

Table 6 shows the market reaction to our 6 major event categories. Interestingly, none of the aggregate event categories are associated with significant returns on the event day, which we would expect to trigger the strongest market reaction. Even among the other days surrounding the event and considering cumulative returns, we find only very few returns that are significant, most only at the 10% level. We consider these results to be spurious. An F-test following Moskowitz and Grinblatt (1999) does not allow us to reject the hypothesis that returns on the event day are not significantly different from zero ($F = 1.42, p = 0.21$) and that they are equal across event types ($F = 1.38, p = 0.25$). However, the market reaction becomes apparent when we distinguish the event categories by sentiment (Table 7). Again, the 6 major event types are examined for both bullish and bearish sentiment separately. Among the resulting 12 event types, 11 are associated with statistically significant returns on the event day, 10 of which are significant at the 5% or 1% level.[91] This finding demonstrates that aggregate news events by topic contain little information about the market reaction and become meaningful only in combination with their sentiment or effect on future prospects.

Generally, bullish news is accompanied by higher returns on the event day than negative news of the same type indicating that these results are not spurious. Table 7 suggests that positive *Restructuring Issues* ($AAR = 0.2705$, $p < 0.01$) and *Financial Issues* ($AAR = 0.2478$, $p < 0.01$) have a greater impact on stock prices than issues related to *Corporate Governance* ($AAR = 0.1655$, $p < 0.05$) and *Operations* ($AAR = 0.1583$, $p < 0.01$). In line with intuition, *Legal Issues* only have a significant impact on the stock price when they are negative ($AAR = -0.1754$, $p < 0.1$). We find that both positive and negative *Technical Trading* signals are accompanied by positive and negative cumulative returns before the event day, respectively. This is in line with a simple momentum strategy followed by the vast majority of technical traders (e.g., Dewally, 2003; Tumarkin & Whitelaw, 2001). For many events, returns seem to "anticipate" the information release with positive returns on some days before bullish events and negative returns on some days before bearish events. Returns then turn into the opposite direction after the event and seem to offset the significant effects before the event. While not all cumulative returns are significant, we find an overall pattern of overreaction. There is one noteworthy distinction: Leaving *Technical Trading* signals aside, all but one bullish news category is associated with significant positive returns before the event day, whereas this is true for only one bearish news category (*Legal Issues*). This supports our aggregate results of bullish news spikes and suggests that positive news is incorporated into stock prices before the information is officially announced. In other words, in contrast to negative news[92], positive news rarely comes as a surprise. This finding is consistent with Schmitz (2007) who found only a short-term post-event drift after good news, but a price drift of several days after bad news suggesting that the market took longer to process negative news. We conclude that an event study needs to distinguish news items by sentiment to derive meaningful conclusions. We find that positive news tends to leak and get incorporated into market prices before the official information release, whereas negative news seem to be less anticipated.

---

[91] The F-test provides further evidence suggesting that returns on the event day are jointly different from zero ($F = 5.61$, $p < 0.01$) and different from each other ($F = 5.00$, $p < 0.01$).

[92] Bearish Legal Issues are the only exception to the rule. This is plausible, however, since the verdicts for most high-profile corporate cases are anticipated with most of the facts being publicly available and experts being able to evaluate likely outcomes.

Hong and Stein (1999) have developed a model suggesting that investors underreact to qualitative information such as news and overreact to pure price-movements. We have included the publication of *Technical Trading* signals as a news category to explore this distinction empirically. According to Hong and Stein (1999), we would expect to find signs of return reversal after *Technical Trading* signals and drift after qualitative news types. However, in contrast to the model developed by Hong and Stein (1999), we find no overreaction to *Technical Trading* signals in the relatively short event window covered by our analysis. Thus the cause of over- and underreaction appear to be more complex than the distinction between qualitative and quantitative information.[93]

Given that almost all of the events defined at the level of event categories are associated with significant effects, we now turn to the level of event details to determine which events are driving these reactions and to distinguish which events have an impact on market prices and which are less important to market participants. We focus the analysis of event details on the event day since we have seen that the market reaction is strongest on this day. Table 8 shows the market reaction for all 16 event details separated by bullish and bearish sentiment. Obviously, at this more granular level of analysis, we cannot expect each and every event type to trigger statistically significant returns.[94] For comparison, note that Antweiler and Frank (2006) find statistically significant cumulative returns for only about 20 of their 48 event types, even though most of these event types contain several thousand observations.[95] In addition, we are dealing with a larger share of comparably less significant minor news items, many of which may not be meaningful enough to move the stock price enough to detect a significant abnormal return on a daily basis. Table 8 illustrates that some event types that are associated with strong market reactions for both positive and negative news, such as *M&A* activity ($AAR = 0.2994$, $p < 0.01$ vs. $AAR = -0.3016$, $p < 0.01$) and *Earnings* ($AAR = 0.3155$, $p < 0.01$ vs. $AAR = -0.1295$, $p < 0.05$). *Technical Trading* signals are also consistently accompanied by substantial stock price changes. Of course, given that our analysis aggregates news items on a daily basis, these messages related to *Technical Trading* may simply follow market movements triggered by other events. Interestingly, there are a number of events for

[93] Akbas, Kocatulum, and Sorescu (2008) support this notion by suggesting that the market appears to overreact to public news following bad past performance and underreact following strong past performance.

[94] In the context of related studies, Morse (1982) has noted this feature as typical in a large-scale event study such as his where “not all of these residuals, however, were significant” (p. 76).

[95] The number of events associated with statistically significant returns varies, depending on the event window for cumulative returns, between 15 and 21. Roughly one quarter of these events are significant only at the 5% level.

which only one of the two sentiments is associated with significant stock price movements. Consistent with findings by McNichols and O'Brien (1997), which suggest that most analyst ratings are positive making investors immune to their advice, it does not come as a surprise that this news category (*Analyst Rating*) is ignored by the market. However, negative *Analyst Ratings* (e.g., downgrades of a stock) do have a significant effect on stock prices ($AAR = -0.2660$, $p < 0.05$). *Product Development* and *Marketing* follow a similar pattern. This suggests that the market does not consider positive product related news items to be new information. Possibly, these types of news are released more slowly (e.g., the launch of a new product) and are incorporated into stock prices over a longer period of time. The official announcement may no longer represent new information. Negative product-related news, however, often comes as a surprise (e.g., a recall). Our results suggest that there are many news categories which truly surprise the market (e.g., *M&A* or *Earnings*) and others (such as *Labor Issues*[96] or *Joint Ventures)*, which rarely contain new information that moves the market.

While the previous findings suggest that market reactions vary across different types of news, we now investigate whether the market reactions of different event types are statistically significant. First of all, the F-statistics suggest that the market reactions are both jointly different from zero and also different from each other. In addition to this aggregate analysis, we conduct mean comparison tests (i.e., t-tests) to explore pairs of different types of news. We start with differences across sentiment, but within the same news type. For simplicity, we focus on the 4 event types for which our results show returns that are statistically different from zero for both bullish and bearish news categories (i.e., news related to *Earnings*, *M&A*, *Stock-Related* and *Market-Related Technical Trading*). All four differences between bullish and bearish news are statistically significant at the 1% level. This provides further evidence suggesting that event studies (even those concerned with a single event type, such as merger or earnings announcements) should distinguish between positive and negative news categories. While many related studies follow this approach for earnings announcements by controlling for earnings surprise (i.e., the difference between forecasts and reported results), it is less common for many other event types (e.g., Ritter, 1991; Agrawal et al., 1992; Michaely et al., 1995).

[96] This is in line with Morse (1982) who also found labor strikes to have no significant market impact on stock prices.

Table 8: Market reaction by event detail and sentiment

Event type			Return		Volume	
Sentiment	**Event detail**	**Obs**	**AAR**	*t-value*	**AAV**	*t-value*
Bullish	**CEO**	96	0.2739 **	*2.16*	3.5160	*0.97*
	Other Executive	347	0.0922	*1.17*	-4.9719 **	*2.10*
	Earnings	905	0.3155 ***	*4.48*	5.9689 ***	*4.12*
	Analyst Rating	322	0.1215	*1.29*	3.2260	*1.36*
	Financial Other	149	0.1500	*1.21*	4.4729	*1.54*
	Labor Issue	108	0.2474 *	*1.87*	3.5629	*0.92*
	Product Development	365	0.0817	*1.08*	4.3984 **	*2.11*
	Operational Performance	349	0.2797 ***	*3.21*	4.7462 **	*2.16*
	Marketing	164	0.1152	*1.19*	2.3084	*0.90*
	Contract	85	0.2639 *	*1.98*	4.0941	*1.07*
	Joint Venture	120	0.1699	*1.63*	-1.3204	*0.42*
	M&A	570	0.2994 ***	*3.65*	8.2264 ***	*4.86*
	Jurisdiction	97	0.3149 **	*2.28*	-1.3016	*0.38*
	Government Authorities	202	-0.0178	*0.17*	5.3230 *	*1.94*
	Stock-Related	2108	0.2230 ***	*7.12*	2.5504 ***	*2.81*
	Market-Related	335	0.2027 **	*2.44*	4.1414 *	*1.79*
Bearish	**CEO**	123	-0.0524	*0.27*	3.0015	*0.68*
	Other Executive	444	-0.1321 **	*2.01*	5.9168 ***	*3.09*
	Earnings	910	-0.1295 **	*2.16*	5.1983 ***	*3.24*
	Analyst Rating	344	-0.2660 **	*2.29*	4.2643 *	*1.87*
	Financial Other	185	-0.0592	*0.56*	-0.1386	*0.04*
	Labor Issue	119	-0.0482	*0.32*	8.4531 **	*2.42*
	Product Development	275	-0.3350 ***	*3.91*	-3.6433	*1.53*
	Operational Performance	286	-0.1393	*1.30*	8.5076 ***	*3.23*
	Marketing	148	-0.3094 **	*2.36*	5.3455	*1.55*
	Contract	97	0.0023	*0.02*	6.3032	*1.64*
	Joint Venture	109	0.0473	*0.31*	11.4638 **	*2.52*
	M&A	498	-0.3016 ***	*2.90*	5.4242 **	*2.50*
	Jurisdiction	137	-0.1252	*1.18*	-3.5197	*1.04*
	Government Authorities	228	-0.2777 *	*1.96*	7.3584 **	*2.17*
	Stock-Related	1778	-0.3163 ***	*8.77*	3.4461 ***	*3.59*
	Market-Related	374	-0.3275 ***	*2.70*	16.1261 ***	*6.87*
	F-statistic (all=0)		6.41 ***		2.40 ***	
	F-statistic (all the same)		6.60 ***		2.00 ***	

Notes: This table shows average abnormal returns (AAR) and volumes (AAV) for the event day only.
**** (**, *) indicates significance at the 1% (5%, 10%) level.*

Of course, the previous finding may primarily be a result of the distinction by sentiment. The more conservative test for different market reactions across event types is to compare different event types with the same sentiment. Most combinations do not show statistically significant differences, even those for which intuition may suggest such as pattern, e.g., positive news about the *CEO* vs. *Other Executives* (*p = 0.27*). This may be a result of a relatively small sample size, given the large variance in returns. However, there are some combinations that show significant differences. The market reaction to positive news related to *Product Development*, for instance, is weaker than for positive news related to *Earnings* (*p = 0.05*) or M&A (*p = 0.07*). We also find significant differences among negative news, even for two event types that both show significant negative returns (i.e., news related to *Product Development* and *Other Executives*, *p = 0.05*). We conclude that not only the sentiment but also the type of a news item can explain the market reaction.

4.4 News types by industry

In this section, we investigate whether the market reaction to news events differs across industries. Before exploring returns, we will first look at the share of the news coverage that our event types represent for different industry groups. We limit the analysis to industries with a sufficient number of high-volume stocks (5), a substantial total number of companies (20) and a significant share of messages in our sample (more than 30,000). These are: Business Services, Depository Institutions, and Industrial Machinery and Equipment. This leaves us with a manageable set, whose combined message volume represents 38.1% of our entire sample. Conveniently, these three industry groups cover three, very different major economic sectors (service firms, financial institutions, and manufacturing). For simplicity, we refer to these industry groups by their sector names.

Table 9 shows the share of news events related to these three industries. We see that the frequency of news items within the three industry groups are not distributed equally over the news categories. Some patterns follow intuition, such as the fact that *Financial Issues* are more closely associated with financial institutions. In addition, there are more subtle, but still comprehensible peaks in the frequency of the news items. In the financial industry, for example, other executives (e.g., traders) who have a greater influence on company performance relative to other industries are much more important than the *CEO*. *Legal Issues* related to *Government Authorities* (i.e., federal regulation, actions taken by the central bank) are on investors' minds more often with respect to the financial industry (6.6%) than service (2.7%) or manufacturing firms (1.2%). On the other hand, *Product Development* is an issue that is associated with business (25.9%) and manufacturing firms (29.5%) more than five times as often as with financial institutions (4.4%). All in all, the content of news coverage varies heavily by industry group indicating that investors attribute varying degrees of importance to different news types depending on the industry that is affected.

4.5 Market impact of different news types across industries

As shown above, the content of news coverage varies across industries. In this section, we will explore whether industry affiliations have an effect on market reactions. Table 10 shows the returns for all event categories across industry groups. Unfortunately, the sample size becomes smaller as we increase the level of detail. Only 10 out of the 36 event types exhibit statistical significance.

Table 10 suggests that both positive and negative news related to *Corporate Governance* have a greater effect on manufacturing firms than the other two industries and that *Restructuring Issues* are more relevant for financial institutions that for Business Services or Industrial Machinery and Equipment. Following the analysis in the previous section, we investigate differences across industries for one and the same event type. However, there is only one event type (*Financial Issues*) that triggers statistically significant returns for two of the three industries. The difference between those returns for Business Services ($AAR = 0.2897, p < 0.01$) and Industrial Machinery and Equipment ($AAR = 0.4812, p < 0.01$) is not significant ($p = 0.33$). Given that even statistically insignificant returns on the event day can carry meaningful insights with respect to the market reaction, we also compared the largest differences among other results across industries. However, most of these differences in returns accompanying these events, such bearish *Restructuring Issues* for financial institutions compared to either manufacturing firms ($p = 0.32$) or service firms ($p = 0.29$) are not significant. Yet, there are a few combinations that are weakly significant at the 10% level. For example, bullish operational issues have a higher impact for financial institutions than for manufacturing firms ($p = 0.09$).[97]

We conclude that while our sample size may be too small to detect a larger number of statistically significant differences across industries, industry classifications seem to matter when explaining market reactions to news. However, the resulting differences appear to be rather small.

[97] The differences for bearish Legal Issues between Business Services and Depository Institutions is also statistically significant (p-value 0.08). However, given that bearish Legal Issues are associated with a positive market reactions, we consider these results to be spurious.

Table 9: News type by industry

Event category	Event detail	Business Services	Depository Institutions	Industrial Machinery & Equipment
Corporate Governance	CEO	2.1%	2.7%	2.0%
	Other Executive	6.7%	9.8%	6.9%
Financial Issues	Earnings	9.8%	18.5%	11.3%
	Analyst Rating	2.3%	6.5%	2.9%
	Financial Other	1.8%	5.0%	2.7%
Operations	Labor Issues	1.1%	2.1%	1.3%
	Product Development	25.9%	4.4%	29.5%
	Operational Performance	9.6%	4.4%	12.0%
	Marketing	1.6%	7.0%	1.8%
	Contract	2.3%	1.0%	2.4%
Restructuring Issues	Joint Venture	5.9%	2.1%	2.6%
	M&A	7.1%	9.1%	6.9%
Legal Issues	Jurisdiction	2.8%	2.4%	2.8%
	Government Authorities	2.7%	6.6%	1.2%
Technical Trading	Stock-Related	10.8%	16.2%	9.3%
	Market-Related	1.6%	7.0%	1.8%
Sample size				
Companies		40	23	32
News items		65,562	33,325	69,255

Notes: This table shows the share of news items for each event type.

Table 10: Market reaction to different events by industry

Event type		Business Services			Depository Institutions			Industrial Machinery& Equipment		
Sentiment	**Event detail**	**Obs**	**AAR**	*t-value*	**Obs**	**AAR**	*t-value*	**Obs**	**AAR**	*t-value*
Bullish	**Corporate Governance**	67	0.1618	*1.43*	42	0.1153	*0.53*	60	0.5119 **	*2.31*
	Financial Issues	134	0.2897 ***	*2.66*	57	0.2019	*0.81*	91	0.4812 ***	*2.72*
	Operations	121	0.2678 **	*2.41*	49	0.4706	*1.29*	93	-0.0506	*0.43*
	Restructuring Issues	83	-0.0479	*0.21*	34	0.4949	*1.33*	102	0.1343	*1.19*
	Legal Issues	43	-0.0918	*0.62*	22	0.1568	*0.44*	38	0.2621	*0.80*
	Technical Trading	174	0.3652 ***	*3.65*	110	0.0713	*0.43*	159	0.1824	*1.55*
Bearish	**Corporate Governance**	72	-0.1663	*1.16*	44	-0.2636	*0.93*	50	-0.4217 **	*2.13*
	Financial Issues	108	-0.1361	*0.97*	66	-0.3055	*1.28*	103	0.0311	*0.19*
	Operations	82	0.1173	*0.76*	45	0.0558	*0.22*	56	-0.2663 *	*1.95*
	Restructuring Issues	59	-0.3245	*1.18*	29	-0.7892 ***	*2.87*	45	-0.1599	*0.33*
	Legal Issues	42	-0.4361	*1.23*	40	0.4493	*1.32*	30	-0.0324	*0.15*
	Technical Trading	141	-0.3633 ***	*3.09*	89	-0.1736	*1.08*	127	-0.3696 ***	*2.82*

Notes: This table shows average abnormal returns (AAR) for the event day.
**** (**, *) indicates significance at the 1% (5%, 10%) level.*

5 Conclusion

5.1 Discussion of results

Our study offers a comprehensive analysis of the market reaction to combinations of different event types, sentiments of these events, and industries and thus provides us with a unique look at the financial market impact of news. Whereas the event study literature has thus far focused on professionally edited sources of news (e.g., newspaper articles, official company press releases), this study uses data from a real-time online stock forum to identify news events from an investor perspective. This data source allows us to accurately investigate the link between news in the eyes of individual market participants, who ultimately determine market prices, and everyday news-events.

Our first research question investigates whether the information published in an online stock forum can be used to detect the types of stock-related news that affect a company on a particular day. Using earnings announcement dates as an external benchmark, we found that the investor discussion in an online stock forum meaningfully reflects real-world news events. This finding helps to establish user-generated online content as a source of company-specific news events and validates the use of stock microblogs for further use in capital market research.

Second, we systematically explored whether and to what extent the distinction between good and bad news matters in the context of an event study with respect to the absolute value of returns. Numerous event studies use news volume (i.e., the number of news articles published about a company as an indicator of information release (e.g., Wysocki, 1998; Chan, 2003). However, we found that news volume as a measure of information arrival is insufficient and misses many nuances that have a significant effect on the results. Thus, our findings imply that event studies need to control for sentiment and even event studies concerned with only a single event type, for instance merger or earnings announcements, need to clearly distinguish between positive and negative news items. Our results indicate that, whereas the price reaction is largely confined to the event day itself for negative news, positive news often leaks and is incorporated into stock prices before the information is officially announced. In other words, in contrast to negative news, positive news may rarely come as a surprise, which indicates more widespread information leakage before positive news. Given that our news items were published by individual investors and represent their perception of news, our results indicate that other market participants, such as institutional investors, were privy to and seem to have already acted on this new information a few days in advance.

Third, while the previous event study literature has largely focused on the analysis of one single event at a time, we examined to what extent the market reaction in terms of returns differs between multiple types of news events. Of course, the price impact of the kind of small news items we have studied is rather modest on a daily basis. Yet, we have observed a sufficient number of differences to conclude that not only the sentiment, but also the type of a news item can explain the market reaction. Our results suggest that there are many news categories, which truly surprise the market (e.g., *M&A* or *Earnings*) and others (such as *Labor Issues or Joint Ventures)*, which rarely contain new information that moves the market. We are unable to find empirical support for the model developed by Hong and Stein (1999) that investors underreact to qualitative information such as news and overreact to pure price-movements (i.e., *Technical Trading* signals).

Finally, we analyzed whether the market reaction to various event types differs across industry groups. Our results show that the content of news coverage varies by industry indicating that the market attributes varying degrees of importance to different news types depending on the industry that is affected. This led us to explore whether the market reaction to these news events differs across industry groups. Previous studies have shown relatively little impact of industries on returns (for an overview, see Moskowitz & Grinblatt, 1999) and do not account for firm specific-news affecting the stock price. Whereas our sample size may have been too small to detect a larger number of statistically significant differences across industries, we are confident to conclude that industry classifications may at least partially explain market reactions to the same type of news.

5.2 Limitations and further research

Our study does not come without limitations. We conducted our analysis based on daily data, even though tweets are time-stamped to the minute. As discussed, some market reactions to small news items may not be detected because their impact may not be strong enough to move daily stock prices. Clarkson et al. (2006) have found that the market reaction to takeover announcements on internet discussion sites have a statistically significant effect in the 10 minute posting interval, but for the larger part this effect reverses within the next 50 minutes. In their study, which is based on a news source that is very similar to ours, many effects would not have been detectable with daily data. Thus future research should seek to investigate intraday stock data.

Our paper focused on quantifying the market impact of different news types across industries. Given the large number of different events and the even larger number of potential reactions when combined with the industry perspective, it is beyond the scope of this study to investigate in detail the reasons leading to each and every one of these effects. Future research should focus on the explanation of these differences to understand why certain types of news are more important for an industry than others.

In an event study of various types of world events, Niederhoffer (1971) concludes that the particular type of event adds little information concerning the subsequent reaction of the stock market index. Our study reveals that the explanatory power of company-specific news is much better and shows that adding sentiment significantly improves our understanding of the information content of a news item. However, there are still many interpretive aspects of news that our approach does not capture. Future research should try to better distinguish the novelty and significance of information.[98] We determined the most relevant event type for a particular day through the number of messages referencing that category. There is good reason to believe that the most impactful news often trigger substantial volume. In fact, this is the logic behind the display of "Trending Topics"[99] on Twitter's website. An increase in related messages, for instance, has been found to identify significant news events such as an earthquake before other official detection methods (Sakaki, Okazaki, & Matsuo, 2010). However, there may be cases in which a few highly significant news items have more impact on the market than a large amount of trivial chatter.

In sum, our study shows that the online chatter in stock microblogging forums is more than just "Noise", even though we are still far from understanding the "News" as easily as the Wall Street Journal.

[98] The study of Groß-Klußmann and Hautsch (2009) is an example of this type of research. The authors find market-wide robust news-dependent responses in volatility and trading volume only if news items are classified as relevant. However, they use data from a commercial product and do not provide details regarding the underlying mechanism by which relevance is determined ("Each news item provides a sentiment and relevance indicator. These indicators are produced based on pattern recognition algorithms", p. 4).

[99] Trending Topics are the most talked about topics currently being discussed on Twitter.

6 Appendix

In this appendix, 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

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

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

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

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

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

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

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

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, & 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 et al. (2008) we use the General Inquirer’s Harvard-IV-4 classification dictionary and add each occurrence of an emotional word to the bag of words for that message. Thus we combine text mining approaches based on pre-defined dictionaries and statistical methods. Sixth, we apply the widely used Porter stemmer in order to remove the morphological endings from words (e.g., “buys” and “buying” are reduced to “buy”; (Porter, 1980). Finally, following established preprocessing procedures (see Rennie et al., 2003), word counts are transformed to a power-law 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). The algorithm, then, treats any remaining distinct sequence of characters separately.

Table A1 shows a few random examples of tweets from both the main data set and the results of the automatic classification. As these examples illustrate the Naïve Bayesian algorithm can classify messages quite well. Table A2 offers an overview of the classification accuracy by event class. The confusion matrix (Table A3) provides further detail by showing the deviations between the manual and automatic classifications in the training set.

Table 11: Sample messages and classification (main data set)

Sample tweets	Automatic classification	
	Event category	**Event detail**
$PFE raised quarterly div by 13% to 18 cents and said more annual increases are likely barring significant unforeseen events	Financial Issues	Earnings
Trader Bots has recently calculated a Bullish Overall Stock Prediction on $NU http://bit.ly/7XONaW	Technical Trading	Stock-Related
i4u: New NVIDIA Video Cards on Way $NVDA - http://bit.ly/8mT4Yv	Operations	Product Development
Stephen R. Covey Grants E-Book Rights to Amazon. $AMZN $CBS #books #media http://bit.ly/4Bu5B6	Operations	Marketing
$MOT Motorola to launch Android application store named: Shop4Apps - Store will allow you to use both PC & phone to download apps.	Operations	Product Development
Deutsche Bank Upgraded Wells Fargo $WFC and BB&T $BBT to Buy http://bit.ly/8FhB1q	Financial Issues	Analyst Rating
Robert J. Bertolini Adds to Board of GENZYME CORP ($GENZ) - http://www.implu.com/story/12828 #fb	Corporate Governance	Other Executive
No respect for $ARRY. The deal with $AMGN nice surprise for this company.	Operations	Contract
$AIG ticks higher after Bernanke said he believes they will payback the Fed	Legal Issues	Government Authorities
$GE CEO: Orders strengthening in 4Q09	Corporate Governance	CEO
Wyeth buyout boosts $PFE dividend http://bit.ly/6af0Wl	Financial Issues	Financial Other
Read EnergyPoint's lastest report on how Halliburton ($HAL) has boosted its performance... http://seekingalpha.com/a/3thv	Operational Issues	Operational Performance
For Wed, two long scalp setups on market and sector confirmation: $LH, $PKG	Technical Trading	Market-Related
OMG, that will get $NKE's attention! RT @n23mc: @optionmonster Elin close to signing deal with Puma. Poor Tiger. http://tinyurl.com/y95lxew	Operations	Contract
$GE CEO Looking To Rebuild After Tough Year looks to its big industrial divisions to navigate out of the deep recession http://bit.ly/6Undmw	Corporate Governance	CEO
http://bit.ly/6b4Ezg $INTC FTC sues Intel, claims company using anticompetitive tactics in CPU and GPU market	Legal Issues	Jurisdiction
http://bit.ly/6mlXzj $KFT $CSG Cadbury Points to Rival Interest as It Rejects Kraft	Restructuring Issues	M&A
$GE orders have strengthened in 4th quarter and worst of finance challenges are over - http://j.mp/7vQFJt	Financial Issues	Earnings
MA crossover not a concern? @TraderFlorida: http://bit.ly/8ato1d $GS - breaks descending trendline on volume - could see a nice move up	Technical Trading	Stock-Related
there could be pressure on $CBS after UBS downgraded it from buy to neutral.	Financial Issues	Analyst Rating

Notes: Tweets were assigned to the event detail class with the largest probability according to the Naïve Bayesian classifier and automatically assigned to the corresponding event category.

Table 12: Classification accuracy (accuracy by class)

Class	True positives	False positives	Precision	Recall	F-Measure	ROC area
CEO	80.0%	0.6%	55.6%	80.0%	65.6%	97.8%
Other Executive	61.1%	1.2%	42.3%	61.1%	50.0%	91.4%
Earnings	65.0%	3.1%	55.2%	65.0%	59.7%	89.7%
Analyst Rating	44.6%	2.5%	29.1%	44.6%	35.2%	87.8%
Financial Other	58.8%	1.7%	48.8%	58.8%	53.3%	89.6%
Labor Issues	54.5%	1.0%	32.4%	54.5%	40.7%	90.0%
Product Development	51.6%	4.5%	48.0%	51.6%	49.7%	90.2%
Operational Performance	49.4%	2.1%	43.3%	49.4%	46.2%	87.8%
Marketing	50.6%	2.7%	38.5%	50.6%	43.8%	87.2%
Contract	38.5%	1.0%	29.4%	38.5%	33.3%	92.2%
Joint Venture	18.2%	0.8%	17.4%	18.2%	17.8%	80.5%
M&A	56.3%	2.9%	45.0%	56.3%	50.0%	87.1%
Jurisdiction	66.7%	0.6%	46.2%	66.7%	54.5%	92.5%
Government Authorities	41.2%	1.7%	33.9%	41.2%	37.2%	90.5%
Stock-Related	76.0%	5.5%	79.7%	76.0%	77.8%	93.1%
Market-Related	43.5%	2.0%	29.0%	43.5%	34.8%	87.8%
Not classified	60.3%	12.1%	76.6%	60.3%	67.5%	83.7%
weighted average	*61.1%*	*7.0%*	*64.6%*	*61.1%*	*62.2%*	*87.8%*

*Notes: This table shows the classification accuracy by class using 10-fold cross validation, in which 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. True positives (or precision) represent, for example, the share of messages classified as CEO, which were labeled as such in the training set. False positives are message classified incorrectly as CEO. 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).*

Table 13: Classification accuracy (confusion matrix)

Manual classification	Total	**Automatic classification** CEO	Other Executive	Earnings	Analyst Rating	Financial Other	Labor Issues	Product Dev.	Ops Performance	Marketing	Contract	JV	M&A	Jurisdiction	Gov't Auth.	Stock-related	Market related	Not classified
CEO	25	**20**	1	0	0	0	0	0	1	1	0	0	2	0	0	0	0	0
Other Executive	33	0	**22**	0	1	1	0	2	0	1	1	2	1	0	1	0	1	3
Earnings	119	0	1	**91**	5	2	0	1	5	0	0	1	3	0	2	3	5	21
Analyst Rating	47	0	0	6	**25**	1	2	1	3	1	2	1	2	0	2	1	0	9
Financial Other	64	1	1	2	4	**40**	1	4	2	1	1	0	4	0	1	2	0	4
Labor Issues	22	1	1	1	1	0	**12**	0	2	0	1	0	1	0	1	1	0	0
Product Dev.	163	1	3	5	2	4	0	**96**	11	18	0	4	10	3	2	2	2	23
Ops Performance	73	1	0	14	2	1	1	9	**39**	1	1	0	2	0	0	1	1	6
Marketing	72	0	3	1	2	2	2	13	0	**42**	1	1	1	1	1	2	0	11
Contract	22	0	1	0	0	0	0	1	0	0	**10**	2	4	2	2	0	0	4
Joint Venture	18	0	1	2	1	0	0	2	1	2	2	**4**	0	1	1	1	0	4
M&A	88	1	0	3	2	6	0	8	2	4	0	1	**58**	0	1	0	2	15
Jurisdiction	17	1	0	2	0	0	0	2	0	0	0	0	0	**12**	0	0	0	1
Gov't Authorities	44	0	1	0	3	2	4	0	1	4	1	0	3	1	**21**	1	2	7
Stock-Related	488	0	3	13	8	0	3	5	4	1	1	0	4	1	5	**424**	16	70
Market-Related	40	0	0	2	3	1	0	0	2	1	0	0	1	0	1	9	**20**	6
Not classified	397	10	14	23	27	22	12	56	17	32	13	7	33	5	21	85	20	**604**
All messages (total)		26	38	142	59	60	25	144	73	77	21	16	96	21	41	447	49	184

Notes: This table shows the classification accuracy within the training set for 10-fold cross validation. In contrast to classification accuracy for the full training set, which bears the risk of crediting overfitting, 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.

7 References

Agrawal, A., J. F. Jaffe, & Mandelker, G. N. (1992). The post-merger performance of acquiring firms: A re-examination of an anomaly. *Journal of Finance*, 47, *1605*–1621.

Akbas, F., E. Kocatulum, & Sorescu, S.M. (2008). Mispricing following public news: Overreaction for losers, underreaction for winners. *Working paper.* Retrieved May 15, 2011 from http://ssrn.com/abstract=1107690

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *Journal of Finance,* 59, 1259-1294.

Antweiler, W., & Frank, M. Z. (2006). Do U.S. stock markets typically overreact to corporate news stories?. *Working paper.* Retrieved May 15, 2011 from http://ssrn.com/abstract=878091

Baginski, S. (1987). Intraindustry information transfers associated with management forecasts of earnings. *Journal of Accounting Research*, 25, 196-216.

Barber, B., & Odean, T. (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.

Bhojraj, S., Lee, C. M. C., & Oler, D. K. (2003). What's my line? A comparison of industry classification schemes for capital market research. *Journal of Accounting Research*, 41, 745-774.

Bradley, M., Jarrell, G. A., & Kim, E.H. (1984). On the existence of an optimal capital structure: Theory and evidence. *Journal of Finance*, 39, 857–878.

Brookfield, D., & Morris, R. (1992). The market impact of UK company news announcements. *Journal of Business Finance & Accounting*, 19, 585-602.

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

BusinessWeek (2009). StockTwits may change how you trade. *BusinessWeek online edition (author Max Zeledon)*, Retrieved May 15, 2011 from http://www.businessweek.com/technology/content/feb2009/tc20090210_875439.htm

Chan, W. (2003). Stock price reaction to news and no-news: Drift and reversal after headlines. *Journal of Financial Economics*, 70, 223-260.

Clarkson, P. M., Joyce, D., & Tutticci, I. (2006). Market reaction to takeover rumour in internet discussion sites. *Accounting and Finance*, 46, 31-52.

Cutler, D., Poterba, J., & Summers, L. (1998). What moves stock prices?. *Journal of Portfolio Management*, 15, 4-12.

De Bondt, W. F. M., & Thaler, R. (1985). Does the stock market overreact?. *Journal of Finance*, 40, 793-805.

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

Dyckman, T., Philbrick, D., & Stephan, J. (1984). A comparison of event study methodologies using daily stock returns: A simulation approach. *Journal of Accounting Research*, 22, 1-30.

Engelberg, J., & Parsons, C. A. (2011) The causal impact of media in financial markets. *Journal of Finance*, 66, 67-97.

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

Fama, E. (1991). Efficient capital markets: II. *Journal of Finance*, 46, 1575–1617.

Fama, E. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, 49, 283-306.

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

Hall, M., Frank, E., Holmes, G., & Pfahringer, B. (2009). The WEKA data mining software: An update. *SIGKDD Explorations*, 11, 10-18.

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

Huberman, G., & Regev, T. (2001). Contagious speculation and a cure for cancer: A nonevent that made stock prices soar. *The Journal of Finance*, 56, 387-396.

Ikenberry, D. L., & Ramnath, S. (2002). Underreaction to self-selected news events: The case of stock splits. *Review of Financial Studies*, 15, 489-526.

Kahle, K. M., & Walkling, R. A. (1996). The impact of industry classifications on financial research. *Journal of Financial & Quantitative Analysis*, 31, 309-331.

Groß-Klußmann, A., & Hautsch, N. (2009). Quantifying high-frequency market reactions to real-time news sentiment announcements. *Working paper.* Retrieved May 15, 2011 from http://ideas.repec.org/p/hum/wpaper/sfb649dp2009-063.html

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

Kothari, S., & Warner, J. (2007). Econometrics of event studies. In B. Espen-Eckbo (Ed.), *Handbook of Corporate Finance: Empirical Corporate Finance* (forthcoming). Amsterdam: Elsevier.

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

McNichols, M., & O'Brien, P. (1997). Self-selection and analyst coverage. *Journal of Accounting Research*, 35, 167-199.

Michaely, R., Thaler, R. H., & Womack, K. L. (1995). Price reactions to dividend initiations and omissions: Overreaction or drift?. *Journal of Finance*, 50, 573–608.

Mitchell, M. L., & Mulherin, J. H. (1994). The impact of public information on the stock market. *Journal of Finance*, 49, 923-950.

Morse, D. (1982). Wall Street Journal announcements and the securities markets. *Financial Analysts Journal*, 38, 69-76.

Moskowitz, T. J., & Grinblatt M. (1999). Do industries explain momentum?. *Journal of Finance*, 54, 1249-1290.

Niederhoffer, V. (1971). The analysis of world events and stock prices. *Journal of Business*, 44, 193-219.

Porter, M. F. (1980). An algorithm for suffix stripping. *Program*, 14, 130−137.

Pritamani, M., & Singal, V. (2001). Return predictability following large price changes and information releases. *Journal of Banking & Finance*, 25, 631-651.

Ramnath, S. (2002). Investor and analyst reactions to earnings announcements of related firms: An empirical analysis. *Journal of Accounting Research*, 40, 1351–1376.

Rennie, J. D., Shih, L., Teevan, J., & Karger, D. R. (2003). Tackling the poor assumptions of Naive Bayes text classifiers. In *Proceedings of the Twentieth International Conference on Machine Learning* (pp. 616-623), Washington DC: AAAI.

Riordan, R., Storkenmaier, A., & Wagener, M. (2010). Public information arrival: Price discovery and liquidity in electronic limit order markets. *Working paper.* Retrieved May 15, 2011 from http://ssrn.com/abstract=1620425

Ritter, J. R. (1991). The long-run performance of initial public offerings. *Journal of Finance*, 46, 3-27.

Roll, R. (1988). R^2. *Journal of Finance*, 43, 541-566.

Ryan, P., & Taffler, R. (2004). Are economically significant stock returns and trading volumes driven by firm-specific news releases?. *Journal of Business Finance & Accounting*, 31, 49-81.

Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes Twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World Wide Web* (pp. 851-860), Raleigh, NC: ACM.

Schmitz, P. (2007) Market and individual investors reactions to corporate news in the media. *Working paper.* Retrieved May 15, 2011 from http://ssrn.com/abstract=1004488

Soroka, S. N. (2006). Good news and bad news: Asymmetric responses to economic information. *Journal of Politics*, 68, 372–385.

Sprenger, T., & Welpe, I. (2010). Tweets and trades: The information content of stock microblogs. *Working paper.* Retrieved May 15, 2011 from http://ssrn.com/abstract=1702854

Storkenmaier, A., Wagener, M., & Weinhardt, C. (2010). Public information in fragmented markets. *Working paper.* Retrieved May 15, 2011 from http://ssrn.com/abstract=1660965

TechCrunch (2011). Twitter tweets some big Q1 stats: 155 million tweets a day now. *TechCrunch Blog*, Retrieved May 15, 2011 from http://techcrunch.com/2011/04/06/twitter-q1-stats/

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

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

Tetlock, P. C. (2011). All the news that's fit to reprint: Do investors react to stale information?. *Review of Financial Studies*, 24, 1481-1512.

Thompson, R. B., Olsen, C., & Dietrich, J. R. (1987). Attributes of news about firms: An analysis of firm-specific news reported in the Wall Street Journal index. *Journal of Accounting Research*, 25, 245-274.

TIME (2009). Turning Wall Street on its head. *TIME magazine online edition (author Douglas McIntyre)*, Retrieved May 15, 2011 from http://www.time.com/time/specials/packages/article/0,28804,1901188_1901207_1901198,00.html

Tumarkin, R., & Whitelaw, R. (2001). News or noise? Internet postings and stock prices. *Financial Analysts Journal*, 57, 41-51.

Wysocki, P. (1998). Cheap talk on the web: The determinants of postings on stock message boards. *Working paper.* Retrieved May 15, 2011 from http://ssrn.com/abstract=160170

II.4 Essay 4

Followers and Foes:
Industry Classification based on Investor Perceptions of Strategic Peer Groups

Abstract

Delineating industry groups of related firms and identifying strategic peers is important for both financial practitioners and scholars. Our study explores whether the degree to which pairs of companies are associated with each other in an online stock forum, where users subscribe to (i.e., follow) news messages posted by other users, is related to the comovement of their stocks. We find that our news-based measure of relatedness can explain stock returns with the same power as the established SIC classification scheme. We investigate, whether our method can serve to define strategic peer groups and conclude that news-based relatedness can help delineate meaningful industry groups and identify a firm's strategic "followers and foes".

JEL Classification: C81; G11

Keywords: Twitter; microblogging; stock market; industry classification; strategic peer group; comovement

Current status: Submitted and currently under review at the *Journal of Accounting Research*

Acknowledgements: This paper contains elements of joint work with Prof. Dr. Isabell M. Welpe.

1 Introduction

"Our ability to define industries is more art than science."

Dranove, Peteraf, and Shanley (1998)

Delineating industry groups of related firms and identifying strategic peers is important for both financial practitioners and scholars. Among financial professionals, industry-related stock market indices, such as Dow Jones' Industrial Average or Internet Composite Index, are the most important indicators of market developments. Investment managers tend to take industry affiliations into account when they structure portfolios and corporate managers often need to identify the most relevant competitors for peer-group comparisons (Chan, Lakonishok, & Swaminathan, 2007). In scholarly publications, industry classifications are used by hundreds of scientific studies and serve 4 primary purposes (Kahle & Walkling, 1996): first, to identify control (matched) firms (48% of studies using industry classifications), second, to describe the industrial composition of the sample (35%), third, to restrict the sample of interest (32%), and fourth, to categorize acquisitions and divestitures as conglomerate or nonconglomerate (9%). In addition, economists, who group firms that supply the same market in industry analysis (Grant, 2010), and strategy researchers, who investigate subgroups within an industry that pursue a similar strategy (e.g., Hunt, 1972; Porter, 1979), have an interest in assigning companies to the most relevant industry classes.

However, identifying related companies and defining industry groups can be challenging. Fan and Lang (2000) have pointed out that "objectively measuring firm relatedness on a large sample is difficult" (p. 629) and "some practitioners even suggest that the selection of comparable firms is essentially an 'art form' that should be left to professionals" (Bhojraj & Lee, 2002, p. 408). There is much empirical evidence illustrating the limited power of established classification schemes. For example, in a cross-sectional analysis of 63 companies, King (1966) finds the industry to explain only about 10% of the variance in quarterly stock returns. Many recent studies have called into question the accuracy of popular methods, such as SIC codes, for industry classification (e.g., Bhojraj, Lee, & Oler, 2003; Clarke, 1989; Fan & Lang, 2000). In particular, Bhojraj et al. (2003), who evaluate the usefulness of industry classifications based on their power to explain cross-sectional stock comovements within an industry, find the SIC to be of limited use.

Yet only a few studies offer alternative methods for industry classification, for example based on commodity flow data (Fan & Lang, 2000) and joint analyst coverage of multiple firms (Ramnath, 2002). However, those methods are either limited to specific types of firms (e.g., companies with a large analyst following; Ramnath, 2002) or vulnerable to subjective judgments (Fan & Lang, 2000). All of these methods provide industry classifications that are fairly stagnant (i.e., they do not reflect changes in firms' relatedness quickly) and provide us with a nominal categorization only (i.e., they do not allow us to interpret the relative strength of firm relatedness).

In this study, we propose an alternative approach to define industry groups based on investor perceptions of the impact of information on various groups of stocks. Ultimately, stock prices are driven by information and the degree to which new information affects two companies (e.g., news stories mentioning both firms) should thus allow us to derive a measure of their relatedness (i.e., the relative frequency with which two stocks are mentioned together). We use the degree to which pairs of companies are associated with each other in an online stock forum to develop a network of relationships and leverage methods from social network analysis in order to extract cohesive groups from an investor perspective. If market participants consider a set of companies closely related, their stocks should experience coincident movement (Chan et al., 2007). Thus, we explore return correlations in order to determine whether our method to delineate industry groups is economically meaningful.

Our study explores the following research questions. First, we investigate whether the degree to which pairs of companies are associated with each other in an online stock forum (i.e., relatedness) corresponds to the comovement of their stocks. Second, we explore, whether our measure of relatedness can serve to define strategic peer groups[100] for individual firms that are meaningful in terms of stock comovement and relative to SIC peers. Third, we analyze whether industry groups defined by our measure of relatedness are a viable alternative to established industry classification schemes.

We find that the degree to which companies are mentioned jointly in an internet stock forum can explain the comovement of their stocks. Our measure of relatedness can help identify a firm's strategic peers from an investor perspective and delineate industry groups, which explain stock returns with the same power as established classification methods, but

[100] Note that we use the term "strategic group" loosely referring to related firms from the perspective of an individual firm. It does not directly follow the definition traditionally used in strategy research referring to intraindustry subgroups of companies following a similar strategy (Hunt, 1972; Porter, 1979).

offers a number of promising advantages. Our approach is, for example, independent of the arbitrary assignment of companies by individual experts by leveraging the insights of hundreds of investors and, in contrast to fixed classification schemes, can quickly reflect changes in firm relatedness.

The main contributions of this study are as follows. First, our study provides empirical evidence supporting the theory that information associated with a set of firms is an indicator of relatedness and the comovement of their stocks (e.g., King, 1966). Second, leveraging methods from social network analysis, we present a novel news-based approach to determine industry classifications and define strategic peer groups from an investor perspective. Third, we demonstrate that our measure offers advantages over established classification schemes in that it provides a continuous measure of relatedness for every pair of two companies, which can be updated in real-time and at arbitrary intervals.

The remainder of the paper is structured as follows. First, we review related work and derive our research questions. Second, we describe our data set and methodology illustrating how online stock forums can be used to define industry groups. Third, we provide results of our analysis of the usefulness of our classification method with respect to the comovement of stocks, the identification of strategic peer groups and as a classification scheme vis-à-vis the established SIC classification system. We conclude that the user perception of strategic peers can be used to delineate meaningful industry groups. Finally, we discuss the implications of our findings and provide suggestions for further research.

2 Related work

There are various industry classification schemes assigning individual companies to a particular industry group (for a comprehensive overview, see Bhojraj et al., 2003). The Standard Industrial Classification (SIC) is the predominant classification system in capital market research, with more than 90% of relevant studies making use of this classification scheme (Bhojraj et al., 2003). SIC codes are based on industry categories defined by the U.S. Census Bureau reflecting similarities between firms with respect to the products they produce or the manufacturing technologies they employ (Clarke, 1989). It has become the primary method for delineating industrial activity in the United States. However, assigning individual companies to a particular industry code falls into the responsibility of data vendors such as CRSP and Compustat. Kahle and Walkling (1996) have shown that, due to competing assignments, even the choice of data vendor can substantially affect the financial research results. Clarke (1989), who has examined whether firms in the same SIC category exhibit similar changes in sales, profit or stock prices, concludes that SIC codes are not particularly successful at identifying firms with such similar characteristics. Responding to changes in the U.S. economy, the North American Industry Classification System (NAICS[101]) was developed, but has, in practice, not yet replaced the well established SIC system. In the meantime, financial practitioners have developed the Global Industry Classifications Standard (GICS[102]) and financial scholars are making adjustments to the SIC categorization for research purposes (Fama & French, 1997)[103]. In sum, due to significant shortcomings of established industry classification systems a considerable amount of time and effort is being spent to divide firms into homogenous groups.

However, few scholars so far have offered viable alternatives. Fan and Lang (2000) use commodity flow data from input-output (IO) tables to construct IO-based measures of interindustry and intersegment relatedness. However, commodity flows are available only for roughly 500 private-sector industries, and are thus not well-suited for firm-level analysis.[104] In a study of analyst reactions to earnings announcements, Ramnath (2002) as well as Zuckerman and Rao (2004) have used an analyst-based definition of industry groups. The authors define an industry as a group of firms having a certain number of security analysts in common. This approach to industry definition is, of course, limited to large companies with a sufficient analyst following. In equity research and valuation, accounting-based multiples (e.g., price-to-earnings, price-to-book ratios) are often used to select comparable firms, but this approach is meaningful only within an industry previously specified by standard classification schemes (Bhojraj & Lee, 2002). Next to these objective approaches, there are a

[101] The NAICS uses a more production-based framework. SIC and NAICS are both maintained by government agencies interested in collecting broad industrial statistics and have a similar hierarchical structure grouping industries according to similarity in the process used to produce goods or services. As a result there is a high degree of correspondence among SIC and NAICS classifications (Bhojraj et al., 2003) and most of the shortcomings of the SIC also apply to the NAICS.

[102] The GICS was developed jointly by Standard&Poor's and Morgan Stanley Capital International (MSCI). It is targeted towards financial professionals and, in contrast to the production-oriented focus on SIC and NAICS, puts greater emphasis on a company's sources of revenues and the market perception of principal business activities as revealed by analyst reports (Bhojraj et al., 2003). However, GICS classifications are a commercial product and not widely available.

[103] Fama and French (1997), for example, have defined 48 industry groups, which are more likely to share common risk characteristics. However, their so-called FF industry classification is basically a reclassification of existing SIC codes.

[104] For firm-level analysis, Fan and Lang (2000) use their own conversion table to link company SIC codes to IO codes and limit the use of their relatedness measure to the analysis of corporate diversification strategies among multi-segment firms.

few studies, particularly related to the literature on firm and portfolio diversification, which use subjective criteria to classify companies into economic sectors (e.g., Rumelt, 1982) or broad categories characterized by growth, cyclical and stable return characteristics (Farrell, 1974) and are thus dependent on expert judgment. In addition, there are attempts to group companies outside the field of financial research that are related to our objective. Strategy researchers identify strategic groups by clustering firms based on firm-level dimensions that characterize strategy (e.g., cost structure, degree of product diversification, formal organization; DeSarbo, Grewal, & Wang, 2009; Dranove, Peteraf, & Shanley, 1998). However, due to the industry-specific definition of firm-level variables these methods are largely limited to defining subgroups within one particular industry and do not allow to group firms across the entire industry landscape. Finally, in econophysics, graph-theoretical methods have been applied to develop networks of financial markets using stocks as nodes and stock comovement as indicators of the strength of their ties (e.g., Bonanno et al., 2004; Mantegna, 1999; Onnela et al., 2003). Using similar methods, practitioners have identified strategic opportunities by drawing maps of competing firms based on semantic clusters of key phrases associated with individual companies in millions of corporate documents (Gourley, 2011). Exploratory and qualitative evidence from these studies suggest that a network perspective of company relationships can be leveraged to derive groups of stocks that are homogeneous with respect to traditional industry classifications and may be used to design stock indices (Tse, Liu, & Lau, 2010). However, research in this field focuses on the analysis of graph-theoretical properties and network topologies without a rigorous, quantitative comparison with existing industry classification systems.

We propose an alternative approach to define industry groups based on the perception of the impact of information on various groups of stocks. Ultimately, stock prices are driven by information. King (1966) has pointed out that “the stock market is subject to a steady inflow of information, much of which will […] fall into various classes according to the scope of its effect on the market” (p. 140). There are some news items, which will have a market-wide impact (e.g., changes in monetary policy), other information, which will affect only a subgroup of stocks (e.g., changes in defense policy affecting the aircraft industry), and a third class of information, which will be relevant only to a particular security (e.g., earnings announcements).[105] The degree to which new information affects two companies should thus

[105] The general theory outlined by King (1966) is widely accepted and incorporated in financial research. On one end of the spectrum, the idea that there are common economic factors (e.g., interest rates, inflation) affecting all

allow us to derive a measure of relatedness. While this line of argument is intriguing, it is usually difficult to observe such as "steady inflow of information" and link it to a particular set of stocks. However, the rise of online stock forums provides us with a unique data source documenting previously unavailable facets of information processing by online investors in real-time (for an example of the use of internet stock message boards in academic research, see Antweiler & Frank, 2004). A working paper investigating the online stock forum Yahoo!Finance indicates that stocks that are associated with each other on internet message boards, exhibit stronger comovement than other stocks (Das & Sisk, 2003). Even though these results are encouraging with respect to our hypothesis, this working paper has two limitations, which we address in our study. First, Das and Sisk (2003) define relatedness as a large share of common users on the message boards of two companies suggesting that "message boards belonging to the same community may result in similarity of opinion reflected in stock trades, ultimately impounded in stock returns" (p. 8). However, users often leave messages at different points in time and thus a common user base among message boards may not necessarily translate into "similarity of opinion reflected in stock trades". Therefore, instead of studying online users as the carriers of static information sets, our study focuses on individual bits and pieces of information that are directly associated with a set of stocks in real-time. Second, Das and Sisk (2003) limit their analysis to the existence of stronger comovement among related stocks without comparing the strength of this effect to objective benchmarks. We leverage our measure of relatedness to delineate industry groups and compare these to existing classification schemes.

In line with previous research (Bhojraj et al., 2003; Chan et al., 2007), we focus the comparison of these classification schemes on their explanatory power for stock comovement as a benchmark of firms' similarity. The literature distinguishes between fundamentals-based and industry-specific comovement (Barberis, Shleifer, & Wurgler, 2005; Pindyck & Rotemberg, 1993). According to the fundamentals-based theory, comovement of stock returns can be linked to similar firm fundamentals (i.e., the assets owned by a firm). The theory of industry-specific comovement attributes comovement of stocks to market frictions and

---

stocks, for instance, are reflected in the arbitrage pricing model theory (Ross, 1976) and have been tested empirically (Chen, Roll, & Ross, 1986). On the other end, much of the event-study literature explores firm-specific news events affecting a particular company. In contrast to that, the empirical evidence exploring the link between news events affecting a specified set of companies and their stock price is scarce, possibly because this data is not easily available. However, empirical evidence of intra-industry information transfer (e.g., a market reaction to news affecting related firms as illustrated by Baginski (1987), Kim, Lacina, and Park (2008) and Ramnath (2002)) suggests investors' awareness of the impact of individual news items on several related stocks.

investor sentiment. Barberis et al. (2005) suggest that industry-specific comovement may result from investor preference for certain industries (habitat view) or an industry focus due to limited processing power (category view). In either case, investors direct their funds on the industry-level resulting in industry-specific comovement. Our approach to industry classification is based on investor perceptions of firms' relatedness, which may reflect both sources of comovement. In other words, investors may mention two stocks jointly because of their perception of similar fundamentals or due to their preference for certain industries.

3 Data set and methodology

3.1 Data set and sample selection

In this section, we describe our data set and detail the methodology used to derive our measure of relatedness between firms. The advent of online stock forums has made observable many aspects of information processing by investors, which were previously unavailable. Internet stock forums allow investors to exchange stock-related information and trading ideas online. Das, Martinez-Jerez and Tufano (2005) have profiled users of these forums and suggest that the majority of them are individual investors and day traders. We chose the microblogging platform Twitter as our data source. Twitter allows users to publish short messages with up to 140 characters, so-called "tweets".[106] 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 millions of messages per day. Many of these messages discuss public companies, trading ideas and current news. Some commentators see in the conversations on this platform "the modern version of traders shouting in the pits" (BusinessWeek, 2009). The investor community has come to call Twitter and related third-party applications "a Bloomberg for the average guy" (BusinessWeek, 2009). Academic researchers were only recently drawn to Twitter as a field for capital market research. A few studies suggest that the information content of Twitter messages may help predict macroeconomic indicators such as the Index of Consumer Sentiment (O'Connor, Balasubramanyan, Routledge, & Smith, 2010), stock market indices such as the Dow Jones Industrial Average (Bollen, Mao, & Zeng, 2010) or the S&P 500 (Zhang, Fuehres, & Gloor, 2010) and even returns and trading volume of individual stocks (Sprenger & Welpe, 2010).

[106] We will refer to these tweets as messages or news items.

Traders have adopted the convention of tagging stock-related messages by a dollar sign followed by the relevant ticker symbol (e.g., "$AAPL"). We focus on this explicit subset of messages. This focus allows us to investigate the most relevant news items and avoid "noise" (i.e., messages that are not related to publicly traded companies). Messages are accessible via the website's application programming interface (API). We study the 6 month period between January 1st and June 30th, 2010, to deal with stable developments on the U.S. financial markets and to avoid potentially distorting repercussions of the subprime mortgage crisis in 2009. During this period, we have collected 439,960 English-language, stock-related microblogging messages containing the dollar-tagged ticker symbol of an S&P 500 company.[107] We focus on the S&P 500 to adequately reflect a wide spectrum of U.S. equities, which permits a cross-industry analysis, while limiting our study to well-known companies that trigger a substantial number of stock microblogs.[108]

3.2 Investor perceptions of strategic peer groups

Table 1 shows several random examples of messages from stock microblogs. In line with King (1966), we find news items that investors relate to one particular stock (e.g., *"$TGT Target Q4 Profits Surge*") as well as others that are associated with multiple firms (e.g., "*Energy doing well. $CHK $OXY*"). Roughly 13.4% of all messages mention more than one stock. Reasons for these joint mentions include the impact of macroeconomic developments (e.g., "*Big banks up or down with Bernanke's re-nomination? $C $BAC $WFC*"), the launch of new products affecting competitors (e.g., "*Crazy Google now building super-high-speed fiber Internet network to scare Comcast and AT&T: http://bit.ly/dvWSzL $GOOG $T $CMCSA"*) and legal actions of one firm against another (e.g., "*Goldman Sachs $GS demands 4 billions from AIG $AIG to cover mortgage securities AIG insured, helped trigger crisis, forced gov to bail out AIG"*). Irrespective of the content of individual messages, they all indicate that one company is associated with another and impacted by the same piece of news. These messages containing joint mentions are the focus of our analysis.

[107] 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 ultimately providing us with an uninterrupted stream of messages for the 6 months covered in this study.
[108] Specifically, we focus on those companies that have been included in the S&P 500 as of January 1, 2010.

Table 1: Sample messages

ID	Text
8309332080	$USU, $MDR, $TOSBF, $GE, $FLR, $SGR, & $HIT are some who could benefit from Obama's call for more nuclear energy. Lets see what happens.
8338276214	Big banks up or down with Bernanke's re-nomination? $C $BAC $WFC
8501641364	Energy doing well. $CHK $OXY
8635566800	$GS below R3, $JPM, $BAC holding at it, $WFC above. waiting for all to drop it to unleash the power of $FAZ
8814470372	$XTO (XTO Energy Inc) $45.75 has crossed its 50 day moving average: $45.55 #empta #stocks http://empirasign.com/l/8l4r.htm
8823022876	Goldman Sachs $GS demands 4 billions from AIG $AIG to cover mortgage securities AIG insured, helped trigger crisis, forced gov to bail out AIG
8838995237	Semis: The Market's Sleeper Stocks http://bit.ly/cuNEjv $ALTR $AVT $BRCM $CSCO $CY $INTC $MRVL $TXN $XLNX $^SOX #StockPicks #StockMarket
8898690783	BofA/Merrill Lynch upgraded Dell $DELL from Neutral to Buy and raised their price target from $16.50 to $18
8909019113	Crazy Google now building super-high-speed fiber Internet network to scare Comcast and AT&T: http://bit.ly/dvWSzL $GOOG $T $CMCSA
9486553527	$FSLR $CSIQ $STP $SOLF $TAN solars break out
9534086413	How To Listen To The Goldman Sachs Tech Conference LIVE Online $GOOG $MSFT $AAPL by @ncsaint http://bit.ly/bDcGSm
9559578919	$GOOG and $AMZN Low Volatility Bearish Flag: http://short-termtrading.blogspot.com/2010/02/google-inc-goog-and-amazon-amzn-low.html
9561185167	$TGT Target Q4 Profits Surge http://bit.ly/ciQFjY
9684326214	$GENZ (Genzyme Corp) $57.61 is trading at a 3 month intra-day high. #emphl #stocks http://empirasign.com/s/2ox

Notes: Tweets were randomly selected and are shown in their original format.

We investigate whether joint mentions can serve as a measure of relatedness between firms. To make the comparison more flexible and the interpretation more straightforward, we focus on the pairwise relationships between companies. Based on the overall probability that any one firm is mentioned in a message, a conditional probability that two firms are mentioned together can be computed. If all combinations were equally likely, this conditional probability should be equal to the observed share of messages mentioning these two firms. Due to different base rates, we divide the observed share of joint mentions by the conditional probability to derive a comparative measure. If *share(AAPL, GOOG)* represents the share of observed joint mentions of these two stocks, the relative frequency *R*, is calculated as follows:

$$(1) \qquad R = \frac{share(AAPL, GOOG)}{\left[P(AAPL|GOOG) + P(AAPL|GOOG)\right]/2}$$

The relative frequency is a measure of *Relatedness* and illustrates how often two stocks are mentioned together relative to the random probability based on the overall "share of voice" of the individual stocks. If R equals 1.5 the share of observed joint mentions is 50% higher than pure chance would suggest (i.e., $R = 1$).[109] This measure has been used successfully in the context of microblogs to derive a measure of the relative frequency of joint mentions of political parties, which was related to their ideological proximity (Tumasjan, Sprenger, Sandner, & Welpe, 2010). We have limited our analysis to firms that were mentioned at least 100 times in our sample period, leaving us with 415 stocks from the S&P 500.[110] Table 7 provides an overview of the ticker symbols and official company names of these stocks. Table 2 displays a list of the most and least related stocks in our sample. Companies that are associated with one another more closely tend to come from the same top-level SIC industry group. The comovement of their stocks is, on average, substantially higher than the comovement of stocks that are mentioned jointly less frequently ($c = 0.66$ *vs.* 0.45 with $p < 0.01$).

[109] To limit the effect of outliers, we use the logged version of relatedness in regression analysis.
[110] We included this filter to focus the analysis on companies, which were mentioned in a substantial number of stock microblogs and thus actually associated with other companies by online users. However, we have confirmed the robustness of our results by repeating our analyses with all stocks of the S&P 500. The results do not change in any material way and we, therefore, do not report them separately.

Table 2: Most and least related firms

	Stock 1		Stock 2		***Relatedness***	**Comovement**	**Same SIC**
	Ticker	Name	Ticker	Name			
Top 10	AYE	Allegheny Energy	FE	FirstEnergy Corp	685.3	0.64	1
	PGN	Progress Energy Inc.	SO	Southern Co.	549.9	0.81	1
	AIV	AIMCO	WMB	Williams Cos.	456.8	0.71	0
	FIS	Fidelity National Information Services	FISV	Fiserv Inc	399.7	0.34	1
	FPL	FPL Group	PGN	Progress Energy Inc.	356.4	0.85	1
	FPL	FPL Group	SO	Southern Co.	340.1	0.76	1
	MDT	Medtronic Inc.	STJ	St Jude Medical	337.5	0.76	1
	BXP	Boston Properties	VNO	Vornado Realty Trust	311.3	0.90	1
	KR	Kroger Co.	SWY	Safeway Inc.	311.1	0.59	1
	CTL	CenturyTel Inc.	Q	Qwest Communications Int.	299.1	0.24	1
	…						
Bottom 10	ADBE	Adobe Systems	BAC	Bank of America Corp.	0.1	0.52	0
	C	Citigroup Inc.	ORCL	Oracle Corp.	0.1	0.57	0
	GOOG	Google Inc.	HAL	Halliburton Co.	0.1	0.44	0
	AAPL	Apple Inc.	KFT	Kraft Foods Inc.	0.1	0.12	1
	MS	Morgan Stanley	MSFT	Microsoft Corp.	0.1	0.44	0
	AAPL	Apple Inc.	WFC	Wells Fargo	0.1	0.56	0
	GOOG	Google Inc.	WFC	Wells Fargo	0.1	0.51	0
	GS	Goldman Sachs Group	ORCL	Oracle Corp.	0.0	0.31	0
	AAPL	Apple Inc.	HBAN	Huntington Bancshares	0.0	0.62	0
	GOOG	Google Inc.	HBAN	Huntington Bancshares	0.0	0.37	0

Notes: This table shows the firms that are most and least related in terms of our measure of relatedness. Results were limited to pairs of stocks with at least 10 joint mentions. The most related firms tend to come from the same SIC industry group and their stocks exhibit, on average, higher comovement.

3.3 Identification of strategic peers and delineation of industry groups

Our measure of relatedness informs us about the news-based proximity of any pair of stocks. In this section, we will introduce the methods used to meet our objective to leverage this information in order to derive a firm's strategic peers and delineate relevant industry groups. The pairwise relationships that our measure of relatedness provides us with essentially create a network of all stocks in our sample. Figure 1 shows a graph for an extract from this network. The network contains information regarding the link (i.e., a line or, in network theory, the edge or tie) between two stocks as well as the strength of this link (i.e., the thickness, which represents *Relatedness*), the SIC industry classification (i.e., the shape of the symbol) and the absolute volume of company mentions (i.e., the size of the stock symbol). In line with the approach taken by Das and Sisk (2003) and researchers in the field of econophysics (e.g., Bonanno et al., 2004; Mantegna, 1999; Onnela et al., 2003), from this network we will extract strategic peers related to a particular stock as well as cohesive subgroups that define an industry.[111]

Taking the perspective of a particular company, we can simply define the peer group as those firms that are most closely related to it, i.e., firms with the highest measure of relatedness.

The above mentioned approach cannot be employed to delineate clear-cut industry groups which clearly assign each company to exactly one category. Thus, to be consistent with established industry classification systems, we also want to partition the network into a pre-determined number of groups. In social network analysis, so-called faction analysis is used to achieve this objective. In short, this clustering algorithm optimizes a cost function, which measures the degree to which a partition forms a cohesive subgroup and takes into account both the ties as well as their strength (for details regarding this method, see Glover, 1990). Cohesiveness can be thought of as the average tie strength within each partition. We use the Ucinet implementation for the faction analysis of our dataset[112] (Borgatti, Everett, & Freeman, 2002).

[111] Note that this approach is related to that in strategy research where strategic groups are identified by clustering firms based on firm-level variables such as R&D expenditures or distribution of sales (e.g., DeSarbo et al., 2009).
[112] In Ucinet 6 the procedure can be found under Tools/Cluster Analysis/Optimization.

3.4 Similarity of stocks

In order to objectively evaluate the quality of our measure of relatedness as a proxy for firms' similarity, we need to define external benchmarks for comparison. In this section, we will describe stock comovement and existing SIC industry classifications as input parameters to construct these benchmarks.

At the level of company pairs, we can compare *Relatedness* to the comovement of stocks. In line with related research (e.g., Bhojraj et al., 2003), we calculate comovement as the correlation of stock returns. To isolate the market-related component of this comovement, we start with a market model

(2) $r_i = \alpha_t + \beta_i \, r_m + \varepsilon_i$,

in which the return of a stock r_i is explained by a firm-specific component α_t and a market-related component, which depends on the stock's sensitivity (β_i) to overall market returns r_m. $\varepsilon_{i,t}$ is a time- and company-specific error term. The S&P 500 serves as our proxy for the market return. The market-related comovement of two stocks i and j can then be computed as

(3) $\sigma_{i,j} = \beta_i \, \beta_j \, (\sigma_m)^2$,

where σ_m is the standard deviation of the market return. We use our entire 6 month sample period to estimate the above mentioned parameters.

At the aggregate level, the most obvious comparison for our industry classification is the widely established SIC system. There are varying levels of granularity in the SIC classification scheme. For the purpose of our study, we use the most common two-digit level of analysis, which Clarke (1989) suggests to have the greatest explanatory power with respect to industry returns.[113] In order to calculate the explanatory power of our classification for the stock returns of the companies in a peer or industry group, we model a firm's stock returns, $r_{i,t,}$ with the simple OLS regression

(4) $r_{i,t} = \alpha_t + \beta \, r_{ind,t} + \varepsilon_{i,t}$,

where r_{ind} is the equally weighted industry return. This permits us to evaluate the explanatory power of alternative industry definitions by comparing adjusted R^2s of various industry definitions, following the methodology use by Bhojraj et al. (2003). All returns are calculated

[113] Others, such as Fertuck (1975), suggest that three-digit SIC codes are the most useful in predicting return variation. Thus we have replicated our analyses at the three-digit level and find our results to be largely in line with those that we report in this paper. However, we prefer the less granular two-digit level of analysis for our relatively small sample of 500 firms to avoid the exclusion of too many industry groups with too few sample companies.

as the log difference of total return to shareholders, which reflects both price changes and dividend payments.

3.5 QAP methodology

When we compare our measure of relatedness to the comovement of stocks, we are dealing with sets of company pairs. The pairwise structure of so-called dyadic data requires special attention because every company is included in multiple sets. We can think of the data as a matrix with every stock listed in one row and one column and the attributes of the resulting pairs (e.g., relatedness, comovement) occupying the cells. Thus, the observations are not independent preventing the use of ordinary regression models if the dependency between the observations is not controlled for. This is a common phenomenon in social network analysis, where the resampling-based nonparametric Quadratic Assignment Procedure (QAP) is used to deal with dyadic data sets (for a comprehensive description of the QAP, see Krackhardt, 1988). Similar to bootstrapping methods, the QAP permutes the rows and columns of the above-mentioned matrix, but maintains rows and columns for individual companies. As a result, the permuted datasets comply with the null hypothesis.[114] Resampling multiple times and running OLS regressions on the “scrambled datasets” permits us to determine the percentile of the original coefficients relative to the empirical distribution of permuted datasets. In sum, we employ the QAP to account for the pairwise nature of our dataset.

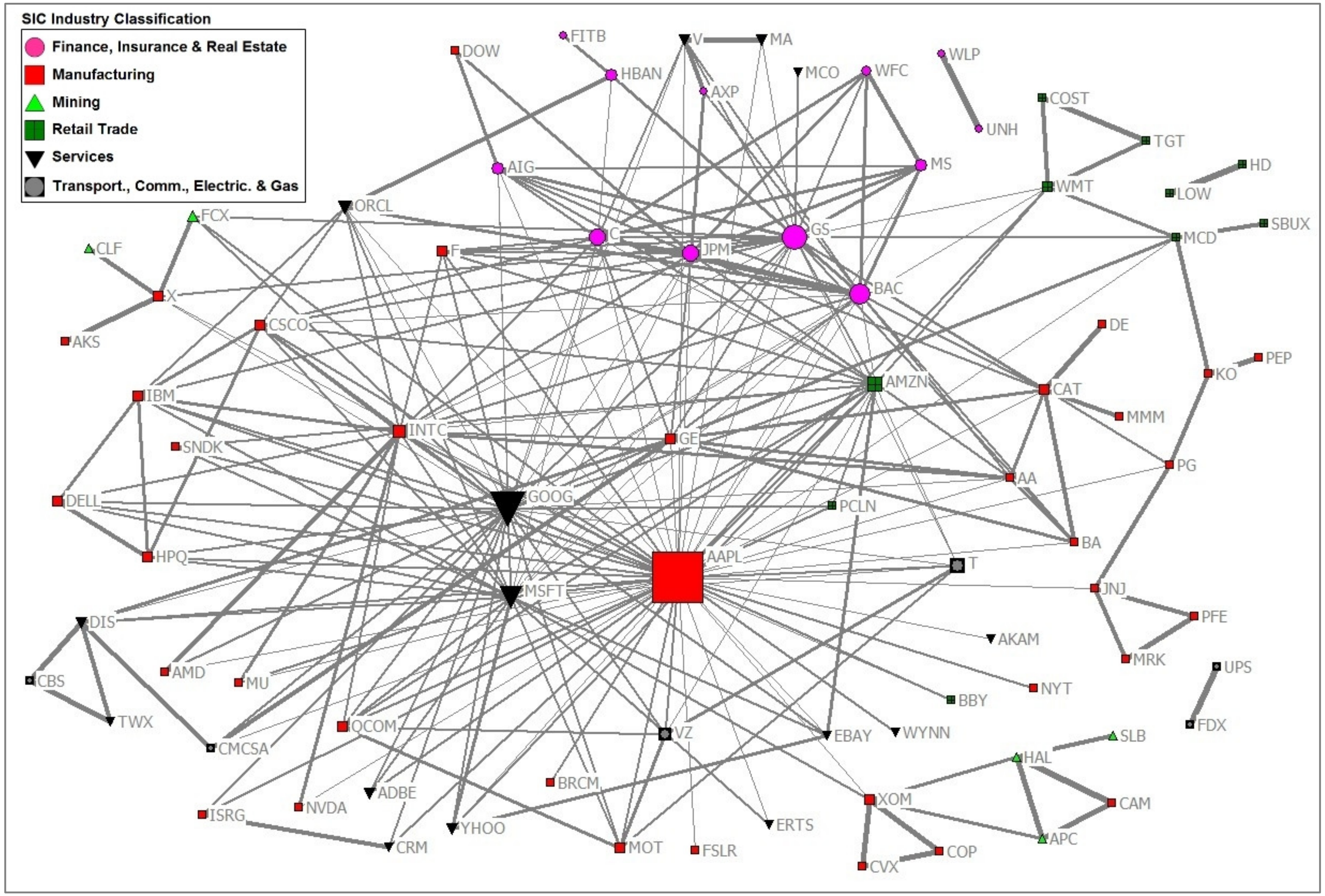
4 Results

4.1 Overview and anecdotal evidence

The results section is structured as follows. We will first highlight noteworthy aspects of the network graph of all stocks introduced in our methodology section in order to provide first descriptive evidence of identifying company groups with our proposed methodology. Then, we will turn to our three research questions and investigate, first, whether relatedness of companies in the eyes of online investors can explain the comovement of their stocks, second, whether relatedness can serve to identify a company’s strategic peers and, third, if we can extract meaningful industry groups from the network of relationships.

[114] The null hypothesis is the hypothesis that there is no relationship in the matrix.

Figure 1: Investor perceptions of the relationship of S&P 500 stocks



Notes: This figure shows the relationship of S&P 500 stocks in terms of joint mentions in stock microblogs. The size of the stock symbol represents the total number of mentions, the thickness of the lines is indicative of the relative frequency of joint mentions (i.e., Relatedness as defined in our methodology section). For better readability the figure was limited to stocks that were mentioned together at least 50 times.

Figure 1 shows the relationship of S&P 500 stocks in terms of joint mentions in stock microblogs. For a rough comparison with the SIC coding, the shape and color of each stock symbol represents the one-digit SIC industry group. The layout of the network graph is not the product of a random process, but was automatically derived in order to maintain roughly an equal distance between nodes while ensuring readability through node repulsion. As a result, apart from tracing direct links, we can interpret proximity of a group of stocks as the degree to which they are related. We find that many subgroups of stocks are consistent with our intuition of classic industry delineations. Financial firms, such as Goldman Sachs (GS), JP Morgen (JPM), and Bank of America (BAC) are closely interconnected. There are also tight-knit smaller groups of stocks, such as the media companies Disney (DIS), CBS Broadcasting (CBS) and Time Warner (TWX). In addition, some subgroups exist that are not connected to the rest of the network, for example the insurance companies WellPoint (WLP) and United Health (UNH), the logistics firms United Parcel Service (UPS) and Fedex (FDX), and the hardware stores Home Depot (HD) and Lowe's (LOW). Their isolated position suggests that these companies form micro industries that are often subject to the same news items, but not frequently associated with other firms. The network graph not only confirms classical industry groupings, but also reveals interesting connections between these industries. For example, while Exxon Mobile (XOM) is obviously closely related to other major energy firms such as Chevron (CVX) and ConocoPhillips (COP), it is also linked to subcontractors such as the exploration firms Anadarko Petroleum (APC) and Halliburton (HAL), which in turn are associated with their equipment suppliers Schlumberger (SLB) and Cameron (CAM). Note that these relationships cut across traditional SIC categories. Online retailer Amazon (AMZN) is another interesting example in this respect, because the company appears to be a hub that is associated with traditional "brick and mortar" retailers, such as Walmart (WMT), Target (TGT) and Costco (COST), online retailers, such as Priceline (PCLN) and Ebay (EBAY), computer soft- and hardware firms, such as Apple (AAPL), Microsoft (MSFT) and Intel (INTC), and communication providers, such as AT&T (T). The network also highlights particularly competitive relationships, such as those between Coke (KO) and Pepsi (PEP) or Visa (V) and Mastercard (MA), as well as issues related to corporate control and ownership, such as the strong link between Disney (DIS) and its parent company General Electric (GE).

We conclude, that the news-based investor perception of strategic peer groups offers unique and rich descriptive insights into the relationship between companies. In the following, we will investigate whether we can leverage this insight to identify meaningful peer groups and establish a link to stock market returns.

4.2 Comovement

In this section, we will explore whether our news-based measure of relatedness is indicative of stock comovement. In order to deal with the dyadic nature of our dataset we employ the QAP introduced in our methodology section. Following related studies that have used a SIC-based variable of relatedness (which equals one if two companies belong to the same SIC code and zero otherwise) and market-related comovement as control variables (e.g., Fan & Lang, 2000), we likewise control for these two measures in our model.

Table 3 shows the results of the QAP regression. Model 1 uses daily comovement as the dependent variable (left panel), whereas model 2 is based on the comovement of weekly returns (right panel). Next to the results for the QAP regressions and as a robustness check, we provide the results for clustered regressions, where robust standard errors are adjusted for intragroup correlation (with one company in each pair defining the group). The independent variables were standardized to compare relative effect strengths. We can see that the same-industry-dummy has the strongest explanatory power for stock comovement ($c = 0.066$). It is roughly 6 times as strong as general market-related comovement ($c = 0.012$). However, beyond these traditional measures of proximity, our measure of relatedness has statistically significant explanatory power for stock comovement ($c = 0.006$). These results, even though based on short-lived news items, are not limited to daily stock comovement. The analysis of weekly comovement shows a very similar pattern (right panel of Table 3). We conclude that relatedness can help explain the comovement of stock returns over and above considering the presence of both companies in the same SIC industry.

The two models in Table 3 were based on relatedness and comovement over our entire sample period. However, our news-based measure of relatedness may have a crucial advantage over static SIC classifications. News-based relatedness could be updated frequently and reflect changes over time. The results from Table 3 do not show whether our measure of relatedness can adjust to changes in firm proximity quickly. Therefore we have conducted OLS-based QAP panel regressions with a time series of *Relatedness* and stock comovement of all company pairs at a monthly interval. In this case, the permutations described in our methodology section ensure that the time series of the observation corresponding to a row or column in the matrix is kept together. Table 4 shows the results of these regressions. Model 2 illustrates that changes in *Relatedness* over time are positively associated with changes in the comovement of daily stock returns. As the change in R^2 shows, *Relatedness* can partially explain comovement even when market-related comovement and SIC-industry dummies are included. Thus, at least at monthly intervals, an increase in *Relatedness* reflects an increase in stock comovement.

Table 3: Explaining stock comovement through *Relatedness*

	Model 1		**Model 2**	
Comovement interval	*Daily*		*Weekly*	
Estimation method	*QAP*	*Clustered*	*QAP*	*Clustered*
Investor perception				
Relatedness	0.006 ***	0.006 ***	0.0068 ***	0.0068 ***
SIC industry component				
same SIC dummy	0.0666 ***	0.0666 ***	0.0978 ***	0.0978 ***
Market component				
Market-related comovement	0.0118 ***	0.0118 ***	0.0131 ***	0.0131 ***
R-squared	0.234	0.234	0.264	0.264
F-value	8838.35 ***	169.1 ***	10385.4 ***	169.1 ***

Notes: This table shows the explanatory power of our measure of relatedness for stock comovements. Stock comovement serves as the dependent variable. The independent variables were standardized to compare relative effect strengths. We can see that relatedness adds to the explanation of comovement beyond market-related comovement and the industry component due to traditional SIC classifications. The sample is based on all pairs of 415 firms (N = 171,810).

Regarding the estimation methods: We have used the QAP with 500 iterations of OLS estimations. In addition, we provide the results for clustered regressions, where robust standard errors are adjusted for intragroup correlation (with one company in each pair defining the group), as a robustness check of our results.

**** indicates significance at the 1% level.*

Table 4: Time-series regression of *Relatedness* and stock comovement

	Model 1	Model 2
Comovement interval	*Daily*	
Estimation method	*QAP*	
Investor perception		
Relatedness		0.0148 ***
SIC industry component		
same SIC dummy	0.0167 ***	0.0141 ***
Market component		
Market-related comovement	0.1383 ***	0.1378 ***
F-value	2189.8 ***	1476.3 ***
R-squared	0.272	0.275
Likelihood-ratio test		
2 × Δ (log likelihood)		48.16 ***

Notes: This table shows the results for an OLS-based QAP panel regressions with a timeseries of Relatedness and stock comovement of all company pairs at a monthly interval. The QAP permutations described in our methodology section ensure that the time series of an observation corresponding to a row or column in the matrix is kept together. The time series consists of monthly observations of the dependent variable stock comovement at daily intervals and the independent variables Relatedness etc. for the same 4 week period. The independent variables were standardized to compare relative effect strengths. The coefficients illustrate the explanatory power of our measure of relatedness for stock comovement over time. We can see that changes in Relatedness are positively associated with changes in stock comovement. Thus, this measure reflects changes in firm relatedness. The sample is based on all pairs of 415 firms (N = 171,810 company pairs).
**** indicates significance at the 1% level.*

4.3 Strategic peer groups

In this section, we will investigate whether relatedness can help identify a firm's strategic peers. For selected companies we compare the peers according to the SIC and our measure of relatedness as well as their explanatory power for industry returns.

Table 5 shows the 10 two-digit SIC groups with more than 20 sample companies. We have selected the company with the highest message volume in our sample for each of these industry groups and show all sample companies with the same SIC code. For comparison, we have chosen the same number of most closely related companies in terms of *Relatedness*. Peers found in both groups are bolded. We can see that the overlap of peers derived by the two methods is around one quarter (23%). Adjusted R^2s are shown for the OLS regressions that try to explain stocks returns through industry returns, where the industry is defined as the specified peer group. While the explanatory power of SIC groups for industry returns is generally higher than that of news-based relatedness (average adjusted R^2 of 0.52 vs. 0.47), this difference in not significant at the 5% level.[115] Thus SIC-defined peer groups are not a significantly better representation of stock-related proximity than peer groups defined by our measure of relatedness.

Next to the return-based quantitative support for *Relatedness* as a meaningful instrument to identify strategic peer groups, the measure also makes sense from a qualitative viewpoint. News-based peer groups reflect many intuitive competitive relationships. Google (GOOG), for instance, is associated with Yahoo (YHOO), Ebay (EBAY) and Microsoft (MSFT) according to both methods. However, beyond a mere reflection of standard classifications, *Relatedness* reveals meaningful relationships between companies, which the SIC scheme does not capture. One of the most strident examples is the frequently referenced rivalry in the high tech industry between Google and Apple. The SIC assigns Google to *Business Services* and classifies Apple as *Technology, Hardware and Equipment*. According to the news, Apple is one of Google's closest competitors along with other technology firms that are assigned to different SIC codes, such as Dell (DELL), Hewlett-Packard (HPQ), and IBM (IBM). Amazon, which is traditionally classified as a retailer, is often mentioned jointly in the news with both Google (GOOG) and Apple (AAPL). While the previous examples are fairly timeless, our news-based measure of relatedness also reflects temporary relationships. In our sample period, insurance firm American International Group, for example, is associated closely with investment bank Goldman Sachs (GS). This reflects suspicions of potential conflicts of interest due to an AIG payment of $12.9 billion to Goldman Sachs, where then-Treasury Secretary Henry Paulson had previously worked as CEO, in the months after AIG was rescued by the government in September 2009. Ongoing legal actions between the two financial institutions followed (see the 6th message from Table 1 for an example). While we can only provide a few examples of this type of anecdotal evidence to illustrate the insights contained in news-based peer groups and cannot explain each and every relationship, there is enough evidence to support the notion that our measure of relatedness can be used to define strategic peer groups. In sum, we find that news-based relatedness can define meaningful strategic peer groups, which exhibit a substantial level of stock comovement.

[115] To check the robustness of our results, we have repeated this analysis using weekly stock returns. As in the case of daily returns, SIC industries are only slightly better at explaining weekly industry returns (average adjusted $R^2 = 0.497$ vs. 0.457), but this difference is not significant at the 10% level.

Table 5: Strategic peer groups (1/2)

Sample firm	SIC				*Relatedness*			
	SIC name	Same SIC	SIC peers	Adj. R^2	Same SIC		Peers (most closely related)	Adj. R^2
GOOG Google)	Business services	36	**ADBE**, ADP, ADSK, AKAM, BMC, CA, CPWR, **CRM**, CSC, **CTSH**, CTXS, **EBAY**, EFX, ERTS, FIS, FISV, INTU, IPG, IRM, JNPR, MA, MCO, **MFE**, **MSFT**, MWW, NOVL, OMC, ORCL, RHI, RHT, SYMC, TSS, V, VRSN, WU, **YHOO**	0.49	7	19%	AAPL, **ADBE**, ALL, AMD, AMZN, BBY, CMCSA, **CRM**, CSCO, **CTSH**, DELL, **EBAY**, EMC, F, FCX, FHN, GE, GS, HPQ, IBM, IGT, INTC, ISRG, **MFE**, MOT, **MSFT**, NVDA, PCLN, PPG, QCOM, STZ, VZ, WPO, XRX, **YHOO**, YUM	0.50
AIG (American Int'l)	Insurance carriers	23	AET, **AFL**, AIZ, **ALL**, **CB**, CI, CINF, CVH, GNW, **HIG**, HUM, L, LNC, LUK, **MET**, PFG, PGR, **PRU**, TMK, TRV, UNH, UNM, WLP	0.56	6	26%	AEE, AEP, AES, **AFL**, **ALL**, APD, AYE, BDK, C, **CB**, CNP, DOW, ECL, FRX, GS, HAR, **HIG**, IPG, MAS, **MET**, MIL, ODP, **PRU**	0.41
C (Citigroup)	Depository institutions	22	**BAC**, BBT, **BK**, **CMA**, **COF**, FHN, FITB, HBAN, HCBK, **JPM**, KEY, **MI**, MTB, **NTRS**, PBCT, PNC, RF, STI, STT, **USB**, WFC, ZION	0.64	8	36%	A, AIG, AXP, **BAC**, **BK**, **CMA**, **COF**, DFS, DNR, DOW, ETFC, F, GS, **JPM**, **MI**, MO, MS, **NTRS**, THC, UNP, **USB**, WFC	0.52
D (Dominion Resources)	Electricity, gas, and sanitary services	39	**AEE**, **AEP**, AES, AYE, CEG, CMS, CNP, DTE, **DUK**, ED, EIX, **EP**, EQT, ETR, EXC, FE, FPL, GAS, NI, NU, PCG, PEG, PGN, PNW, POM, PPL, RSG, SCG, SE, SO, SRCL, SRE, STR, TE, TEG, WEC, **WM**, WMB, XEL	0.63	5	13%	A, ADM, **AEE**, **AEP**, AFL, APA, AVP, BMY, CA, CAG, CHK, CLX, CNX, **DUK**, EMR, **EP**, FE, FPL, GT, HES, K, KEY, L, LLY, MMM, MO, NVDA, NWL, PGN, PXD, SLE, SO, SYY, TSN, TSO, UTX, **WM**, WY, XTO	0.50
A (Agilent Technologies)	Instruments and related products	24	AGN, BAX, **BCR**, **BDX**, BSX, CFN, COL, **DHR**, EK, FLIR, ISRG, KLAC, MDT, MIL, PKI, RTN, STJ, SYK, **TER**, TMO, WAT, XRAY, XRX, ZMH	0.40	4	17%	AEP, APD, ARG, **BCR**, **BDX**, BEN, CNP, D, DHI, **DHR**, EXPD, FIS, JBL, K, L, M, MFE, NWL, PPG, R, STR, **TER**, VAR, WDC	0.47

Table 5: Strategic peer groups (2/2)

Sample firm	**SIC**				***Relatedness***			
	SIC name	Same SIC	SIC peers	Adj. R^2	Same SIC		Peers (most closely related)	Adj. R^2
CSCO (Cisco)	Electrical and electronic equipment	30	ADI, ALTR, AMD, APH, **BRCM**, EMR, FSLR, GE, HAR, HRS, **INTC**, **JDSU**, LLL, LLTC, LSI, MCHP, MOLX, MOT, MU, NSM, NTAP, NVDA, NVLS, QCOM, QLGC, TLAB, TXN, WFR, WHR, **XLNX**	0.56	4	13%	AIZ, AKAM, BDK, **BRCM**, CLX, CMCSA, CSC, CTXS, DD, EMC, FO, GLW, HPQ, IBM, **INTC**, IP, JBL, **JDSU**, JNJ, JNPR, KSS, LLL, MMM, MU, NOVL, NWSA, PBCT, PLD, WFMI, **XLNX**	0.49
AAPL (Apple)	Industrial machinery and equipment	30	AMAT, BDK, BHI, CAM, CAT, CMI, DE, **DELL**, DOV, **EMC**, ETN, FLS, FTI, **HPQ**, **IBM**, IGT, ITT, JBL, LXK, MMM, NOV, PBI, PH, PLL, ROK, SII, SNDK, TDC, VAR, WDC	0.57	4	13%	ADBE, AMZN, AVY, BBY, BRCM, CPWR, CRM, CSCO, **DELL**, DIS, EK, **EMC**, F, GOOG, **HPQ**, **IBM**, INTC, ISRG, MOT, MSFT, NYT, PCP, QCOM, T, TAP, TXN, VZ, WIN, WMT, WYNN	0.44
PFE (Pfizer)	Chemicals and allied products	34	**ABT**, **AMGN**, APD, AVP, BIIB, **BMY**, CELG, CEPH, CF, CL, CLX, DD, DOW, ECL, EL, EMN, FMC, FRX, GENZ, **GILD**, HSP, IFF, **JNJ**, **KG**, LIFE, **LLY**, MON, MRK, MYL, PG, PPG, PX, SIAL, WPI	0.44	7	21%	**ABT**, ADM, AET, **AMGN**, BAX, BDK, BMS, **BMY**, BSX, CI, DUK, FRX, **GILD**, GIS, **JNJ**, **KG**, **LLY**, LM, LO, MCK, MDT, MRK, PPL, R, STJ, SYK, SYY, THC, TRV, UTX, VMC, WAT, WIN, WLP	0.38
K (Kellogg)	Food and kindred products	19	ADM, **CAG**, CCE, **CPB**, DPS, **GIS**, **HNZ**, HRL, HSY, **KFT**, KO, MJN, MKC, PEP, SJM, **SLE**, STZ, TAP, TSN	0.41	6	32%	AVP, **CAG**, CL, CLX, **CPB**, DNB, FAST, **GIS**, **HNZ**, IP, **KFT**, NOV, RAI, SJM, **SLE**, TSN, VAR, WAT, WY	0.48
APC (Ampco-Pittsburgh)	Oil and gas extraction	19	**APA**, **BJS**, **CHK**, COG, DNR, **DO**, **DVN**, **EOG**, ESV, **HAL**, NBL, NBR, OXY, PXD, RDC, RRC, SLB, SWN, XTO	0.64	7	37%	**APA**, BHI, **BJS**, CAM, **CHK**, CNP, COP, **DO**, **DVN**, **EOG**, FTI, **HAL**, MJN, NBL, PCL, RRC, SYY, WAT, XOM	0.56
Average		**27.6**		**0.53**	**5.8**	**23%**		**0.47**

Notes: This table shows sample companies (selected by highest message volume for SIC industry groups with at least 20 companies) and their SIC peers as well as the most closely related companies in terms of Relatedness. Peers found in both groups are bolded. Adjusted R^2s are shown for the OLS regression $r_{i,t} = \alpha_t + \beta\, r_{ind,t} + \varepsilon_{i,t}$. The explanatory power of SIC groups for industry returns is generally higher than that of news-based relatedness, but the difference in not significant at the 5% level. We can see that there is an overlap of 23% between the two methods to define peer groups.

Table 6: Industry classification (1/2)

SIC		*Relatedness*	
Industry Group	**Constituents**	**Industry Group**	**Constituents**
Oil and gas extraction	APA, APC, BJS, CHK, COG, DNR, DO, DVN, EOG, ESV, HAL, NBL, NBR, OXY, PXD, RDC, RRC, SLB, SWN, XTO	**Energy**	APA, APC, BHI, CAM, CHK, CMI, COG, DNR, DO, DVN, EOG, EP, EQT, FTI, HAL, NBL, NBR, NOV, OXY, PXD, RDC, RRC, SII, SLB, STR, SWN, WMB, XTO
Food and kindred products	ADM, CAG, CCE, CPB, DPS, GIS, HNZ, HRL, HSY, K, KFT, KO, MJN, MKC, PBG, PEP, SJM, SLE, STZ, TAP, TSN	**Pharmaceutical and food products**	ABT, ADM, ADP, AET, AFL, AGN, AMGN, BAX, BDX, BIIB, BMY, CAG, CAT, CELG, CEPH, CI, CL, CLX, CPB, CVH, DD, DOW, EMR, GILD, GIS, HNZ, HSY, HUM, JNJ, K, KFT, KO, LLY, MMM, MRK, PEP, PFE, PG, SJM, SLE, SRCL, SYK, UNH, WHR, WLP, WM
Chemicals and allied products	ABT, AMGN, APD, AVP, BIIB, BMY, CELG, CEPH, CF, CL, CLX, DD, DOW, ECL, EL, EMN, FMC, FRX, GENZ, GILD, HSP, IFF, JNJ, KG, LIFE, LLY, MON, MRK, MYL, PFE, PG, PPG, PX, SIAL, WPI	**Biotechnology**	AVP, CCE, CF, CRM, CTSH, DHR, DPS, EL, FLS, FSLR, HSP, ISRG, LIFE, MIL, MJN, PLL, TMO, WAT
Industrial machinery & equipment	AAPL, AMAT, BDK, BHI, CAM, CAT, CMI, DE, DELL, DOV, EMC, ETN, FLS, FTI, HPQ, IBM, IGT, ITT, JBL, LXK, MMM, NOV, PBI, PH, PLL, ROK, SII, SNDK, SWK, TDC, VAR, WDC	**Internet**	AAPL, ADBE, AKAM, CPWR, GOOG, IPG, JBL, MCO, MOT, MSFT, MWW, ORCL, YHOO
Electrical and electronic equipment	ADI, ALTR, AMD, APH, BRCM, CSCO, EMR, FSLR, GE, HAR, HRS, INTC, JDSU, LLL, LLTC, LSI, MCHP, MOLX, MOT, MU, NSM, NTAP, NVDA, NVLS, QCOM, QLGC, TLAB, TXN, WFR, WHR, XLNX	**Computer soft- and hardware**	ADI, ADSK, ALTR, AMAT, AMD, BMC, BRCM, CA, CSCO, CTXS, DELL, EBAY, EMC, HPQ, IBM, INTC, INTU, JNPR, KLAC, LLTC, LSI, MCHP, MU, NOVL, NSM, NTAP, NVDA, QCOM, RHT, SNDK, SYMC, TER, TXN, VRSN, XLNX
Instruments & related products	A, AGN, BAX, BCR, BDX, BSX, CFN, COL, DHR, EK, FLIR, ISRG, KLAC, MDT, MIL, PKI, RTN, STJ, SYK, TER, TMO, WAT, XRAY, XRX, ZMH	**Medical technology**	BSX, DE, FRX, GE, GENZ, KG, MDT, MON, MYL, PPG, STJ, STZ, TAP, TSN, WPI
Electric gas and sanitary services	AEE, AEP, AES, AYE, CEG, CMS, CNP, D, DTE, DUK, ED, EIX, EP, EQT, ETR, EXC, FE, FPL, GAS, NI, NU, PCG, PEG, PGN, PNW, POM, PPL, RSG, SCG, SE, SO, SRCL, SRE, STR, TE, TEG, WEC, WM, WMB, XEL	**Utilities and energy**	AEE, AEP, AES, APD, AYE, CEG, CMS, CNP, D, DTE, DUK, ECL, ED, EXC, FE, FPL, LLL, PCG, PGN, PPL, SO

Table 6: Industry classification (2/2)

SIC		Relatedness	
Industry Group	**Constituents**	**Industry Group**	**Constituents**
Depository institutions	BAC, BBT, BK, C, CMA, COF, FHN, FITB, HBAN, HCBK, JPM, KEY, MI, MTB, NTRS, PBCT, PNC, RF, STI, STT, USB, WFC, ZION	**Financial instititutions**	APH, BAC, BBT, BK, C, CMA, ETN, FHN, FITB, HBAN, IRM, JPM, KEY, MI, MTB, NTRS, PBCT, PH, PNC, RF, STI, STT, USB, WFC
Insurance carriers	AET, AFL, AIG, AIZ, ALL, CB, CI, CINF, CVH, GNW, HIG, HUM, L, LNC, LUK, MET, PFG, PGR, PRU, TMK, TRV, UNH, UNM, WLP, XL	**Insurance and high tech suppliers**	AIG, ALL, CB, EK, ERTS, FIS, FISV, GNW, HAR, HIG, ITT, JDSU, LNC, LXK, MET, MFE, PRU, TLAB, TRV, WFR, WU, XL
Business services	ACS, ADBE, ADP, ADSK, AKAM, BMC, CA, CPWR, CRM, CSC, CTSH, CTXS, EBAY, EFX, ERTS, FIS, FISV, GOOG, INTU, IPG, IRM, JNPR, MA, MCO, MFE, MSFT, MWW, NOVL, OMC, ORCL, RHI, RHT, RX, SYMC, TSS, V, VRSN, WU, YHOO	**Miscellaneous**	A, AIZ, BCR, COF, CSC, DOV, FLIR, IGT, L, MA, RTN, UNM, V, VAR, WDC, XRX
Adj. R^2:	**0.546**	**Adj. R^2:**	**0.540**

Notes: This table shows the 10 largest SIC groups by number of sample companies and the 10 groups resulting from a cluster analysis of these firms based on our measure of relatedness. We can see that the network of joint mentions generates plausible industry groups, which differ from traditional classification schemes. These groups have the same explanatory power for industry returns as the well-established SIC.

4.4 Industry classification

In this section, we will explore whether our news-based measure of relatedness can serve to delineate meaningful industry groups. In the previous section we defined strategic peer groups as the most closely related companies from the perspective of one sample company and permitted firms to belong to multiple strategic peer groups. In the following, we will leverage social network analysis to enable the direct comparison of SIC- versus news-based industry groups.

We limit our analysis to the sample companies assigned to the 10 largest two-digit SIC groups by number of companies. This sample, along with the official SIC names, is shown in the left panel of Table 6. As in the case of traditional industry classification, we have reassigned each and every one of these firms to exactly one news-based industry group. For a direct comparison with the 10 SIC groups, we employ social network cluster analysis to delineate exactly 10 clusters of stocks. The right panel of Table 6 shows the resulting news-based industry groups.[116] While there are some similarities to traditional SIC codings, the news-based classification shows distinct emphases. The cluster analysis, for instance, isolated biotech companies from the more broadly defined SIC group *Chemicals and allied products*, but combined pharmaceutical companies (e.g., Abbott (ABT), Pfizer (PFE), and Merck (MRK)) and food companies (e.g., Kraft (KFT), PepsiCo (PEP), and ConAgra Foods (CAG)). Internet-related firms, such as Google (GOOG) and Apple (AAPL) were assigned to one group as well as computer soft- and hardware companies, which the SIC system splits into *Industrial machinery and equipment* (e.g., Hewlett-Packard (HPQ), Cisco (CSCO) and Dell (DELL)) and *Business services* (e.g., Novell (NOVL), Juniper Networks (JNPR)). Given ever faster changes in the industrial landscape over the past decades, our methodology may be better suited to identify current industry lines than the increasingly outdated SIC scheme. One needs to keep in mind that our approach slightly favors the SIC classification, since we have used a clearly defined SIC sample as the starting point for news-based rearrangement. Even so, the explanatory power of news-based industry groups for stock returns is almost identical to SIC industry groups (average adjusted R^2 of 0.546 vs. 0. 540).[117] We conclude that news-based industry groups can be a plausible alternative to traditional industry classifications.

---

[116] The names for these industry groups were assigned by the authors and seek to loosely describe the constituents.

[117] Similar to our analysis of strategic peer groups, we have repeated this analysis using weekly stock returns to check for robustness. As in the case of daily returns, SIC industries are only insignificantly better at explaining weekly industry returns (adjusted R^2 = 0.516 vs. 0.506).

5 Conclusion

5.1 Discussion of results

Given the limitations of popular methods for industry classification (e.g., Bhojraj et al., 2003; Clarke, 1989; Fan & Lang, 2000), this study set out to investigate in how far the investor perception of strategic peer groups can help determine the relatedness of firms and be used to identify homogenous groups of firms. We have employed a unique dataset and leveraged methods from social network analysis to compute the frequency with which pairs of companies are mentioned together in news-related messages in an online stock community. We explored the following three research questions: first, whether our measure of relatedness can help explain the comovement of stocks, second, whether the relationships can help identify a firm's strategic peers, and, third, whether meaningful industry groups can be extracted from the network of relationships.

With respect to our first research question, we conclude that news-based relatedness can help explain the comovement of stock returns and adds explanatory power for comovement to traditional measures of proximity such as SIC classifications. An increase in *Relatedness* over time quickly reflects an increase in stock comovement. Our results support the theory that information associated with a set of firms is an indicator of relatedness and the comovement of their stocks (e.g., King, 1966). As far as the identification of strategic peers is concerned, we find that relatedness can define meaningful strategic peer groups, which exhibit a substantial level of stock comovement. This approach provides an alternative to clustering firms based on a multitude of firm-level dimensions frequently employed by strategy researchers (e.g., DeSarbo et al., 2009; Dranove et al., 1998). Finally, our results suggest that news-based industry groups can be a plausible alternative to traditional industry classifications. Overall, our results support the view that a news-based measure of firm relatedness can be used to reliably define strategic peer and industry groups. These findings confirm previous exploratory evidence that a network perspective of the stock market can be used to derive groups of stocks that are homogeneous with respect to traditional industry classifications (e.g., Bonanno et al., 2004; Mantegna, 1999; Onnela et al., 2003) and may be used to design stock indices (Tse et al., 2010).

Our innovative approach to measure firm relatedness has multiple advantages over traditional methods of industry classification: First, our approach is transparent and does not depend on the arbitrary assignment of companies by experts or the census bureau. Our method leverages the insights of hundreds of investors who associate one firm with another. Our study focuses on the context of stock microblogs which offer a large number of news items that users associate with one or several firms. However, our approach is not limited to this particular data source and could easily be adapted to other media such as newspaper articles or financial news wires. Second, our definition of relatedness provides a quantitative measure of the proximity of firms. In contrast to the nominal categorization of standard industry classifications, it offers a continuous measure of relatedness for all pairs of companies. Third, news-based relatedness can be calculated for various time horizons and may thus reflect changes in firm relatedness quickly compared to fairly stagnant manual classification schemes such as the SIC.

There are various promising applications for the accurate measurement of firm relatedness. As laid our in our section covering related research, both academics and practitioners spent considerable time and effort on the delineation of industry groups (e.g., Fama & French, 1997; Fan & Lang, 2000; Ramnath, 2002). Research and practical applications include the design of meaningful stock indices that are composed of a coherent group of companies and the calculation of industry-specific costs of capital. In addition, financial professionals, such as analysts and investors, may find multiple applications ranging from the identification of relevant competitors to the selection of diversified portfolios (Chan et al., 2007).

5.2 Limitations and further research

Our study does not come without limitations. First, our quantitative analysis of firm relatedness was mainly limited to the similarity of stock returns. Stock returns are largely driven by new information and thus lend themselves in particular to our news-based definition of relatedness. However, there are many other dimensions of relatedness and future research should investigate in how far a news-based industry classification can delineate homogenous groups with respect to these dimensions (e.g., accounting figures, overlap in customer base, coverage by the same analyst, ownership by the same investors, etc.). Second, our method to delineate industry groups may prove to be helpful to the strategic groups literature. Dranove et al. (1998) have pointed to a shortcoming in much of the existing research, which groups firms based on firm-level factors that do not capture strategic interactions. Our approach to delineate industry groups through a network of relationships may address this shortcoming and serve to capture the patterns of interactions within industries. Third, we drew on basic methods from social network analysis. However, there is a rich repository of analytical resources in the realm of social network analysis that may help refine the delineation of industry groups. For instance, while we have defined strategic peers to be the companies that are most closely related to a particular company directly (i.e., a direct link or joint mention had to be present) to be included in the investigation of *Relatedness* and comovement on the level of company pairs, other methods (such as so-called cliques or k-plexes) may be better suited to the analysis of larger networks. Exploring these methods may help identify even more relevant subgroups of strategic peers. Finally, our analysis focused on the mere joint mentioning of two companies. However, our method opens the door to more nuanced analysis of relatedness. Leveraging information regarding the context in which two stocks are associated with each other (e.g., a news item referring to a joint venture vs. a legal action) may enable us to define the type of relatedness and the competitive relationship in further detail (see Gourley, 2011). While the present study is limited to identifying industry groups including both strategic followers *and* foes, this extension may help to distinguish between strategic partners and competitors and characterize a company's peers as either follower *or* foe.

6 Appendix: Overview of company tickers

Ticker	Company name	Ticker	Company name	Ticker	Company name	Ticker	Company name
A	Agilent Technologies	CTAS	Cintas Corporation	IPG	Interpublic Group	PLD	ProLogis
AA	Alcoa Inc	CTL	CenturyTel Inc	IRM	Iron Mountain Incorporated	PLL	Pall Corp.
AAPL	Apple Inc.	CTSH	Cognizant Technology Solutions	ISRG	Intuitive Surgical Inc.	PM	Philip Morris International
ABC	AmerisourceBergen Corp.	CTXS	Citrix Systems	ITT	ITT Corporation	PNC	PNC Financial Services
ABT	Abbott Laboratories	CVH	Coventry Health Care Inc.	JBL	Jabil Circuit	PPG	PPG Industries
ADBE	Adobe Systems	CVS	CVS Caremark Corp.	JCI	Johnson Controls	PPL	PPL Corp.
ADI	Analog Devices	CVX	Chevron Corp.	JCP	Penney (J.C.)	PRU	Prudential Financial
ADM	Archer Daniels Midland	D	Dominion Resources	JDSU	JDS Uniphase Corp.	PTV	Pactiv Corp.
ADP	Automatic Data Processing Inc.	DD	Du Pont (E.I.)	JEC	Jacobs Engineering Group	PXD	Pioneer Natural Resources
ADSK	Autodesk Inc.	DE	Deere & Co.	JNJ	Johnson & Johnson	Q	Qwest Communications Int
AEE	Ameren Corporation	DELL	Dell Inc.	JNPR	Juniper Networks	QCOM	QUALCOMM Inc.
AEP	American Electric Power	DF	Dean Foods	JPM	JPMorgan Chase & Co.	R	Ryder System
AES	AES Corporation	DFS	Discover Financial Services	JWN	Nordstrom	RAI	Reynolds American Inc.
AET	Aetna Inc.	DGX	Quest Diagnostics	K	Kellogg Co.	RDC	Rowan Cos.
AFL	AFLAC Inc.	DHI	D. R. Horton	KEY	KeyCorp	RF	Regions Financial Corp.
AGN	Allergan Inc.	DHR	Danaher Corp.	KFT	Kraft Foods Inc-A	RHT	Red Hat Inc.
AIG	American International Group	DIS	Walt Disney Co.	KG	King Pharmaceuticals	RL	Polo Ralph Lauren Corp.
AIV	AIMCO	DNB	Dun & Bradstreet	KIM	Kimco Realty	ROST	Ross Stores Inc
AIZ	Assurant Inc	DNR	Denbury Resources Inc.	KLAC	KLA-Tencor Corp.	RRC	Range Resources Corp.
AKAM	Akamai Technologies Inc	DO	Diamond Offshore Drilling	KMB	Kimberly-Clark	RSH	RadioShack Corp
AKS	AK Steel Holding Corp.	DOV	Dover Corp.	KO	Coca Cola Co.	RTN	Raytheon Co.
ALL	Allstate Corp.	DOW	Dow Chemical	KR	Kroger Co.	SBUX	Starbucks Corp.
ALTR	Altera Corp.	DPS	Dr Pepper Snapple Group	KSS	Kohl's Corp.	SCHW	Charles Schwab
AMAT	Applied Materials	DRI	Darden Restaurants	L	Loews Corp.	SHLD	Sears Holdings Corporation
AMD	Advanced Micro Devices	DTE	DTE Energy Co.	LEN	Lennar Corp.	SHW	Sherwin-Williams
AMGN	Amgen	DTV	DIRECTV Group Inc.	LIFE	Life Technologies	SII	Smith International
AMP	Ameriprise Financial Inc.	DUK	Duke Energy	LLL	L-3 Communications Holdings	SJM	Smucker (J.M.)
AMT	American Tower Corporation	DV	DeVry, Inc.	LLTC	Linear Technology Corp.	SLB	Schlumberger Ltd.
AMZN	Amazon Corp.	DVA	DaVita Inc.	LLY	Lilly (Eli) & Co.	SLE	Sara Lee Corp.
AN	AutoNation Inc.	DVN	Devon Energy Corp.	LM	Legg Mason	SLM	SLM Corporation
ANF	Abercrombie & Fitch Co.	EBAY	eBay Inc.	LMT	Lockheed Martin Corp.	SNDK	SanDisk Corporation
APA	Apache Corp.	ECL	Ecolab Inc.	LNC	Lincoln National	SO	Southern Co.
APC	Anadarko Petroleum Corporation	ED	Consolidated Edison	LO	Lorillard Inc.	SPG	Simon Property Group Inc
APD	Air Products & Chemicals	EK	Eastman Kodak	LOW	Lowe's Cos.	SPLS	Staples Inc.
APH	Amphenol Corp A	EL	Estee Lauder Cos.	LSI	LSI Corporation	SRCL	Stericycle Inc
APOL	Apollo Group	EMC	EMC Corp.	LTD	Limited Brands Inc.	STI	SunTrust Banks
ARG	Airgas Inc	EMR	Emerson Electric	LUV	Southwest Airlines	STJ	St Jude Medical
ATI	Allegheny Technologies Inc	EOG	EOG Resources	LXK	Lexmark Int'l Inc	STR	Questar Corp.
AVB	AvalonBay Communities	EP	El Paso Corp.	M	Macy's Inc.	STT	State Street Corp.
AVP	Avon Products	EQR	Equity Residential	MA	Mastercard Inc.	STZ	Constellation Brands
AVY	Avery Dennison Corp.	EQT	EQT Corporation	MAR	Marriott Int'l.	SUN	Sunoco Inc.
AXP	American Express	ERTS	Electronic Arts	MAS	Masco Corp.	SVU	Supervalu Inc.
AYE	Allegheny Energy	ESRX	Express Scripts	MAT	Mattel Inc.	SWN	Southwestern Energy
AZO	AutoZone Inc.	ETFC	E-Trade	MCD	McDonald's Corp.	SWY	Safeway Inc.
BA	Boeing	ETN	Eaton Corp.	MCHP	Microchip Technology	SYK	Stryker Corp.
BAC	Bank of America Corp.	EXC	Exelon Corp.	MCK	McKesson Corp.	SYMC	Symantec Corp.
BAX	Baxter International Inc.	EXPD	Expeditors Int'l	MCO	Moody's Corp	SYY	Sysco Corp.
BBBY	Bed Bath & Beyond	EXPE	Expedia Inc.	MDT	Medtronic Inc.	T	AT&T Inc.
BBT	BB&T Corporation	F	Ford Motor	MEE	Massey Energy	TAP	Molson Coors Brewing
BBY	Best Buy Co. Inc.	FAST	Fastenal Co	MET	MetLife Inc.	TER	Teradyne Inc.
BCR	Bard (C.R.) Inc.	FCX	Freeport-McMoran	MFE	McAfee	TGT	Target Corp.
BDX	Becton Dickinson	FDO	Family Dollar Stores	MHP	McGraw-Hill	THC	Tenet Healthcare Corp.
BEN	Franklin Resources	FDX	FedEx Corporation	MHS	Medco Health Solutions Inc.	TIE	Titanium Metals Corp
BHI	Baker Hughes	FE	FirstEnergy Corp	MI	Marshall & Ilsley Corp.	TIF	Tiffany & Co.
BIG	Big Lots Inc.	FHN	First Horizon National	MIL	Millipore Corp.	TJX	TJX Companies Inc.
BIIB	BIOGEN IDEC Inc.	FIS	Fidelity National IS	MJN	Mead Johnson Nutrition Co	TLAB	Tellabs Inc.
BK	Bank of New York Mellon Corp.	FISV	Fiserv Inc	MMM	3M	TMO	Thermo Fisher Scientific
BLL	Ball Corp.	FITB	Fifth Third Bancorp	MO	Altria Group Inc.	TRV	The Travelers Companies Inc.
BMC	BMC Software	FLIR	FLIR Systems	MON	Monsanto Co.	TSN	Tyson Foods
BMS	Bemis	FLR	Fluor Corp.	MOT	Motorola Inc.	TSO	Tesoro Petroleum Co.
BMY	Bristol-Myers Squibb	FLS	Flowserve Corporation	MRK	Merck & Co.	TWC	Time Warner Cable Inc.
BRCM	Broadcom Corporation	FO	Fortune Brands Inc.	MRO	Marathon Oil Corp.	TWX	Time Warner Inc.
BSX	Boston Scientific	FPL	FPL Group	MS	Morgan Stanley	TXN	Texas Instruments
BTU	Peabody Energy	FRX	Forest Laboratories	MSFT	Microsoft Corp.	TXT	Textron Inc.
BXP	Boston Properties	FSLR	First Solar Inc	MTB	M&T Bank Corp.	UNH	United Health Group Inc.
C	Citigroup Inc.	FTI	FMC Technologies Inc.	MU	Micron Technology	UNM	Unum Group
CA	CA, Inc.	FTR	Frontier Communications	MUR	Murphy Oil	UNP	Union Pacific
CAG	ConAgra Foods Inc.	GCI	Gannett Co.	MWW	Monster Worldwide	UPS	United Parcel Service
CAH	Cardinal Health Inc.	GD	General Dynamics	MYL	Mylan Inc.	USB	U.S. Bancorp
CAM	Cameron International Corp.	GE	General Electric	NBL	Noble Energy Inc	UTX	United Technologies
CAT	Caterpillar Inc.	GENZ	Genzyme Corp.	NBR	Nabors Industries Ltd.	V	Visa Inc
CB	Chubb Corp.	GILD	Gilead Sciences	NDAQ	NASDAQ OMX Group	VAR	Varian Medical Systems
CBG	CB Richard Ellis Group	GIS	General Mills	NEM	Newmont Mining Corp. (Hldg. Co.)	VLO	Valero Energy
CBS	CBS Corp.	GLW	Corning Inc.	NKE	NIKE Inc.	VMC	Vulcan Materials
CCE	Coca-Cola Enterprises	GME	GameStop Corp.	NOC	Northrop Grumman Corp.	VNO	Vornado Realty Trust
CCL	Carnival Corp.	GNW	Genworth Financial Inc.	NOV	National Oilwell Varco Inc.	VRSN	Verisign Inc.
CEG	Constellation Energy Group	GOOG	Google Inc.	NOVL	Novell Inc.	VZ	Verizon Communications
CELG	Celgene Corp.	GPS	Gap (The)	NSC	Norfolk Southern Corp.	WAG	Walgreen Co.
CEPH	Cephalon Inc	GS	Goldman Sachs Group	NSM	National Semiconductor	WAT	Waters Corporation
CF	CF Industries Holdings Inc	GT	Goodyear Tire & Rubber	NTAP	NetApp	WDC	Western Digital
CHK	Chesapeake Energy	GWW	Grainger (W.W.) Inc.	NTRS	Northern Trust Corp.	WFC	Wells Fargo
CHRW	C. H. Robinson Worldwide	HAL	Halliburton Co.	NUE	Nucor Corp.	WFMI	Whole Foods Market
CI	CIGNA Corp.	HAR	Harman Int'l Industries	NVDA	Nvidia Corporation	WFR	MEMC Electronic Materials
CL	Colgate-Palmolive	HAS	Hasbro Inc.	NWL	Newell Rubbermaid Co.	WHR	Whirlpool Corp.
CLF	Cliffs Natural Resources Inc	HBAN	Huntington Bancshares	NWSA	News Corporation	WIN	Windstream Corporation
CLX	Clorox Co.	HD	Home Depot	NYT	New York Times Cl. A	WLP	WellPoint Inc.
CMA	Comerica Inc.	HES	Hess Corporation	NYX	NYSE Euronext	WM	Waste Management Inc.
CMCSA	Comcast Corp.	HIG	Hartford Financial Svc.Gp.	ODP	Office Depot	WMB	Williams Cos.
CME	CME Group Inc.	HNZ	Heinz (H.J.)	ORCL	Oracle Corp.	WMT	Wal-Mart Stores
CMI	Cummins Inc.	HOG	Harley-Davidson	ORLY	O'Reilly Automotive	WPI	Watson Pharmaceuticals
CMS	CMS Energy	HON	Honeywell Int'l Inc.	OXY	Occidental Petroleum	WPO	Washington Post Co B
CNP	CenterPoint Energy	HOT	Starwood Hotels & Resorts	PBCT	People's United Bank	WU	Western Union Co
CNX	CONSOL Energy Inc.	HPQ	Hewlett-Packard	PCAR	PACCAR Inc.	WY	Weyerhaeuser Corp.
COF	Capital One Financial	HRB	Block H&R	PCG	PG&E Corp.	WYNN	Wynn Resorts Ltd
COG	Cabot Oil & Gas	HSP	Hospira Inc.	PCL	Plum Creek Timber Co.	X	United States Steel Corp.
COH	Coach Inc.	HST	Host Hotels & Resorts	PCLN	Priceline.com Inc	XL	XL Capital
COP	ConocoPhillips	HSY	The Hershey	PCP	Precision Castparts	XLNX	Xilinx Inc
COST	Costco Co.	HUM	Humana Inc.	PCS	MetroPCS Communications Inc.	XOM	Exxon Mobil Corp.
CPB	Campbell Soup	IBM	International Bus. Machines	PEP	PepsiCo Inc.	XRX	Xerox Corp.
CPWR	Compuware Corp.	ICE	IntercontinentalExchange Inc.	PFE	Pfizer Inc.	XTO	XTO Energy Inc.
CRM	Salesforce.com	IGT	International Game Technology	PG	Procter & Gamble	YHOO	Yahoo Inc.
CSC	Computer Sciences Corp.	INTC	Intel Corp.	PGN	Progress Energy Inc.	YUM	Yum! Brands Inc
CSCO	Cisco Systems	INTU	Intuit Inc.	PH	Parker-Hannifin	ZION	Zions Bancorp
CSX	CSX Corp.	IP	International Paper	PHM	Pulte Homes Inc.		

7 References

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *Journal of Finance*, 59, 1259-1294.

Baginski, S. (1987). Intraindustry information transfers associated with management forecasts of earnings. *Journal of Accounting Research*, 25, 196-216.

Barberis, N., Shleifer, A., & Wurgler, J. (2005). Comovement. *Journal of Financial Economics*, 75, 283-317.

Bhojraj, S., & Lee, C. M. (2002). Who is my peer? A valuation-based approach to the selection of comparable firms. *Journal of Accounting Research*, 40, 407-439.

Bhojraj, S., Lee, C. M., & Oler, D. K. (2003). What's my line? A comparison of industry classification schemes for capital market research. *Journal of Accounting Research*, 41, 745-774.

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

Bonanno, G., Caldarelli, G., Lillo, F., Micciché, S., Vandewalle, N., & Mantegna, R. N. (2004). Networks of equities in financial markets. *The European Physical Journal B - Condensed Matter and Complex Systems*, 38, 363-371.

Borgatti, S.P., Everett, M.G., & Freeman, L. C. (2002). *Ucinet for Windows: Software for social network analysis*. Harvard, MA: Analytic Technologies.

BusinessWeek (2009, February 11). StockTwits may change how you trade. *BusinessWeek (online edition)* Retrieved May 15, 2011 from http://www.businessweek.com/print/technology/content/feb2009/tc20090210_875439.htm

Chan, L., Lakonishok, K. C., & Swaminathan, B. (2007). Industry classifications and return comovement. *Financial Analysts Journal*, 63, 56-70.

Chen, N., Roll, R., & Ross, S. A. (1986). Economic forces and the stock market. *The Journal of Business*, 59, 383-403.

Clarke, R. N. (1989). SICs as delineators of economic markets. *Journal of Business*, 62, 17-31.

Das, S. R., & Sisk, J. (2003). Financial communities. *Working paper*. Retrieved May 15, 2011 from http://ssrn.com/abstract=404621

Das, S. R., Martinez-Jerez, A., & Tufano, P. (2005). eInformation: A clinical study of investor discussion and sentiment. *Financial Management*, 34, 103–137.

DeSarbo, W. S., Grewal, R., & Wang, R. (2009). Dynamic strategic groups: deriving spatial evolutionary paths. *Strategic Management Journal*, 30, 1420-1439.

Dranove, D., Peteraf, M., & Shanley, M. (1998). Do strategic groups exist? An economic framework for analysis. *Strategic Management Journal*, 19, 1029-1044.

Fama, E. F., & French, K. R. (1997). Industry costs of equity. *Journal of Financial Economics*, 43, 153-193.

Fan, J. P. H., & Lang, L. H. P. (2000). The measurement of relatedness: An application to corporate diversification. *Journal of Business*, 73, 629-660.

Farrell, J. (1974). Analyzing covariation of returns to determine homogeneous stock groupings. *The Journal of Business*, 47, 186-207.

Fertuck, L. (1975). A test of industry indices based on SIC codes. *Journal of Financial & Quantitative Analysis*, 10, 837-848.

Glover, F. (1990). Tabu search - part II. *Journal on* Computing, *2*, 4–32.

Gourley, S. (2011). Vision Statement: Locating Your Next Strategic Opportunity*, Harvard Business Review*, 89, 34-35.

Grant, R. (2010). *Contemporary Strategy Analysis*, 7th ed., Oxford: Wiley.

Hunt, M. (1972). Competition in the major home appliance industry, 1960-1970, Ph.D. dissertation, Harvard University.

Kahle, K. M., & Walkling, R. A. (1996). The impact of industry classifications on financial research. *Journal of Financial & Quantitative Analysis*, 31, 309-331.

Kim, Y., Lacina, M., & Park, M. S. (2008). Positive and negative information transfers from management forecasts. *Journal of Accounting Research*, 46, 885–908.

King, B. F. (1966). Market and industry factors in stock price behavior. *Journal of Business*, 39, 139-190.

Krackhardt, D. (1988). Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social Networks*, 10, 359–381.

O’Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. (2010). From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of the International Conference on Weblogs and Social Media* (pp. 122-129), Washington, DC: AAAI.

Porter, M. (1979). The structure within industries and companies' performance. *The Review of Economics and Statistics*, 61, 214-227.

Pindyck, R. S., & Rotemberg, J. J. (1993). The comovement of stock prices. *The Quarterly Journal of Economics*, 108, 1073-1104.

Ramnath, S. (2002). Investor and analyst reactions to earnings announcements of related firms: An empirical analysis. *Journal of Accounting Research*, 40, 1351–1376.

Rumelt, R. P. (1982). Diversification strategy and profitability. *Strategic Management Journal*, 3, 359–369.

Onnela, J., Chakraborti, A., Kaski, K., Kertész, J., & Kanto, A. (2003). Dynamics of market correlations: taxonomy and portfolio analysis. *Physical Review E*, 68, 056110.

Sprenger, T., & Welpe, I. (2010). Tweets and trades: The information content of stock microblogs. *Working paper.* Retrieved May 15, 2011 from http://ssrn.com/abstract=1702854

Tse, C. K., Liu, J., & Lau, F. C. (2010). A network perspective of the stock market. *Journal of Empirical Finance*, 17, 659-667.

Tumasjan, A., Sprenger, T. Sandner, P., & Welpe, I. (2010). Election forecasts with Twitter: How 140 characters reflect the political landscape. *Social Science Computer Review*. Advance online publication. doi: 10.1177/0894439310386557

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

Zuckerman, E. W., & Rao, H. (2004). Shrewd, crude or simply deluded? Comovement and the internet stock phenomenon. *Industrial and Corporate Change*, 13, 171-212.

II.5 Essay 5

TweetTrader.net:
Leveraging Crowd Wisdom in a Stock Microblogging Forum

Abstract

TweetTrader.net is a stock microblogging forum that leverages the wisdom of crowds to aggregate the information contained in stock-related tweets. Based on insights from academic research on stock microblogs, the application integrates inputs from text classification, user voting and a proprietary stock game in order to extract the sentiment (i.e., the bullishness) of online investors with respect to all publicly traded companies of the S&P 500. A demo of TweetTrader.net is available at http://TweetTrader.net.

JEL Classification: C88

Keywords: Twitter; microblogging; stock market; industry classification

Current status: Accepted for presentation at the ICWSM 2011 (July 2011) and publication as a demo paper in the *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, Barcelona, Spain.

Acknowledgements: This paper contains elements of joint work with Jan Stanzel, Philip Frank, Albert Feller, and Matthias Kuhnert.

1 Background of stock microblogging

> *"[Twitter] has become a marketplace for ideas that lets amateurs mix it up with former hedge fund managers and celebrity traders. With just a Twitter username and password, users are immersed in a virtual trading floor. […] You like what one trader is doing? Simply press follow; you don't like what he had to say about IBM or you find his trading style too risky? Don't follow him. It's that simple. Underperformers will be ignored, and rightly so – trading is a zero-sum game and bad advice is a waste of time and money."*
>
> BusinessWeek (2009)

1.1 Popularity of stock microblogs

Twitter has become a vibrant platform to exchange trading ideas and other stock-related information. Traders have adopted the convention of tagging stock-related messages (i.e., stock microblogs) with a dollar sign followed by the relevant company's ticker symbol (e.g., "$AAPL" for tweets related to Apple Inc.). The business press describes the conversations on Twitter as "the modern version of traders shouting in the pits" (BusinessWeek, 2009). There are investors who attribute their trading success to the information they find on social media websites and, moreover, financial professionals have developed Twitter-based trading systems to identify sentiment-based investment opportunities. As a result, the investor community has come to call Twitter and related third-party applications "a Bloomberg for the average guy" (BusinessWeek, 2009).

1.2 Related academic research

A few recent studies suggest that the information content of general Twitter messages, including those without a specific reference to the stock market, may help predict macroeconomic indicators such as the Index of Consumer Sentiment (O'Connor, Balasubramanyan, Routledge, & Smith, 2010) or stock market indices such as the Dow Jones Industrial Average (Bollen, Mao, & Zeng, 2011) and the S&P 500 (Zhang, Fuehres, & Gloor, 2010). Regarding the information content of microblogs with respect to individual stocks, Sprenger and Welpe (2010) find robust relationships between the sentiment of stock microblogs (i.e., their bullishness) and abnormal returns as well as message volume and trading volume. In addition, the authors offer an explanation for the mechanism leading to the efficient aggregation of information in microblogging forums by showing that users who provide above average investment advice are retweeted (i.e., quoted) more often and have more followers. In other words, retweets and followership may represent the Twittersphere's "currency" and provide it with a mechanism to weigh information.

The stock microblogging forum TweetTrader.net leverages the insights from this research and enables online investors to cut through the noise of thousands of daily messages. In contrast to a few related third-party applications such as StockTwits.com, which are limited to filtering the message stream by ticker symbol, TweetTrader.net taps the wisdom of crowds and aggregates the information in a meaningful fashion.

2 Features of TweetTrader.net

2.1 Tapping the wisdom of crowds

TweetTrader.net uses the Twitter Search API to provide a *Livestream* of all tweets related to S&P 500 stocks (see Figure 1). Users have a choice between all tweets or a subset of selected indices or industries. Bullish words (e.g., buy, upgrade, or growth) are highlighted in green, bearish words (e.g., sold, downgrade, or decline) are marked in red. Embedded within these basic functionalities are three main features that tap the collective wisdom of thousands of online investors: automatic text classification, user-driven sentiment voting and a *Stock Game*.

2.1.1 Text classification

First and foremost, every stock-related tweet is automatically classified as either a buy, hold, or sell signal based on a manually coded training set of 2,000 tweets and the multinomial Naïve Bayesian classifier of the Weka machine learning package (Hall et al. 2009). 10-fold cross validation of the model classifies 64.2% of all messages correctly (for details regarding the classification approach, see Sprenger and Welpe, 2010). The accuracy by class further validates the automatic classification of sentiment because the worst type of misclassification (i.e., of buy signals as sell signals and vice versa) occurs only rarely (less than 2% of all messages). Backtesting results illustrate that a trading strategy based on these signals would have earned profits of up to 15% over the first half of 2010.

2.1.2 Sentiment voting

The classification results demonstrate that the input from human coders can help correctly classify the sentiment of stock-related microblogs. Thus, next to every tweet users find two sentiment buttons (i.e., up and down) to vote the message as either a buy or a sell signal. This constant flow of manually coded messages can be used to update the automated classifier in real-time. Once users vote on a tweet, they can see how other online investors have rated the same piece of information. In addition, they have the opportunity to retweet this message and thereby give the information greater weight on the public timeline (e.g., because it will represent another data point for the text classifier).

Figure 1: *Livestream* with retweet function and sentiment voting

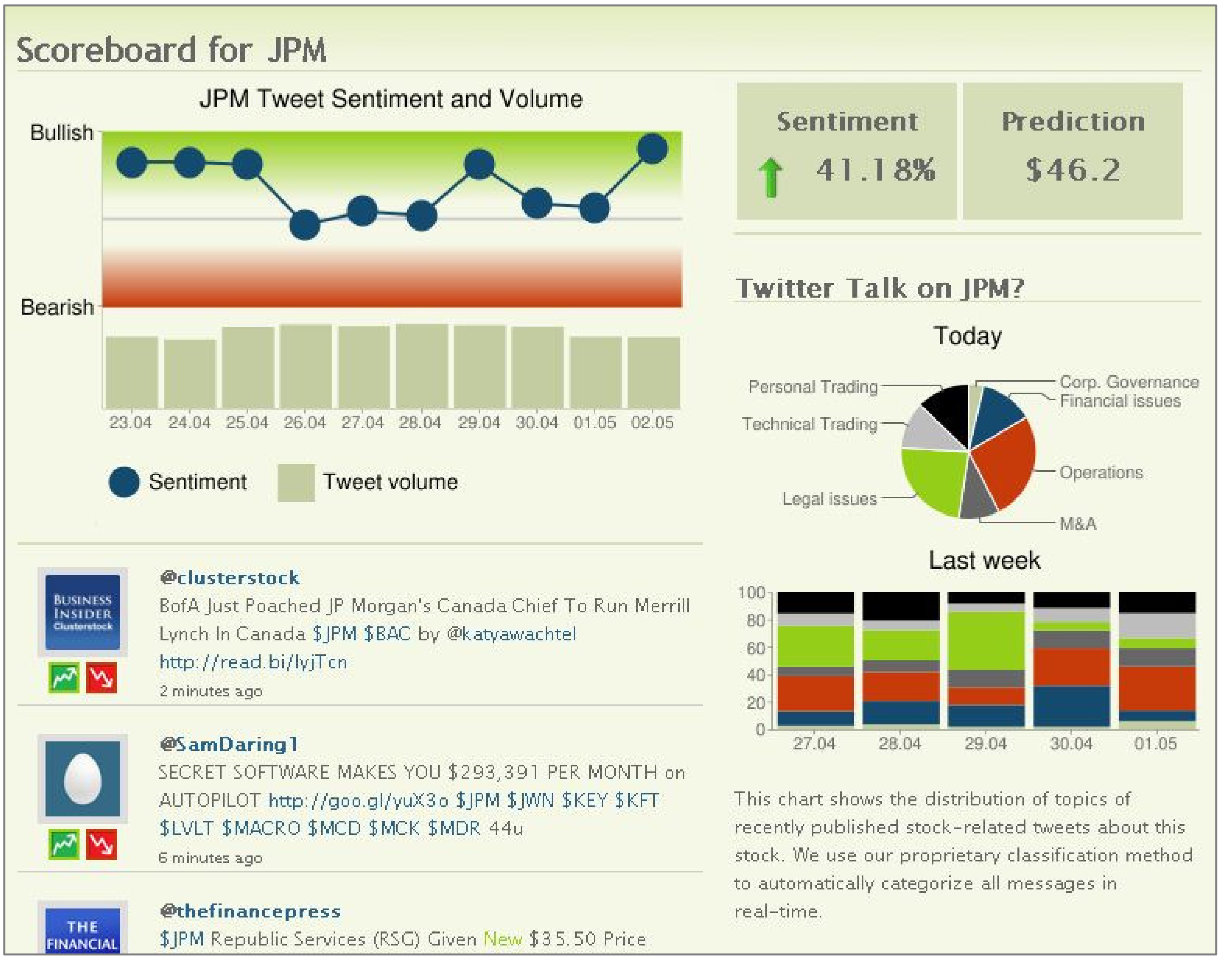
2.1.3 Stock game

The transformation into a retweet slightly alters the syntax of the original message. If a user votes a message as a buy (sell) signal, related ticker symbols are appended by a plus (minus) sign (e.g., “$AAPL+” for a buy signal). These appended tags represent predictions for the end of day stock price relative to the stock price at the time of the prediction. A point is awarded for every correct prediction. Obviously, besides retweeting messages, users can simply enter *Stock Game* predictions directly by following the syntax either on the TweetTrader.net website or through any other Twitter client. TweetTrader.net keeps track of these predictions, evaluates them in real-time and shows a ranking of all participating users. This allows players to monitor their predictions and provides transparency for other stock microbloggers to identify and follow the best investment advisers in the Twittersphere.

2.2 Aggregating stock-related information

The *Scoreboard* presents information related to a given stock (see Figure 2). In particular, the buy and sell signals from the three above-mentioned inputs are combined into one sentiment signal, which shows how bullish or bearish (i.e., positive or negative) investors are with respect to a stock. In addition, because Sprenger and Welpe (2011) have shown that not only the sentiment but also the type of stock-related news events (e.g., financial issues vs. operations) affects subsequent returns, text classification methods, like those used to extract the sentiment, are employed to show the distribution of recently discussed topics for a stock (e.g., corporate governance, financial issues, or operations).

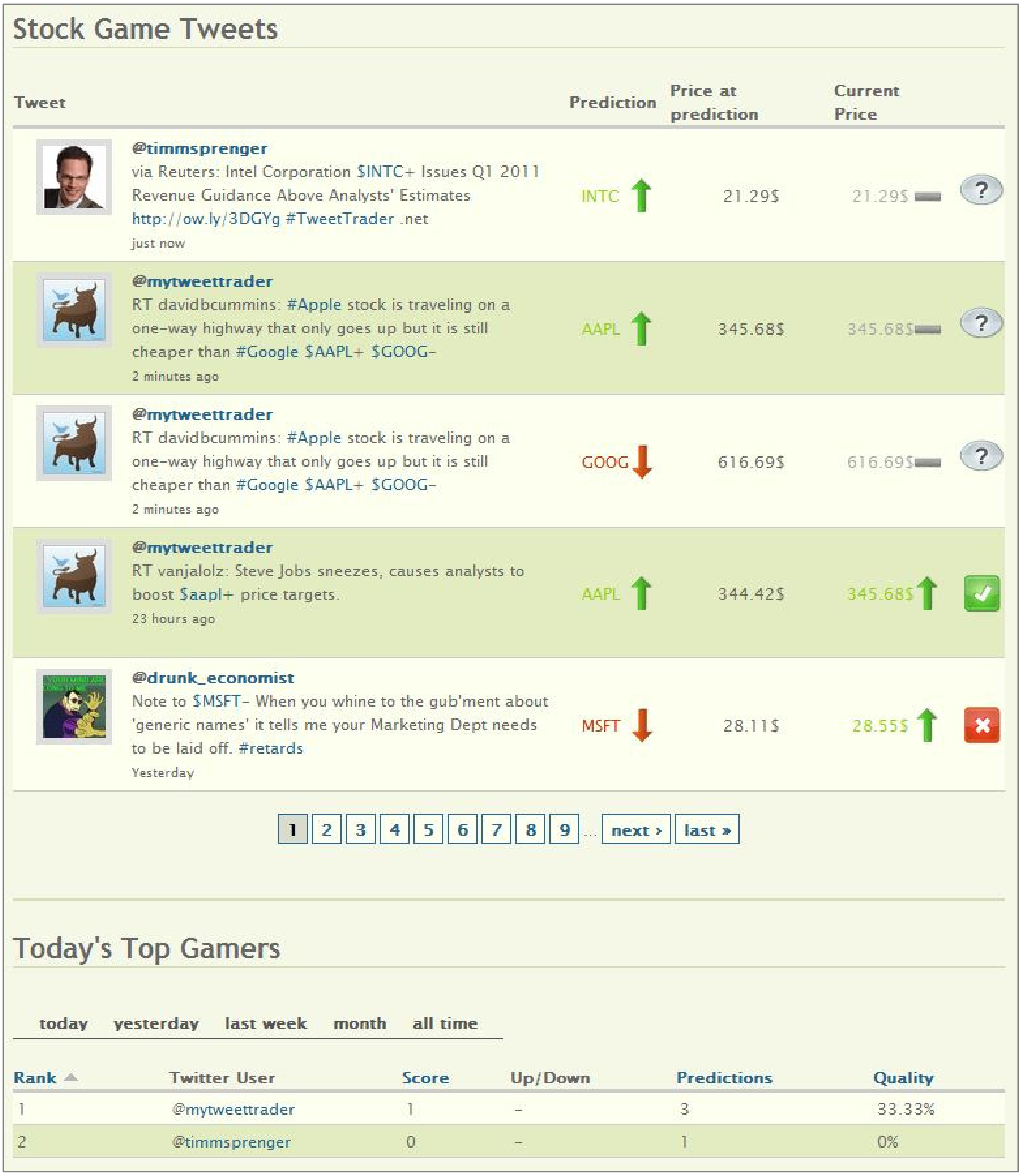
Figure 2: *Scoreboard* with sentiment and recently discussed topics (example JPM)



While TweetTrader.net is still in the early stages of development, this application shows promising ways to aggregate the collective wisdom of online investors in a meaningful fashion and enable stock microblogging forums to become to qualitative (i.e., textual) information what market mechanisms are to quantitative price signals.

3 Appendix

Figure 3: *Stock Game* with recent game tweets and ranking of players



4 References

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

BusinessWeek (2009, February 11). StockTwits may change how you trade. *BusinessWeek (online edition)* Retrieved May 15, 2011 from http://www.businessweek.com/print/technology/content/feb2009/tc20090210_875439.htm

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten I. H. (2009). The WEKA data mining software: An update. *SIGKDD Explorations*, 11, 10-18.

O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. (2010). From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of the International Conference on Weblogs and Social Media* (pp. 122-129), Washington, DC: AAAI.

Sprenger, T., & Welpe, I. (2010). Tweets and trades: The information content of stock microblogs. *Working paper.* Retrieved May 15, 2011 from http://ssrn.com/abstract=1702854

Sprenger T., and Welpe, I. 2011. News or Noise? The stock market reaction to different types of company-specific news events. *Working paper*. Retrieved May 15, 2011 from http://ssrn.com/abstract=1734632

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

III. Conclusion

1 Summary of results

This dissertation set out to determine whether the insights extracted from microblogging forums can serve as an indicator of real-world events and investigate the mechanism that can explain the efficient aggregation of information in microblogging forums.

Using the context of a national election, Essay 1 has illustrated that microblogging forums are used extensively for political deliberation. Related microblogs reflect the political preferences of the general population and may even serve to predict election results. The analysis of party-affiliated accounts suggests a bias of tweet sentiment (i.e., positive and negative emotions associated with a politician) and volume (i.e., the number of messages dedicated to a particular candidate or party) corresponding to users' political preferences. Given this bias in party orientation among individual users and the fact that the political debate was dominated by a few heavy users, it is all the more surprising that these heavy users were unable to impose their views on the discussion and affect the accuracy of aggregate results. This finding is in contrast to previous studies of political deliberation online (e.g., Jansen & Koop, 2005). However, empirical evidence suggests that these results did not come about by chance of this particular dataset and that, despite individual biases, errors can cancel each other out. In sum, the results support the notion that the size of the followership and the rate of retweets provide microblogging forums with a mechanism for weighing information.

While previous studies suggest that the information content of general microblogs may help predict macroeconomic market indicators such as the Dow Jones Industrial Average (DJIA) or the S&P 500 (O'Connor et al., 2010; Zhang et al., 2010; Bollen et al., 2011), they do not allow us to draw conclusions about the information content of stock microblogs with respect to individual stocks. Filling this research gap, Essay 2 shows that the sentiment (i.e., bullishness) of tweets is correlated with and can even predict abnormal stock returns and that message volume can predict next-day trading volume. In addition and in contrast to theoretical arguments that question the ability of blogs to aggregate dispersed information (e.g., Sunstein, 2008), it provides empirical evidence supporting the idea that followership relationships and retweets represent the Twittersphere's "currency" for weighing information. The 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.

Most previous event-studies use professionally edited news content to identify event days (e.g., Mitchell & Mulherin, 1994). Essay 3 shows that the information published in a stock microblogging forum can be used to detect not only investor sentiment, but also which types of stock-related news affect a company on a particular day from an investor perspective (e.g., news related to corporate governance, operations, or legal issues). Next to this methodological contribution, the essay illustrates how stock microblogs can be used in capital market research, in this particular case to discern genuine news that moves the market from insignificant noise without market reaction. The results indicate more widespread information leakage before good news, suggesting that future event studies control for news sentiment, and shows that the market reaction differs substantially across various types of news events, supporting the notion that there are certain event types to which investors attribute greater importance and other events, which rarely contain new information that moves the market. In addition, a cross-industry comparison indicates that industry classification may partially explain the market reaction to the same type of event.

Given the importance of industry classifications suggested by Essay 3 and limitations of traditional classification schemes (e.g., Bhojraj et al., 2003), Essay 4 proposes an alternative approach to defining industry groups based on the investor perception of the relatedness of stocks. The results show that the degree to which companies are mentioned jointly in an internet stock forum can explain the comovement of their stocks. The proposed measure of relatedness can help identify a firm’s strategic peers and delineate industry groups, which explain stock returns as well as established classification methods and offer a number of promising advantages (e.g., availability, timeliness).

In line with Kurt Lewin’s (1890-1947) famous quote “Nothing is as practical as a good theory”, Essay 5 illustrates how the above-mentioned research results were transferred to the stock microblogging forum TweetTrader.net, a fully functional Web 2.0 application, which leverages crowd wisdom by integrating inputs from text classification, user voting and a proprietary stock game in order to extract the sentiment (i.e., the bullishness) of online investors with respect to publicly traded stocks.

2 Limitations and further research

This section will elaborate on the most critical challenges to the overarching notion that microblogging forums my serve as information markets with mechanisms to weigh and ultimately aggregate information.[118] Microblogging forums were originally designed with a minimum number of features to provide personal status updates and are thus far from perfect in fulfilling this function. The “market mechanism”, including followership relationships, retweets and hashtags[119] has for the larger part emerged naturally by consensus among the user community in an attempt to structure and organize the message stream (Dann, 2010). Future research should explore possibilities to improve the market mechanism by design. This includes three major areas: The prevention of manipulation, the empowerment of users to weigh information, and improvements in the aggregation and presentation of information.

One of the most critical challenges is the threat of manipulation. Increasing awareness among microbloggers that influential media sources and analysts are paying attention to the content on Twitter may provide incentives to manipulate the message stream. For example in the case of political surveys, party affiliates may manipulate the public timeline.[120] Likewise in the stock market domain, there have been recent reports of Twitter abuse for a “pump & dump” stock fraud touting worthless penny stocks and leading to a $3 million profit on part of the suspect (McCool, 2010). As in many cases where technological advances trigger fraudulent behavior, these attempts of manipulation will create an incentive for researchers and practitioners to better understand user behavior and potentially debias social media content. Along these lines, Asur and Huberman provide an example by showing that “sentiments extracted from Twitter can be further utilized to improve the forecasting power of social media [mentions]” (2010, p. 1). In the financial domain, the stock microblogging forum TweetTrader.net, for instance, includes a sophisticated spam-filter, which enables the automatic detection of irrelevant messages and allows a moderator to ban fraudulent users. These examples illustrate that methodological challenges can be overcome in order to

---

[118] For specific limitations of the individual studies, we refer to the section covering limitations of the respective essays.

[119] Note that this includes both traditional hash tags (e.g., #CDU) as well as financial hash tags (e.g., $AAPL).

[120] These attempts may follow the behavior of internet savvy members of the Pirates Party before the federal election in Germany, who generated a substantial buzz on Twitter that was certainly not representative of the general population. According to a direct response to the publication of Essay 1 in the Social Science Computer Review (Jungher, Jürgens and Schoen, 2011), roughly 35% of the political microblogs posted prior to the election were related to the Pirate’s Party. The sample in Essay 1 was limited to the major parties represented in the German parliament before the election. As a result, the Pirates Party, which campaigned almost exclusively on internet-related topics, was not included in the sample.

leverage the data sources, methods, and results presented in this dissertation in further research endeavors.

With respect to weighing information, practical applications have taken the lead by implementing various related functions to enable users. The commercial stock microblogging forum StockTwits.com, for example, allows users to “Like” or “Flag” individual messages in order to “recognize greatness as they see it on the StockTwits stream” and to alert the provider of “spam, pumping, penny stocks, or any other malicious behavior.”[121] TweetTrader.net includes various features (e.g., user votings and stock game), which allow participants to express their opinions on individual messages (see Essay 5 for details). However, these are only simple practice-oriented examples. The academic exploration and development of suitable market designs may help tap the wisdom of the user crowd more effectively. Note that this type of scientific design has long provided ground for academic research in the related fields of information markets (for an overview, see Hahn & Tetlock, 2006) and online feedback systems (e.g., Garcin, Faltings, & Jurca, 2009).

Finally, the aggregation of information remains crucial. The display of word clouds and the ranking of “Trending Topics”[122] on Twitter’s website illustrate popular attempts of information aggregation. The aggregation methods used in this dissertation, such as the mere count of party mentions in Essay 1, are fairly simple. Even the classification of messages by sentiment and news topics (Essays 2 and 3) has limitations when it comes to understanding the information content of a particular news item. There are many interpretive aspects of news that this approach does not capture. Future research should, for instance, try to better distinguish the novelty and significance of information (Groß-Klußmann & Hautsch, 2009). Progress in these research areas may help strengthen the role of microblogging forums as information markets.

3 Implications

The results in this dissertation provide evidence supporting the theory that followership relationships and retweets provide microblogging forums with a mechanism for weighing information and enables them to function as information markets. Interestingly, business models are already being created that support the notion that access to one’s followers and retweets are scarce and valuable resources not unlike commercial airtime in public

[121] http://blog.stocktwits.com/introducing-like-and-flag-features/ (last accessed, May 15, 2011)
[122] Trending Topics are the most talked about topics currently being discussed on Twitter.

broadcasting. Retweets can, for example, be used as a digital currency to pay for products online[123] and followers are being sold much like other intangible assets.[124]

As we move increasingly larger parts of our lives online, it becomes more and more important to read these digital footprints and understand them relative to their real-world counterparts.[125] This dissertation shows how the combination of tools from fields such as computational linguistics and social network analysis with challenges from a social science context can help to address research questions that are hard to tackle with traditional methods. Among these tools, sentiment analysis is the most established area of research (Pang & Lee, 2008). This dissertation has illustrated how more nuanced aspects of social media content can be used in academic research ranging from news topics to relationships between frequently mentioned entities (e.g., political parties or publicly traded companies). The political and financial domains were used as the specific contexts in this dissertation, but related research suggests that findings may be transferrable to other domains (Asur & Huberman, 2010; O'Connor et al., 2010). Overall, these findings demonstrate that Twitter can be considered a valid indicator of real-world events.

The results clearly suggest that Twitter may complement traditional methods of forecasting and public opinion research (e.g., polls or surveys in the political arena and analyst models in capital market research). There are multiple advantages of extracting public opinion from microblogging content including cost, speed, timeliness, and a greater variety of topics (O'Connor et al., 2010). In sum, this dissertation hopes to establish microblogging forums as an information market so that they can become for qualitative (e.g., textual) information what the market mechanisms represents to quantitative price signals.

[123] http://www.paywithatweet.com/

[124] http://buyafollower.com/

[125] Note that the FuturICT flagship proposal is one of the finalists currently competing for a 10 year 1 billion dollar research grant from the EU Commission (for details see http://www.futurict.ethz.ch, last accessed May 15, 2011). It aims to leverage real-time data from information systems worldwide to develop a sophisticated simulation, visualization and participation platform („Living Earth Simulator“) in order to detect and mitigate crises (e.g., financial crises) and support the decision-making of policy-makers, business people and citizens.

4 References (Introduction & Conclusion)

Asur, S., & Huberman, B. (2010). Predicting the future with social media. *Working paper*. Retrieved May 15, 2011 from http://arxiv.org/pdf/1003.5699

Bhojraj, S., Lee, C. M. C., & Oler, D. K. (2003). What's my line? A comparison of industry classification schemes for capital market research. *Journal of Accounting Research*, 41, 745-774.

Böhringer, M., & Richter, A. (2009). Adopting social software to the intranet: A case study on enterprise microblogging. In *Proceedings of the 9th Mensch & Computer Conference*, 293-302. Berlin.

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

Boyd, D., Golder, S., & Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. In *Proceedings of the 43rd Hawaii International Conference on System Sciences* (pp. 1-10), Hawaii.

Brustein, J. (2010, October 31). Nation's political pulse, taken using net chatter, *New York Times online edition*. Retrieved May 15, 2011 from http://www.nytimes.com/2010/11/01/technology/01sentiment.html

Dann, S. (2010). Twitter content classification. *First Monday*, 15, 1-11.

Cahan, R., Luo, Y., Jussa, J., & Alvarez, M.-A. (2010). Signal processing: Beyond the headlines, *Deutsche Bank Global Markets Research - Quantitative Strategy Research.* Retrieved May 15, 2011 from http://online.thomsonreuters.com/newsscopereports/

Edison Research (2010). *Twitter usage in America: 2010*. Retrieved May 15, 2011 from http://www.edisonresearch.com/home/archives/2010/04/twitter_usage_in_america_2010_1.php

Garcin, F., Faltings, B., & Jurca, R. (2009). Aggregating reputation feedback. In *Proceedings of the First International Conference on Reputation: Theory and Technology* (pp. 62-74), Gargonza, Italy.

Groß-Klußmann, A., & Hautsch, N. (2009). Quantifying high-frequency market reactions to real-time news sentiment announcements. *Working Paper.* Retrieved May 15, 2011 from http://ideas.repec.org/p/hum/wpaper/sfb649dp2009-063.html

Hahn, R., & Tetlock, P. (Eds.) (2006). *Information markets – a new way of making decisions*. Washington: The AEI Press.

Grosseck G., & Holotescu, C. (2008). Can We use Twitter for educational activities?. In *Fourth International Scientific Conference eLearning and Software for Education*, Bucharest, Romania.

Hayek, F. V. (1945). The use of knowledge in society. *American Economic Review*, 35, 519-530.

Jansen, H. J., & Koop. R. (2005). Pundits, ideologues, and ranters: The British Columbia election online. *Canadian Journal of Communication*, 30, 613-632.

Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, 60, 1-20.

Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why we twitter: Understanding microblogging usage and communities. In *Proceedings of the Ninth WebKDD and First KDD Workshop on Web Mining and Social Network Analysis* (pp. 56-65). San Jose, CA: ACM.

Jordan, J. (2010). Hedge fund will track Twitter to predict stock moves, *Bloomberg (online edition)*. Retrieved May 15, 2011 from http://www.bloomberg.com/news/2010-12-22/hedge-fund-will-track-twitter-to-predict-stockmarket-movements.html

Kwak, H., Lee C., Park, H., & Moon S. (2010). What is Twitter, a social network or a news media?. In *Proceedings of the 19th International Conference on World Wide Web* (pp. 591-600), New York: ACM.

McCool, G. (2010, October 5). Facebook & Twitter used in stock fraud. *Reuters online edition*. Retrieved May 15, 2011 from http://www.reuters.com/article/idUSTRE69453L20101005

Mitchell, M. L., & Mulherin, J. H. (1994). The impact of public information on the stock market. *Journal of Finance*, 49, 923-950.

O'Connor, B., Balasubramanyan, R., Routledge, B., & Smith, N. (2010). From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of International Conference on Weblogs and Social Media* (pp. 122-129). Washington, DC: AAAI.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*, 1–135.

Pew Research Center (2010). *8% of online Americans use Twitter*. Retrieved June 1, 2011 from http://www.pewinternet.org/Reports/2010/Twitter-Update-2010.aspx

Reinhardt, W., Ebner, M. Beham, G., & Costa C. (2009). How people are using Twitter during conferences. In *Proceeding of Fifth EduMedia Conference* (pp. 145-156), Salzburg.

Sunstein, C. (2007). Neither Hayek nor Habermas. *Public Choice*, 134, 87-95.

Surowiecki, J. (2004). *The Wisdom of crowds: Why the many are smarter than the few and wow collective wisdom shapes business, economies, societies, and nations*. New York: Random House.

TechCrunch (2010). Twitter seeing 90 million tweets per Day, 25 percent contain Links. *TechCrunch Blog*. Retrieved May 15, 2011 from http://techcrunch.com/2010/09/14/twitter-seeing-90-million-tweets-per-day/

TechCrunch (2011). Twitter tweets some big Q1 stats: 155 million tweets a day now. *TechCrunch Blog*, Retrieved May 15, 2011 from http://techcrunch.com/2011/04/06/twitter-q1-stats/

Zhang, W., & Skiena, S. (2010). Trading strategies to exploit blog and news sentiment. In *Proceedings of the International Conference on Weblogs and Social Media* (pp. 275-378), Washington, DC: AAAI.

Zhao, D., & Rosson M. B. (2009). How and why people Twitter: The role that micro-blogging play in informal communication at work. In *Proceedings of the International Conference on Supporting Group Work* (pp. 243-252), Sanibel Island, Florida: ACM.