Working Paper
Series

FINRISK
National Centre of Competence in Research
Financial Valuation and Risk Management

National Centre of Competence in Research
Financial Valuation and Risk Management

Working Paper No. 719

# News Sensitivity and the Cross-Section of Stock Returns

Michal Dzielinski

First version: July 2011
Current version: July 2011

This research has been carried out within the NCCR FINRISK project on
"Behavioural and Evolutionary Finance"

# News sensitivity and the cross-section of stock returns.*

Michal Dzielinski
University of Zurich, Department of Banking and Finance

July 2011

**Abstract**

The paper is the first one outside the high-frequency domain to use sentiment-signed news to directly compare news and no-news stock returns. This is done by estimating whether returns on positive, neutral and negative news days are significantly different from the average daily return for a large sample of US stocks over the period from January 2003 to August 2010. The general results show that positive news days indeed have above-average returns and negative news days returns are below average, while the neutral news days are economically barely distinguishable from the average. The market also proves to be fast and accurate at pricing new information, as there are no signs of drift shortly after news days. On the contrary, a directionally correct and statistically significant movement can be found on the day before the news day. The cross-sectional analysis reveals significant differences in the strength of market reactions between stocks ranked on size, book-to-market or news coverage. The general results however hold across all subsamples and are also not driven by earnings announcements or past stock returns. Moreover, the average news sensitivity is itself a priced source of risk. A portfolio of stocks with high sensitivity to news outperforms a portfolio of stocks with low sensitivity by a statistically and economically significant 0.84% per month. This news premium seems to primarily relate to the high impact of news in situations of general uncertainty.

Keywords: company news, sentiment, news analytics, public information
JEL: C33, G12

*I would like to thank Thorsten Hens, Marc Paolella, Kerstin Kehrle, Sven Steude, Marie Mikl, other participants of the BBL Seminar at the University of Zurich and the participants of the 2011 Warsaw International Economic Meeting for useful comments. The financial support from "LGT and Science" is gratefully acknowledged. Financial support by the University Research Priority Programme Finance and Financial Markets of the University of Zurich and the national center of competence in research Financial Valuation and Risk Management is gratefully acknowledged. The national centers in research are managed by the Swiss National Science Foundation on behalf of the federal authorities. As always, I honor my parents for bringing me to where I am now.

The stock market is one of the most information-intensive environments and trading is largely information-driven. There is however less accord concerning the source of this information, is it rather public or private. The question is especially relevant for company-specific information, where the potential for insider relationships is the greatest. Maintaining equal access to information has been advocated as an important aspect of the level playing field for investors and with ad-hoc publicity rules in place and numerous communication channels there are presently thousands of news announcements released around the world every day.

The natural way to resolve this issue would be to test the price impact, whenever a new piece of private or public information arrives. This paper gives a partial answer by analyzing public information in a comprehensive way. Such information can be better identified, so direct tests of its price impact are in principle possible. The ideal test requires the knowledge of: (1) whether the *ex ante* content of the information was positive or negative, and (2) what were the market expectations. A measure of expectations is important because in standard asset pricing theory only information different from expected can impact the price. In practice, directly measuring the expectations of the market as a whole proves impossible and proxies are needed.

For quantitative information, like earnings, it is straightforward to satisfy the first requirement, insofar as earnings growth and anything that contributes to it is good. The arrival of most of such quantitative information is scheduled in advance, so there are also readily available, up to date proxies of expectations, like consensus estimates of stock analysts. Measures of the unexpected fraction of earnings, like SUE, are also available. A similar approach also works for macroeconomic releases. Consequently, the literature analyzing the impact of quantitative information is immense. A survey by Yu (2009) cites more than twenty studies of the post-earnings announcements drift (PEAD) alone.

However, it has been long suggested that qualitative information matters too, even if it only accompanies quantitative information, like the actual text of an earnings announcement. Much more such qualitative information is transmitted in the form of analysts reports and ad hoc announcements about company events. The problem here is that there is no immediate value for good and bad anymore, because such information is transmitted as unstructured text. Such a value must first be computed and borrowing from computer linguistics one could talk of opinion mining or sentiment analysis where the aim is to determine whether the author of the text meant to express a positive or negative opinion about whatever the subject of the text was. As such it has long been the focus of communication and content analysis, relatively new is the widespread use of automatic algorithms to scan the texts. Research on automated sentiment analysis has recently been very prolific, helped by increasing computational power and the demand to incorporate the vast universe of online communication[1], and different methods have been developed. An excellent and largely still up to date survey is provided by Pang and Lee (2008).

The approach that first made its way into the finance literature treats a text as a "bag of words" and the sentiment score is based on the frequency certain words, defined by the researcher, appear in it. Li (2006) focuses on the words "risk" and "uncertain" in company annual reports. Davis et al. (2006) count words from texts of earnings announcements which fall into certain categories, defined in the DICTION dictionary, associated with positive and negative language. Similarly, Tetlock (2007) uses the Harvard-IV-4 dictionary to determine the fraction of negative words in a popular *Wall Street Journal* column and Tetlock et al. (2008) extend this approach to all company-specific news from the *WSJ* and *Dow Jones Newswire* archive. All of these paper show that measures of language are significant in explaining returns and changes in fundamentals, independently of quantitative factors.

[1] Early adopters of this technology included Holywood producers interested in online film reviews.

The "bag of words" approach, though praised for its simplicity and reliability, has some important limitations. From an information processing viewpoint, it is not clear that the frequency of words really determines the sentiment of text. On one hand, Pang et al. (2002) show that the *presence* of certain words can be a better indicator than frequency. Otherwise, it is also easy to construct examples where the *order* of the words is decisive. Even more importantly perhaps, treating a text as a collection of words tells nothing about the relationships between them. The next generation text reading algorithms therefore aim also at extracting these relationships, identifying words within the subject-predicate-object syntax. Such algorithms can be either *deductive*, meaning they follow explicit rules on how to parse the text which have to be defined in advance, or *inductive* in which case a learning set of evaluated texts has to be supplied from which the algorithm attempts to read the rules applied to all subsequent items. Both approaches can also be combined. In a finance setting it might be reasonable to predefine the list of potential subjects, company names for instance, and let the algorithm determine what is being said about each of them. In any case, improvements in accuracy achievable by introducing syntax are significant. Also, contrary to what is sometimes said, syntactic approaches are not any more subjective than "bag of words". In fact, surveys of methods of content analysis assign all of them to the family of "supervised approaches", indicating human involvement in their design. This is because the dictionaries, which are behind any "bag of words" analysis have to be created by humans. Even inductive algorithms offer a fair degree of inter-subjective reliability, because the learning sets are always evaluated by more than one person and the results of learning are only accepted when the agreement between the instructors and the machine but also among the instructors themselves reaches a certain, appropriately high threshold. The sophistication of syntactic algorithms is in a sense limiting in that they rarely work outside the topic area for which they were originally developed, simply due to the differences in language used e.g. in political debate and online product reviews. This should not be a major concern for analyzing company news however, because they are a rather well defined group and thanks to established styles of reporting also not prone to sudden changes of language over time.

In the financial industry, Thomson Reuters was the first to apply a state-of-the-art text reading algorithm to its own extensive archive of company-specific newswire messages, as they appeared on traders' screens, going back to January 2003. It is arguably representative of the timely, public information available at least to professional investors, i.e. public news. Each message is assigned an exact timestamp and a list of companies it mentions. The subsequent content analysis takes place at company level, so there are as many scores per message, as there were mentioned companies. The first score is the relevance of the message for each company, which gives and indication of how often and how prominently it was mentioned. The centerpiece of the analysis is the probability of the author sentiment being positive, neutral or negative. In a finance setting it is important to stress the advantages of focusing on the *author* rather than the *recipients* of the message. Focusing on the recipients, financial investors in that case, would in the extreme lead to classifying as good those news, which moved the market up and as bad news those, which moved the market down. Such a news measure would be obviously endogenous. To the extent that the authors of newswire messages, some three thousands Reuters journalists, are not purely driven by past stock performance, author sentiment has a claim on exogeneity.

Earlier studies using this data have covered three topics. Groß-Klußman and Hautsch (2010) focus on the instantaneous impact of public news using high-frequency data. They find strong responses in volatility, volume and liquidity measures and a mixed picture for returns. Storkenmaier et al. (2011) also construct measures of intraday liquidity and trading activity but focus on comparing the averages for negative and positive news days to no-news days. They find that liquidity is significantly lower on negative news days but not significantly higher on positive ones. Trading activity always increases on news days, regardless of the sentiment. Finally, Sinha (2010) uses news sentiment to calculate monthly sentiment scores for individual stocks, which are then used to construct "news momentum" portfolios. Interestingly, the information contained in past news sentiment is shown to subsume information contained in past returns.

I take a different approach based on constructing a daily history of public news for a representative sample of US stocks and focusing on daily returns. Using the Thomson Reuters News Analytics archive, each day for every stock in the sample is assigned to one of four categories: no-news day, positive-news day, neutral-news day or negative-news day. In case of multiple news items on the same day, the prevailing sentiment is computed as a weighted average of the individual sentiments. Then, a panel regression is estimated to test whether public news systematically impacts returns, that is whether returns on news days are significantly different from returns on no-news days, taking into account stock-specific differences. It turns out that they are and the estimated differences between the average news day and no-news day return, which I will call news sensitivities, are both large and in line with the news sentiment: +45bps for positive news days and -43bps for negative news days. This is not trivial, since it is neither given that public news indeed conveys new information to the markets nor that author sentiment is a good indicator of market reaction.

Contemporaneous returns have also been neglected in a string of papers on news impact (Chan (2003), Gutierrez and Kelley (2008), Tetlock (2010) and Tetlock (2011)), which all use returns in news-related time periods (months, weeks or days) to predict future returns. At the same time it would be important to know whether the content of the news had any immediate impact on the stock price in the first place. One likely reason for this ommission is that the news data employed in those studies was not sentiment-signed, so there was no obvious way to design a test of market reaction, given that positive and negative news should have different impact. Using absolute returns is a dangerous way around this problem, since it is biased towards finding significant results. To see how, assume that positive news should have positive market impact. In empirical data one is also likely to find cases, where positive news was associated with negative returns.[2] When using raw returns these cases would be correctly classified as "wrong", i.e. contradicting the initial hypothesis. An analogous argument can be constructed for negative news. When using absolute returns however, all returns are put on one side and the "wrong" cases mixed with the "right" ones.

The issue of news content and contemporaneous returns was mentioned by Tetlock et al. (2008), who measure the news content by the standardized fraction of negative words in the text. An increase of one standard deviation in this measure corresponds to an 8bps drop in return. Once only news about earnings are considered, the impact grows to around 32bps. These results are conditional on there being news on a given day in the first place. Separately, they show an event-study type plot of abnormal returns around news days but do not comment on the statistical significance of these returns.

The only study, which to my best knowledge directly attempted to compare news and no-news returns dates back to well before the age of text-reading algorithms. Thompson et al. (1987) perform the monumental task of hand-collecting one-year's worth of messages from the Wall Street Journal Index and assigning them to companies and topic groups. The problem with their results is that they are based on parametric $t$-test, which rely on the independence and normality of observations, both of which are most likely not true when comparing news and no-news returns. They also did not have any means of determining news sentiment.

Apart from contemporaneous reactions I also look at an event window before and

[2] This must not be a weakness of the sentiment scoring tool but rather evidence of factors different from the language content playing a role in determining the market reaction.

after the news day in order to measure the extent news anticipation and the speed of its incorporation. It turns out that prices move up (down) already one day before a positive (negative) news day but only a fraction of the total impact is incorporated. There is no identifiable move before a neutral news day. Investors are therefore able to correctly assess the directional impact of upcoming news but not its magnitude. On the other hand, the post-news event window does not reveal any significant drift or reversal, except for a few isolated cases, which suggests that markets are fast and accurate in incorporating public news. It also means that news is not just short-term noise, where prices promptly return to their earlier level, but has persistent impact.

The contemporaneous effect persist across a range of cross-sectional breakdowns. The approach here is to divide stocks into quintiles based on market capitalization, book-to-market ratio, trading volume, volatility and news coverage and run the panel regression in each quintile separately. The most important finding is that the news reactions found in the whole sample are not confined to any special subset. In fact, because the the distribution of news is highly skewed towards a relatively narrow group of companies, most cross-sectional subsamples have larger average news sensitivities that the whole sample. Notwithstanding, the analysis uncovers a fair deal of heterogeneity. The news sensitivities are greatest among smaller, less liquid and more volatile stocks, where the magnitude of the average daily news sensitivities is well above 1%. Interestingly, infrequent news is also associated with stronger reactions, largely independently of other company characteristics. This fits nicely with limited attention models in that investors bombarded with news become less sensitive to it.

In this context it is particularly interesting to look at how news impact changes during the financial crisis. It turns out that investors react more strongly to news during this period, indicating that they pay more attention to them. Interestingly, the crisis also increases the polarity in the sense that positive news days are associated with even higher returns and the opposite is true for negative news days.

However, reactions to news are not only relevant for same-day returns. Differentiating stocks based on their average news sensitivity produces a significant premium, with high news sensitivity stocks earning 84bps per month in excess of low news sensitivity stocks, after traditional risk factors are controlled for. Intuitively, stocks which react more strongly in the face of uncertain future news, are less predictable and thus riskier for investors. The intuition is confirmed by the fact that the news sensitivity premium is the strongest among stocks with already high uncertainty, proxied by idiosyncratic volatility. Interestingly, past news sentiment is largely irrelevant for the news premium, even though it would be reasonable to assume that investors are mostly concerned with sensitivity to the downside.

The paper contributes to the literature in several ways. It shows in a comprehensive fashion that even in the presence of market expectations and private channels of information, both of which are impossible to control for, public news still plays a significant role in the stock market, impacting both short and long-term returns. In the short-term framework it is the first to directly compare returns on positive and negative news days and explore the heterogeneity across company characteristics, industries as well as the special case of the financial crisis. In the context of long-term returns it is the second one after Fang and Peress (2009) to find a risk premium measure directly based on the media. Compared to their paper it goes one step further in that not only the occurrence of the news is considered but also its impact. Moreover, the paper confirms that a "face value" assessment of news content, based only on an analysis of language, is already a good indicator of the market interpretation.

The remainder of the paper is organized as follows. Section I describes the data in greater detail. Sections II and III deal with the empirical findings on short and long-term impact respectively. Section IV is devoted to robustness checks, while Section V looks specifically at the impact of the financial crisis. Section VI concludes.

# I Data

The data on news come from Thomson Reuters News Analytics, which is an algorithmic text-reading tool. Its two main components are: a real-time news and news metadata feed, suitable for automated news-driven trading, and a systematically updated archive, at the time of writing spanning the period January 2003 - August 2010 and containing more than 12 million news item for about 16'000 stocks worldwide. The inputs are all company-specific news items coming over the Reuters newswire, which are then described with more than 40 pieces of metadata. The most important of those are:

- identifier of the company mentioned in the news
- timestamp, indicating the news arrival to millisecond precision
- sentiment, a discrete variable indicating whether the news was positive (+1), neutral (0) or negative (-1). This is determined based on a purely linguistic analysis of text and does not in particular contain any form of market feedback
- relevance, a continuous variable on the (0;1] interval indicating how prominently the company was mentioned in the news. A relevance score of one generally indicates the company was mentioned directly in the headline
- sentiment probabilities, this set of three variables adding up to one shows the probability of assigning each of the possible sentiments to the news item (the one eventually assigned is the one with the highest probability) and thus, intuitively, a positive news with a probability of 0.8 can be seen as more clearly positive than one with a probability of 0.4

- novelty, which shows whether there have been preceding news in any of the five predefined time windows, ranging from six hours to five days

- topic code, a set of predefined Thomson Reuters codes (the full list contains more than 4'000 items), which together give an idea of the subject matter or, in the parlance of news analytics, "aboutness" of the news

It is important to stress that the meaning of news sentiment in this setting is not related to investor mood as in the usual sense. In particular, a news item might be assigned positive sentiment, because it indeed conveys important favorable information about a company *or* because it uses enthusiastic language. The latter "over-reactionist" component is however mitigated by comprehensive rules being in place at Thomson Reuters, which govern the wording of news announcements. Therefore, I tend to assume that significant reactions to news are due to their informational content and not just language.

As the equity universe I use the constituents of the Datastream US Total Market Index, which consists of both NYSE and NASDAQ stocks, representing roughly 90% of market capitalization. Datastream is used because it is most compatible with news database, also provided by Thomson Reuters. Ince and Porter (2006) point out certain issues related to using equity returns data from Datastream, most notably the fact that Datastream equity constituent lists contain other instruments than common stock. Fortunately, these concerns mostly impact small stocks, which do not enter into my sample (the smallest stocks in my sample correspond roughly to the fourth decile in Fama-French dataset), however, I still attempt to address them. Similarly to what has been suggested, I manually filter the list of companies, eliminating those that contain terms such as "ADR" or "Pref" in the NAME datatype.

Most importantly, I only retain stocks which had at least one news day in any given year. The sample contains 780 stocks in the first year and grows to almost 950 towards the end reflecting improving news coverage over this period. It is biased towards large cap stocks but this was necessary to ensure they have sufficient news. For each stock in the sample I retrieve its history of news announcement, recording the timestamp, the sentiment and the three sentiment probabilities. To be included, the news had to arrive at least two hours before the close and be "novel", i.e. have no precedents in any of the predefined time windows. No differentiation was made relating to whether the news mentioned single or multiple stocks, what kind of release it was or the topic. Depending on news specification I either use a relevance filter of 1 (leaving in only headline news) or no filter at all.

Subsequently, I multiply the sentiment of each news item with its probability and in the case of multiple news per day, compute a weighted average (where weights are the sentiment probabilities) to obtain the prevailing sentiment for that day:

$$avg\_sent = \frac{\sum 1 \cdot prob_{pos} + \sum(-1) \cdot prob_{neg}}{n_{pos} + n_{neut} + n_{neg}} \in [-1; 1] \tag{1}$$

A news day is considered as positive if its weighted average sentiment is greater than 0.33, as negative if it is less than -0.33 and as neutral if it is in between. Therefore, although neutral news items do not enter the sentiment calculation in the numerator of the above expression (due to their sentiment of 0) their number shows up in the denominator pushing the overall score downward, so that a day with many neutral news items would still be classified as neutral. This procedure generally replicates rules applied to individual news items. Repeating it for each stock I obtain a complete history of news days and their respective sentiment classification.

For each stock in the sample I also collect the daily return, the daily market capitalization, the daily trading volume, the market to book ratio as of end of each year and the industry code based on the IBC ten industry groups classification. Panel A of Table I shows the relevant summary statistics in yearly breakdown. As mentioned before, the median stock is fairly big but this is important to guarantee that from the beginning of the sample period most stocks have at least some news days. Over the last three years this is the case for basically all of them. The average number of news days per stock also increases with time, meaning that the informational environment becomes more dense. Looking at the sentiment breakdown, the majority of companies feature news days of each type in any given year, though neutral news days are slightly more frequent, followed by positive ones. Even during the crisis years (2008 and 2009) negative news days are consistently the smallest group in terms of number of stocks receiving them. However, the average per stock is higher for negative than for positive news days during those years, meaning that some companies (presumably financial ones) had noticeably skewed proportions of news days at that time.

Table I around here

It is also worth noting, that news coverage is far from being uniform across stocks. Looking ahead at the topmost panel of Table III, there are large difference between quintiles of stocks ranked by total news coverage. The average stock in the top quintile has about ten times as much news days per year as in the lowest one. Although the companies in the top news coverage quintile are bound to be bigger and of more interest to investors, this fact also raises questions about the information content of such frequent news. I return to this issue in the middle part of the empirical section.

Besides using all news, I also consider different news specifications, which are summarized in Panel C of Table I. The left hand side reports the average number of news day per stock and per year for each sentiment group and the right hand side the corresponding average sentiments, which are conditional on the news day belonging to the respective sentiment group. Thus the possible range of sentiment for a positive news day is between 0.33 and 1. One dimension I use is the relevance of the news, where I distinguish between highly relevant (and visible) "headline" news and the rest. For some news days, those which previously featured both headline and non-headline news, this might mean a reclassification to a different sentiment category, as now only headline news on those days are used to calculate the weighted sentiments. That is why the average news numbers in rows 2 and 3 only add up to the number in row 1 for the "Total" column. In general, there are more news days with headlines than without, so concentrating only on the former does not lead to a dramatic loss of data (roughly one third). On the second dimension of specification I focus only on news days with extreme sentiment, which is defined as sentiment in excess of 0.5 in magnitude. This leads to a further loss of data (in particular neutral news is dropped, since it would be awkward to define "extremely neutral" sentiment) but especially for headline news it is rather limited. It seems that a vast majority of headline news days have a clear cut sentiment anyway. The corresponding average sentiments also change only slightly. It is also important to note that neutral news days make up a large fraction of the news day population in all specifications, in which they are included. Therefore, isolating neutral news, matters greatly for any assessment of news impact.

# II Short term impact of news on stock prices

In order to examine the impact of news on returns on a short-term basis, I estimate a panel regression of daily returns on the news day dummies:

$$r_{i,t} = \beta_M \cdot r_{M,t} + \sum_{k=-3}^{3} \delta_k^{pos} \cdot News_{i,t-k}^{pos} + \sum_{k=-3}^{3} \delta_k^{neut} \cdot News_{i,t-k}^{neut} + \sum_{k=-3}^{3} \delta_k^{neg} \cdot News_{i,t-k}^{neg} + \epsilon_{i,t} \quad (2)$$

The incidence of news anticipation and post-news drift can be judged based on an event window of three days before and after the news day. Additionally, I control for the contemporaneous market return to eliminate any beta effect. The panel approach is best suited for this type of analysis, because it estimates the incremental impact of news, on top of the given company's average return. Its merits for analyzing daily differences have been shown in the study by Patton and Verardo (2010) on daily betas around earnings announcements, to which my study is methodologically similar. The resulting parameter estimates on the dummy variables should be interpreted as differences between the average return on a given day during the event window and the average daily return outside the event window. I will refer to them as news sensitivities. Another advantage of this approach is that systematic differences in returns between stocks are absorbed by the panel.

The significance of the estimates can be judged by the related t-statistics, which are robust to clustering across the time and stock dimensions. This is important because both the persistence in the news reactions of particular stocks (firm effect) and spillovers in the sense of news about one stock also affecting other stocks (time effect) can lead to correlated residuals. Moreover, neither of these effects needs to be uniform, meaning that the firm effect may be time-varying and the spillovers can only affect a subgroup of stocks rather than all of them. Including dummy variables for stocks and time periods is not sufficient in that case, since it still leads to biased standard errors (Petersen (2009)). To avoid these issues, I employ the sandwich estimator of Cameron et al. (2006) [3], which

[3] I would like to thank the authors for making their code available online

allows for arbitrary two-dimensional clustering. In the first dimension each company is an individual cluster, in the second, time, dimension, the size of the cluster is set to one day.

Working with daily aggregates means that whatever happens intraday is not looked at. Although the high-frequency analysis of Groß-Klußman and Hautsch (2010) shows that stock prices move after news, there might also be significant intra-day anticipation (Ranaldo (2007)). In general, anticipation does not invalidate my results as long as the measured returns are still attributable to the news and not to other factors. In other words, the analysis makes sense as long as the news is instrumental to and not just coincidental with the returns. For earlier studies, using monthly or weekly time windows to relate news and returns, it is arguably a serious problem. Daily window is less prone to such concerns. A more serious issue might arise if news would only ex post describe what has already happened in the market, so called "no content" news. Although I cannot guarantee eliminating such news (it is anyway a contentious issue, what constitutes content, given market expectations) but I try to minimize their influence. On one hand, relying on newswire items, as opposed to print media or online news, is in itself a partial remedy, since the main purpose of the newswire is to deliver real-time factual information to investors. Otherwise, filtering for maximum novelty should eliminate repetitive items, while focusing on high relevance should eliminate collective articles, which for instance merely quote the performance of some stocks. Finally, truncating the news flow before the market close filters out the wave of daily market commentaries and reports, which tend to be clustered in the afternoon.

## A Baseline results

The results summarized in Table II first of all show the perils of analyzing news without regard to their sentiment. In the first column, where all news days are pooled together, the results are hardly meaningful, despite being statistically significant. Focusing only on headlines does not improve the situation. The estimated news sensitivities are in fact very similar to those for neutral news days. This is most likely due to the fact that as was shown before, the positive and negative news days are roughly equal in numbers.

Table II around here

Only after separating between positive and negative news days the results become truly meaningful (Figure 1). Doing this reveals that news sensitivities agree with the prevailing sentiment of the news and are highly statistically and economically significant. Headlines are associated with clearly larger news sensitivities, while non-headlines are not very important. This means that results for all news are mainly driven by the most visible items, so the salience of news seems to play a big role. On the other hand, extreme sentiment does not seem to add much to the initial sensitivities on positive or negative news days. Furthermore, the magnitude of the impact is roughly equal for positive and negative news days. The "bad is stronger than good" effect, frequently cited in psychology, is not found here. Given the large and roughly equal number of positive and negative news days and the fact that the average sentiment is roughly equal in both cases, it would be hard to suggest systematically different content of positive and negative news. The results indeed are a strong indication that the "negativeness" as such does not affect the pricing of news.

Looking at the event window before the news days shows that the market is able to correctly anticipate the direction of the news but only a fraction of its total impact. This is evidence of either only partial information leakage or the fact that the population of better-informed insiders, at least those that have a lead of a whole trading day, is relatively small. It also seems that the market is quick and accurate in pricing new information, as there are basically not significantly different than average returns after the news days, apart from a few instances of reversal of rather minor magnitude.

## B Cross-sectional analysis

To account for the heterogeneity of news impact, I re-estimate the panel regression in (2) for quintiles of stocks sorted according to various company characteristics. A secondary goal of the cross-sectional analysis is to verify that the results from the previous section are not driven by any specific subset of stocks, thus the sorts follow the risk factor literature. Apart from the most commonly used size and book-to-market ratio, they also include liquidity, volatility and news coverage. Separately, I do an industry comparison based on the IBC industry group classification, which is the highest level classification consisting of ten broad categories. For the sake of compactness, I only use the headline news specification in the cross-sectional analysis, since headlines seem to drive the overall news impact.

Table III around here

At the end of each year stocks are sorted into quintiles based on the respective characteristic. The assignments are pooled across years and used to estimate the news impact. This effectively means adopting a backward-looking perspective, which I do to make sure that stocks do not change into a different quintile during the estimation period. The results are very similar if the year following quintile formation is used instead. Due to using end-of-year breakpoints, the analysis covers the period 2003 - 2009. Finally, only the estimates for the news day itself are reported to conserve space, given that no substantial effects could be found in the rest of the event window[4].

[4]Full results are available upon request.

Sorting on news coverage as a risk factor is motivated by the findings of Fang and Peress (2009) that stocks neglected by the media earn higher average returns. They suggest that their argument only applies to the general media, like newspapers, which do not transmit what they call ”genuine news”, however coverage is arguably also the feature of channels like the newswire. For instance, Caterpillar and Alcoa are companies of similar size and both are constituents of the Dow Jones Industrial Average index, yet Alcoa had roughly 50% more news days over the sample period. It seems that the differences in the number of newswire reports are not entirely due to differences in actual news events. Therefore, I use the total number of news days in a given year to group stocks into quintiles of news coverage (Panel A in Tables III and IV). The quintile means show that the distribution of news days is indeed highly skewed. In fact, stocks in the upper quintile receive almost as many news days as all other quintiles combined (47 vs 57 per stock per year on average). Although no direct measure is available, I tend to interpret infrequent news as having higher informational content. Indirect evidence in provided by the much stronger market reactions to infrequent news, which are roughly four times larger in the bottom quintile than in the top one (1.45% vs 0.35% for positive and -1.55% vs -0.39% for negative news days). However, frequent news still has meaningful impact.

Table IV around here

Sorting on size is based on the average daily market capitalization, computed as the product of closing price and number of shares outstanding. Small stocks have significantly less news days than large ones and the effect is also quite persistent, with the second largest quintile having just over half the number of the top quintile. To the extent that the number of news days increases monotonically with size, it raises the question whether sorting on size is any different to sorting on news coverage but in fact only 38% of the stocks are assigned to the same quintile in both cases. Nevertheless, the fact that only big news about small stocks tend to get published likely plays a role in explaining while small stocks react so much more strongly on news days (1.65% vs 0.36% for positive and -1.36% vs -0.40% for negative news days). On the other hand, it is also easy to imagine that for a small specialized company a single event, captured in the news, can have a much greater impact on future prospects than for a large diversified one. Large companies, however, also react significantly on news days.

When sorting on liquidity, proxied by the average daily trading volume, i.e. the product of the number of shares traded and the average price during the day, the main conclusion is that the news impact is not an artifact of trading fictions. Though less liquid stocks react stronger than more liquid ones (1.27% vs 0.29% for positive and -1.26% vs -0.31% for negative news days), the spread is actually smaller than in the case of size, which suggests that liquidity is a less important factor. There might again be some confusion with news coverage but it is rather limited as just under 40% of stocks have the same assignment in this case.

For the remaining two sorts, the number of news does not play a significant role, so the differences in news sensitivities are most likely not influenced by the content of the news. Therefore, it seems that stocks with low book-to-market ratio (calculated by dividing the book value of equity reported for the end of a given year through the average market capitalization during that year) and high total volatility (average daily squared return) are indeed more sensitive to news. In the case of low book-to-market stocks, this might be due to the fact that they are effectively more leveraged, so any news affecting the expected value of their assets will have a greater impact on their valuation. The case of high volatility stocks is more puzzling. According to Merton (1987) model of market equilibrium with incomplete information, high volatility should be a property of "neglected stocks" that investors know little about. To the extent that the degree of awareness can be proxied by news coverage, my findings do not confirm these predictions. In fact, news seems to contribute to volatility. A hint at the reasons is perhaps given by the fact that high volatility stocks have the most negative news days, roughly equal to the number of positive news days, whereas all other volatility quintiles have more positive news days. The news might thus reflect the changing fortunes of such stocks, which make them more volatile.

In sum, the cross-sectional analysis demonstrates that news sensitivities are greatest among smaller, less liquid stocks, where it might also relate to the higher informational content of infrequent news and among low book-to-market and volatile stocks, regardless of the news frequency. Most importantly however, the results are a strong indication that significant reactions to news are a general phenomenon, not confined to any special subset of stocks.

## C Industry comparison

The industry comparison also yields interesting insights (Table V). First of all, there is no significant variation in the average number of news days among industries, despite substantial differences in market capitalization. Furthermore, all industries react significantly to news, though there is considerable heterogeneity, which appears weakly correlated with size or news coverage. For instance, utility stocks which are among the smallest on average have also the smallest news sensitivities, while technology stock, among the largest, react much stronger. The heterogeneity also does not seem to follow the lines on any of the common industry divisions like cyclical vs. defensive or innovative vs. traditional. Clearly, some industries just have higher news sensitivity than others.

Table V around here

Another interesting feature of the results is that some industries appear to react more strongly to negative news (financials, telecom), other to positive ones (healthcare, oil and gas), while the rest has pretty symmetric reactions.[5] Unless they are an artefact of the observation period (a nice case would be to see what happens in the oil and gas sector after the Macondo spill), these differences possibly reveal how investors weigh chances and risks in the respective sectors.

# III News sensitivity as a risk factor

The results of the analysis so far encourage the investigation whether differences in short-term reactions to news translate into systematic differences in returns. In other words, this section explores the possibility of framing news sensitivity as a risk factor, affecting the cross-section of returns. Of course, news sensitivity can be reasonably defined only for stocks that receive news in the firs place. Therefore, at the end of each month only stocks with at least four news days of each type over the preceding 12-month period are selected. This reduces the available sample size substantially, especially early in the sample period, however between 25% and 63% of the stocks are still included. To the extent that these stocks have more news coverage and are also likely the larger ones and both of these features have been shown to reduce news impact, the selection process biases against finding significant results.

In the next step, I regress the returns of each eligible stock on the market return and the three news day dummies. The separate regressions are necessary to obtain individual parameter estimates for each stock. The average magnitude of the parameters on the dummy variables is then the news sensitivity score of that stock. Repeating the above procedure every month yields a monthly ranking of the eligible stocks based on their

[5] The financial crisis does not seem to be a decisive explanatory factor of this effect.

average news sensitivity over the preceding rolling 12-month window. To the extent that the same news days can flow into the regression equation for a given stock over several months, these rankings (and the resulting portfolio returns) will be autocorrelated. Conceptually, this does not appear any different from the standard price-based momentum case, where overlapping returns are used, so similarly to momentum studies I apply a Newey-West correction for up to 12 lags to mitigate this effect.

The final step is to build equally weighted portfolios by taking quintiles of the news sensitivity ranking. These portfolios are held for one month and then rebalanced. The risk premium in question is simply the monthly return on a portfolio that goes long high-news sensitivity stocks (top quintile) and short the low-news sensitivity (bottom quintile) stocks. The monthly return on this portfolio is significantly positive and quite persistent (mean: 0.95%, Newey-West t-stat: 2.67). It is also quite insensitive to the inclusion of well-known risk factors (Table VI). The alpha from the four-factor model controlling for the market excess return, size, book-to-market and momentum is only slightly lower than the time series mean.

Table VI around here

The factor loadings provide interesting insights. The fact that the returns of the long-short portfolio co-move with the market returns is not obvious, given that the largest stocks are likely to be in the short leg. Rather, the news-sensitive stocks respond more to news that also move the market. The positive loading on size and negative (though mostly insignificant) loading on book-to-market confirm earlier findings that it is smaller and growth stocks that react most strongly to news. The significantly negative loading on momentum is interesting. It appears that low-news sensitivity stocks, which form the short leg of the portfolio, tend to exhibit momentum behavior. Given that such stocks react, by definition, weakly on news days this is suggestive of the "underreaction to news" explanation of the momentum puzzle. However, it might also be an artefact of the frequently cited short-term reversal of momentum stocks. Some evidence of it is contained in the (unreported) fact that the loading on momentum loses significance once longer holding periods are used.

Table VII around here

Using different holding periods is a way to robustify the initial result and this is reported in Table VII. The fact that the risk premium is significant across holding periods ranging from 1 month (as in the baseline specification) to 12 months is evidence that the high-news sensitivity effect is not just a short-term phenomenon. In other words, it is not likely to be due to short-term reactions to the news events themselves but rather a stable cross-sectional effect, consistent with the risk factor interpretation. Further support of the risk factor interpretation is given by the fact that the effect is not confined to the extremes. Using tercile or median breakpoints essentially leads to smaller estimated risk premia, which however remain significant across all types of factor models and holding periods.

## A Explaining the news sensitivity effect

Given the persistence of the news premium and its relative insensitivity to common risk factors, the most interesting question is why it arises in the first place. Attempts to answer this question are summarized in Table VIII. One explanation could be that it is driven by certain firm characteristics. To check for this I compute news premia for subsamples of stocks sorted on size, book-to-market and past 12-month return. For these and all subsequent double-sorts I use median breakpoints for both the average news sensitivity and the respective second dimension. It is on one hand the most conservative approach, given that median breakpoints lead to the smallest news premium, and allows to maintain the largest sample size on the other. Using more extreme breakpoints would, if anything, lead to more pronounced results.

Similarly to the news coverage effect identified by Fang and Peress (2009), the news sensitivity effect is stronger in the lower size group if mean estimates are considered. However, it also appears to be much more volatile among such stocks, because the statistical significance is given only in one of the five cases. In the book-to-market case, the news sensitivity effect is significant in both groups, however it is much larger for value (i.e. high book-to-market) stocks. Along the past return dimension, the news sensitivity effects is clearly concentrated among loser (i.e. low past return) stocks.

Table VIII around here

The greater news sensitivity effect among smaller stocks points to illiquidity as a likely driver. To the extent that less liquid stocks are also less able to absorb shocks, which could stem from the news, it is also a plausible explanation. In that case, one should observe a larger news premium among less liquid stocks. Sorting stocks on average trading volume during the past 12 months produces exactly this kind of result. The news sensitivity premium is about three times as large among less liquid stocks, thus illiquidity plays a significant role and can in particular explain the persistence of the news sensitivity effect.

The fact that the news premium is more pronounced among value and loser stocks gives a different hint. Value stocks are usually associated with poor earnings growth and high distress risk, while loser stocks have by definition a poor record in their stock prices. Both is pretty bad news to investors and to the extent that they anticipate more bad news in the future, the will penalize stocks with high news sensitivity, because such stocks would be hardest hit. This effect could be dubbed "fear of bad news hypothesis" and should lead to a stronger news sensitivity effect among stocks with higher likelihood of bad news. To capture the likelihood of future bad news directly, I use two measures of *past* news sentiment, which hinges on the high autocorrelation typically exhibited by news sentiment. The first measure is the ratio of positive to negative news days and the second one is the fraction of positive news days in the total. Both are computed monthly based on the past 12-month window. They are highly correlated but not identical, differing in the treatment of neutral news. The ratio measure disregards them completely, while the fraction measure includes them as an attenuation factor in the denominator. In both cases, a stronger news sensitivity effect would be expected in the group with low values of the measure. As it turns out, the differences are marginal and completely insignificant for both measures. Either the "fear of bad news" does not matter or the proposed measures fail to capture it adequately.

A final possible explanation relates to the uncertainty surrounding a stock. If there is little agreement concerning its prospects, then future news might have a disproportionately high impact, especially so for stocks already having a high news sensitivity. Therefore, news sensitivity should be perceived as particularly risky in combination with uncertainty. To proxy for uncertainty I use idiosyncratic volatility, measured relative to the Fama-French three factor model, as in Ang et al. (2006), over a rolling 12-month window prior to the observation month. This approach is reasonably successful with the news sensitivity effect almost twice as large in the higher uncertainty group. It is however also significant among stocks with lower uncertainty.

The overall conclusion is that the news sensitivity effect is quite pervasive and not driven away by any other effect identified here. It seems however to arise as a results of high uncertainty surrounding certain stocks and its persistence is made possible by their relative illiquidity.

# IV Robustness checks

The primary purpose of this section is to confirm that the results on daily news sensitivities are not spurious and also not too reliant on the particular specification. Subsections A and B deal with the impact on the results of changing some aspects of the original framework. Subsections C to E in turn investigate the possibility that regressing returns on the news day dummy is in fact implicitly regressing on something else. This is done by expanding the original panel regression with interaction terms between the news day dummy and other potential sources of impact on returns.

$$\begin{aligned} r_{i,t} = \beta_M \cdot r_{M,t} &+ \sum_{k=-3}^{3} \delta_k^{pos} \cdot News_{i,t-k}^{pos} + \sum_{k=-3}^{3} \delta_k^{neut} \cdot News_{i,t-k}^{neut} + \sum_{k=-3}^{3} \delta_k^{neg} \cdot News_{i,t-k}^{neg} + \\ &+ \sum_{k=-3}^{3} \gamma_k^{pos} \cdot News_{i,t-k}^{pos} \cdot Rob_{i,t-k} + \sum_{k=-3}^{3} \gamma_k^{neut} \cdot News_{i,t-k}^{neut} \cdot Rob_{i,t-k} + \\ &+ \sum_{k=-3}^{3} \gamma_k^{neg} \cdot News_{i,t-k}^{neg} \cdot Rob_{i,t-k} + \epsilon_{i,t} \end{aligned} \tag{3}$$

The idea of the interaction terms is to take out any influence the additional factors could have had on the initial news sensitivities. If the $\delta$ in the regression remain significant, then this means the initial results were not driven by the factor in question. The factors investigated are earnings announcements, past stock returns and the market return.

## A Including afternoon news

News arriving after 2pm on a trading day, which I shall call afternoon news for short, were eliminated from the original analysis. The decision was motivated by concern about the high volume of ”no-content” news during that time and the subsequent positive bias it might introduce to the results. It is nonetheless revealing to test the impact of such news directly. To this end the history of news days has to be reconstructed. News which arrived during the last two trading hours are thus assigned to the same day, while those which arrived after the close are assigned to the next. The estimated news sensitivities turn out to be smaller than in the original specification. This indicates that much the afternoon newsflow is actually noise with no clear direction and, in the absence of a better way to identify "no-content" news, the decision to cut it off was reasonable.

Table IX around here

## B Changing the event window

Another potentially contentious feature of the model is the arbitrary length of the event window. Its choice was motivated by the will to investigate the extent of pre- and post-news activity on one hand and to limit overlaps between subsequent news events on the other. The exact length of the event window should have no impact on the results, in the sense that e.g. the estimate on the news day itself should not depend on if and how many days before or after were also included, so it is instructive to let this parameter vary. Looking at a range between 0 and +/- 5 days around the news day shows that the estimated news sensitivities indeed are unaffected. This applies first of all to estimates on the news days itself but also to the surrounding days as the event window is extended. Therefore, the choice of the event window is not a salient feature of the model from the econometric point of view and can be safely set to any convenient value.

## C Earnings announcements

Earnings announcements are probably the most closely watched among company news and their impact on the stock price is obvious and has been extensively studied.

It is therefore possible that only this type of news matters and the rest is noise. To control for this I interact the original news day dummy with another dummy variable representing a 3-day event window around earnings announcements days as indicated in I/B/E/S data. Whenever an earnings announcement takes place after market close, it is counted for the next day. The interaction term effectively isolates the impact of those news days, which are also earnings announcements days.

As was to be expected, earnings announcements are very important, as reported in the second section of Table **??**, although other news days remain statistically and economically significant. Positive earnings announcements days have an impact, which is almost 1% higher than other news days and the difference is even larger for negative news days. Interestingly, neutral earnings announcements days have a clearly negative tilt compared to other neutral news days. Another striking feature is the huge amount of discovery, which takes place the day before the earnings get announced and accounts for around 40% of the total impact.

Table X around here

## D Past returns

Controlling for past return of a stock is in fact another way of dealing with the "no content" news mentioned before. This is because in the simplest setting, positive news days would just follow positive returns and negative news days would follow negative returns. To check whether this is indeed the driving force behind news day returns, I add the previous day return interacted with the news day dummy to the regression. The results in section 4 of Table **??** indicate that a small part of the previous day's return is reversed on positive and neutral news days and there seems to be some continuation after neutral news days but the magnitude of these effects is minor. Overall, the previous day return does not seem to be an important factor in explaining the impact of news.

## E Market return

Including an interaction term between the market return and the news day dummy might seem superfluous, given that market return is already present in the regression. However, there is a difference between the two. The parameter estimate on the market return variable accounts for the beta of the *stock* over the entire sample period. The parameter estimate on the interaction term by contrast accounts for the beta of the *news days* and thus controls for the possibility that returns on news days are just outsized reactions to the market return. Due to the fact that both variables are present in the regression, the parameter estimate on the interaction term in fact gives the *difference* between the beta of news days as compared to the average beta of the stock.

The results show that news days (and to some extent the surrounding days) do indeed have higher betas. This effect is particularly pronounced for negative news days where it is also spread over the whole event window. This indicates that the incidence of negative news is more correlated with the general market development than in the case of positive news.

# V Impact of the financial crisis

The purpose of this section is to investigate the impact of the recent financial crisis on the results, given that it covers a substantial part of the sample period. It is also interesting to see how an event of this magnitude affected the news environment. To this end I re-estimate all the panel regression including an interaction term between the respective news dummy and the financial crisis variable, represented as a dummy variable for the period January 2008 - December 2009. When exactly the financial crisis took place (or even whether it is already over) is of course a matter of contention and any choice of time window will be by necessity arbitrary. A conservative approach would be to choose a period where the effects of the crisis were most likely to bias my results. The period of steep market fall and subsequent rally thus seems suitable.

Table XI around here

The general conclusion based on Table XI is that the financial crisis does not change the qualitative conclusions drawn in the previous section. The significance of returns on news days is not limited to the crisis period, although the impact of the crisis is substantial in some cases. First of all, what seems to happen is that the gap between positive and negative news increases. It is on one hand reasonable for a period of extreme market conditions to be associated with more pronounced reactions to news. At the same time it might be surprising that positive news were not being dismissed by investors. Neutral news by contrast are basically unaffected, which confirms the initial intuition that they are of relatively little interest.

# VI Conclusions

Ever since the early study of Mitchell and Mulherin (1994) literature on news impact has been struggling with the inability to distinguish between positive and negative news, without raising suspicions of data snooping or running into data volumes impossible to handle. Modern text-reading algorithms have created the possibility of processing textual information on a virtually unlimited scale. In particular, this includes evaluating the sentiment of text, based solely on its language, which is at the center of a whole field called news analytics. Having an ex ante measure of the news sentiment, not contaminated by any form of market feedback, is ideal for testing the impact of news on stock returns and thus opens up a whole new frontier for research.

Employing a comprehensive database of all company-specific announcements aired on the Thomson Reuters newswire in the period January 2003 - August 2010 this paper takes full advantage of these possibilities. Each announcement in the dataset is linked to a company and evaluated on its sentiment, relevance and several other dimensions. Focusing on a broad selection of US stocks, I construct a daily history of positive, neutral and negative news days for each stock in the sample, based either on all news or just headlines. In the first step, the idea is to test whether returns on news days are significantly different than returns on no-news days and whether the direction of the difference agrees with the observed news sentiment. The results from a robust panel regression confirm both points. Returns on positive news days are significantly above the average across stocks, while those on negative news days are significantly below. The effect is surprisingly symmetric, indicating that investors do not pay more attention to negative news. Neutral news days are associated with statistically significant positive returns, which are however of minor economic importance. This part of the analysis also shows that pooling news days together regardless of their sentiment is effectively like using only on neutral news days.

The cross-sectional analysis confirms that news are an important factor irrespective of various company characteristics or industry classification. There are, however, substantial differences along these dimensions, some of which link back to news coverage (size), while others point out to a greater news sensitivity of certain stocks (e.g. growth or technology stocks).

Most importantly, news sensitivity itself appears to be an important company characteristic with a risk factor interpretation. Stocks with high news sensitivity consistently outperform stocks with low sensitivity by a statistically and economically significant amount, even after traditional risk factors are accounted for. The most likely explanation places the source of this risk premium with the fact that for highly news sensitive stocks future and uncertain news can have a significant impact on their future prospects. Supporting evidence shows a higher news sensitivity premium among stocks with higher general uncertainty.

Further research should focus most of all on cementing the general results shown in this paper, for example by breaking down the impact by different news types, news topics or conditioning on different market environments. The news analytics methods are also spilling over to other asset classes, like commodities and bonds, which opens up new markets to this novel research.

# References

Ang, A., R. Hodrick, Y. Xing, and H. Zhang, 2006: The cross-section of volatility and expected returns. *Journal of Finance*, **61 (1)**, 259–299.

Cameron, A. C., J. B. Gelbach, and D. L. Miller, 2006: Robust inference with multi-way clustering. *NBER Working Paper*.

Chan, W., 2003: Stock price reactions to news and no-news: Drift and reversal after headlines. *Journal of Financial Economics*, **70 (2)**, 223–260.

Davis, A., J. Piger, and L. Sidor, 2006: Beyond the numbers: An analysis of optimistic and pessimistic language in earnings press releases. *Working paper, Federal Reserve Bank of St. Louis*.

Fang, L. and J. Peress, 2009: Media coverage and the cross-section of stock returns. *Journal of Finance*, **64 (5)**, 2023–2052.

Groß-Klußman, A. and N. Hautsch, 2010: When machines read the news: using automated text analysis to quantify high frequency news-implied market reactions. *Journal of Empirical Finance*, **18 (2)**, 321–340.

Gutierrez, R. and E. Kelley, 2008: The long-lasting momentum in weekly returns. *Journal of Finance*, **61 (1)**, 415–447.

Ince, O. and R. Porter, 2006: Individual equity return data from thomson datastream: handle with care! *Journal of Financial Research*, **29 (4)**, 463–479.

Li, F., 2006: Do stock market investors understand the risk sentiment of corporate annual reports? *Working paper, University of Michigan*.

Merton, R., 1987: A simple model of capital market equilibrium with incomplete information. *Journal of Finance*, **42 (3)**, 483–510.

Mitchell, M. and J. Mulherin, 1994: The impact of public information on the stock market. *Journal of Finance*, **49 (3)**, 923–950.

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

Pang, B., L. Lee, and S. Vaithyanathan, 2002: Thumbs up? sentiment classification using machine learning techniques. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Patton, A. and M. Verardo, 2010: Does beta move with news? firm-specific information flows and learning about profitability. *Working Paper*.

Petersen, M., 2009: Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies*, **22 (1)**, 435–480.

Ranaldo, A., 2007: Intraday market dynamics around public information arrivals. *Stock Market Liquidity: Implications for Market Microstructure and Asset Pricing*, G. N. Gregoriou and F. S. Lhabitant, Eds., John Wiley & Sons.

Sinha, N., 2010: News articles and momentum. *SSRN Working Paper*.

Storkenmaier, A., M. Wagener, and C. Weinhardt, 2011: Public information in fragmented markets. *SSRN Working Paper*.

Tetlock, P., 2007: Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*, **62 (3)**, 1139–1168.

Tetlock, P., 2010: Does public financial news resolve asymmetric information? *Review of Financial Studies*, **23 (9)**, 3520–3557.

Tetlock, P., 2011: All the news that's fit to reprint: Do investors react to stale information? *Review of Financial Studies*, forthcoming.

Tetlock, P., M. Saar-Tsechansky, and S. Macskassy, 2008: More than words: Quantifying language to measure firms' fundamentals. *Journal of Finance*, **63 (6)**, 1437–1467.

Thompson, R., C. Olsen, and R. Dietrich, 1987: Attributes of news about firms: an analysis of firm-specific news reported in the wall street journal index. *Journal of Accounting Research*, **25 (2)**, 245–274.

Yu, L., 2009: A survey on post-earnings announcement drift (PEAD). *6th International Conference on Service Systems and Service Management*.

Figure 1: Short term news impact

The figure presents non-cumulated daily returns from 3 days before to 3 days after the news day (day 0). Green lines represent positive news days, blue lines neutral news days and red lines negative news days. Furthermore, following specifications are shown:
all - all news items
h - only headlines
ext - only news days with extreme sentiment (either based on all news or only headlines)

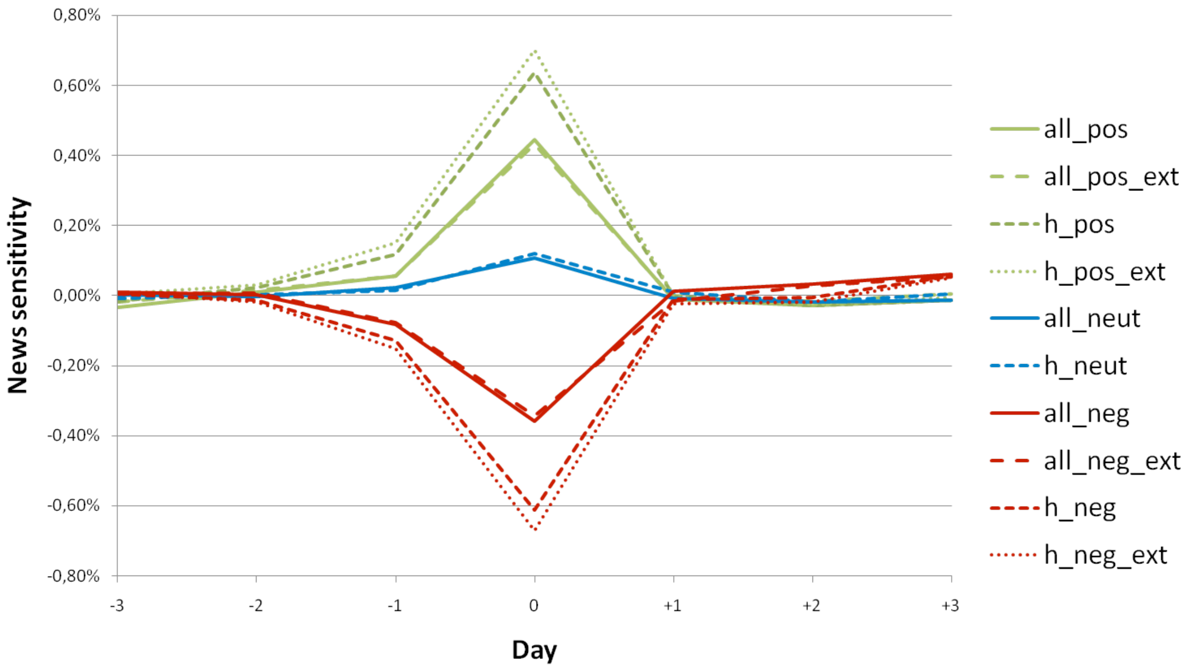

Table I: Summary statistics: whole sample

Panel A of this table gives, for every year in the sample, (2) the number of stocks in the sample, (3) the median market capitalization in billions of USD, (4) the median book-to-market ratio, (5) the number of stocks, which had at least one news day during that year, (6) - (8) the number of stocks which had at least one news day with the respective sentiment during that year, and finally (9) - (11) the average number of news days per stock during that year, broken down by sentiment.

Panel B presents the average number of news days per year and per stock depending on the news specification used and broken down by sentiment. ”Headlines” includes only days, which featured headline news about a stock (but also if there were other, ”non-headline” news on the same day), ”non-headlines” includes only days without a headline. Extreme sentiment is understood as the weighted sentiment on a given news day exceeding 0.5 in magnitude. The right-hand side presents the corresponding average sentiments.

Panel A: company-specific news days by year

| Year | No. of stocks | Market cap ($ bln) | Book-to-market | Avg. number of news days per stock Positive | Neutral | Negative |
|---|---|---|---|---|---|---|
| 2003 | 780 | 2.74 | 0.54 | 8.71 | 10.56 | 10.84 |
| 2004 | 806 | 3.57 | 0.46 | 8.91 | 10.63 | 9.55 |
| 2005 | 848 | 4.18 | 0.42 | 10.15 | 14.34 | 9.25 |
| 2006 | 882 | 4.67 | 0.39 | 11.18 | 15.44 | 9.49 |
| 2007 | 913 | 5.46 | 0.36 | 11.43 | 18.61 | 9.63 |
| 2008 | 933 | 4.43 | 0.43 | 11.42 | 20.16 | 13.39 |
| 2009 | 945 | 3.46 | 0.59 | 12.19 | 21.93 | 12.88 |
| 2010 (till Aug.) | 946 | 4.63 | 0.44 | 9.03 | 16.30 | 8.29 |

Panel B: average number of news days and average sentiment by news specification

| News | Avg. number of news days | | | | Avg. sentiment | | |
|---|---|---|---|---|---|---|---|
| specification | Total | Positive | Neutral | Negative | Positive | Neutral | Negative |
| All news | 33.7 | 9.6 | 15.1 | 9.0 | 0.61 | 0.03 | -0.63 |
| Headlines | 20.9 | 6.6 | 9.3 | 5.0 | 0.65 | 0.04 | -0.68 |
| Non-headlines | 12.8 | 4.1 | 3.9 | 4.8 | 0.61 | 0.03 | -0.65 |
| Extreme sentiment | 13.2 | 6.5 | | 6.7 | 0.68 | | -0.70 |
| Headlines + extreme sentiment | 9.4 | 5.2 | | 4.1 | 0.71 | | -0.74 |

## Table II: Short term news impact: whole sample

The table presents average daily returns around news arrivals computed as the difference to the average non-news daily return for all stocks in the sample (news sensitivities). The news sensitivities are estimated from a panel regression of daily returns on the news dummy, for positive, neutral or negative news respectively, allowing for arbitrary clustering of standard errors along the firm and time dimensions. Robust *t*-statistics in parentheses. Contemporaneous market return is included as a control variable (not reported). Day 0 is the news arrival day.
Significance levels: * - 0.05, ** - 0.01

| | | | By sentiment | | | By sentiment intensity | |
|---|---|---|---|---|---|---|---|
| News specification | Day | All news days | Positive news days | Neutral news days | Negative news days | Extreme positive | Extreme negative |
| All news | -3 | -0.01% * | -0.03% ** | 0.00% | 0.01% | -0.02% | 0.00% |
| | | ( -2.38 ) | ( -3.77 ) | ( 0.45 ) | ( 0.59 ) | ( -1.80 ) | ( 0.38 ) |
| | | 0.00% | 0.01% | 0.00% | 0.00% | 0.01% | 0.01% |
| | | ( -0.35 ) | ( 1.14 ) | ( -0.52 ) | ( -0.03 ) | ( 1.39 ) | ( 0.58 ) |
| | | 0.01% | 0.06% ** | 0.02% ** | -0.09% ** | 0.06% ** | -0.08% ** |
| | | ( 0.76 ) | ( 5.92 ) | ( 2.87 ) | ( -7.58 ) | ( 4.79 ) | ( -6.01 ) |
| | 0 | 0.12% ** | 0.45% ** | 0.10% ** | -0.41% ** | 0.43% ** | -0.35% ** |
| | | ( 11.25 ) | ( 21.36 ) | ( 9.42 ) | ( -21.11 ) | ( 20.01 ) | ( -17.12 ) |
| | | -0.01% | -0.01% | -0.01% | 0.01% | -0.01% | -0.02% |
| | | ( -0.84 ) | ( -1.22 ) | ( -0.85 ) | ( 0.81 ) | ( -0.81 ) | ( -1.31 ) |
| | | -0.03% ** | -0.03% ** | -0.02% * | 0.03% ** | -0.02% | 0.03% * |
| | | ( -4.31 ) | ( -3.42 ) | ( -2.49 ) | ( 2.92 ) | ( -1.55 ) | ( 2.37 ) |
| | +3 | 0.01% | -0.01% | -0.01% | 0.06% ** | 0.00% | 0.05% ** |
| | | ( 1.12 ) | ( -1.50 ) | ( -1.74 ) | ( 6.02 ) | ( 0.46 ) | ( 4.70 ) |
| Headlines | -3 | -0.01% ** | -0.01% | -0.01% | 0.00% | 0.00% | 0.01% |
| | | ( -2.83 ) | ( -1.38 ) | ( -0.81 ) | ( 0.01 ) | ( 0.38 ) | ( 0.55 ) |
| | | 0.00% | 0.02% * | 0.00% | -0.02% | 0.03% ** | -0.02% |
| | | ( -0.90 ) | ( 2.15 ) | ( 0.16 ) | ( -1.26 ) | ( 2.67 ) | ( -1.28 ) |
| | | 0.00% | 0.12% ** | 0.01% | -0.14% ** | 0.15% ** | -0.15% ** |
| | | ( -0.71 ) | ( 9.77 ) | ( 1.68 ) | ( -7.70 ) | ( 11.00 ) | ( -7.60 ) |
| | 0 | 0.10% ** | 0.64% ** | 0.12% ** | -0.69% ** | 0.70% ** | -0.68% ** |
| | | ( 11.79 ) | ( 22.37 ) | ( 8.02 ) | ( -21.64 ) | ( 22.86 ) | ( -18.09 ) |
| | | -0.01% * | 0.00% | 0.01% | -0.01% | 0.00% | -0.02% |
| | | ( -2.10 ) | ( 0.06 ) | ( 1.00 ) | ( -1.00 ) | ( 0.01 ) | ( -1.37 ) |
| | | -0.02% ** | -0.03% * | -0.02% | -0.01% | -0.03% * | -0.02% |
| | | ( -3.23 ) | ( -2.57 ) | ( -1.93 ) | ( -0.67 ) | ( -2.23 ) | ( -1.15 ) |
| | -3 | 0.00% | -0.01% | 0.00% | 0.05% ** | -0.01% | 0.05% ** |
| | | ( 0.72 ) | ( -1.29 ) | ( 0.47 ) | ( 3.71 ) | ( -0.57 ) | ( 3.24 ) |
| Non-headlines | -3 | 0.00% | -0.01% | 0.03% ** | -0.02% | | |
| | | ( -0.19 ) | ( -0.93 ) | ( 2.64 ) | ( -1.28 ) | | |
| | | 0.00% | 0.00% | 0.02% | -0.02% | | |
| | | ( 0.02 ) | ( 0.34 ) | ( 1.57 ) | ( -1.22 ) | | |
| | | -0.01% | 0.02% | 0.01% | -0.04% ** | | |
| | | ( -1.10 ) | ( 1.34 ) | ( 0.38 ) | ( -2.94 ) | | |
| | 0 | 0.03% ** | 0.17% ** | 0.06% ** | -0.17% ** | | |
| | | ( 3.32 ) | ( 10.44 ) | ( 4.03 ) | ( -10.51 ) | | |
| | | -0.01% | -0.01% | -0.01% | 0.00% | | |
| | | ( -1.07 ) | ( -0.71 ) | ( -0.45 ) | ( -0.22 ) | | |
| | | 0.02% | 0.00% | 0.01% | 0.04% ** | | |
| | | ( 1.86 ) | ( -0.14 ) | ( 0.76 ) | ( 2.75 ) | | |
| | | 0.00% | 0.02% | -0.03% * | 0.02% | | |
| | +3 | ( 0.45 ) | ( 1.23 ) | ( -2.08 ) | ( 1.51 ) | | |

Table III: Summary statistics: cross section

The table presents summary statistics for quintiles of stocks grouped by different company characteristics. Each panel gives gives the average value of the respective characteristic for each quintile as well as the average number of news days per year and per stock in that quintile, broken down by news sentiment.

| | Mean | Number of news days per stock | | |
|---|---|---|---|---|
| Group | value | Positive | Neutral | Negative |
| Panel A: news coverage | | | | |
| 1 (low) | 5.2 | 1.7 | 2.3 | 1.2 |
| 2 | 11.1 | 3.7 | 4.8 | 2.6 |
| 3 | 16.2 | 5.2 | 7.4 | 3.6 |
| 4 | 24.8 | 7.7 | 11.3 | 5.8 |
| 5 (high) | 46.9 | 14.1 | 20.2 | 12.6 |
| Panel B: market capitalization ($ bln) | | | | |
| 1 (small) | 1.26 | 3 | 4.5 | 2.9 |
| 2 | 2.45 | 4.1 | 5.8 | 3.2 |
| 3 | 4.11 | 4.9 | 7 | 4 |
| 4 | 8.66 | 6.9 | 10 | 5.3 |
| 5 (large) | 45.8 | 12.6 | 17 | 10 |
| Panel C: daily trading volume ($ mln) | | | | |
| 1 (low) | 9.4 | 2.5 | 3.6 | 1.9 |
| 2 | 21.2 | 3.6 | 5.3 | 2.7 |
| 3 | 40.7 | 4.9 | 7 | 3.7 |
| 4 | 79.7 | 6.8 | 9.6 | 5.3 |
| 5 (high) | 273.7 | 11.8 | 15.5 | 9.6 |
| Panel D: book-to-market ratio | | | | |
| 1 (growth) | 0.08 | 7.3 | 9.1 | 5 |
| 2 | 0.29 | 6.7 | 8.8 | 5.1 |
| 3 | 0.41 | 6.3 | 8.5 | 4.8 |
| 4 | 0.56 | 6.3 | 9.2 | 5.2 |
| 5 (value) | 0.88 | 5.5 | 9 | 5.7 |
| Panel E: total volatility (% p.a.) | | | | |
| 1 (low) | 22.8% | 7 | 9.5 | 4.7 |
| 2 | 32.3% | 6.5 | 8.9 | 4.6 |
| 3 | 36.1% | 6.4 | 8.7 | 5 |
| 4 | 45.7% | 6 | 8.4 | 5 |
| 5 (high) | 62.6% | 6 | 9.2 | 6.2 |

Table IV: Short term news impact: cross section

The table presents average news sensitivities in the cross-section, using only headline news items. News sensitivities are the difference of the average return on news days to the average no-news daily return. Stocks are grouped into quintiles according to various cross-sectional features and all quintiles are updated annually. The news sensitivities are estimated separately for each quintile from a panel regression of daily returns on the news day dummy, for positive, neutral or negative news days respectively, allowing for arbitrary clustering of standard errors along the firm and time dimensions. Robust $t$-statistics in parentheses. Contemporaneous market return is included as a control variable (not reported).
Significance levels: * - 0.05, ** - 0.01

| | News day sentiment | | | | | |
|---|---|---|---|---|---|---|
| Group | Positive | | Neutral | | Negative | |
| Whole sample | 0.64% ** | ( 22.37 ) | 0.12% ** | ( 8.02 ) | -0.69% | ( -21.64 ) |
| Panel A: news coverage | | | | | | |
| 1 (low) | 1.45% ** | ( 14.9 ) | 0.24% ** | ( 3.18 ) | -1.55% ** | ( -10.9 ) |
| 2 | 1.17% ** | ( 15.1 ) | 0.24% ** | ( 3.80 ) | -1.17% ** | ( -10.4 ) |
| 3 | 1.04% ** | ( 17.2 ) | 0.17% ** | ( 3.84 ) | -1.23% ** | ( -14.2 ) |
| 4 | 0.77% ** | ( 17.9 ) | 0.10% ** | ( 3.19 ) | -0.89% ** | ( -14.0 ) |
| 5 (high) | 0.35% ** | ( 11.9 ) | 0.05% ** | ( 2.64 ) | -0.39% ** | ( -10.4 ) |
| Panel B: market capitalization | | | | | | |
| 1 (small) | 1.65% ** | ( 17.21 ) | 0.29% ** | ( 3.79 ) | -1.36% ** | ( -10.2 ) |
| 2 | 1.10% ** | ( 14.0 ) | 0.26% ** | ( 3.95 ) | -1.23% ** | ( -12.7 ) |
| 3 | 0.88% ** | ( 16.4 ) | 0.11% * | ( 2.37 ) | -1.09% ** | ( -13.7 ) |
| 4 | 0.71% ** | ( 16.1 ) | 0.09% ** | ( 2.76 ) | -0.73% ** | ( -11.4 ) |
| 5 (large) | 0.36% ** | ( 12.4 ) | 0.05% ** | ( 2.78 ) | -0.40% ** | ( -10.5 ) |
| Panel C: trading volume | | | | | | |
| 1 (low) | 1.20% ** | ( 13.8 ) | 0.29% ** | ( 4.37 ) | -1.07% ** | ( -8.33 ) |
| 2 | 1.00% ** | ( 14.7 ) | 0.22% ** | ( 3.85 ) | -1.19% ** | ( -12.7 ) |
| 3 | 0.92% ** | ( 15.7 ) | 0.12% * | ( 2.39 ) | -1.04% ** | ( -12.7 ) |
| 4 | 0.76% ** | ( 15.2 ) | 0.09% ** | ( 3.14 ) | -0.81% ** | ( -12.2 ) |
| 5 (high) | 0.44% ** | ( 11.8 ) | 0.05% ** | ( 2.62 ) | -0.48% ** | ( -10.3 ) |
| Panel D: book-to-market | | | | | | |
| 1 (growth) | 0.87% ** | ( 12.4 ) | 0.15% ** | ( 4.29 ) | -0.85% ** | ( -9.86 ) |
| 2 | 0.80% ** | ( 13.5 ) | 0.07% * | ( 2.06 ) | -0.80% ** | ( -10.3 ) |
| 3 | 0.72% ** | ( 14.9 ) | 0.10% ** | ( 2.92 ) | -0.73% ** | ( -12.1 ) |
| 4 | 0.58% ** | ( 12.4 ) | 0.06% * | ( 2.07 ) | -0.69% ** | ( -9.38 ) |
| 5 (value) | 0.51% ** | ( 9.31 ) | 0.17% ** | ( 3.79 ) | -0.68% ** | ( -8.63 ) |
| Panel E: total volatility | | | | | | |
| 1 (low) | 0.29% ** | ( 11.9 ) | 0.06% ** | ( 3.43 ) | -0.31% ** | ( -9.28 ) |
| 2 | 0.49% ** | ( 15.2 ) | 0.13% ** | ( 5.43 ) | -0.51% ** | ( -9.61 ) |
| 3 | 0.75% ** | ( 18.6 ) | 0.09% ** | ( 3.08 ) | -0.68% ** | ( -13.2 ) |
| 4 | 0.85% ** | ( 16.8 ) | 0.09% * | ( 2.17 ) | -0.90% ** | ( -13.9 ) |
| 5 (high) | 1.27% ** | ( 14.2 ) | 0.18% ** | ( 2.81 ) | -1.26% ** | ( -11.2 ) |

## Table V: Short term news impact: industry comparison

The table presents average daily returns on news days computed as the difference to the average non-news daily return for stocks grouped into industries based on the IBC classification. The news-related returns are estimated from a panel regression of daily returns on the news dummy, for positive, neutral or negative news respectively, allowing for arbitrary clustering of standard errors along the firm and time dimensions. Robust $t$-statistics in parentheses. Contemporaneous market return is included as a control variable (not reported).
Significance levels: * - 0.05, ** - 0.01

| | Basic Materials | Consumer Goods | Consumer Services | Financials | Healthcare | Industrials | Oil and Gas | Technology | Telecom | Utilities |
|---|---|---|---|---|---|---|---|---|---|---|
| **Panel A: summary statistics** | | | | | | | | | | |
| No. of stocks | 51 | 87 | 129 | 179 | 89 | 151 | 88 | 103 | 15 | 55 |
| Mean capitalization ($ bln) | 6.9 | 12.9 | 13.0 | 11.9 | 15.6 | 10.1 | 14.2 | 17.6 | 23.2 | 7.4 |
| Number of news days | | | | | | | | | | |
| Positive | 5.7 | 6.2 | 6.4 | 5.4 | 5.5 | 5.6 | 6.1 | 7.1 | 7.5 | 7.4 |
| Neutral | 7.7 | 9.6 | 9.1 | 8.7 | 8.5 | 7.8 | 8 | 8.5 | 9.5 | 8.9 |
| Negative | 4.1 | 5.3 | 5.3 | 4.9 | 5.2 | 3.6 | 5 | 5.4 | 6.4 | 4.3 |
| **Panel B: news sensitivities** | | | | | | | | | | |
| Positive | 0.87% ** | 0.63% ** | 0.89% ** | 0.40% ** | 1.02% ** | 0.67% ** | 0.52% ** | 1.10% ** | 0.30% ** | 0.24% ** |
| | ( 7.32 ) | ( 7.22 ) | ( 11.0 ) | ( 6.56 ) | ( 7.52 ) | ( 9.24 ) | ( 6.38 ) | ( 9.23 ) | ( 2.73 ) | ( 4.22 ) |
| Neutral | -0.01% | 0.16% ** | 0.11% ** | 0.12% ** | 0.12% * | 0.13% ** | 0.13% * | 0.08% | 0.10% | 0.04% |
| | ( -0.11 ) | ( 2.79 ) | ( 2.42 ) | ( 2.08 ) | ( 1.97 ) | ( 3.37 ) | ( 2.10 ) | ( 1.53 ) | ( 0.93 ) | ( 0.99 ) |
| Negative | -0.93% ** | -0.66% ** | -0.79% ** | -0.73% ** | -0.81% ** | -0.76% ** | -0.28% ** | -1.11% ** | -0.49% ** | -0.28% * |
| | ( -7.63 ) | ( -5.82 ) | ( -8.39 ) | ( -6.62 ) | ( -6.73 ) | ( -8.13 ) | ( -2.11 ) | ( -8.14 ) | ( -3.41 ) | ( -4.10 ) |

## Table VI: Risk factor regressions

In this table monthly returns of a portfolio going long the top quintile of most news-sensitive stocks and short the bottom quintile of least news-sensitive stocks are regressed on common risk factors. News sensitivity is defined over the preceding rolling 12-month window. Both the long and short positions are equally weighted, and held for 1 month after portfolio formation. Portfolios are rebalanced monthly. Newey-West $t$-statistics in parentheses, * and ** indicate statistical significance at the 10% and 5% level respectively.

| | Time series mean | CAPM | Fama-French three-factor | Carhart four-factor | Pastor-Stambaugh five-factor |
|---|---|---|---|---|---|
| Intercept | 0.95 ** | 0.89 ** | 0.85 ** | 0.84 ** | 0.84** |
| | ( 2.67 ) | ( 3.49 ) | ( 3.69 ) | ( 3.82 ) | (3.96) |
| Mkt-Rf | - | 0.47 ** | 0.39 ** | 0.35 ** | 0.35** |
| | | ( 10.3 ) | ( 10.2 ) | ( 9.22 ) | (9.55) |
| SMB | - | - | 0.49 ** | 0.51 ** | 0.50** |
| | | | ( 4.63 ) | ( 5.00 ) | (4.56) |
| HML | - | - | -0.10 | -0.17 * | -0.19 |
| | | | ( -1.07 ) | ( -1.75 ) | (-1.40) |
| UMD | - | - | - | -0.12 ** | -0.12** |
| | | | | ( -2.69 ) | (-2.34) |
| LIQ_V | - | - | - | - | -0.03 |
| | | | | | -0.62 |
| No. obs. | 80 | 80 | 80 | 80 | 80 |
| Adj. $R^2$ | - | 0.4 | 0.55 | 0.67 | 0.67 |

Table VII: Different breakpoints and holding periods

In this table monthly returns of a portfolio going long the most news-sensitive stocks and short the least news-sensitive stocks are evaluated with respect to different breakpoints and holding periods. News sensitivity is defined over the preceding rolling 12-month window. Both the long and short positions are equally weighted. The numbers are the times series mean and intercepts (alphas) from common risk factors models. Newey-West $t$-statistics in parentheses, * and ** indicate statistical significance at the 10% and 5% level respectively.

| Holding period | Time series mean | CAPM intercept | Fama-French intercept | Carhart intercept | Pastor-Stambaugh intercept |
|---|---|---|---|---|---|
| | | | Panel A: quintile breakpoints | | |
| 1 month | 0.95 ** | 0.89 ** | 0.85 ** | 0.84 ** | 0.84 ** |
| | ( 2.67 ) | ( 3.49 ) | ( 3.69 ) | ( 3.81 ) | ( 3.96 ) |
| 3 months | 0.72 ** | 0.67 ** | 0.62 ** | 0.62 ** | 0.57 ** |
| | ( 2.11 ) | ( 2.74 ) | ( 2.98 ) | ( 2.97 ) | ( 3.13 ) |
| 6 months | 0.65 * | 0.59 ** | 0.56 ** | 0.56 ** | 0.50 ** |
| | ( 1.86 ) | ( 2.37 ) | ( 2.71 ) | ( 2.69 ) | ( 2.83 ) |
| 12 months | 0.52 | 0.47 * | 0.46 ** | 0.46 ** | 0.36 ** |
| | ( 1.51 ) | ( 1.81 ) | ( 2.17 ) | ( 2.17 ) | ( 2.08 ) |
| | | | Panel B: tercile breakpoints | | |
| 1 month | 0.67 ** | 0.66 ** | 0.63 ** | 0.62 ** | 0.60 ** |
| | ( 2.47 ) | ( 2.5 ) | ( 3.3 ) | ( 3.87 ) | ( 3.99 ) |
| 3 months | 0.57 ** | 0.53 ** | 0.50 ** | 0.50 ** | 0.47 ** |
| | ( 2.05 ) | ( 2.73 ) | ( 3.68 ) | ( 3.63 ) | ( 4.00 ) |
| 6 months | 0.52 * | 0.48 ** | 0.46 ** | 0.46 ** | 0.41 ** |
| | ( 1.85 ) | ( 2.48 ) | ( 3.29 ) | ( 3.22 ) | ( 3.31 ) |
| 12 months | 0.45 | 0.41 ** | 0.40 ** | 0.40 ** | 0.33 ** |
| | ( 1.60 ) | ( 2.25 ) | ( 2.96 ) | ( 3.04 ) | ( 2.72 ) |
| | | | Panel C: median breakpoints | | |
| 1 month | 0.53 ** | 0.49 ** | 0.46 ** | 0.46 ** | 0.44 ** |
| | ( 2.46 ) | ( 2.59 ) | ( 3.52 ) | ( 4.32 ) | ( 4.26 ) |
| 3 months | 0.42 ** | 0.38 ** | 0.35 ** | 0.35 ** | 0.35 ** |
| | ( 2.03 ) | ( 2.98 ) | ( 4.01 ) | ( 4.15 ) | ( 4.36 ) |
| 6 months | 0.36 * | 0.32 ** | 0.30 ** | 0.30 ** | 0.28 ** |
| | ( 1.77 ) | ( 2.52 ) | ( 3.39 ) | ( 3.34 ) | ( 3.32 ) |
| 12 months | 0.30 | 0.26 ** | 0.24 ** | 0.24 ** | 0.20 ** |
| | ( 1.51 ) | ( 2.20 ) | ( 3.00 ) | ( 3.04 ) | ( 2.46 ) |

Table VIII: Explanations of the news sensitivity effect

The table examines the news sensitivity premium in subsamples of stocks sorted on various firm characteristics (Panels A to C) and proxies for possible explanations: liquidity (Panel D), likelihood of bad news (Panels E and F) and uncertainty (Panel G). The numbers are the times series mean and intercepts (alphas) from common risk factors models. Newey-West $t$-statistics in parentheses, * and ** indicate statistical significance at the 10% and 5% level respectively.

| | Time series mean | CAPM intercept | Fama-French intercept | Carhart intercept | Pastor-Stambaugh intercept |
|---|---|---|---|---|---|
| Panel A: By size | | | | | |
| Small | 0.33 | 0.31 | 0.33 | 0.32 | 0.35 * |
| | ( 1.60 ) | ( 1.18 ) | ( 1.54 ) | ( 1.61 ) | ( 1.79 ) |
| Big | 0.28 | 0.24 ** | 0.24 ** | 0.24 ** | 0.25 ** |
| | ( 1.23 ) | ( 2.05 ) | ( 2.84 ) | ( 2.59 ) | ( 2.97 ) |
| Panel B: By book-to-market | | | | | |
| Low | 0.40 * | 0.37 ** | 0.38 ** | 0.38 ** | 0.33 ** |
| | ( 1.83 ) | ( 2.50 ) | ( 3.76 ) | ( 3.63 ) | ( 2.93 ) |
| High | 0.65 ** | 0.60 ** | 0.55 ** | 0.54 ** | 0.56 ** |
| | ( 2.53 ) | ( 2.69 ) | ( 2.84 ) | ( 4.06 ) | ( 4.56 ) |
| Panel C: By past 12-month return | | | | | |
| Low | 0.63 ** | 0.60 ** | 0.55 ** | 0.55 ** | 0.56 ** |
| | ( 2.56 ) | ( 2.23 ) | ( 2.52 ) | ( 3.24 ) | ( 2.91 ) |
| High | 0.25 | 0.22 | 0.24 | 0.24 ** | 0.20 |
| | ( 1.18 ) | ( 1.53 ) | ( 1.63 ) | ( 2.01 ) | ( 1.59 ) |
| Panel D: By average trading volume | | | | | |
| Low | 0.53 ** | 0.50 ** | 0.48 ** | 0.48 ** | 0.54 ** |
| | ( 2.32 ) | ( 2.04 ) | ( 2.46 ) | ( 2.68 ) | ( 3.05 ) |
| High | 0.30 | 0.25 * | 0.26 ** | 0.25 ** | 0.19 ** |
| | ( 1.09 ) | ( 1.89 ) | ( 2.75 ) | ( 2.90 ) | ( 2.10 ) |
| Panel E: By ratio of positive to negative news days | | | | | |
| Low | 0.58 ** | 0.54 ** | 0.50 ** | 0.50 ** | 0.48 ** |
| | ( 2.45 ) | ( 2.75 ) | ( 2.46 ) | ( 3.57 ) | ( 3.27 ) |
| High | 0.52 ** | 0.48 ** | 0.45 ** | 0.45 ** | 0.46 ** |
| | ( 2.11 ) | ( 2.26 ) | ( 2.75 ) | ( 3.40 ) | ( 3.60 ) |
| Panel F: By fraction of positive news days in total | | | | | |
| Low | 0.53 ** | 0.49 ** | 0.47 ** | 0.46 ** | 0.44 ** |
| | ( 2.39 ) | ( 2.60 ) | ( 3.01 ) | ( 3.37 ) | ( 2.98 ) |
| High | 0.51 ** | 0.47 ** | 0.44 ** | 0.44 ** | 0.45 ** |
| | ( 2.16 ) | ( 2.17 ) | ( 2.93 ) | ( 3.54 ) | ( 3.87 ) |
| Panel G: By idiosyncratic volatility | | | | | |
| Low | 0.32 ** | 0.31 ** | 0.30 ** | 0.29 ** | 0.24 ** |
| | ( 1.97 ) | ( 2.76 ) | ( 3.00 ) | ( 2.98 ) | ( 2.51 ) |
| High | 0.51 ** | 0.50 ** | 0.48 ** | 0.48 ** | 0.51 ** |
| | ( 2.96 ) | ( 2.12 ) | ( 2.47 ) | ( 3.10 ) | ( 2.82 ) |

Table IX: Robustness checks: part 1

The first robustness check covered in this table (Panel A) relates to the sensitivity of the results to including news arriving after the previously imposed threshold of two hours before the close. Such news is counted for the trading next day. The second check examines the sensitivity of the parameter estimates on news days to changing the length of the event window (Panel B). Only headline news items are used in both cases.

Panel A: results including afternoon news

| Day | Positive news days | | Neutral news days | | Negative news days | |
|---|---|---|---|---|---|---|
| -3 | 0.01% | ( 1.44 ) | 0.01% | ( 1.51 ) | -0.04% ** | ( -3.81 ) |
| | 0.04% ** | ( 4.01 ) | -0.01% | ( -0.85 ) | -0.05% ** | ( -3.92 ) |
| | 0.11% ** | ( 10.49 ) | 0.00% | ( 0.54 ) | -0.16% ** | ( -10.09 ) |
| 0 | 0.57% ** | ( 26.19 ) | 0.09% ** | ( 8.37 ) | -0.56% ** | ( -21.37 ) |
| | 0.08% ** | ( 7.34 ) | 0.05% ** | ( 4.31 ) | -0.09% ** | ( -5.88 ) |
| | -0.02% | ( -1.66 ) | -0.02% ** | ( -2.84 ) | -0.05% ** | ( -4.11 ) |
| +3 | 0.01% | ( 0.95 ) | 0.00% | ( 0.53 ) | 0.02% | ( 1.51 ) |

Panel B: sensitivity to the length of the news window

| News | Length of event window | | | | | |
|---|---|---|---|---|---|---|
| days | 5-days | 4-days | 3-days | 2-days | 1-day | None |
| Positive | 0.69% ** | 0.69% ** | 0.68% ** | 0.68% ** | 0.68% ** | 0.68% ** |
| | ( 25.83 ) | ( 25.82 ) | ( 25.73 ) | ( 25.67 ) | ( 25.56 ) | ( 25.49 ) |
| Neutral | 0.13% ** | 0.13% ** | 0.12% ** | 0.12% ** | 0.12% ** | 0.12%** |
| | ( 9.28 ) | ( 9.28 ) | ( 9.15 ) | ( 9.15 ) | ( 8.97 ) | ( 8.86 ) |
| Negative | -0.68% ** | -0.68% ** | -0.69% ** | -0.69% ** | -0.69% ** | -0.70% ** |
| | ( -20.92 ) | ( -21.01 ) | ( -21.15 ) | ( -21.26 ) | ( -21.45 ) | ( -21.86 ) |

# Table X: Robustness checks: part 2

The table presents average daily returns around news arrivals computed as the difference to the average non-news daily return, estimated from a panel regression of daily returns on the news dummy, for positive, neutral or negative news days respectively, and including interactions with the following control variables:

- $Earn_{i,t-day}$ - dummy variables for earnings announcements event window (+/- 3 days)
- $r_{i,t-(day-1)}$ - previous day return for all days in the event window
- $r_{M,t-day}$ - contemporaneous market return for all days in the event window

The first section of the table contains parameter estimates on the news day dummies (news sensitivities), while sections 2-4 contain parameter estimates on the interactions between the news day dummy and the respective control variable. Robust $t$-statistics in parentheses are computed allowing for arbitrary clustering of standard errors along the firm and time dimensions. Day 0 is the news arrival day. Only headline news items are used.
Significance levels: * - 0.05, ** - 0.01

| | Day | Positive news | | Neutral news | | Negative news | |
|---|---|---|---|---|---|---|---|
| $\delta_{i,t-day}$ | -3 | 0.00% | ( -0.11 ) | 0.02% ** | ( 2.64 ) | -0.01% | ( -1.20 ) |
| | | 0.03% ** | ( 3.83 ) | 0.00% | ( 0.70 ) | -0.03% ** | ( -2.90 ) |
| | | 0.05% ** | ( 5.77 ) | 0.02% ** | ( 2.86 ) | -0.08% ** | ( -7.49 ) |
| | 0 | 0.38% ** | ( 21.77 ) | 0.09% ** | ( 8.91 ) | -0.29% ** | ( -16.63 ) |
| | | 0.02% ** | ( 2.60 ) | 0.00% | ( 0.58 ) | -0.04% ** | ( -4.10 ) |
| | | 0.00% | ( -0.17 ) | -0.01% * | ( -2.14 ) | -0.02% * | ( -2.52 ) |
| | +3 | 0.00% | ( -0.58 ) | -0.01% | ( -1.76 ) | -0.01% | ( -0.65 ) |
| $Earn_{i,t-day}$ | -3 | 0.20% | ( 0.92 ) | -0.12% | ( -0.76 ) | 0.00% | ( -0.01 ) |
| | | 0.20% | ( 1.07 ) | -0.03% | ( -0.21 ) | -0.38% | ( -1.93 ) |
| | | 0.67% ** | ( 4.88 ) | -0.19% | ( -1.56 ) | -1.00% ** | ( -6.32 ) |
| | 0 | 0.94% ** | ( 7.63 ) | -0.18% * | ( -2.19 ) | -1.36% ** | ( -8.29 ) |
| | | 0.00% | ( -0.02 ) | -0.02% | ( -0.16 ) | -0.17% | ( -0.88 ) |
| | | -0.13% | ( -0.71 ) | -0.24% | ( -1.63 ) | 0.22% | ( 1.19 ) |
| | +3 | 0.17% | ( 1.01 ) | -0.10% | ( -0.59 ) | 0.28% | ( 1.28 ) |
| $r_{i,t-(day-1)}$ | -3 | 0.011 | ( 1.58 ) | 0.006 | ( 0.76 ) | 0.000 | ( 0.06 ) |
| | | -0.010 | ( -1.29 ) | -0.002 | ( -0.25 ) | 0.017 * | ( 2.00 ) |
| | | 0.000 | ( 0.04 ) | 0.007 | ( 1.00 ) | 0.011 | ( 1.08 ) |
| | 0 | -0.023 ** | ( -2.90 ) | -0.019 * | ( -2.09 ) | 0.013 | ( 1.49 ) |
| | | 0.007 | ( 1.28 ) | 0.016 ** | ( 3.09 ) | 0.011 | ( 1.43 ) |
| | | 0.011 | ( 1.54 ) | -0.002 | ( -0.29 ) | 0.010 | ( 1.05 ) |
| | +3 | -0.010 | ( -1.26 ) | 0.018 ** | ( 2.76 ) | 0.002 | ( 0.22 ) |
| $r_{M,t-day}$ | -3 | 0.010 | ( 0.77 ) | 0.017 | ( 1.42 ) | 0.044 ** | ( 3.42 ) |
| | | 0.013 | ( 1.07 ) | -0.015 | ( -1.32 ) | 0.056 ** | ( 3.78 ) |
| | | 0.040 ** | ( 3.31 ) | 0.026 * | ( 2.34 ) | 0.057 ** | ( 4.24 ) |
| | 0 | 0.039 ** | ( 3.02 ) | 0.031 * | ( 2.48 ) | 0.110 ** | ( 7.26 ) |
| | | -0.010 | ( -0.74 ) | -0.031 ** | ( -3.02 ) | 0.055 ** | ( 3.50 ) |
| | | -0.021 | ( -1.79 ) | -0.033 ** | ( -2.99 ) | 0.039 ** | ( 2.97 ) |
| | +3 | 0.004 | ( 0.29 ) | -0.018 | ( -1.90 ) | 0.039 ** | ( 2.73 ) |

## Table XI: Impact of the financial crisis

Panel A of the table presents the estimates of news sensitivities outside ($\delta$) and during the financial crisis ($\delta \cdot cris$), defined as the period Jan 2008 - Dec 2009.

| | | | | | Extreme sentiment | |
|---|---|---|---|---|---|---|
| News days | | All news | Headlines | Non-headlines | All news | Headlines |
| Positive | $\delta$ | 0.44% ** | 0.64% ** | 0.17% ** | 0.44% ** | 0.69% ** |
| | | (21.2) | (22.4) | (10.6) | (20.1) | (22.5) |
| | $\delta \cdot cris$ | 0.09% ** | 0.11% ** | -0.01% | -0.04% | 0.09% ** |
| | | ( 2.95 ) | ( 3.39 ) | ( -0.17 ) | ( -1.07 ) | ( 2.44 ) |
| Neutral | $\delta$ | 0.10% ** | 0.12% ** | 0.06% ** | | |
| | | (9.41) | (8.04) | (4.23) | | |
| | $\delta \cdot cris$ | -0.03% | 0.00% | -0.02% | | |
| | | ( -1.46 ) | ( -0.12 ) | ( -0.42 ) | | |
| Negative | $\delta$ | -0.36% ** | -0.63% ** | -0.13% ** | -0.35% ** | -0.70% ** |
| | | (-18.4) | (-18.5) | (-8.69) | (-17.2) | (-18.7) |
| | $\delta \cdot cris$ | -0.14% ** | -0.17% ** | -0.11% ** | -0.06% | -0.20% ** |
| | | ( -4.04 ) | ( -3.55 ) | ( -3.06 ) | ( -1.64 ) | ( -3.69 ) |