University of Konstanz
Department of Economics

Mechanically Extracted Company Signals and their Impact on Stock and Credit Markets

Ferdinand Graf

Working Paper Series
2011-18

http://www.wiwi.uni-konstanz.de/workingpaperseries

Mechanically Extracted Company Signals and their Impact on Stock and Credit Markets

Ferdinand Graf[1]

University of Konstanz

May 31, 2011

Abstract: I analyze company news from Reuters with the 'General Inquirer' and relate measures of positive sentiment, negative sentiment and disagreement to abnormal stock returns, stock and option trading volume, the volatility spread and the CDS spread. I test hypotheses derived from market microstructure models. Consistent with these models, sentiment and disagreement are strongly related to trading volume. Moreover, sentiment and disagreement might be used to predict stock returns, trading volume and volatility. Trading strategies based on positive and negative sentiment are profitable if the transaction costs are moderate, indicating that stock markets are not fully efficient.

Keywords: Content Analysis, Company News, Market Microstructure
JEL-classification: G12, G14

[1]Department of Economics, Box D-147, 78457 Konstanz, Germany, Phone: +49 7531 88 3620, E-Mail: ferdinand.graf@uni-konstanz.de

1 Introduction

Investors read daily newspapers, internet articles, watch TV news and listen to the radio. The information obtained might affect their trading decision and, hence, market prices, trading volume and volatility. Barber and Odean (2008) show that the number of news releases by Dow Jones News Service is related to the trading behavior of individual investors, but not institutional investors. Engelberg and Parsons (2011) find a causal relationship between financial news articles in local newspapers and the trading volume of local retail investors. However, news have many dimensions. The number of relevant news articles for a company is a very restrictive measure and ignores much information that might be important for financial markets, e.g. the sentiment. Tetlock (2007) and Groß-Klußmann and Hautsch (2011) find that the sentiment of news articles predicts daily index returns, and intraday liquidity and volatility, respectively. The sentiment of chat-room postings, which could contain news as well, may have predictive power for financial markets too, see Antweiler and Frank (2004), Das, Matinez-Jeres and Tufano (2005) and Das and Chen (2007). I build on these studies and construct a flexible content analysis algorithm and analyze company news from Reuters.

Reuters company news usually describe and interpret a wide range of facts and events which might be relevant for companies. The author's interpretation and her word choice may provide valuable information for financial markets. The author's view might account for the economic environment, the firm's industry position, the management quality and much more aspects which are rather hard to measure quantitatively. If the author concludes that some fact is positive news for a company, she will use friendly and positive words to write the news story. If the facts are considered as negative, alarmed and sad words will probably characterize the news story. Of course, the quality of the author's comments depends on her background. This makes the analysis of chat-room postings and their impact on the market difficult, since everybody can post her opinion, rumors or lies without reputation damage. Another advantage of Reuters company news is that it allows to study the impact of heterogeneous events on financial markets simultaneously.

I use the 'General Inquirer' to measure the sentiment of a news story with respect to a company and disagreement among news stories mechanically. The 'General Inquirer' assigns words to word categories. The word categories 'positive' and 'negative' are used to measure positive and negative sentiment of news stories. Also, I use the word categories 'strong' and 'weak' to measure the uncertainty of a news story. I test if positive sentiment, negative sentiment and disagreement of Reuters company news articles impact financial markets. The data cover 62 large U.S. companies listed at the NYSE or the Nasdaq with liquid stock option and CDS markets for the time period June 01, 2007 to December 31, 2010.

First, I investigate the impact of sentiment and disagreement on the abnormal stock return derived from the three factor Fama-French model, stock and option trading volume, the volatility spread and the CDS spread using daily data. This analysis allows to test implications given by market microstructure models where investors interpret public signals individually. My results are consistent with these models. Second, I show that sentiment and disagreement have predictive power for (abnormal) stock returns, stock trading volume and the volatility spread. Positive sentiment is frequently followed by positive (abnormal) returns and disagreement tends to lower the (abnormal) return on the following day. The volatility spread increases after negative sentiment and disagreement. Stock trading volume is significantly higher after news with positive sentiment, but disagreement reduces stock trading volume at the following day. The latter finding is surprising and might be due to immediate execution of scheduled orders, giving contradicting news articles.

Finally, I test the economic relevance of positive and negative sentiment by analyzing trading strategies based on sentiment. Even with realistic transaction costs of 10 bps per round-trip, the trading strategies are comparable to approximate arbitrage opportunities, indicating that the stock market is not fully efficient. For transaction costs of 20 bps, the trading strategies are on average still profitable, but bear a substantial loss potential. The strategies cannot compensate transaction costs of 30 bps and more.

The contribution of this paper is manyfold. (1) I consider a large number of companies with liquid stock and derivative markets and analyze the relationship between news articles and abnormal stock returns, stock and option trading volume, the volatility spread and the CDS spread company individually. Hence, I do not aggregate returns, etc., at the same day across companies. This distinguishes this study from Tetlock (2007), who considers index returns, and from Das et. al. (2005), who analyze four representative companies individually. (2) I analyze a comprehensive and hand-collected dataset of news stories, downloaded from the homepage of Reuters with a flexible procedure, and extend Groß-Klußman and Hautsch (2011), who relate pre-calculated dummy variables for positive and negative sentiment to the stock market, using a continuous sentiment score. (3) The Reuters company news are highly credible. This distinguishes this analysis from Antweiler and Frank (2004) and Das and Chen (2007), who study chat-room postings. Das et. al. (2005) analyze chat-room postings, too, claiming that these postings disseminate public information. This paper analysis might contribute to those study since I analyze news articles which might be closer to public information and, hence, less noisy. (4) To my best knowledge, this study is the first that analyzes the relationship between sentiment respectively disagreement of general news articles and the CDS spread.

The rest of the paper is organized as follows. Section 2 gives a literature review. Section 3 derives testable hypotheses from market microstructure models. Thereafter, I explain how market activity is measured. I describe my hand-collected news database in section 4. Section 5 describes the content analysis and defines measures for sentiment and disagreement. Thereafter, I relate these measures to the market variables and develop trading strategies based on sentiment. Section 8 concludes and gives an outlook for further research.

2 Related Literature

Several papers investigate the relationship between a company's publicity and the stock market. Publicity often refers to the number of news articles on the company. In an early study, Mitchell and Mulherin (1994) relate the number of news releases by Dow Jones & Company to the absolute value of the market return, the absolute value of firm-specific return and the trading volume. By controlling for macroeconomic announcements and weekday effects, the study documents a significant relationship between news activity and market activity. Barber and Odean (2008) define attention grabbing stocks as stocks with high abnormal trading volume, extreme returns or news coverage. They show that individual investors are more likely to purchase attention-grabbing stocks than other stocks. Engelberg and Parsons (2011) address the causality between news articles and investors' behavior. They identify articles on earnings announcement in local newspapers. Local news coverage predicts trading volume of local investors and gives strong support to a causal relationship from news coverage to trading. Fang and Peress (2009) study the cross-section of stock returns. They find that stocks with media coverage, measured by the number of articles on the company in the four major US newspapers (New York Times, USA Today, Wall Street Journal, Washington Post), underperform stocks without media coverage.

Of course, the number of news per day ignores the content of the news article. Tetlock (2007) identifies weak or negative words in the daily article 'Abreast of the Market' in the Wall Street Journal with a content analysis algorithm, the 'General Inquirer'. He finds that the number of negative or weak words predicts the return of the Dow Jones Industrial Average on the following day. This effect is offset within the subsequent five days and disappears after one week. Groß-Klußmann and Hautsch (2011) show that the sentiment of news articles and their relevance for stocks listed at the LSE predict high frequency returns, volatility and liquidity. The sentiment of a news article is calculated by Reuters and can take on only the values +1, 0 and -1. The relevance of the news story determines the sensitivity of the market with respect to the news article. Tetlock et. al. (2008) show that print news can predict fundamental value as well as market value. However, trading strategies based on these forecasts generate profits only if transaction costs are excluded. Carretta et. al. (2010) study the Italian stock market and its reaction to corporate governance news. News stories are analyzed with respect to content and tone, revealing that the content of news on profitable corporations is important to explain stock returns.

Several studies use a more general definition of news and analyze chat-room postings. However, this kind of information is presumably more noisy and, hence, less credible than regular news articles. Antweiler and Frank (2004) relate measures for bullishness and disagreement in chat-room postings and chat-room activity to market activity. Their main finding is that chat-room postings predict realized volatility and trading volume, given high frequency data. Das, Martinez-Jerez and Tufano (2005) analyze chat-room postings of four representative companies from different industries and find a contemporaneous relationships between the sentiment in investors' conversations and market returns, but no predictive power. This motivates their conclusion that investors first trade and then talk. Das and Chen (2007) apply a wide spectrum of text analysis algorithms to chat-rooms postings and develop measures for sentiment and disagreement. Relating these measures to the stock market return of a company shows that market activity is related to small investors' sentiment. Tumarkin and Whitelaw (2001) analyze chat-room postings, too. However, their findings on the interdependence between market observations and posted news are inconclusive.

By using a narrow definition of news / events, the number of articles might be reduced significantly and a mechanical content analysis might be redundant. Brooks, Patel and Su (2003) analyze stock responses to rare, negative surprises like the Exxon Valdes catastrophe, plane crashes or the sudden death of a CEO. They find that stocks respond with a delay to fully unanticipated news, but overreact, see also Brourn and Derwall (2010), who study terrorist attacks and earthquakes, respectively. Yu (2011) uses the dispersion in analyst forecasts to measure disagreement. A portfolio of stocks with high disagreement underperforms compared to a portfolio with low disagreement. Boyd, Hu and Jagannathan (2005) do not focus on firm specific news, they analyze unemployment reports and find that stock markets respond to unemployment news conditional on the state of the economy.

Not only stock prices seem to respond to textual information, there is evidence that the price of credit derivatives and fixed income securities do so as well. Norden (2008) studies the relationship between the credit spread of credit default swaps and rating announcements. He finds that the rating downgrade of a company is anticipated by the company's major lenders, concluding that information spills over from the major lenders to the market, see also Hull, Predrescu and White (2004). Hess et. al. (2008) study the impact of macroeconomic news on commodity future price indices. The index return responds to news about the inflation rate or real activity only in a recession. Hautsch and Hess (2002) analyze the U.S. employment report impact on the mean and the volatility of T-bond futures returns. The mean's reaction is related to surprises and the volatility's reaction is related to uncertainty in the announcements. Besides of liquidity patterns, the study documents asymmetries in the T-bond future price reaction to positive and negative news. Coval and Shumway (2001) propose a very remarkable measure of information arrival, the ambient noise in the CBOT trading pit. This measure predicts returns, liquidity and the customer order flow of the 30 year U.S. treasury bond for several minutes.

3 Market Reactions

3.1 Hypotheses

The efficient market hypothesis says that market prices adjust immediately to public information. I test this hypothesis. Hence

Hypothesis 1: Market prices adjust immediately to public information, leaving no predictive power for public company news.

Assuming homogeneous beliefs, the absence of private information and homogeneous preferences, investors do not trade if new information becomes public, starting at an equilibrium, see Milgrom and Stokey (1982). However, this is inconsistent with the empirical studies cited before. Harris and Raviv (1993) and Kandel and Pearson (1995) drop the assumption of homogeneous beliefs. They assume that investors observe noisy, public signals and update their beliefs consistent with their individual interpretation. Different levels of confidence with respect to the noisy, public signal across investors (difference of opinion) might cause heterogenous changes in the demand for risky assets and, hence, trading. Furthermore, Cao and Ou-Yang (2009) extend this framework and show that public signals and heterogeneous priors may also cause trading in stock options. Banerjee and Kremer (2010) relate a time-varying magnitude of difference of opinion to trading volume and price volatility and find that 'periods of major disagreement are periods of higher volume and also of higher absolute price changes'. The latter might be used as measure for volatility.

Company news might be closely related to public signals. Therefore, I approximate the intensity of public company signals by the sentiment of relevant news stories. The degree of differences of opinion is approximated by the variation in the sentiment of relevant news articles within on trading day, hereafter called disagreement. Hence

Hypothesis 2: Trading volume of stocks and options increases with positive sentiment and negative sentiment.

Hypothesis 3: An increase in disagreement raises trading volume of stocks and options.

Hypothesis 4: The stock return volatility increases with disagreement.

According to Hong and Stone (2007), heterogenous priors of investors are one explanation why disagreement affects the stock market. Others are limited attention or gradual information flow. However, these explanations have similar implications on the relationship between the stock market and disagreement.

Another strand of literature explains trading patterns by information asymmetries across investors. Blume, Easley and O'Hara (1994) show theoretically that trading volume might contain valuable information to determine the precision of noisy, private information and might be useful for stock pricing, see also Suominen (2001). Tetlock (2010) analyzes market data around company announcements and finds pattern which are consistent with information asymmetries. Sarwar (2005) and Kyriacou and Sarno (1999) study option trading volume and market volatility. Both studies find a strong predictive power of option trading volume for volatility and vice versa. Adjusting hedged portfolios to changes in volatility might explain why volatility predicts option trading volume. Also, investors with private information might exploit their informational advantage aggressively with options and use the leverage effect or bet on volatility via derivatives. Hence, trading volume might predict volatility. By analyzing the ratio of traded put and call options, Pan and Poteshman (2006) find that stocks with low ratios significantly outperform stocks with high ratios. Again, this indicates that informed traders use derivates to benefit from their informational advantage.

Empirically, it is likely that evidence for - at least parts of - both strands of literature, i.e. difference of opinion and asymmetric information, appears jointly. I test the implications given by the difference of opinion theory, but allow for inter-temporal dependencies between trading volume, stock volatility and returns to account for information asymmetries. Furthermore, I include the CDS spread of a company for two reasons: (1) Structure models for credit derivatives like Merton (1974) imply that the equity market and the credit market are closely linked. Cremers et. al. (2006) and Zhang, Zhou and Zhu (2009) document a close relationship between credit markets and equity markets. Hence, I control for information spillovers from debt to equity markets and vice versa. (2) I test if the CDS spread is related to the degree of difference of opinion and to public signals. Since equity volatility and the unobservable asset volatility in structural models are positively related, the CDS spread might also respond to a change in difference of opinion, given that the equity volatility reacts. Therefore

Hypothesis 5: The CDS spread increases with disagreement.

Furthermore, the CDS spread represents the market price of a traded derivative. Predictability of the CDS spread might be related to market inefficiency and to Hypothesis 1.

3.2 Measures of Market Reactions

The daily close-to-close excess stock return of company i at day t, denoted $r_{i,t}$, might be used as a measure for the stock market's response to news releases. More appropriate and in line with many other studies is the abnormal stock return, measured by the residuum in the three factor Fama-French model (hereafter FF model / factors / residuum), see Fama and French (1993). The residual measures the stock price movements that are not due to common market risk factors but might be due to firm-specific risk respectively news. The FF factors and the risk-free interest rate are downloaded from the homepage of Kenneth French (see `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/index.html`), the dividend adjusted stock prices are downloaded from Thomson Reuters Datastream. I estimate

$$r_{i,t} = \alpha_i + \beta_{i,\text{Market}} X_{\text{Market},t} + \beta_{\text{SMB},i} X_{\text{SMB},t} + \beta_{\text{HML},i} X_{\text{HML},t} + \varepsilon_{i,t}, \tag{1}$$

where $\beta_{i,\cdot}$ denotes the factor loadings of the corresponding factor $X_{\cdot,t}$ (Market, **S**mall **M**inus **B**ig market capitalization, **H**igh **M**inus **L**ow book to market ratio). The estimated residuum is defined by $\hat{\varepsilon}_{i,t} := r_{i,t} - \hat{r}_{i,t}$, where $\hat{r}_{i,t}$ is the explained stock return. If $\hat{\varepsilon}_{i,t}$ is large in absolute value, it is likely that important firm-specific information arrives. However, a residuum close to zero does not necessarily indicate a calm trading day.

I measure the stock trading volume by the daily turnover volume, divided by the average turnover volume in the preceding 3 months. This measure is denoted $T_{i,t}$. To address the study of Sarwar (2005) and Kyriacou and Sarno (1999), and to test the model of Cao and Ou-Yang (2009), I also include the cumulated option trading volume of all outstanding options on stock i at day t, divided by its 3-month moving average. This measure is denoted $O_{i,t}$. All time series are provided by Thomson Reuters Datastream. I use the 3-month volatility spread, defined by

$$V_{i,t} = IV_{i,t} - RV_{i,t},$$

to measure the investor expectations on volatility relative to realized volatility. $IV_{i,t}$ is the at-the-money implied volatility of 3-month constant maturity options, calculated by Thomson Reuters Datastream. According to Martens and van Dijk (2007), the 3-month realized volatility is well approximated by

$$RV_{i,t} = \sqrt{\frac{1}{2} \sum_{s=t-60}^{t} [(\ln H_{i,s} - \ln L_{i,s})^2 - (2\ln 2 - 1)(\ln R_{i,s})^2]},$$

where $H_{i,s}$ is the highest intraday stock price within day s and $L_{i,s}$ is the lowest intraday stock price. $R_{i,s}$ is the close-to-close gross stock return of day s[2]. Finally, I use the 5-year CDS spread on senior debt as an indicator for the company's default risk. The CDS spreads is denoted $C_{i,t}$, the data are provided by CMA.

4 Company News

4.1 Data Description

My hand-collected database consists of mainly fundamental and unscheduled news stories on companies in the S&P500, FTSE100 or EuroStoxx50 for the time period June 01, 2007 to December 31, 2010. Given a date (*mmddyyyy*) and a company, identified with its *RIC* (= Reuters instrument code), the domain `Reuters.com` returns a list of up to ten news articles on the uniform resource locator `http://www.reuters.com/finance/stocks/companyNews?symbol=`*RIC*`&date=`*mmddyyyy*. All company news stories are downloaded mechanically. Long news stories might span

[2]The volatility spread measures the expected excess volatility relative to the realized volatility. It might measure the risk aversion of the market, too. However, the time series V is non-stationary for many companies and is, hence, differentiated. Since the realized volatility moves very slowly, it has only little impact on the first difference of V such that the results do not depend on the realized volatility.

over more than one internet page. However, the download routine recognizes this and controls for it.

A news story consists of a headline, the full text or body, a time stamp (date and time), keywords and a list of companies indicating for whom the news story might be important. In the following, this list is called 'related RICs'. The assignment of keywords and related RICs to a news story is done by Reuters. Keywords provide a rough, standardized categorization of the news story (e.g. Major Breaking News, Debt ratings news, Corporate Results, Mergers and Acquisitions). In a nutshell, company news inform about rating adjustments, analyst reports and changes for the stock price target, give summarizing figures on quarterly and annual reports and general news (e.g. macro-economic indicators, political events, articles in the Washington Post, New York Times, Wall Street Journal, etc.). Corrected or updated news are not excluded to capture the information flow correctly. Even though the list of company news on the homepage of Reuters is limited to ten, the number of daily news articles per company, e.g. identified by searching for the company's RIC in 'related RICs', is not bounded in my database because there are many news articles that mention a company or have it in the field 'related RICs' and do not appear in the list for the company on the homepage of Reuters. For the observation period June 01, 2007 to December 31, 2010, there are more than 350,000 unique news stories with respect to the url. The average news article consists of 301.29 words (including numerical expressions and symbols) with a standard deviation of 239.64 words. The median of words per news article is 272 and indicates that the distribution is skewed to the right. The 99% quantile is 961 words. On average, a news article consists of 14.02 sentences with a standard deviation of 33.20. The median is 11, again, indicating that the distribution of sentences per news story is skewed to the right, and the 99% quantile is 44.

Table 1 provides descriptive statistics for the number of news articles per day for all S&P500 companies jointly, for the index components of the Dow Jones Industrial Average by January 01, 2011 and for some frequently used keywords. I have 210,495 news articles for all S&P500 companies on 1311 days. Hence, the daily, average number of news articles for all S&P500 companies is 160.56 with a standard deviation of 94.01. Ignoring Saturdays and Sundays, the average number of news releases per day increases to 212.63 with standard deviation 52.94. On October 22, 2009, 354 news stories were published, this is the maximum number of news stories per day in the observation period. There are 17,525 news stories labeled with the keywords 'Corporate Result', 'Result Forecast' or 'Warnings', this gives a daily average of 13.36 with a standard deviation of 20.13. The total number of news stories with 'Broker research and recommendation' is 1,867, the daily average is, hence, 1.42 and the standard deviation is 2.26. News stories on e.g. Bank of America (BAC.N), identified by searching for 'BAC.N' in 'related RICs', sum up to 11,974, with a mean of 9.13 news stories per day and standard deviation of 8.23.

Figure 1, upper plot, shows the time series of the daily number of news stories with the keywords 'Bankruptcy' or 'Insolvency' (blue curve) and its 3 day moving average (red curve). Since there are no such news stories prior to November 23, 2007, the plot excludes the period June 1, 2007 to November 23, 2007. The large number of news stories in the middle of September 2008 indicates the bankruptcy of Lehman Brothers and the peaks in 2009 and 2010 are mainly due to the sovereign debt crisis in Europe. The lower plot shows the time series of the daily number of news stories for the Bank of America and the corresponding 3-day moving average. This time series starts on June 1, 2007. The time series displays a weekly cyclicality caused by the low number of news articles during the weekend. Again, the default of Lehman Brothers at September 15, 2008 can be clearly identified. The peak in January 2009 is caused by the arranged acquisition of Merrill Lynch by Bank of America.

[Table 1 about here.]

[Figure 1 about here.]

4.2 News Coverage

To improve the understanding of the news database, in the following I investigate the company characteristics that expose a company to news coverage. I consider 62 large companies in the S&P500 with liquid option and CDS market. Table 9 lists these companies. The news exposure of a company is measured by its average number of news articles per day, identified with the company's RIC in 'related RICs'. This measure is denoted Q_i. Alternatively, news coverage is measured by the number of days with at least one news story. This measure is denoted Y_i. Companies are characterized by the average market capitalization in the observation period, CAP_i, the average price-to-book ratio, $P2B_i$, the stock return during the observation period, Ret_i, and the corresponding realized volatility, $\sigma(Ret_i)$.

The average company has an average market capitalization during June 01, 2007 to December 31, 2010 of 8,3223 billions USD and an average price-to-book ratio of 2.67. The average stock market performance in this period and across companies is -9.41% and the average stock return volatility is 42.63. I estimate an ordinary linear regression model, i.e.

$$Q_i = \alpha + \beta_1 P2B_i + \beta_2 \ln(CAP_i) + \beta_3 Ret_i + \beta_4 \sigma(Ret_i) + \eta_i. \tag{2}$$

Table 2 shows the regression estimates for (2) and some straightforward adjustments. a indicates significance at the 1% confidence level, b at the 5% level and c at the 10% level. Even though this analysis excludes small and mid-sized companies, the company size is still a significant, positive determinant of the news coverage. The price-to-book ratio is significant and negative in all regressions. This indicates that companies with high ratios are less often in the news than companies with low ratios. One reason for this pattern might be that the latter companies have a higher potential for stock price increases. The stock return is weakly significant and negative. The stock return volatility is significant, too, and positive. Both indicate that troubled companies are frequently in the media. However, this result might also be due to the financial crisis.

All results are qualitatively the same if Y_i is considered instead of Q_i. Hence, large companies with low price-to-book ratios and volatile stock returns have a high media coverage, or conversely, companies with a high media coverage are large, have a low price-to-book ratio and their stock price is rather volatile.

[Table 2 about here.]

5 Content Analysis and Variable Construction

A company, a news article is relevant for a company if

(a) contains the company's RIC in the field 'related RICs', or

(b) mentions the company name or its nickname in the headline and has the company's RIC in the field 'related RICs'.

Definition (a) is, of course, a broader definition than (b), and sensitive to news regarding the whole industry or direct competitors. The term company name in definition (b) and in the following refers to the shortest fraction of the full company name that clearly identifies the company, e.g. 'Disney' instead of 'Walt Disney Co' or 'Conoco' instead of 'ConocoPhillips'. For most companies I am able to identify the company's nickname very accurately. For example, Bank of America is frequently called BofA, Johnson & Johnson is called J&J and American Express is AmEx. Texas Instruments, often called TI, and General Electrics, shortened GE, can only be identified with a small error rate. Filtering for related RICs in definition (b) ensures that a news story with a headline such as '*BofA cuts Google price target*' is assigned to Google, but not to Bank of America.

Even though a news article is relevant for a company, it is unlikely that all words in the full text are important for the company as well. Hence, I define which passages in the full text of a relevant news story have to be analyzed. Given a relevant news story according to definition (a) or (b), I define the relevant text by:

(c) All words in a sentence are relevant if the company name or nickname is mentioned within the same sentence, or

(d) All words with a distance of at most 5 words to the company name or nickname are relevant.

Words that contain numerical expressions (e.g. 'B787', 'A330-200', '$35') are not counted for the word distance since they are not related to the sentiment of a news story.

Given a company, I analyze the content of the relevant text of a relevant news story and assign a numerical value to it. The approach relies on the 'General Inquirer' (`http://www.webuse.umd.edu:9090/`). The 'General Inquirer' is a dictionary based content analysis algorithm. It assigns words to word categories and reports the number of hits in each cluster, relative to all analyzed words, see Stone et. al. (1966). There are more than 80 word categories. However, I restrict myself to the categories 'positive', 'negative', 'strong' and 'weak'. Even though being very popular, the 'General Inquirer' is not perfect. Many words have more than one meaning and might be incorrectly assigned to a word category, see for example Loughran and McDonald (2011), who test the performance of the 'General Inquirer' by analyzing annual and 10-K reports, and find that a substantial fraction of negative words (about 60%) is misinterpreted. However, the content of the Reuters news articles is very general and hardly comparable to annual reports, hence I expect a low error rate.

Consider, for example, the following news stories:

> Feb. 29, 2008, Northrop-EADS beats Boeing to built U.S. tanker
>
> WASHINGTON, Feb. 29, 2008 - The U.S. Air Force said on Friday it had picked a transatlantic team led by Northrop Grumman, instead of Boeing, to start building a new aerial refueling fleet in a surprise choice worth about $35 billion. Northrop Grumman Corp (NOC.N) and its European partner, Airbus parent EADS (EAD.PA), "clearly provided the best value to the government," Sue Payton, the Air Force's chief weapons buyer, told reporters at a briefing. The contract is to supply up to 179 tanker aircraft in a deal valued at about $35 billion over the next 15 years, the Air Force said in a statement. The aircraft will replace [...].

> Sept. 29, 2009, Kenya Airways eyes Airbus A330-200s sources
>
> NAIROBI, Sept. 29, 2009 - Kenya Airways (KQNA.NR) is in talks with Airbus (EAD.PA) about buying several A330-200 planes after delays to Boeing's (BA.N) much-anticipated B787 Dreamliner jet, senior officials at the airline said on Tuesday. The carrier's Chief Executive Officer Titus Naikuni said on Friday the company was in talks with Airbus [...].

Clearly, the news stories are rather positive for Northrop and EADS, respectively, and negative for Boeing. According to the 'General Inquirer' dictionary, there are several positive words in the second sentence of the first news story ('clearly', 'provide', 'best'). There, Northrop and EADS are mentioned, but not Boeing. Regarding the second news story, approach (c) might fail since Airbus and Boeing are mentioned in the same sentence. However, the five word distance around Boing covers the word 'delay', which clearly signals negative sentiment for Boeing, but there are no negative words within the five word distance around EADS. Of course, the performance of both approaches depends on the structure of the news story. If a news article describes complex events where many companies interact, both approaches might fail to measure the correct sentiment. However, both approaches perform very well for simple or well structured news. Furthermore, the company name is sometimes replaced by a synonym, e.g. 'planemaker' instead of 'Boeing'. Such cases are not recognized.

To homogenize market data and news stories, I assign news stories which were released after 4 p.m. New York time to the following trading day. News stories published between Friday, 4 p.m. and Monday, 4 p.m. are assigned to Monday. Assume there are $Q_{i,t} \in \mathbb{N}$ news stories for company i on day t according to definition (a) or (b). Given the relevant text following definition (c) or (d), let $Pos_{i,t,j}$ [$Neg_{i,t,j}$] denote the number of positive [negative] words relative to the total number of words in the relevant text of news story $j = 1, \ldots, Q_{i,t}$. Then, the average, relative number of positive words and the average, relative number of negative words are used to measure positive signals, $P_{i,t}$, and negative signals, $N_{i,t}$, at t and for company i, i.e.

$$\begin{aligned} P_{i,t} &= \max\left\{ \frac{1}{Q_{i,t}} \sum_{j=1}^{Q_{i,t}} (Pos_{i,t,j} - Neg_{i,t,j}), 0 \right\}, \\ N_{i,t} &= \max\left\{ \frac{1}{Q_{i,t}} \sum_{j=1}^{Q_{i,t}} (Neg_{i,t,j} - Pos_{i,t,j}), 0 \right\}. \end{aligned} \tag{3}$$

$P_{i,t}$ and $N_{i,t}$ might be interpreted as positive and negative public signals in the style of Harris and Raviv (1993). High values of $P_{i,t}$ or $N_{i,t}$ indicate strong signals.

It is likely that there is a monotone relationship between the average of net sentiment of a day, i.e. $\frac{1}{Q_{i,t}} \sum_{j=1}^{Q_{i,t}} (Pos_{i,t,j} - Neg_{i,t,j})$, and the abnormal stock returns or the CDS spread, but trading volume and volatility are presumably not monotonically related to the net sentiment. Therefore, positive and negative signals are disentangled. I do not exclude days without news releases since these days are important as well and might indicate 'neutral' or 'calm' days. The sentiment for these days is set to zero.

Furthermore, I define two disagreement scores. Usually, the investors' view on a company is influenced, perhaps driven, by public information. If news stories disagree heavily, it is likely that investors disagree as well. Hence, the degree of difference of opinion among investors might well be approximated by the variation in the sentiment of news stories. I define

$$D^{\text{std}}_{i,t} = \sigma\left((Pos_{i,t,j} - Neg_{i,t,j})_{j \in Q_{i,t}} \right), \tag{4}$$

where $\sigma(\cdot)$ is the standard deviation. I set $D^{\text{std}}_{i,t} = 0$ if $Q_{i,t} \leq 1$.

Inspired by Das and Chen (2007), I construct a second measure for disagreement. Define the auxiliary variable $A_{i,t,j} := \mathbf{1}(Neg_{i,t,j} < Pos_{i,t,j}) - \mathbf{1}(Neg_{i,t,j} > Pos_{i,t,j})$. $\mathbf{1}(\cdot)$ is the indicator function. It is one if and only if the argument is true. Hence, $A_{i,t,j} = 1$ if the net sentiment of news j is positive, it is zero if $Pos_{i,t,j} = Neg_{i,t,j}$ and -1 otherwise. A might be interpreted as a buy- or sell-signal for investors.

Disagreement is alternatively measured by

$$D_{i,t}^{\text{pol}} = \frac{\max\left\{\sum_{j=1}^{Q_{i,t}} |A_{i,t,j}|, 1\right\}}{\max\left\{\left|\sum_{j=1}^{Q_{i,t}} A_{i,t,j}\right|, 1\right\}}. \tag{5}$$

If all news stories on day t and for company i have a positive sentiment or all news stories have a negative sentiment, $D_{i,t}^{\text{pol}} = 1$. This might indicate no disagreement. For days without news stories ($Q_{i,t} = 0$) I set $D_{i,t}^{\text{pol}} = 1$, too. For all other days $D_{i,t}^{\text{pol}} > 1$. $D_{i,t}^{\text{pol}}$ is high if there are many news stories with positive or negative sentiment (numerator is large) and the number of positive and negative news stories is balanced (denominator is small). These days might be associated with high disagreement across investors. Whereas D^{pol} measures the polarity of $(A_{i,t,j})_{j \in Q_{i,t}}$ and ignores the magnitude of the net sentiment, D^{std} is sensitive to variations in the net sentiment even though the sign of the net sentiment might be the same for all news stories.

6 Regression Results

6.1 Contemporaneous Analysis

I analyze the contemporaneous relationship between sentiment respectively disagreement and the financial market. This analysis is motivated by Das, Martinez-Jeres and Tufano (2005) and the literature on difference of opinion. This analysis allows to test Hypotheses 2 to 5 on the co-movement of market variables and public signals respectively the degree of difference of opinion. The analysis does not allow to conclude on market efficiency and the predictability of market returns. Even though the news stories are unscheduled, a significant relationship between sentiment respectively disagreement and market prices or returns on a daily frequency might be consistent with efficient markets if the market anticipates the news.

According to Groß-Klußmann and Hautsch (2011), the relevance of a news article for a company determines the strength of the relationship between sentiment of the news and the market. Hence, I apply the more restrictive definition of relevance, i.e. definition (b), and use definition (c) to identify the relevant words. The other definitions are used for robustness tests.

6.1.1 Company Individual Analysis

As shown in Blume, Easley and O'Hara (1994), volatility and historical stock prices might be valuable information for future stock returns. Pan and Poteshman (2006) document that option trading contains relevant information for stock returns, too, and according to Sarwar (2006) and Kyriacou and Sarno (1999), option trading volume and volatility interact. Cremers et. al. (2008) report a significant relationship between equity markets and credit markets. Chordia, Sarkar and Subrahmanyam (2005) study the intertemporal association between liquidity, volatility and returns by applying a vector autoregressive model. Also, the difference of opinion literature implies positive autocorrelation in trading volume, negative autocorrelation in returns and positive correlation between trading volume and volatility. To control for these associations and to determine the relationship between the financial market and sentiment and disagreement, respectively, accurately, I choose the most parsimonious regression model that allow for the aforementioned pattern, a vector autoregressive process with one lag. I estimate

$$\begin{aligned}&\Big[\hat{\varepsilon}_{i,t}\ T_{i,t}\ \Delta V_{i,t}\ O_{i,t}\ \Delta C_{i,t}\Big]' \\ &= \Lambda_i \Big[\hat{\varepsilon}_{i,t-1}\ T_{i,t-1}\ \Delta V_{i,t-1}\ O_{i,t-1}\ \Delta C_{i,t-1}\Big]' + \beta_i [P_{i,t}\ N_{i,t}\ D_{i,t}]' + K_i U_t + \eta_{i,t}.\end{aligned} \tag{6}$$

D stands for the disagreement score and refers to D^{std} or D^{pol}. Frequently, the augmented Dicky-Fuller test cannot reject the unit-root hypothesis for the CDS spread and for the volatility spread. Hence, I replace these time series by the first difference for all companies. ΔV and ΔC denote the first difference of the volatility spread and the CDS spread, respectively. $\hat{\varepsilon}_{i,t}$, $T_{i,t}$ and $O_{i,t}$ are always stationary. Λ_i is a 5×5 matrix and captures possible inter-temporal associations between the abnormal returns, trading volume in stock and options and the change in the volatility spread and the CDS spread, respectively. β_i is a 5×3 matrix and measures the association between sentiment respectively disagreement and the market. U_t is 5×1 vector with weekday dummies and K_i's dimension is 5×5. $\eta_{i,t}$ is white noise.

[Table 3 about here.]

[Table 4 about here.]

[Table 5 about here.]

Tables 3, 4 and 5 show the estimates for β_i. The companies listed in these tables are chosen because I have a sufficient number of daily observations for all time series jointly to calculate reliable regression coefficients and p-values. A list of company names and RIC is provided in the appendix. The option data is available for most companies since June 2008 and determine the beginning of the estimation period, whereas the CDS spread is available until October 2010 and determine the end. With the exception of Intel Technology (INTC.O) and Travelers Companies (TRV.N), I have 597 days without missing observations for each company. There are 479 observations for Intel and 660 observations for Travelers. The FF residuum is estimated in-sample using the time span June 01, 2008 to September 30, 2010. The average FF-R^2 across all companies is about 54%, indicating that the general market movements explain a substantial fraction of the variation in the stock returns. The average correlation between the abnormal return and the change in the volatility spread across all companies is -0.2859. Stock and option trading volume are on average correlated by 0.3201 and the change in the volatility spread and the CDS spread are on average correlated by 0.2219. All other correlations between the market variables are close to zero. On average, positive and negative sentiment are correlated by -0.0725, positive sentiment and D^{std} are correlated by 0.2613 and negative sentiment and D^{std} by 0.1727. The average correlation between D^{pol} and P respectively N is insignificantly higher.

Table 3 gives the estimated, contemporaneous relationship between positive sentiment and abnormal returns, stock trading volume, the change in the volatility spread, option trading volume and the change in the CDS spread (this is the first column of $\hat{\beta}_i$), as well as the number of days with positive sentiment, $\#(P > 0)$, the mean of positive sentiment, given all days with positive sentiment, $m(P|P > 0)$, and the standard deviation, $\sigma(P|P > 0)$. Table 4 shows the second column of $\hat{\beta}_i$, this is the estimated relationship between negative sentiment and the market, and the corresponding descriptive statistics. Table 5 shows the estimated regression coefficients of disagreement. In all tables, a indicates significance at the 1% confidence level, b at the 5% level and c at the 10% level. I do not show the regression estimates for Λ_i and K_i.

As can be seen in the first column of Tables 3 and 4, positive sentiment and negative sentiment are frequently significant for the FF residuum. Often, the coefficient of positive sentiment is positive and the coefficient of negative sentiment is negative, indicating that positive news are associated with positive abnormal returns and negative news with negative abnormal returns. This suggests that the General Inquirer and the relevant text identification procedure approximate the 'true' sentiment or signal accurately. Disagreement is frequently significant, but the sign of the significant coefficients varies among companies. There are 9 significant, positive coefficients and 10 significant, negative coefficients. Hence, it is unclear which effect dominates on average. The average R^2 across all companies with respect to the abnormal return is 4.57%. Compared to an average R^2 of 2.99% in regression model (6) and omitting $\beta_i[P_{i,t}N_{i,t}D_{i,t}]$, shows that positive and negative sentiment and disagreement account on average for 1.58 percentage points in the R^2. This significant increase by more than 50% is exclusively due to the content analysis and highlights its accuracy. Even though the news articles are usually unscheduled and fundamental, the significant link between returns and news does not allow conclusion on market efficiency.

The average R^2 of regression model (6) with respect to stock trading volume is 38.48% and the average R^2 of (6) and without the regressors $[P_{i,t}N_{i,t}D_{i,t}]$ is 34.92%, indicating that the content analysis explains about 4 percentage points. Regarding option trading volume, the average R^2 increases from 15.43% without the content analysis to 16.46%. To test Hypothesis 2, I use positive sentiment and negative sentiment to approximate the public signal's intensity and study its association with trading volume on the same day. As shown in the second column of Tables 3 and 4, the estimated coefficient of positive sentiment on stock trading volume is positive and significant for 19 out of 62 companies. The coefficient of negative sentiment is positive and significant for 12 companies. Option trading volume shows similar patterns, but the dependencies are less pronounced. However, the signal's intensity seems to be positively related to trading volume, as stated in Hypothesis 2. Hypothesis 3 relates stock and option trading volume to disagreement. Table 5, columns 2 and 4, show the estimated relationship between disagreement across news and trading volume. High disagreement is associated with significantly higher stock trading volume for 49 companies out of 62. There is no company with a significant, negative regression coefficient. Regarding the relationship between option trading volume and disagreement, I find 23 out of 62 positive and significant relationships. Hence, I have strong support for Hypotheses 3. Investors seem to trade on public signals and the degree of disagreement accelerates the trading volume.

The relationship between the change in the volatility spread and disagreement is inconclusive, see Table 5, column 3. The number of significant regression coefficients is low, and the number of significantly negative regression coefficients and significantly positive regression coefficients are almost balanced. Hence, it is infeasible to draw robust conclusions on the relationship between volatility and disagreement. However, Table 4 indicates that the volatility spread widens subsequent to days with negative sentiment (12 positive and significant coefficients in Table 4, column 3). This finding is consistent with evidence on negative correlation between index returns and volatility, since days with negative sentiment are also associated with negative abnormal returns. The average R^2 of the full regression model with respect to the change in the volatility spread is 6.80% and the contribution of the content analysis in terms of average R^2 is 1.13 percentage points. Nevertheless, Hypothesis 4 is not supported.

The change in the CDS spread is often negatively correlated with positive sentiment and positively correlated with negative sentiment. This is consistent with the relationship between the abnormal stock returns and sentiment, and with the response of the volatility spread. Given a negative signal, the value of equity decreases and the equity volatility goes up. Consistent with structural models, the distance to default is reduced and the expected default loss, measured by the CDS spread, increases. As can be seen in Table 5, column 5, disagreement has frequently a significant, positive regression coefficient and supports Hypothesis 5. The content analysis increases the average R^2 of the change in the CDS spread from 5.91% to 7.11%

Most results remain qualitatively unchanged if I consider D^{pol} as a measure of disagreement instead of D^{std}. Therefore, the results are not shown but only discussed briefly. The relationship between abnormal returns and positive sentiment becomes slightly stronger and the coefficient of disagreement is now frequently significant, negative for abnormal returns. This is consistent with Yu (2011), who shows that stocks with high analyst forecast dispersion underperform relative to stocks with low forecast dispersion. Another noteworthy change is that the CDS spread increases with the alternative disagreement measure for many companies. This gives further support to Hypothesis 5.

6.1.2 Pooled Analysis

Next, I analyze all companies jointly. The purpose of this analysis is to investigate the dominating effects between the financial market and sentiment and disagreement, respectively, for all companies. It simplifies the interpretation of the regression coefficients. I estimate

$$\begin{aligned}&\Big[s\hat{\varepsilon}_{i,t}\ sT_{i,t}\ s\Delta V_{i,t}\ sO_{i,t}\ s\Delta C_{i,t}\Big]' \qquad (7)\\ &= \Lambda\Big[s\hat{\varepsilon}_{i,t-1}\ sT_{i,t-1}\ s\Delta V_{i,t-1}\ sO_{i,t-1}\ s\Delta C_{i,t-1}\Big]' + \beta[sP_{i,t}\ sN_{i,t}\ sD_{i,t}]' + KU_t + \eta_{i,t}.\end{aligned}$$

The regression coefficients Λ, β and K are now independent of the company index i. Hence, I make the strong assumption that the relationship between the market variables, measured by Λ, and between the market variables and the information extracted from company news, measured by β, is described by the same coefficients for all firms. I standardize and pool all time series (with the exception of the weekday dummies, which are pooled without further manipulation) by subtracting the time series' individual mean and dividing by the time series' standard deviation. The standardized, pooled time series are marked with the prefix s. As an example, the pooled, standardized stock trading volume is given by the vectors

$$\begin{aligned} sT_{-1} &= \left[\left[\frac{T_{1,t}-m(T_{1,\cdot})}{\sigma(T_{1,\cdot})}\right]_{t=1,\dots,G_1-1},\dots,\left[\frac{T_{L,t}-m(T_{L,\cdot})}{\sigma(T_{L,\cdot})}\right]_{t=1,\dots,G_L-1}\right]' \quad \text{and}\\ sT &= \left[\left[\frac{T_{1,t}-m(T_{1,\cdot})}{\sigma(T_{1,\cdot})}\right]_{t=2,\dots,G_1},\dots,\left[\frac{T_{L,t}-m(T_{L,\cdot})}{\sigma(T_{L,\cdot})}\right]_{t=2,\dots,G_L}\right]',\end{aligned}$$

where $m(\cdot)$ denotes the mean, $\sigma(\cdot)$ is the standard deviation, L is the number of companies and G_i is the number of observations for company i. Then, the estimates in the pooled regression model are given by

$$\begin{aligned}\{\hat{\Lambda},\hat{\beta},\hat{K}\} &= \operatorname{argmin}_{\Lambda,\beta,K}\Big\{\mathbf{1}_{1\times G}\Big([s\hat{\varepsilon}\ sT\ s\Delta V\ sO\ s\Delta C] \qquad (8)\\ &- [s\hat{\varepsilon}_{-1}\ sT_{-1}\ s\Delta V_{-1}\ sO_{-1}\ s\Delta C_{-1}]\,\Lambda - [sP\ sN\ sD]\,\beta - UK\Big)^2\mathbf{1}_{5\times 1}\Big\},\end{aligned}$$

where $G = \sum_{i=1}^{L}(G_i - 1)$, $\mathbf{1}_{a\times b}$ is a matrix of dimension $a \times b$ with 1s everywhere and U is the pooled matrix of weekday dummies. The square symbol in (8) refers to each component in the vector of residuals and is not a matrix operator.

Pooling all company-specific observations gives in total 36,229 company-day observations. Table 6 shows $\hat{\Lambda}$, $\hat{\beta}$ and $\hat{K}$. Disagreement is measured by D^{std} in the upper panel and by D^{pol} in the lower panel. The regression estimates of Λ and K are very similar for D^{std} and D^{pol}. Stock trading volume displays positive autocorrelation, which is consistent with the models of Harris and Raviv (1993) and Banerjee and

Kremer (2010). Furthermore, there are several pattern which are consistent with information asymmetry. Trading volume predicts abnormal stock returns, as discussed in Blume, Easley and O'Hara (1994). Also option trading volume predicts abnormal returns, which might be related to the results of Pan and Poteshman (2006), even though I do not study the ratio of traded put and call options, but the sum.

[Table 6 about here.]

Abnormal stock returns, the change in the volatility spread (which is closely related to the absolute return) and the change in the CDS spread positively predict stock trading volume. This finding is consistent with Barber and Odean (2008), who identify attention-grabbing stocks also with large stock price movements, and find that these stocks have a higher turnover volume than stocks that do not attract attention. However, attention might also be gained by large movements in the CDS spread and an increase in volatility. Consistent with structural models on credit derivatives, the CDS spread increases given an increase in volatility. Surprisingly, it also increases given a large abnormal return. This might be due to analyzing abnormal returns instead of gross returns. The weekday dummies are frequently significant, indicating the presence of weekday effects.

Positive and negative sentiment are still highly significant for abnormal returns. Consistent with the results in the previous section, positive sentiment is positively related to abnormal returns and negative sentiment negatively. The coefficient of D^{std} is insignificant, see upper panel. This does not necessarily mean that disagreement is not relevant for the abnormal return. The insignificance might be due to the heterogeneous relationship between stock prices and disagreement among news articles, e.g. Table 5 shows 9 significant, positive and 10 significant, negative relationships. Hence, both effects are likely to cancel out in the pooled regression. Furthermore, the alternative disagreement score D^{pol} detects a significant, negative relationship between abnormal returns and disagreement in the pooled analysis, see Table 6, lower panel. Again, this is consistent with Yu (2011) and Das et. al. (2005).

The R^2s in Table 6 with respect to the abnormal return are lower than the average R^2 of the firm individual regressions. So, the R^2 of $s\hat{\varepsilon}$ is 0.4% and 0.51%, respectively, whereas the average R^2 of $\hat{\varepsilon}_i$ is 4.57%. This decrease might be due to the restrictive assumption of identical regression coefficients for all companies. The contribution of the content analysis to the R^2 of $s\hat{\varepsilon}$ is about 0.2 percentage point and doubles the explained variation in abnormal returns.

The average R^2 of stock trading volume and allowing for company individual regression coefficients is 38.48% and reduces to 34.99% respectively 35.52% in the pooled analysis. The R^2 of standardized option trading volume is 10.67% and 10.97%, respectively, whereas the average R^2 of the company individual analysis is 16.46%. This moderate decrease in terms of R^2 might indicate that the assumption of identical regression coefficients is not too restrictive for trading volume. Investors' trading behavior seems to be similarly related to information such as sentiment, lagged volatility, etc. for all companies. The contribution of the text analysis to the R^2 is about 3 percentage points for stock trading volume and about 0.65 percentage points for option trading volume. In the upper panel, standardized stock and option trading volume increase with positive and negative sentiment and disagreement. The relationship is highly significant and consistent with the company individual analysis and Hypotheses 2 and 3. Surprisingly, negative sentiment is negatively related to trading volume in the lower panel of Table 6. As discussed in Barber and Odean (2008), investors might be subject to investment restrictions such as short-selling restrictions. Then, negative signals are only relevant for investors who already own the stock. On the other hand, positive signals are relevant for all investors. Therefore, negative signals might reduce trading whereas positive signals increase trading.

The volatility spread narrows with positive sentiment and it increases with negative sentiment. However, the estimated relationship between the volatility spread and disagreement is inconclusive. Whereas the coefficient of D^{std} is insignificant in the upper panel of Table 6, the coefficient of D^{pol} is negative and weakly significant. Both results are inconsistent with Hypothesis 4. Nevertheless, and consistent with Hypothesis 5, the CDS spread increases with high disagreement. This increase is presumably due to the decrease in the equity value, given high disagreement, and not due to an increase in the equity volatility and asset volatility, respectively. Moreover, the CDS spread increases with negative sentiment and, at least in the lower panel of Table 6, decreases with positive sentiment as expected.

The regression results in this section and the previous sections show that simple measures of sentiment and disagreement based on company news articles of Reuters add useful information to standard factors which are frequently used to explain market activity. However, the predictive power of sentiment and disagreement is ambiguous, even though the news articles may be fundamental news and unscheduled.

6.2 Predicting Market Activity

Now, I use the pooled regression model to study the predictive power of sentiment and disagreement. Hence, I do not analyze the contemporaneous relationship between market activity and sentiment respectively disagreement, but the relationship between the market and sentiment respectively disagreement from the previous trading day. Hence, the regression model changes to

$$\begin{aligned}\left[s\hat{\varepsilon}_{i,t}\ sT_{i,t}\ s\Delta V_{i,t}\ sO_{i,t}\ s\Delta C_{i,t}\right]' &= \Lambda\left[s\hat{\varepsilon}_{i,t-1}\ sT_{i,t-1}\ s\Delta V_{i,t-1}\ sO_{i,t-1}\ s\Delta C_{i,t-1}\right]' \\ &+ \beta[sP_{i,t-1}\ sN_{i,t-1}\ sD_{i,t-1}]' + KU_t + \eta_{i,t}.\end{aligned} \tag{9}$$

Now, the residual in the objective function (8) is

$$\begin{aligned}\left[s\hat{\varepsilon}\ sT\ s\Delta V\ sO\ s\Delta C\right] - \left[s\hat{\varepsilon}_{-1}\ sT_{-1}\ s\Delta V_{-1}\ sO_{-1}\ s\Delta C_{-1}\right]\hat{\Lambda} \\ -\left[sP_{-1}\ sN_{-1}\ sD_{-1}\right]\hat{\beta} - U\hat{K}.\end{aligned} \tag{10}$$

Table 7 shows the regression estimates for Λ, β and K. The results in the upper panel are based on the disagreement measure D^{std} and the results in the lower panel on D^{pol}. The estimates for Λ and for K are very similar to the contemporaneous analysis. The R^2s decrease compared to the contemporaneous analysis of sentiment, disagreement and market activity. Positive sentiment is still highly significant and predicts positive abnormal returns on the following day. Both disagreement measures predict negative abnormal returns on the following trading day. Negative sentiment is insignificant. This might be due to incorrect assignments of words to the word category 'negative' by the 'General Inquirer', see Loughran and McDonald (2011). However, the relationship between abnormal stock returns and sentiment and disagreement, respectively, are still unexpected and might be inconsistent with Hypothesis 1. Assuming efficient markets, prices should respond to new information immediately. However, the significance of positive sentiment and disagreement hints towards market inefficiencies even on a daily frequency. These results become even stronger if I consider the excess stock return instead of the abnormal stock return. Then, the R^2 of the excess return is 1.16%, positive sentiment is positive, significant and negative sentiment and disagreement are significant, negative. The results are not shown.

There is no significant relationship between the change in the CDS spread and the one day lagged sentiment and disagreement, respectively. Hence, the credit market seems to be efficient with respect to the information extracted from the Reuters company news and in this framework. However, the company individual analysis (the results are not shown) indicates that negative sentiment and disagreement predict the change in the CDS spread for several companies. The sign of the company-individual regression coefficients varies across firms and might destroy a significant relationship in the pooled regression model.

[Table 7 about here.]

Furthermore, positive sentiment predicts stock trading volume on the following day, indicating that positive signals have a long-lasting impact on trading volume. However, negative sentiment is insignificant. This heterogeneity might be due to investment restrictions, see Barber and Odean (2008). Surprisingly, the regression coefficients of both disagreement measures are significantly negative. One possible explanation might be that investors tend to over-react to disagreement. Then, the stock trading volume might be lower during the following days.

The volatility spread increases significantly after negative sentiment and after disagreement, measured by D^{std}. D^{pol} is insignificant. Compared to the contemporaneous relationship between disagreement and the volatility spread, which is inconclusive, the result for the one day lagged D^{std} is more consistent with Hypothesis 4. The delayed response of the volatility spread could be due to a rather slow information processing and might also hint at market inefficiencies. Positive and negative sentiment and disagreement have only little predictive power for option trading volume. These results are confirmed by the company individual analysis.

6.3 Robustness

A further extension of these simple analyses is to weight the sentiment with its degree of uncertainty. I measure uncertainty with two approaches. (1) Uncertainty in news articles might be measured with the 'General Inquirer' word categories 'strong' and 'weak'. However, there is a substantial overlap between the categories 'positive' and 'strong' respectively 'negative' and 'weak'. This might bias the results. Nevertheless, I measure the uncertainty attached to a news article by

$$H_{i,t,j}^{(1)} = \frac{Z_{i,t,j} + \vartheta}{W_{i,t,j} + Z_{i,t,j} + 2\vartheta},$$

where $Z_{i,t,j}$ denotes the number of strong words and $W_{i,t,j}$ is the number of weak words. ϑ is a small, positive constant that ensures the existence of $H_{i,t,j}$ even though there are neither strong nor weak words in the news story j. Then, $H_{i,t,j} = 0.5$. If there are only strong words, $H_{i,t,j} \approx 1$ and if there are only weak words $H_{i,t,j}$ is close to zero. (2) Alternatively, if the author of a news article uses many positive words and negative words in the relevant text for a company, she might be unsure about the final consequences. Therefore, uncertainty is measured by

$$H_{i,t,j}^{(2)} = \frac{|Pos_{i,t,j} - Neg_{i,t,j}|}{\max\{Pos_{i,t,j} + Neg_{i,t,j}, \vartheta\}}.$$

If positive and negative words are almost balanced, $H_{i,t,j}^{(2)}$ is close to zero. If either positive words clearly dominate negative words or negative words clearly dominate positive words, $H_{i,t,j}^{(2)}$ is close to 1.

Now, by multiplying the net sentiment $Pos_{i,t,j} - Neg_{i,t,j}$ in (3) with $H_{i,t,j}^{(1)}$ or $H_{i,t,j}^{(2)}$, the uncertainty that is related to a news article can be taken into account. The results with respect to both measures of uncertainty stay qualitatively the same compared to the results discussed above and, hence, are not shown.

Moreover, zooming into the news story and analyzing words within the close neighborhood of the company name or nickname, as described in definition (d), gives very similar results. The results are not shown, too, but indicate that a small fraction of the full news text already contains valuable information for the financial market.

7 Trading Strategies

According to the previous section, positive sentiment and disagreement are statistically significant to predict abnormal stock returns and excess returns. However, this does not allow to conclude on the economic significance and on market efficiency. Therefore, I trading strategies based on positive and negative sentiment to gain insights on the economic significance of news articles for the stock market.

Assume that an investor trades in the 62 stocks simultaneously. The investor has no initial endowment. She observes the signal

$$X_{i,t} = \mathbf{1}(P_{i,t} > N_{i,t}) - \mathbf{1}(P_{i,t} < N_{i,t}).$$

$X_{i,t}$ can take on the value $+1, 0$ and -1. $X_{i,t} = 1$ might be interpreted as a buy-signal and $X_{i,t} = -1$ as a sell-signal[3]. $X_{i,t} = 0$ might indicate a neutral position. Since $X_{i,t}$ incorporates news from 4 p.m. at $t-1$ to 4 p.m. at day t, trading on $X_{i,t}$ might imply a substantially delayed response to new information. Whenever $X_{i,t} = 1$, the investor borrows one USD at the risk-free rate and purchases (a fraction of) stock i at the closing price of day t. At the following day, the stock is sold at the closing price of day $t+1$ and the loan is repaid, if the signal changes to neutral or to sell. Otherwise, the position is not closed until the buy-signal disappears. If $X_{i,t} = -1$, the investor short-sells one USD in stock i, invests this one USD at the risk-free rate R^f and holds the position until the signal changes. Profits and losses, due to trading are collected in her money account, which is grossed up with the risk-free rate. Furthermore, the investor has to pay transaction costs for one round-trip. For simplicity, I assume that the risk-free rate for lending and borrowing is the same and that the transaction costs are payed when the position is closed.

More precisely, let M_t denote the value of the money account at day t, $S_{i,t}$ the closing price of stock i at day t and R^f_t is the daily gross risk-free interest rate, taken from the data library of Kenneth French. By assumption, $M_0 = 0$. The money account at t is given by

$$\begin{aligned} M_t &= M_{t-1}R^f_t + \sum_{i=1}^{62}\left(Long_{i,t} + Short_{i,t}\right), && (11)\\ Long_{i,t} &= \left(\prod_{s=\tau(t)+1}^{t} \frac{S_{i,s}}{S_{i,s-1}} - \prod_{s=\tau(t)+1}^{t} R^f_s - TC\right)\mathbf{1}(X_{i,t-1} = 1 \vee X_{i,t} \neq 1), && (12)\\ Short_{i,t} &= \left(\prod_{s=\rho(t)+1}^{t} R^f_s - \prod_{s=\rho(t)+1}^{t} \frac{S_{i,s}}{S_{i,s-1}} - TC\right)\mathbf{1}(X_{i,t-1} = -1 \vee X_{i,t} \neq -1), && (13) \end{aligned}$$

where $\tau(t) = \max\{s < t | X_{i,s-1} \neq 1 \vee X_{i,s} = 1\}$ denote the most recent 'buy'-signal and $\rho(t) = \max\{s < t | X_{i,s-1} \neq -1 \vee X_{i,s} = -1\}$ the most recent 'sell'-signal. TC denotes the transaction costs. The indicator function in (12) and (13) is one if and only if a position is closed. Then, the profit is assigned to the money account.

Alternatively, I test this trading strategy against the market. This means that the investor does not finance trades at the risk-free rate and invest at the risk free-rate if a stock is short-sold, respectively, but at the market return. Then, R^f in (12) and (13) is replaced by R_{Market}, both benchmarks are downloaded from the homepage of Kenneth French.

[3]The variables X and A differ since A is defined for each news story individually whereas X refers to the average net sentiment of a trading day.

In addition to $X_{i,t}$, I consider trading strategies that are based on the signals $X_{i,t}^{+} = \max\{X_{i,t}, 0\}$ and $X_{i,t}^{-} = \min\{X_{i,t}, 0\}$. Whereas $X_{i,t}^{+}$ consists only of buy-signals, $X_{i,t}^{-}$ incorporates only sell-signals. I do not consider trading strategies that are based on disagreement to avoid conflicting signals between sentiment and disagreement. Furthermore, I do not incorporate the signal intensity, i.e. $P_{i,t} - N_{i,t}$, nor the trading volume in the corresponding stock, the stock volatility or the company's CDS spread. Those trading strategies might depend on parameter values and, hence, require an in-sample optimization and an out-of-sample performance evaluation. However, the short time span of my data sample is insufficient for this approach.

The full observation period June 01, 2007 to December 31, 2010 covers 56110 company-day observations (62 companies $\times$ 905 days). Using definition (b) and (c) to calculate $P_{i,t}$ and $N_{i,t}$ results in 8757 buy-signals and 3816 sell-signals. This yields 6062 long-positions and 3042 short positions with an average duration of 1.44 days and 1.25 days, respectively. Excluding transaction costs and refinancing costs, the average gain of a long-position is 29 bps with a standard deviation of 272 bps and the average gain of a short-position is 51 bps with standard deviation 365 bps. Hence, trades on sell-signals are more profitable and less frequent. The lower number of sell-signals and their shorter duration compared to buy-signals is somewhat surprising since the observation period covers the financial crisis. Furthermore, transaction and financing costs of 30 bps and more would render trading on buy-signals, on average, non-profitable. Sell-signals seem to be more robust against transaction costs. Moreover, the profits of the daily, aggregated long and short trades are correlated by -0.45. Therefore, the trading strategy on $X_{i,t}^{+}$ might be an efficient hedge for the strategy on $X_{i,t}^{-}$.

Table 8 shows summary statistics for the money accounts of the trading strategies based on the signal $X_{i,t}$, $X_{i,t}^{+}$ and $X_{i,t}^{-}$ and for the benchmarks risk-free rate and market return, assuming different levels of transaction costs. Without transaction costs and by benchmarking against the risk-free rate, the money account of the trading strategy that incorporates buy- and sell-signals increases from 0 USD by June 01, 2007, to 33.32485 USD by December 31, 2010. The money account's minimum is -0.0406 USD and it turns negative only for one day. Hence, there is almost no risk of losing money, indicating that the strategy might be interpreted as an approximate arbitrage opportunity. The trading strategies based on buy- respectively sell-signals exclusively have similar gain-loss profiles and might be seen as approximate arbitrage opportunities as well. The gain-loss profiles of the trading strategies are almost unchanged if the market return is used as a benchmark. However, the short-positions suffer slightly presumably due to long-investments in the poorly performing stock market during the financial crisis.

By assuming 10 bps transaction costs[4] per round-trip, the terminal values of the money account of the joint trading strategy on buy- and sell-signals are 24.0594 USD respectively 19.8684 USD, depending on the benchmark, and the gain-loss ratios

[4]The transaction costs might also cover the bid-ask spread and different rates for borrowing and lending.

are still very attractive and comparable to an approximate arbitrage opportunity. The 5% quantile, $q_{0.05}(M_t)$, is positive for both strategies, and the money accounts turn negative for only 3 respectively 4 days with a minimum value of -0.1286 USD respectively -0.1429 USD. However, trading on buy-signals only, financed at the risk-free rate becomes quite risky compared to the scenario without transaction costs. The 5% quantile of the money account is -0.5732 USD and the money account is negative for 155 days. The reason might be that long-signals generate only little profits in the financial crisis. These profits hardly cover the transaction costs and increase the probability that the money account turns negative. Also, trading on sell-signals only and investing into the stock market bear some shortfall risk now.

Figure 2 depicts the value of the money accounts of the three strategies when the risk-free rate is used as benchmark. The blue, solid curve shows the money account of trading on $X_{i,t}$, the green, dashed curve is the money account of trading on $X_{i,t}^+$ and the red, dotted curve of $X_{i,t}^-$. The money account of $X_{i,t}$ increases almost monotonically. During the heydays of the financial crisis (June 2007 to April 2010), the trading strategy on buy-signals generates significant losses, but the performance of trades on sell-signals is excellent and compensates the losses of the buy-signals fully. However, in spring 2009, governments and central banks successfully calmed down the financial markets and the stock market recovered. In the aftermath, the trading strategy on sell-signals fails to generate profits and becomes unprofitable. At the same time, buy-signals work very well. This underlines the hedging quality of trading on both, buy- and sell-signals, jointly. Furthermore, Figure 2 shows the strongest decrease in the value of the money account of $X_{i,t}$ (black line, 1.81 USD in May and June 2009) and the longest waiting period to establish a new high watermark (light blue line, 112 days during Spring and Summer 2010). Both figures are moderate[5].

Increasing the transaction costs to more than 10 bps reduces the performance of all trading strategies and increases the likelihood of a negative money account value significantly. The assumption of 20 bps transaction costs per round-trip reduces the terminal value of the money account of the joint trading strategy on buy- and sell-signals to 14.8703 USD, including 197 days with a negative value and a minimum of -1.5261 USD. This trading strategy might be still an attractive investment opportunity, but it now bears a substantial shortfall risk. Transaction costs of 30 bps and more imply that the investor looses money on almost every buy-signal and on many sell-signals. Hence, the terminal values of the money accounts of $X_{i,t}$ and $X_{i,t}^+$ are negative. However, $X_{i,t}^-$ might still be profitable.

[Table 8 about here.]

[Figure 2 about here.]

[5] The worst case, i.e. the strongest downturn and the longest waiting period to exceed the high watermark appear jointly at day zero, might be an indication for the minimum equity buffer in the approximate arbitrage portfolio.

8 Conclusion and Outlook

Das et. al. (2005) analyzed chat-room postings and conclude that ‘investors first trade and then talk’. I analyze company news of Reuters. These news are more reliable than chat room postings which, at best, disseminate company news. Simple dictionary based content analysis algorithms with rather high error rates might be applied to measure sentiment and disagreement of those news articles. Both contain valuable information for financial markets.

My results are mostly consistent with models on difference of opinion, i.e. investors are more likely to trade stocks and options after observing public signals. Disagreement across news articles is also positively correlated with stock and option trading volume and expected stock volatility. Moreover, sentiment and disagreement are statistically significant to predict returns, volatility and trading volume. With moderate transaction costs, it might be possible to exploit market inefficiencies by trading on buy- and sell-signals based on the mechanical evaluation of company news. However, transaction costs of more than 10 bps destroy this approximate arbitrage opportunity. Therefore, only institutional investors might be able to take advantage of this inefficiency. For transaction costs in the range of 10 bps up to 30 bps, the expected profits of the trading strategies are still positive, but the strategies become risky, i.e. there might be a substantial probability that the terminal value of money account is negative. Even higher transaction costs render the strategies useless.

I consider the following extensions. (1) Classifier: The ’General Inquirer’ dictionary is very general. The dictionary is not designed for analyzing financial news. Hence, it is likely that the results will improve significantly if I adjust the dictionary to account for important characteristics in financial and economic news, which might be misinterpreted right now. Furthermore, it might be interesting to determine the sentiment / disagreement by applying different methods that are not based on a dictionary approach. Bayesian classifiers, adjective-adverb classifiers and vector distance classifiers could be used as well. The grammar, text lengths or readability might be incorporated to evaluate the sentiment. (2) Industry Portfolios: Companies within the same industry might have a similar exposures to news. Hence, by studying industry portfolios instead of all companies separately or the overall pooled sample, the results might become even stronger. Furthermore, it might be possible to compare information processing among industries.

Appendix

[Table 9 about here.]

References

[1] Antweiler, Werner and Murry Z. Frank, 2004, "Is all that just Noise? The Information Content of internet Stock Message Boards", *Journal of Finance* 59, 1259-1294.

[2] Banerjee, Snehal and Ilan Kremer, 2010, "Disagreement and Learning: Dynamic Pattern of Trading Volume", *Journal of Finance* 65, 1269-1302.

[3] Barber, Brad M. and Terrance Odean, 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.

[4] Blume, Lawrence, David Easley, Maureen O'Hara, 1994, "Market Statistics and Technical Analysis: The Role of Volume", *Journal of Finance* 49, 153-181.

[5] Boyd, John H., Jian Hu and Ravi Jagannathan, 2005, "The Stock Market's Reaction to Unemployment News: Why Bad News Is Usually Good for Stocks" *Journal of Finance* 60, 649-672.

[6] Brooks, Raymond M., Ajay Patel and Tie Su, 2003, "How the Equity Market Responds to Unanticipated Events", *Journal of Business* 76, 109-133.

[7] Brounrn, Dirk and Jeroen Derwall, 2010, "The Impact of Terrorist Attacks on International Stock Markets", *European Financial Management* 16, 585-598.

[8] Carretta, Alessandro, Vincenzo Farina, Duccio Martelli, Franco Fiordelisi and Paola Schwizer, 2010, "The Impact of Corporate Governance Press News on Stock Market Return", *European Financial Management* 17, 100-119.

[9] Cao, H. Henry and Hui Ou-Yang, 2009, "Differences of Opinion of Public Information and Speculative Trading in Stocks and Options", *Review of Financial Studies* 22, 299-335.

[10] Charida, Tarum, Asani Sarkar and Avanidhar Subrahmanyam, 2005, "An Empirical Analysis of Stock and Bond Market Liquidity", *Review of Financial Studies* 18, 85-129.

[11] Coval, Joshua D. and Tyler Shumway, 2001, "Is Sound Just Noise?", *Journal of Finance* 56, 1887-1910.

[12] Cremers, Martijn, Joost Driessen, Pascal Maenhout and David Weinbaum, 2008, "Individual stock-option prices and credit spreads", *Journal of Banking and Finance* 32, 2706-2715.

[13] Das, Sanjiv R. and Mike Y. Chen, 2007, "Yahoo! for Amazon; Sentiment Extraction form Small Talk on the Web", *Management Science* 53, 1375-1388.

[14] Das, Sanjiv R., Asis Martinez-Jerez and Peter Tufano, 2005, "e-Information: A Clinical Study of Investor Discussion and Sentiment", *Financial Management* 34, 103-137.

[15] Engelberg, Joseph E. and Christopher A. Parsons, 2011, "The Causal Impact of Media in Financial Markets", *Journal of Finance* 66, 67-98.

[16] Fama, Eugene F., 1998, "Market Efficiency, Long-term Returns, and Behavioral Finance", *Journal of Financial Economics* 49, 283-306.

[17] Fama, Eugene F. and Kenneth French, 1993, "Common Risk Factors in the Returns on Stocks and Bonds", *Journal of Financial Economics* 33, 3-56.

[18] Fang, Lily and Joel Peress, 2009, "Media Coverage and the Cross-section of Stock Returns", *Journal of Finance* 64, 2023-2052.

[19] Groß-Klußmann, Axel and Nikolas Hautsch, 2011, "When machines read the news: Using automated text analysis to quantify high frequency news-implied market reactions", *Journal of Empirical Finance* 18, 321-340.

[20] Harris, Milton and Artur Raviv, 1993, "Differences of Opinion Make a Horse Race", *Review of Financial Studies* 6, 473-506.

[21] Hautsch, Nikolaus and Dieter Hess, 2002, "The Processing of Non-Anticipated Information in Financial Markets: Analyzing the Impact of Surprises in the Employment Report", *European Financial Review* 6, 133-161.

[22] Hess, Dieter, 2008, "How Do Commodity Futures Respond to Macroeconomic News?", *Financial Markets and Portfolio Management* 22, 127-146.

[23] Hong, Harrison and Jeremy C. Stein, 2007, "Disagreement and the Stock Market", *Journal of Economic Perspectives* 21, 109-128.

[24] Hull, John, Mirela Predescu and Alan White, 2004, "The relationship between credit default swap spreads, bond yields, and credit rating announcements", *Journal of Banking and Finance* 28, 2789-2811.

[25] Kandel, Eugene and Neil D. Pearson, 1995, "Interpretation of Public Signals and Trade in Speculative Makets, *Journal of Political Economy* 103, 831-872.

[26] Kyriacou, Kyriaos and Lucio Sarno, 1999, "The Temporal Relationship between Derivatives Trading and Spot Market Volatility in the U.K.: Empirical Analysis and Monte Carlo Evidence", *Journal of Futures Markets* 19, 245-270.

[27] Loughran, Tim and Bill McDonald, 2011, "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks", *Journal of Finance* 66, 35-66.

[28] Milgrom, Paul and Nancy Stokey, 1982, "Information, Trade and Common Knowledge", *Journal of Economic Theory* 26, 17-27.

[29] Mitchell, Mark L. and J. Harold Mulherin, 1994, "The Impact of Public Information on the Stock Market", *Journal of Finance* 49, 923-950.

[30] Merton, Robert C., 1974, "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates", *Journal of Finance* 29, 449-70

[31] Norden, Lars, 2009, "Credit Derivatives, Corporate News, and Credit Ratings", *working paper.*

[32] Pan, Jun and Allen M. Poteshman, 2006, "The Information in Option Volume for Future Stock Prices", *Review of Financial Studies* 19, 871-908.

[33] Sarwar, Ghulam, 2005, "The Informational Role of Option Trading Volume in Equity Index Options Markets", *Review of Quantitative Finance and Accounting* 24, 159-176.

[34] Stone, Philip J., Dexter C. Dunphy, Marshall S. Smith, Daniel M. Ogilvie, and associates. "The General Inquirer: A Computer Approach to Content Analysis," *The MIT Press*, 1966.

[35] Suominen, Matti, (2001), "Trading Volume and Information Revelation in Stock Markets", *Journal of Financial and Quantitative Analysis* 36, 545-565.

[36] Tetlock, Paul C., 2007, "Giving Content to Investor Sentiment: The Role of Media in the Stock Market", *Journal of Finance* 62, 1139-1167.

[37] Tetlock, Paul C., 2010, "Does Public Financial News Resolve Asymmetric Information?", *Review of Financial Studies* 23, 3520-3557.

[38] Tetlock, Paul C., Maytal Saar-Tsechansky and Sofus Macskassy, 2008, "More than Words: Quantifying Language to Measure Firms' Fundamentals", *Journal of Finance* 63, 1437-1467.

[39] Tumarkin, Robert and Robert F. Whitelaw, 2001, "News or Noise? Internet Posting and Stock Prices", *Financial Analysts Journal* 57, 41-51.

[40] Yu, Jialin, 2011, "Disagreement and return predictability of stock portfolios", *Journal of Financial Economics* 99, 162-183.

[41] Zhang, Benjamin Yibin, Hou Zhou and Haibin Zhu, 2009, "Explaining Credit Default Swap Spreads with the Equity Volatility and Jump Risks of Individual Firms", *Review of Financial Studies* 22, 5099-5131.

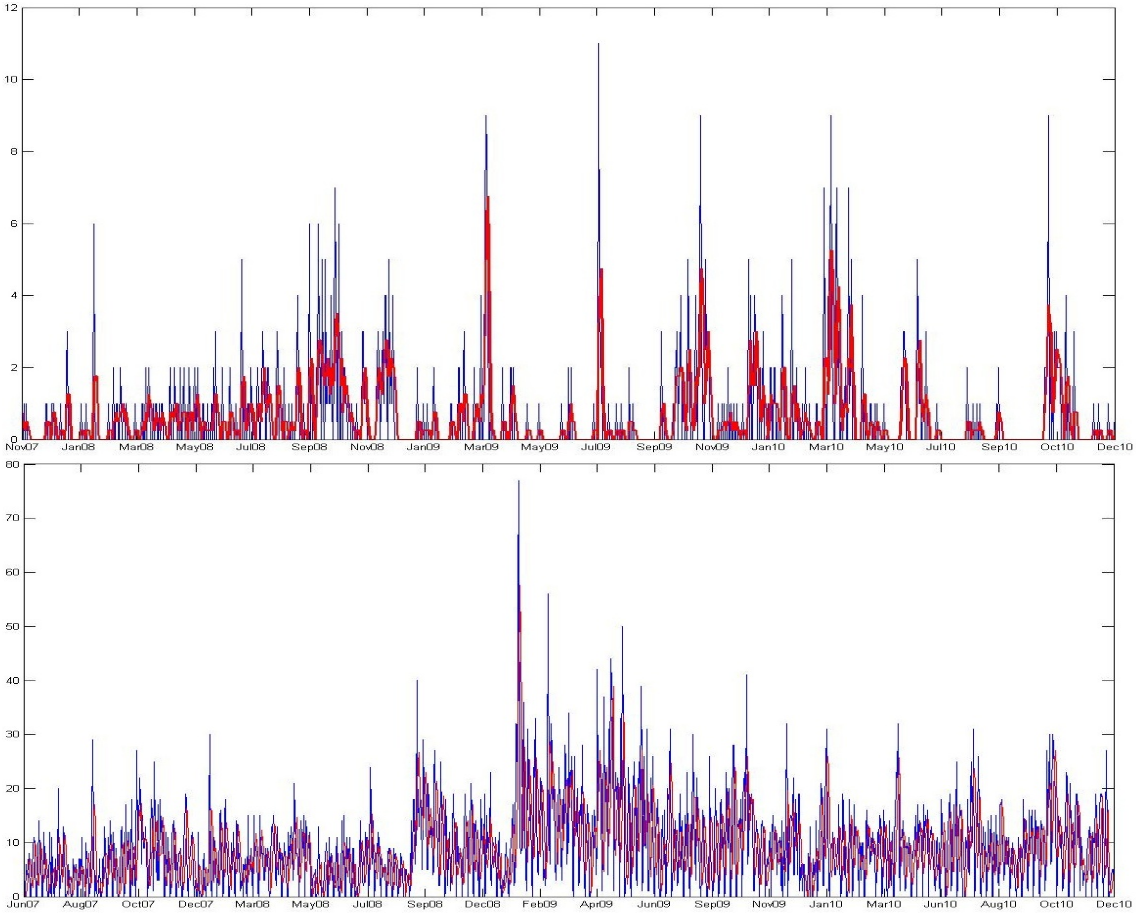

Figure 1: The upper figure shows the daily number of news stories with keywords 'Bankruptcy' or 'Insolvency' (blue curve) and the 3 day moving average (red curve), starting in June 01, 2007 to December 31, 2010. The lower figure shows the daily number of news stories for the Bank of America.

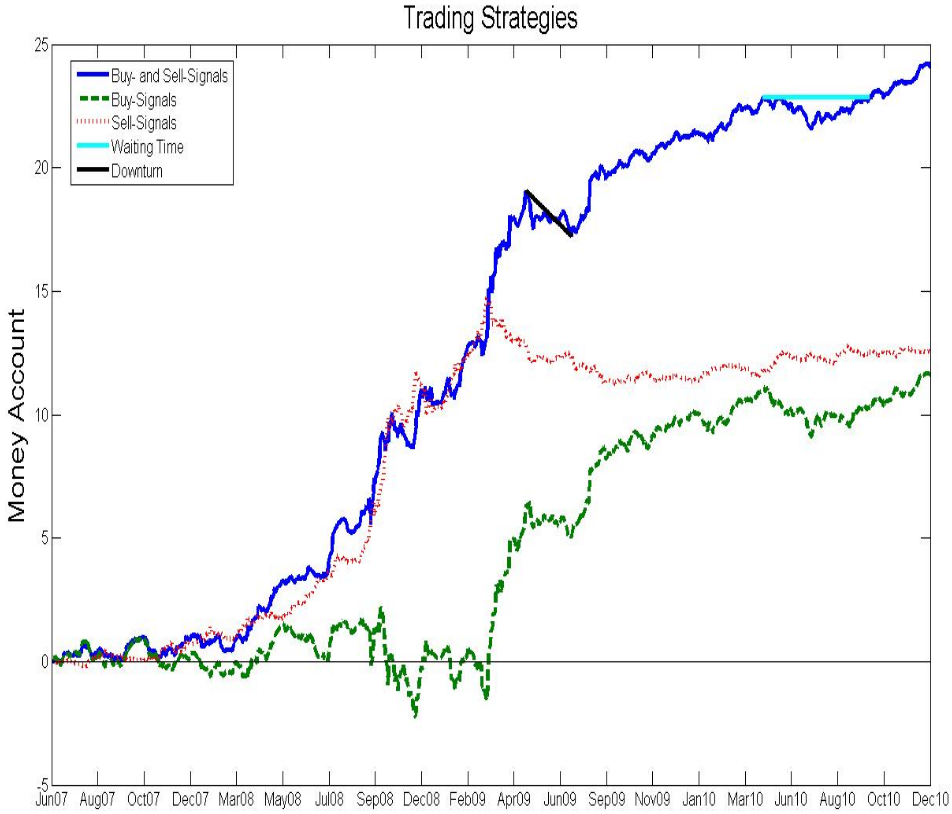

Figure 2: The figure shows the value of the money accounts of trading on buy- and sell-signals (blue, solid curve), buy-signals (green, dashed curve) and sell-signals (red, dotted curve), assuming 10 bps transaction costs per round-trip and the risk-free interest rate as benchmark. For the trading strategy on buy- and sell-signals, the black line marks the strongest downturn, realized in May and June 2009, and the light blue line marks the longest waiting period to establish a new high watermark, observed in Summer 2010.

Summary statistics for Reuters news	Sum	Mean	Std.	Max.
All	210495	160.56	94.01	354
All w/o Saturdays / Sundays	199238	220.64	52.91	354
Economic news / Macroeconomics	56531	43.12	38.61	196
General News	28085	21.42	21.00	118
Debt ratings / Credit Market News	2461	1.88	2.65	22
Society / Science / Nature	2426	1.85	4.17	30
Major Breaking News	5042	3.85	9.82	65
Bankruptcy / Insolvency	671	0.51	1.21	11
Broker Research and Recommendation	1867	1.42	2.26	17
Corporate Results / Results Forecasts / Warnings	17525	13.37	20.13	150
Mergers / Acquisitions / Takeovers	13598	10.37	11.47	60
AA.N	2770	2.11	4.16	49
AXP.N	2341	1.79	3.55	40
BA.N	4163	3.18	3.65	21
BAC.N	11974	9.13	8.23	77
CAT.N	2442	1.86	3.65	50
CSCO.O	3085	2.35	3.83	35
CVX.N	5687	4.34	3.67	29
DD.N	980	0.75	1.86	22
DIS.N	5326	4.06	3.86	26
GE.N	10236	7.81	5.98	42
HD.N	1546	1.18	2.99	34
HPQ.N	4593	3.50	4.31	31
IBM.N	4790	3.65	4.64	40
INTC.O	5219	3.98	5.26	40
JNJ.N	2860	2.18	3.01	29
JPM.N	11723	8.94	7.62	45
KFT.N	2171	1.66	3.43	35
KO.N	2080	1.59	2.54	18
MCD.N	2120	1.62	3.02	38
MMM.N	1043	0.80	2.32	28
MRK.N	3223	2.46	3.40	41
MSFT.O	10495	8.01	7.21	68
PG.N	2096	1.60	2.86	42
PFE.N	3803	2.90	3.61	52
T.N	4559	3.47	3.95	33
TRV.N	407	0.31	1.22	19
UTX.N	2026	1.55	2.59	20
VZ.N	3435	2.62	3.32	28
WMT.N	6676	5.09	5.19	45
XOM.N	8096	6.18	4.98	33

Table 1: This table gives summary statistics (sum, mean, standard deviation and maximum) for the number of news articles per day. The upper panel classifies news on the S&P500 companies by keywords and the lower panel show the statistics for all members of the Dow Jones Industrial Average separately. A news story is considered as relevant for a company if the company's RIC is mentioned in the field 'related RICs'.

	Q_i		Y_i	
constant	-20.8872^a	-23.0588^a	-2.0450^a	-2.0569^a
$P2B$	-0.3051^a	-0.2101^a	-0.0251^a	-0.0225^a
$\ln(CAP)$	2.2701^a	2.3374^a	0.2481^a	0.2501^a
Ret	-	-1.9734^b	-	-0.1547^b
$\sigma(Ret)$	-	6.9538^a	-	0.2916^b
R^2	38.73%	68.26%	54.58%	65.89%
# Obs.	61	61	61	61

Table 2:
The table shows the regression estimates for model (2). The subscript a, b and c indicate significance at the 1%, 5% and 10% confidence level. Q_i denotes the average number of news per day of company i, and Y_i is the average number of days with at least on news story. A news story is relevant for a company if the company's RIC is mentioned in the field 'related RICs'.

Optimism	$\hat{\varepsilon}$	T	V	O	C	$\#(P>0)$	$m(P\|P>0)$	$\sigma(P\|P>0)$
$AA.N$	0.0017^{b}	0.0182	−0.0002	0.0343^{b}	−0.0609	79.0	3.2504	2.4254
$ABT.N$	0.0007	0.0249^{b}	0.0007	0.0046	−0.0322	64.0	2.9495	1.9430
$AIG.N$	−0.0002	0.0011	0.0001	−0.0107	−0.3763	173.0	3.6032	2.3316
$AMGN.O$	−0.0001	0.0125	0.0003	0.0773^{a}	−0.0121	60.0	2.4562	1.4674
$APC.N$	0.0002	0.0164	−0.0004	−0.0057	−0.2079	56.0	4.2889	2.6812
$AXP.N$	−0.0002	0.0011	0.0001	0.0285^{a}	−0.4751	64.0	3.8703	4.7347
$BA.N$	0.0002	−0.0057	0.0004	0.0087	−0.1926	172.0	2.3197	1.7603
$BAC.N$	−0.0007	0.0104	0.0003	−0.0031	0.0503	229.0	2.6885	2.0884
$BAX.N$	0.0004	0.0296	−0.0010	−0.0017	−0.0543	27.0	3.1350	2.3361
$BMY.N$	0.0010^{b}	0.0273^{c}	−0.0009	0.0177	-0.0914^{c}	68.0	3.5177	2.1955
$BSX.N$	0.0017	−0.0332	−0.0001	−0.0070	−0.1993	28.0	2.4508	1.9893
$C.N$	−0.0003	0.0132	-0.0038^{c}	0.0059	−0.5522	302.0	3.1279	2.0583
$CAT.N$	0.0031^{a}	0.0587^{a}	−0.0008	0.0223	-0.6997^{b}	52.0	3.0266	1.8953
$COP.N$	0.0004	−0.0024	−0.0002	0.0037	−0.1246	130.0	2.9440	1.5714
$CSC.N$	−0.0009	0.0238	−0.0010	0.0358	−0.0119	7.0	5.2307	3.1978
$CSCO.O$	0.0004	−0.0006	0.0002	−0.0071	-0.1287^{c}	120.0	3.6380	2.6329
$CVX.N$	-0.0006^{c}	0.0031	−0.0002	−0.0244	0.0150	136.0	2.7571	2.0611
$DD.N$	0.0001	−0.0094	−0.0007	−0.0253	−0.0287	42.0	3.4745	2.2462
$DELL.O$	0.0006	0.0122	0.0006	0.0468^{b}	0.0704	99.0	2.7777	2.1504
$DIS.N$	−0.0002	−0.0023	0.0009	0.0082	−0.0449	123.0	3.1999	2.1781
$DOW.N$	0.0010	0.0029	0.0011	0.0137	−0.5548	28.0	4.2514	2.3947
$DVN.N$	0.0005	0.0191	0.0008	−0.0012	−0.1463	32.0	3.4291	2.1876
$F.N$	0.0008	0.0268^{b}	−0.0022	−0.0008	1.1325	198.0	3.0296	2.2643
$FDX.N$	0.0040^{a}	0.0758^{a}	−0.0001	0.1048^{a}	-0.5887^{b}	40.0	2.7887	2.2183
$GE.N$	−0.0008	0.0266^{c}	0.0033^{a}	0.0092	0.7432	59.0	3.8617	2.7574
$GLW.N$	0.0018^{b}	0.0226^{c}	−0.0008	0.0625^{b}	−0.5179	19.0	5.2515	3.9589
$GR.N$	-0.0015^{a}	0.1334^{a}	−0.0012	0.0872^{c}	-0.2023^{b}	20.0	4.7729	5.7485
$GS.N$	0.0003	−0.0049	−0.0003	−0.0003	−0.3873	235.0	2.9275	2.1619
$HON.N$	0.0000	0.0282^{b}	0.0014	0.0273	−0.0772	45.0	3.3022	2.1268
$HPQ.N$	0.0010^{b}	0.0023	0.0012^{c}	0.0029	0.2170^{a}	119.0	3.0873	2.2698
$IBM.N$	0.0002	0.0023	−0.0004	0.0024	0.0542	138.0	3.6778	2.9647
$INTC.O$	0.0001	−0.0014	0.0003	−0.0013	−0.0736	112.0	2.9896	2.2001
$JNJ.N$	0.0000	0.0196^{b}	0.0009^{c}	−0.0046	−0.0076	102.0	3.1333	2.6229
$JPM.N$	0.0010^{c}	0.0041	−0.0001	0.0002	0.1774	183.0	2.7744	2.2258
$KFT.N$	0.0004	0.0223	−0.0002	−0.0271	0.0710	94.0	2.6039	1.6123
$KO.N$	0.0009^{c}	0.0328^{b}	−0.0001	0.0166	−0.0389	50.0	3.3672	2.3124
$MCD.N$	0.0010^{b}	0.0310^{b}	−0.0006	0.0001	−0.0065	70.0	3.4556	2.3593
$MDT.N$	0.0003	−0.0236	0.0000	−0.0129	0.0866	38.0	3.5020	2.7540
$MO.N$	0.0002	0.0080	-0.0022^{a}	0.0058	−0.1456	27.0	2.9978	2.6200
$MON.N$	0.0010^{c}	−0.0034	0.0006	0.0109	0.0525	65.0	3.9448	3.6470
$MMM.N$	0.0022^{a}	0.0532^{a}	−0.0010	0.0687	−0.0954	27.0	3.5051	2.1892
$MRK.N$	0.0001	0.0282^{a}	−0.0003	−0.0122	0.0518	99.0	3.4270	2.6114
$MS.N$	0.0008	0.0167	-0.0036^{c}	0.0259^{c}	0.7892	201.0	2.8148	2.1083
$MSFT.O$	0.0005	0.0056	0.0003	0.0038	−0.0363	226.0	2.6009	1.6957
$LLY.N$	-0.0013^{b}	0.0223	0.0005	−0.0185	0.0591	56.0	2.9219	1.7841
$LMT.N$	0.0002	0.0122	-0.0008^{c}	0.0413^{c}	0.0365	105.0	3.5112	2.4568
$ORCL.O$	0.0006	0.0599^{a}	−0.0010	0.0971^{a}	−0.0725	70.0	2.8478	2.2948
$OXY.N$	−0.0002	0.0172	−0.0006	0.0324^{b}	0.2416^{b}	23.0	4.5061	3.3075
$PFE.N$	0.0002	−0.0021	0.0036	−0.0088	0.0384	122.0	3.0920	2.3710
$PG.N$	0.0001	0.0289^{c}	−0.0016	−0.0161	0.0033	52.0	2.5772	1.7435
$SLB.N$	-0.0018^{b}	0.0143	−0.0018	0.0229	0.1169	34.0	3.2163	2.5419
$T.N$	0.0003	0.0104	0.0003	0.0066	0.0119	102.0	3.0401	1.8108
$TRV.N$	0.0105^{a}	0.1612^{b}	-0.0140^{c}	0.1683	0.1582	9.0	1.8596	0.9999
$TWX.N$	0.0000	0.0091	0.0010	0.0239	−0.0975	75.0	3.0055	2.1811
$TXN.N$	0.0006	0.0124	0.0003	0.0119	0.0478	21.0	5.6115	4.4692
$UTX.N$	0.0021^{a}	0.0053	−0.0018	−0.0035	−0.2170	8.0	5.0200	3.0939
$VZ.N$	0.0004	−0.0003	0.0003	0.0057	0.1712	121.0	2.8905	2.0383
$WFC.N$	0.0011	0.0387^{b}	-0.0066^{c}	0.0086	−0.2366	84.0	3.4648	2.4066
$WMT.N$	0.0003	0.0094	−0.0003	−0.0408	−0.0443	180.0	2.7180	2.0239
$WLP.N$	−0.0011	0.0387^{b}	−0.0003	0.0257	0.4564^{b}	36.0	3.7594	2.3134
$XOM.N$	−0.0002	0.0006	−0.0001	−0.0216	−0.0209	160.0	3.2283	2.1963
Pos. & sig.	13	19	3	11	3			
Neg. & sig.	4	0	6	0	5			

Table 3: The table shows the estimated, company-individual co-movement of positive sentiment $P_{i,t-1}$ and abnormal returns ($\hat{\varepsilon}$), stock trading volume (T), the first difference of the volatility spread (V), cumulated option trading volume (O) and the first difference of the CDS spread (C), according to regression model (6). The subscript a, b and c indicate significance at the 1%, 5% and 10% confidence level. Column 7, 8 and 9 show the number of days with positive sentiment, the conditional mean of positive sentiment and its standard deviation.

Pessimism	$\hat{\varepsilon}$	T	V	O	C	$\#(N>0)$	$m(N\|N>0)$	$\sigma(N\|N>0)$
AA.N	-0.0035^a	-0.0356^c	0.0088^a	-0.0501^c	2.9180^a	48.0	2.7442	2.0826
ABT.N	0.0001	−0.0258	−0.0010	0.0271	−0.0018	20.0	2.3062	1.1289
AIG.N	−0.0008	-0.0327^c	−0.0048	0.0035	−0.0403	136.0	3.9970	3.0133
AMGN.O	−0.0007	0.1162^a	0.0048^a	0.1479^a	0.3611^c	27.0	1.9520	1.8325
APC.N	-0.0030^b	0.0184	0.0050^a	0.0395^b	3.0280^a	25.0	3.7985	2.7814
AXP.N	−0.0001	0.0115	0.0022	0.0160	2.1222^a	65.0	3.1979	2.4588
BA.N	−0.0001	0.0162	−0.0004	−0.0045	−0.1733	138.0	2.4733	2.0844
BAC.N	−0.0010	0.0154	−0.0026	0.0108	0.0427	107.0	2.4113	2.1768
BAX.N	-0.0240^a	1.2724^a	0.0108^b	0.2341	0.0491	6.0	1.3261	1.0626
BMY.N	0.0012	0.0125	-0.0025^c	−0.0037	0.0375	20.0	2.0438	1.6211
BSX.N	−0.0039	0.7366^a	−0.0073	0.0303	0.8928	17.0	1.3504	0.9238
C.N	−0.0015	−0.0022	0.0010	−0.0014	−1.4546	99.0	1.9581	1.6824
CAT.N	−0.0011	−0.0276	−0.0009	−0.0301	−0.6743	16.0	1.7822	1.3524
COP.N	−0.0002	0.0016	0.0004	0.0067	-0.1752^c	84.0	2.8550	2.0177
CSC.N	0.0091^a	0.1801^b	-0.0080^c	−0.0537	−0.7654	2.0	2.8250	2.5809
CSCO.O	−0.0010	0.0783^a	−0.0034	0.0588^c	−0.1064	19.0	2.0095	1.6628
CVX.N	−0.0000	0.0117	−0.0002	−0.0156	0.0176	94.0	3.0215	1.9161
DD.N	−0.0008	0.0323^c	−0.0015	−0.0388	1.8261^a	23.0	2.8427	1.9362
DELL.O	-0.0045^a	0.0706^a	0.0008	0.0472	0.5376^c	42.0	1.8228	1.7795
DIS.N	-0.0024^b	−0.0076	0.0040	0.0174	0.1655	31.0	1.9420	1.3118
DOW.N	0.0000	−0.0144	−0.0001	−0.0133	−0.0967	31.0	3.9294	2.7804
DVN.N	-0.0088^a	0.0188	0.0220^a	0.0137	−0.6530	8.0	1.8642	1.5859
F.N	−0.0020	−0.0082	0.0017	−0.0090	11.9571	88.0	2.4773	2.0080
FDX.N	0.0005	−0.0053	−0.0001	−0.0261	0.1863	30.0	3.2106	3.6588
GE.N	0.0007	0.0529^b	0.0000	0.0107	0.9876	37.0	3.0648	1.9041
GLW.N	-0.0073^a	0.2269^a	0.0026	0.0348	0.6894	15.0	2.6617	3.9609
GR.N	0.0159^c	−0.0328	−0.0025	1.4221^c	0.0063	2.0	1.4533	0.5704
GS.N	-0.0019^b	0.0480^b	0.0034^b	0.0338^c	0.1399	116.0	2.3064	1.9284
HON.N	0.0005	0.0029	−0.0033	−0.0353	−0.4228	9.0	2.5706	1.5175
HPQ.N	-0.0022^a	0.0652^a	0.0012	0.0419^c	0.3178^b	52.0	2.1668	1.8391
IBM.N	−0.0003	0.0054	0.0014^c	0.0092	0.1860	37.0	2.7328	2.5395
INTC.O	−0.0012	−0.0199	0.0021	−0.0652	-0.7197^a	48.0	2.2750	1.5690
JNJ.N	−0.0001	0.0025	0.0000	−0.0145	0.0012	46.0	2.8465	2.7153
JPM.N	0.0003	0.0017	0.0032^b	0.0040	0.1679	106.0	2.9178	2.6319
KFT.N	-0.0019^c	0.0219	−0.0005	0.1888	−0.1957	14.0	2.4609	1.9463
KO.N	0.0001	0.0169	0.0005	−0.0139	0.0210	20.0	5.2119	4.8020
MCD.N	0.0006	0.0104	0.0005	0.0140	0.0247	31.0	2.0148	1.5340
MDT.N	−0.0019	0.1235^a	0.0051^b	0.0897	0.4595	21.0	1.6765	1.1224
MO.N	0.0026	−0.0224	−0.0024	−0.0074	−0.0677	6.0	1.0153	1.2205
MON.N	−0.0003	−0.0204	0.0006	−0.0103	−0.0741	30.0	2.3710	2.2904
MMM.N	0.0019	−0.0005	0.0050	−0.0135	−0.3744	5.0	2.2973	0.9381
MRK.N	-0.0023^a	0.0267	0.0003	0.0119	0.1090	48.0	2.3156	1.9083
MS.N	-0.0028^c	0.0269	0.0043	−0.0106	9.7139	75.0	2.4378	1.9433
MSFT.O	−0.0004	0.0088	−0.0009	-0.0615^c	0.0550	72.0	1.8498	1.4563
LLY.N	−0.0003	0.0022	0.0002	−0.0225	−0.0258	25.0	3.2295	2.9857
LMT.N	0.0013	−0.0065	0.0005	−0.0117	−0.0058	44.0	2.1761	1.8434
ORCL.O	−0.0001	0.0128	0.0001	0.0119	0.0585	51.0	2.9248	2.5158
OXY.N	−0.0019	−0.0035	−0.0135	0.2157	−0.7786	2.0	1.3525	0.6824
PFE.N	−0.0007	0.0185	0.0015	−0.0478	0.0943	62.0	2.3397	1.5395
PG.N	0.0002	−0.0186	-0.0023^c	-0.1008^c	0.0270	48.0	2.2890	1.4934
SLB.N	−0.0026	−0.0102	0.0084^a	0.0284	0.5508^b	12.0	2.5346	1.9957
T.N	−0.0002	0.0007	−0.0010	−0.0013	−0.1212	36.0	2.0412	1.8221
TRV.N	0.0075^a	0.0632	−0.0072	0.0580	0.3904	8.0	3.3083	2.2270
TWX.N	−0.0010	0.0136	−0.0006	−0.0571	0.0371	21.0	1.8660	2.7775
TXN.N	0.0028	0.0238	0.0011	−0.0238	4.5247^a	5.0	2.4740	1.1307
UTX.N	−0.0019	0.0468	−0.0006	−0.0560	0.1638	7.0	3.9548	0.7271
VZ.N	−0.0021	0.0348	0.0045	0.0133	−0.3960	26.0	1.5650	0.8817
WFC.N	-0.0033^b	0.0730^b	0.0019	0.0207	0.1401	54.0	2.3091	1.6347
WMT.N	-0.0012^b	0.0121	0.0016^b	−0.0899	0.0103	62.0	2.6808	2.1770
WLP.N	−0.0001	0.0251	−0.0008	−0.0011	−1.2920	9.0	1.3725	1.2788
XOM.N	−0.0002	−0.0002	−0.0003	0.0032	0.0288	89.0	2.6779	1.9955
Pos. & sig.	3	12	11	6	9			
Neg. & sig.	14	2	3	2	2			

Table 4: The table shows the estimated, company-individual co-movement of negative sentiment $P_{i,t-1}$ and abnormal returns ($\hat{\varepsilon}$), stock trading volume (T), the first difference of the volatility spread (V), cumulated option trading volume (O) and the first difference of the CDS spread (C), according to regression model (6). The subscript a, b and c indicate significance at the 1%, 5% and 10% confidence level. Column 7, 8 and 9 show the number of days with negative sentiment, the conditional mean of negative sentiment and its standard deviation.

Disagreement	$\hat{\varepsilon}$	T	V	O	C	$\#(D>0)$	$m(D\|D>0)$	$\sigma(D\|D>0)$
$AA.N$	-0.0001	0.1077^a	-0.0028	0.0469^c	-0.5025	54.0	2.4188	2.2592
$ABT.N$	-0.0022^b	0.0682^b	-0.0005	0.0195	0.1882	31.0	1.6257	1.6121
$AIG.N$	0.0017	0.0653^a	0.0029	0.0437^b	-2.3837	209.0	3.2710	2.1547
$AMGN.O$	0.0085^a	0.2571^a	-0.0051^a	0.1350^a	-0.6805^a	33.0	1.7816	1.6824
$APC.N$	0.0027	0.1083^a	-0.0009	0.0215	-0.3629	32.0	1.8993	1.7249
$AXP.N$	-0.0004	0.0749^a	-0.0019	0.0277^c	-1.0286	54.0	2.7427	2.8239
$BA.N$	-0.0018^a	0.0514^a	0.0013^b	0.0091	0.2428	202.0	2.0917	1.7702
$BAC.N$	0.0000	0.0628^a	0.0028	0.0367^a	-0.2434	216.0	2.2327	1.9952
$BAX.N$	-0.0024^c	0.3327^a	-0.0022	0.2572^a	-0.0154	13.0	2.8396	2.5411
$BMY.N$	0.0002	0.0427	-0.0006	0.0091	0.0492	43.0	1.7008	1.7800
$BSX.N$	-0.0124^a	0.3925^a	0.0169^a	0.0894	3.3971^a	19.0	1.9882	2.4059
$C.N$	0.0003	0.0515^a	0.0021	0.0234^b	0.5848	276.0	2.2967	1.6895
$CAT.N$	0.0002	0.1059^a	-0.0024	0.0315	-0.4019	27.0	2.1223	2.3512
$COP.N$	-0.0004	0.0207^c	0.0016	0.0008	0.2143^c	103.0	2.3616	1.5052
$CSC.N$	0.0128^b	0.3420^b	0.0217^a	-0.4354	7.2457^a	2.0	1.8420	0.9650
$CSCO.O$	-0.0013^b	0.0349^b	0.0001	0.0169	-0.0436	73.0	1.7396	2.1665
$CVX.N$	0.0008^c	0.0021	0.0002	0.0041	0.1201	133.0	2.1836	1.8328
$DD.N$	0.0007	0.1190^a	0.0026	-0.0062	-0.4010	21.0	1.1783	1.1597
$DELL.O$	-0.0021	0.0898^a	0.0013	0.1456^a	0.2407	78.0	1.3372	1.3688
$DIS.N$	0.0017^b	0.0797^a	-0.0068	0.0144	-0.3150^c	86.0	1.3702	1.4605
$DOW.N$	0.0002	0.2729^a	0.0011	0.1078^a	0.7003	29.0	3.1736	2.5388
$DVN.N$	0.0061^a	0.0848^a	-0.0015	0.1263^a	-0.0593	20.0	2.4095	2.5574
$F.N$	-0.0001	0.0440^b	0.0028	0.0411^a	-10.0795	177.0	1.6024	1.5340
$FDX.N$	-0.0029^b	0.2532^a	-0.0009	0.0986^a	0.9380^a	31.0	2.4819	2.4811
$GE.N$	0.0021^c	0.0072	-0.0058^b	0.0584	-1.1426	42.0	2.3322	1.9647
$GLW.N$	-0.0017	0.0610^a	-0.0006	-0.0232	0.3987	17.0	3.2953	3.3540
$GR.N$	0.0096^a	0.3587^a	-0.0019	0.4344^c	0.3461	8.0	1.8870	1.7545
$GS.N$	0.0006	0.0229	0.0008	0.0098	1.0095^c	207.0	2.1838	1.7548
$HON.N$	0.0001	0.0832^b	-0.0046^c	-0.0144	0.2443	27.0	1.2727	1.2025
$HPQ.N$	-0.0002	0.0813^a	-0.0025^b	0.0591^a	0.1156	88.0	2.1268	1.8925
$IBM.N$	0.0002	0.0592^a	-0.0009	0.0982^a	-0.0697	111.0	1.6985	2.2262
$INTC.O$	0.0007	0.0745^a	-0.0015	0.0864^b	0.2383^c	94.0	1.8811	1.9041
$JNJ.N$	-0.0006	0.0272^c	-0.0003	0.0177	0.0037	72.0	2.1668	1.9909
$JPM.N$	-0.0010^c	0.0080	-0.0025	-0.0056	-0.2332	162.0	2.4532	2.5032
$KFT.N$	-0.0013	0.0557	0.0003	-0.5110	-0.0836	82.0	1.3556	1.1954
$KO.N$	0.0007	0.0599^a	-0.0022^b	0.0502	0.1193	23.0	3.1491	2.7196
$MCD.N$	0.0006	0.0846^a	0.0008	0.0246	0.0502	53.0	1.8974	1.3850
$MDT.N$	-0.0045^a	0.2090^a	0.0001	0.0732	-0.2749	32.0	1.7002	1.7557
$MO.N$	-0.0033^c	0.1571^a	0.0023	0.0858	0.8789	13.0	1.7484	0.9833
$MON.N$	-0.0003	0.0819^a	-0.0006	0.0666^a	0.1395	35.0	2.3683	3.0473
$MMM.N$	-0.0020	0.1856^a	-0.0028	0.3179^c	0.6947^c	12.0	1.4302	1.3572
$MRK.N$	0.0009	0.0479^b	0.0019^c	0.0110	0.0390	69.0	2.2193	1.8914
$MS.N$	-0.0009	0.0284	0.0037	-0.0044	-1.4121	151.0	2.1024	1.8497
$MSFT.O$	-0.0004	0.0423^a	-0.0013	0.0135	-0.0051	212.0	1.8450	1.5528
$LLY.N$	-0.0008	0.0413^b	0.0017	-0.0208	0.2304^b	33.0	2.3060	1.8998
$LMT.N$	-0.0016	0.0861^a	-0.0010	0.0193	0.0244	55.0	1.6729	1.3771
$ORCL.O$	0.0009	0.0151	0.0012	0.0388	-0.0426	65.0	2.2632	2.4857
$OXY.N$	0.0022	0.0646	0.0012	0.1249^b	-0.2937	7.0	2.1770	1.7105
$PFE.N$	-0.0009	0.0358^a	0.0038	0.0817	0.3807^a	87.0	1.9404	1.7370
$PG.N$	-0.0011	0.0588^b	0.0007	0.1997^a	-0.1966	49.0	1.8380	1.7458
$SLB.N$	0.0064^a	0.0920^b	-0.0071^b	0.0708^c	-0.8166^a	22.0	1.7485	1.7624
$T.N$	-0.0000	0.0526^a	-0.0019	0.0084	0.5071^b	77.0	1.2798	1.3578
$TRV.N$	-0.0081^a	0.0590	0.0027	-0.0531	-0.7709	8.0	2.2566	2.0838
$TWX.N$	-0.0016	0.0070	-0.0008	-0.0466	-0.1464	59.0	1.1914	1.5567
$TXN.N$	-0.0026	0.2319^a	-0.0001	0.2729^a	-0.1736	15.0	2.6765	1.2774
$UTX.N$	0.0052	-0.0015	-0.0026	0.1113	-0.5210	3.0	1.4775	1.5095
$VZ.N$	0.0019^b	0.0550^a	0.0006	0.0204	-0.0501	76.0	1.4279	1.4429
$WFC.N$	0.0019	0.1155^a	0.0016	0.0590^c	0.0182	62.0	1.6387	1.5371
$WMT.N$	0.0007	0.0763^a	-0.0014^c	0.3018^a	0.0805	137.0	1.7030	1.5085
$WLP.N$	0.0023	0.1351^a	-0.0017	-0.0064	1.1201^c	18.0	1.7603	1.6106
$XOM.N$	-0.0007	0.0050	0.0000	0.0645	0.0096	149.0	2.0730	1.6049
Pos. & sig.	9	49	4	23	11			
Neg. & sig.	10	0	7	0	3			

Table 5: The table shows the estimated, company-individual co-movement of disagreement, measured by D^{std}, and abnormal returns ($\hat{\varepsilon}$), stock trading volume (T), the first difference of the volatility spread (V), cumulated option trading volume (O) and the first difference of the CDS spread (C), according to regression model (6). The subscript a, b and c indicate significance at the 1%, 5% and 10% confidence level. Column 7, 8 and 9 show the number of days with disagreement, the conditional mean of disagreement and its standard deviation.

Pooled Analysis	$s\hat{\varepsilon}$	sT	sV	sO	sC
$s\hat{\varepsilon}_{-1}$	-0.0010	0.0088^b	-0.0034	0.0063	0.0151^a
sT_{-1}	0.0221^a	0.5380^a	-0.0528^a	0.0854^a	-0.0049
sV_{-1}	0.0349^a	0.0734^a	-0.1591^a	0.0184^a	0.1232^a
sO_{-1}	-0.0190^a	0.0262^a	0.0037	0.2701^a	-0.0163^a
sC_{-1}	-0.0082	0.0274^a	0.0035	-0.0018	0.0876^a
sP	0.0273^a	0.0455^a	-0.0111^a	0.0237^a	-0.0069
sN	-0.0321^a	0.0374^a	0.0128^a	0.0088^c	0.0194^a
sD^{std}	0.0008	0.1187^a	-0.0058	0.0599^a	0.0142^a
Monday	0.0493^a	-0.2202^a	0.1149^a	-0.0089	-0.1259^a
Tuesday	0.0349^b	0.0616^a	-0.0621^a	0.0489^a	-0.1059^a
Wednesday	-0.0085	-0.0541^a	0.0538^a	-0.0091	-0.0880^a
Thursday	0.0203	0.0095	0.1196^a	0.0184	-0.0408^b
Constant	-0.0199	0.0387^a	-0.0445^a	-0.0097	0.0713^a
R^2	0.0041	0.3499	0.0340	0.1067	0.0297
$s\hat{\varepsilon}_{-1}$	-0.0017	0.0103^b	-0.0033	0.0063	0.0159^a
sT_{-1}	0.0237^a	0.5345^a	-0.0522^a	0.0849^a	-0.0043
sV_{-1}	0.0353^a	0.0731^a	-0.1605^a	0.0175^a	0.1220^a
sO_{-1}	-0.0169^a	0.0228^a	0.0044	0.2683^a	-0.0159^a
sC_{-1}	-0.0077	0.0250^a	0.0046	-0.0034	0.0870^a
sP	0.0429^a	0.0124^a	-0.007	0.0016	-0.0125^b
sN	-0.0146^b	-0.0137^a	0.0176^a	-0.0230^a	0.0115^b
sD^{pol}	-0.0388^a	0.1650^a	-0.0131^b	0.0969^a	0.0237^a
Monday	0.0472^a	-0.2201^a	0.1138^a	-0.0120	-0.1259^a
Tuesday	0.0359^b	0.0568^a	-0.0616^a	0.0454^a	-0.1043^a
Wednesday	-0.0058	-0.0561^a	0.0528^a	-0.0119	-0.0893^a
Thursday	0.0206	0.0091	0.1199^a	0.0170	-0.0416^b
Constant	-0.0194	0.0402^a	-0.0443^a	-0.0076	0.0716^a
R^2	0.0051	0.3552	0.0341	0.1097	0.0299
#Obs.	36229	36229	36229	36229	36229

Table 6: The table shows the regression estimates for Λ, β and K in the pooled regression model with contemporaneous relationships between the market variables and sentiment and disagreement, respectively, i.e. (7). The upper panel measures disagreement with D^{std} and the lower panel with sD^{pol}. a denotes significance at the 1% confidence level, b at the 5% confidence level and c at the 10% level.

Pooled Analysis	$s\hat{\varepsilon}$	sT	sV	sO	sC
$s\hat{\varepsilon}_{-1}$	-0.0005	0.0079^c	-0.0028	0.0054	0.0155^a
sT_{-1}	0.0239^a	0.5475^a	-0.0548^a	0.0910^a	-0.0044
sV_{-1}	0.0347^a	0.0751^a	-0.1605^a	0.0187^a	0.1224^a
sO_{-1}	-0.0182^a	0.0350^a	0.0033	0.2749^a	-0.0147^a
sC_{-1}	-0.0081	0.0272^a	0.0041	-0.0022	0.0871^a
sP_{-1}	0.0105^a	0.0093^b	-0.0058	0.0043	-0.0011
sN_{-1}	-0.0042	-0.0027	0.0149^a	0.0070	0.0074
sD^{std}_{-1}	-0.0163^a	-0.0248^a	0.0097^c	-0.0104^a	0.0071
Monday	0.0501^a	-0.2338^a	0.1152^a	-0.0190	-0.1275^a
Tuesday	0.0342^b	0.0687^a	-0.0615^a	0.0524^a	-0.1021^a
Wednesday	-0.0064	-0.0366^a	0.0513^a	-0.0013	-0.0874^a
Thursday	0.0200	0.0248^c	0.1189^a	0.0252	-0.0401^b
Constant	-0.0196	0.0331^a	-0.0440^a	-0.0116	0.0707^a
R^2	0.0025	0.3291	0.0340	0.1018	0.0291
$s\hat{\varepsilon}_{-1}$	-0.0010	0.0069	-0.0026	0.0052	0.0154^a
sT_{-1}	0.0242^a	0.5487^a	-0.0544^a	0.0906^a	-0.0033
sV_{-1}	0.0345^a	0.0746^a	-0.1605^a	0.0187^a	0.1222^a
sO_{-1}	-0.0179^a	0.0356^a	0.0033	0.2749^a	-0.0144^a
sC_{-1}	-0.0080	0.0275^a	0.0041	-0.0022	0.0872^a
sP_{-1}	0.0131^b	0.0151^a	-0.0055	0.0042	0.0015
sN_{-1}	0.0007	0.0067	0.0142^b	0.0082	0.0096
sD^{pol}_{-1}	-0.0179^a	-0.0319^a	0.0058	-0.0069	-0.0020
Monday	0.0500	-0.2342^a	0.1150^a	-0.0188	-0.1279^a
Tuesday	0.0337^b	0.0676^a	-0.0616^a	0.0524	-0.1026^a
Wednesday	-0.0064	-0.0367^a	0.0513^a	-0.0012	-0.0874^a
Thursday	0.0201	0.0251^b	0.1189^a	0.0252	-0.0399^a
Constant	-0.0193	0.0334^a	-0.0440^a	-0.0116	0.0708^a
R^2	0.0025	0.3292	0.0340	0.1017	0.0291
#Obs.	36229	36229	36229	36229	36229

Table 7: The table shows the regression estimates for Λ, β and K in the pooled regression model without contemporaneous relationships, i.e. (9). The upper panel measures disagreement with D^{std} and the lower panel with sD^{pol}. a denotes significance at the 1% confidence level, b at the 5% confidence level and c at the 10% level.

Trading Strategy	Risk-free rate $X_{i,t}$	Risk-free rate $X_{i,t}^{+}$	Risk-free rate $X_{i,t}^{-}$	Market rate $X_{i,t}$	Market rate $X_{i,t}^{+}$	Market rate $X_{i,t}^{-}$
No transaction costs						
M_T	33.2485	17.7226	15.5405	29.0575	15.4238	13.6477
$\max_{t\in[0,T]}\{M_t\}$	33.3755	17.7738	16.3134	29.1761	15.6800	13.7278
$\min_{t\in[0,T]}\{M_t\}$	-0.0406	-0.1168	-0.0917	-0.0549	0.0000	-0.2398
$q_{0.05}(M_t)$	0.8345	0.4743	0.2734	1.1774	1.0912	0.0000
$\sum_t \mathbf{1}(M_t < 0)$	1	1	12	1	0	43
10 bps						
M_T	24.0594	11.6034	12.4646	19.8684	9.3046	10.5718
$\max_{t\in[0,T]}\{M_t\}$	24.2044	11.6657	14.7080	20.2108	10.5371	10.6630
$\min_{t\in[0,T]}\{M_t\}$	-0.1286	-2.2423	-0.1940	-0.1429	-0.0399	-0.3421
$q_{0.05}(M_t)$	0.2535	-0.5732	0.0557	0.3913	0.4594	-0.1355
$\sum_t \mathbf{1}(M_t < 0)$	3	155	22	4	1	134
20 bps						
M_T	14.8703	5.4843	9.3888	10.6793	3.1855	7.4959
$\max_{t\in[0,T]}\{M_t\}$	15.3437	5.0716	13.1026	14.0891	6.6094	8.6155
$\min_{t\in[0,T]}\{M_t\}$	-1.5261	-4.8499	-0.3238	-1.6210	-0.9771	-0.6888
$q_{0.05}(M_t)$	-0.9640	-3.0753	-0.1363	-0.8408	-0.3894	-0.4805
$\sum_t \mathbf{1}(M_t < 0)$	197	413	82	156	116	212
30 bps						
M_T	5.6812	-0.6349	6.3129	1.4902	-2.9337	4.4201
$\max_{t\in[0,T]}\{M_t\}$	8.954	1.1072	11.4972	9.1693	3.2732	7.0032
$\min_{t\in[0,T]}\{M_t\}$	-3.4651	-7.6135	-0.6654	-3.5721	-3.1743	-1.2930
$q_{0.05}(M_t)$	-2.6450	-5.7681	-0.3723	-2.4114	-2.7774	-0.9527
$\sum_t \mathbf{1}(M_t < 0)$	321	687	176	311	510	285
$\sigma(\Delta M_t)$	0.1667	0.1819	0.1224	0.1774	0.2018	0.1340
Number of trades	9104	6062	3042	9104	6062	3042
Average duration	1.39 days	1.46 days	1.26 days	1.39 days	1.46 days	1.26 days

Table 8: The table shows the terminal value, the maximum and the minimum value, the 5% quantile and the number of days with a negative value of the money account for trading strategies on company signals and different levels of transaction costs. $X_{i,t}$ incorporates of buy- and sell-signals, $X_{i,t}^{+}$ only buy-signals and $X_{i,t}^{-}$ only sell-signals. The benchmark in the left half is the risk-free rate. The benchmark in the right half is the stock market. The lower panel shows the volatility of the change in the value of the money account, the number of trades and the average duration per trade.

RIC	Company Name
$AA.N$	Alcoa Incorporated
$ABT.N$	Abbott Laboratories
$AIG.N$	American International Group Inc
$AMGN.O$	Amgen Inc
$APC.N$	Anadarko Petroleum Corp
$AXP.N$	American Express Co
$BA.N$	The Boeing Company
$BAC.N$	Bank of America Corp
$BAX.N$	Baxter International Inc
$BMY.N$	Bristol Myers Squibb Co
$BSX.N$	Boston Scientific Corp
$C.N$	Citigroup Inc
$CAT.N$	Caterpillar Inc
$COP.N$	ConocoPhillips
$CSC.N$	Computer Sciences Corp
$CSCO.O$	Cisco Systems Inc
$CVX.N$	Chevron Corp
$DD.N$	E I Du Pont De Nemours And Company
$DELL.O$	Dell Inc
$DIS.N$	Walt Disney Co
$DOW.N$	The Dow Chemical Co
$DVN.N$	Devon Energy Corp
$F.N$	Ford Motor Co
$FDX.N$	Fedex Corp
$GE.N$	General Electric Co
$GLW.N$	Corning Inc
$GR.N$	Goodrich Corp
$GS.N$	The Goldman Sachs Group Inc
$HD.N$	The Home Depot Inc
$HON.N$	Honeywell International Inc
$HPQ.N$	Hewlett Packard Co
$IBM.N$	International Business Machines Corp
$INTC.O$	Intel Corp
$JNJ.N$	Johnson & Johnson
$JPM.N$	Jpmorgan Chase & Co
$KFT.N$	Kraft Foods Inc
$KO.N$	The Coca Cola Co
$MCD.N$	McDonald's Corp
$MDT.N$	Medtronic Inc
$MO.N$	Altria Group Inc
$MON.N$	Monsanto Co
$MMM.N$	3m Co
$MRK.N$	Merck and Co Inc
$MS.N$	Morgan Stanley
$MSFT.O$	Microsoft Corp
$LLY.N$	Eli Lilly And Co
$LMT.N$	Lockheed Martin Corp
$ORCL.O$	Oracle Corp
$OXY.N$	Occidental Petroleum Corp
$PFE.N$	Pfizer Inc
$PG.N$	Procter & Gamble Co
$SLB.N$	Schlumberger NV
$T.N$	AT&T Inc
$TRV.N$	Travelers Companies Inc
$TWX.N$	Time Warner Inc
$TXN.N$	Texas Instruments Inc
$UTX.N$	United Technologies Corp
$VZ.N$	Verizon Communications Inc
$WFC.N$	Wells Fargo and Co
$WMT.N$	Wal Mart Stores Inc
$WLP.N$	WellPoint Inc
$XOM.N$	Exxon Mobil Corp

Table 9: The table gives the list of companies that are included in the analyses and matches the company name with the company's RIC (= Reuters instrument code)