Imperial College London
Business School

Essays in asset pricing: information and stock markets

Rastislav Molnar

Submitted in partial fulfillment of the requirements for the degree of
Master of Philosophy of Imperial College London

Abstract

The research in this thesis has been split into two chapters. Both deal with textual information and its impact on stock markets. Chapter 1 contains a discussion of the development of a database of textual news that are processed and creates a number of news characteristic variables to measure sentiment, novelty, relevance, ambiguity as well as the topics and entities mentioned in the news. The main contribution of this part is the database itself. We briefly show the descriptive statistics of the data set and also discuss the potential use of data.

The second chapter, based on a draft written jointly with Pedro Saffi, builds on the author's work in the first chapter by testing news variables around earnings announcements. It finds that they indeed have an effect on abnormal returns after them. It focuses on post-earnings announcement drift, trying to evaluate the information value of news around these scheduled events. We look, in particular, at whether our news data can shed some light on information asymmetry and heterogeneous beliefs.

Contents

List of Figures

List of Tables

Declaration of Originality

'I herewith certify that this thesis constitutes my own work and that all material, which is not my own work, has been properly acknowledged'

Rastislav Molnar

Copyright

Rastislav Molnar

Acknowledgments

I would like to express my sincere gratitude to all that supported my journey at Imperial College London. I would like to thank both my supervisors for constructive criticism on my research as well as their support through whole time. My family for being supportive during all those years as well as to my friends in London, Oxford, Zurich and in Bratislava.

Namely:

- Prof Robert Kosowski (supervisor)
- Prof Walter Distaso (supervisor)
- Dr. Pedro Saffi
- Prof James Sefton
- Dr Eugen Molnar
- Davide Benedetti
- Lukas Platinsky
- Peter Krajnak
- Dr Martin Klauco
- Prof Jozef Komornik

Thank you all.
Ras

Chapter 1

News and Stock Markets - A News Database

This chapter on the news database is the sole work of the author.

I. Introduction

We understand that research in finance is driven by next-generation models or the application of more or less exotic methods. This is especially interesting for today's highly competitive, high-frequency trading. We believe that data plays an important role, as researchers try to find new data sources or look at old data from a new perspective. The goal of this research is to look at alternative data, created in such a manner that it can compete with more robust and pricey solutions, while being available at no cost and having a methodology used to derive variables of interest that is readily understandable.

In this research we focus on textual news. Researchers in the past did one of two things: they either created their own variables from news datasets available to them or they used proprietary tools that could directly provide them with news characteristic variables. The first approach is represented by Antweiler and Frank (2004), Tetlock (2007), Fang and Peress (2009) , Tetlock (2011) and Garcia (2013), while the latter is represented by Groß-Klußmann and Hautsch (2011), von Beschwitz, Keim, and Massa (2013), Boudoukh, Feldman, Kogan, and Richardson (2013) , Kelley and Tetlock (2013) and Smales (2014). We have decided to follow the first approach and improve upon the aforementioned papers. We created our own news database for two reasons: first, we wanted to use a wide range of news characteristic variables, and this could not be easily done using proprietary software. Second, we wanted our methodology to be straightforward and our news characteristic variables to be based on methods that we can explain and describe. Second, this requirement would not be fulfilled by proprietary software, as companies are often secretive about the methods underlying the variables they create and often limit the description to 'machine learning' or 'machine learning enhanced' or a similar, relatively generic description.

In order to be able to measure the effects of textual news on financial markets, we first need to transform text into measurable variables. This effectively means using statistical, rule-based and machine-learning algorithms to capture a number of news characteristic variables. A number of researchers have decided to create their own measurements. The best examples of this approach are probably the works of Paul Tetlock, who studied textual news from a number of perspectives. In his works, he focuses on a single aspect of news: in (Tetlock, 2007) and (Tetlock, SAAR-TSECHANSKY, and Macskassy, 2008) he focuses on sentiment, while in (Tetlock, 2011) he focuses on novelty. We believe that this single news characteristic approach is somewhat limited because it focuses on one aspect of news and ignores others. An example of this is looking at sentiment to see whether news is positive or negative, but not looking at whether this news is old, whether it had been previously published, or was anticipated. This inevitably means the results of such analysis are biased for news novelty.

Another stream of literature, represented by Groß-Klußmann and Hautsch (2011), Boudoukh, Feldman, Kogan, and Richardson (2013), von Beschwitz, Keim, and Massa (2013), Kelley and Tetlock (2013), and Smales (2014), uses black-box proprietary software to obtain news characteristic variables. We believe that this approach is limited and not very theoretically sound because, despite being able to look at news from a number of angles, it does not provide any insights on how individual variables were extracted from text. One argument is that if the information is extracted using machine learning, one cannot say more than to simply state the algorithm one used to extract the information. This motivated us to use relatively simpler and straightforward methods that had been used with success by other authors in the past. Our approach allows us to clearly state how and what our variable measures, even as we employ a number of tests to measure a wider range of news characteristic variables. We believe that this research represents the best of both, the clear approach of papers utilizing own methodology to extract information from text, as well as the complexity of black-box software offering a number of news characteristic variables. And overview of variables we extract from news is provided in Table 1.1. The following subsections describe our news database and the variables we created based on textual news in greater detail.

[Insert Table 1.1 about here]

A. Motivation

One could argue about why we do not utilize one of existing data sets such as Factiva, RavenPack or services like Reuters NewsScope and instead rely on creating our own database. We see several advantages of creating our own data set. First of all, we have a relatively large collection of high-frequency news. While we are missing social network data (which are, for example, available in RavenPack) or news from sources other than which is published by Reuters, we believe it is a good and representative sample. Moreover, it corresponds to news published in Reuters Eikon, which is arguably same news investors that use Eikon see.

This alone only shows that our sample is good enough; the real advantage is what we can do and what we have done with raw news data. We have tested and implemented a number of methods extracting a number of news characteristic variables. The depth with which we analyse each news story is comparable to enterprise solutions by companies like Reuters or Bloomberg. On the top of this, we are able to test alternative measures as well as possibly implement new news sources into our framework.

On the top of the complexity of identified variables as well as modularity in terms of the methodology we used to extract news data, the third important aspect is the openness. As we created news characteristic variables and measures of news, we are able to share and use them as we please. This way, we can possibly contribute to the research community by letting them utilize the news database for their research.

B. Contribution

We recognize the following contributions of our database:

1 A large collection of news, with a broad range of topics and themes.

2 Complex framework of news characteristic variables, proxies contributing to our understanding not only of whether the news is positive or negative but going deeper into how ambiguous or uncertain they are as well as splitting into themes, how novel or relevant the news is.

3 Expandable in the future as new news arrives, right now more than 8 million news stories, stored in raw text, with the possibility of defining alternative or updated measures.

4 Modularity & contribution to research community: a database that can be shared and used in other fields and also expanded to include other sources, such as social media.

We understand that, while contributions 1 and 2 are present in some commercially available databases, our goal was to make a database that is publicly available and accessible to all researchers. We believe that, compared to other 'free' datasets, our data have advantages, especially thanks to 3 and 4. The ability of the database to expand as time progresses, as well as the ability to revisit old data and re-run tests on them. Following the completion of this thesis, we plan to make the database publicly available to all researchers in near future.

II. News data

We used the Reuters News Web Archive as the source for our news characteristic variables database. We downloaded news articles published from January 2007 through December 2015. To collect the news stories, we constructed both downloading and parsing scripts. Downloading scripts took care of downloading raw news and parsing made sure the articles were cleaned to contain raw text only, without any additional texts or parts that could bias the analysis. After initial downloading and parsing, the news was stored in the database for further processing.

The raw textual news covers a wide range of themes, topics and regions. We have decided to include the whole universe of retrieved news in our analysis. This allows for a great range of possible research projects based on this database and news characteristic variables. For each news story, we calculate the sentiment, ambiguity, novelty and then identify both topics and companies mentioned in individual news stories. We also recover the date and time when news were released, which is reported with minute precision allowing to both distinguish news released at certain times, for example after-hours news, as well as to perform a high-frequency study using intraday financial data and news. An example of the way raw news stories look is displayed in Figure 1.1.

[Insert Figure 1.1 about here]

In order to put our raw news database in perspective, the whole universe of news from January 2007 until December 2015 is more than 8 million news stories. From those, we were able to identify around 5 million articles in which listed companies were mentioned at least once in the story. Further processing narrows the set of listed-company-relevant news further, but as the resulting group is in the hundreds of thousands of observations, we can still speak about a relevant sample. This still excludes news that might be relevant for the market or industry, but in which no companies are mentioned in the text. Examples of the coverage is illustrated in Figure 1.2 where the portion of S&P 500 companies mentioned on average every day in a given week in news in general is represented by blue dots and a fraction of S&P 500 companies mentioned in news identified as earnings announcements on average every day in a given week is represented by red dots. As visible from Figure 1.2 , a fraction of S&P 500 companies mentioned per day is 20% until 2008; it then rises to 40-45% from 2008 and then increases again to 40-60% of companies being mentioned every day from 2012 onward. This implies that our news database contains a representative sample of news stories being published during the period at least for large US companies.

[Insert Figure 1.2 about here]

III. News Characteristic Variables

In previous sections we have described existing research as well as the motivation to create our own news database. In Table 1.1 we briefly reported on variables we identify in news. As presented in the table, we use a number of different approaches and methods to extract variables of interest. All these were developed for the purposes of creating this database and required careful research on which method is best suitable for which variables and what researchers did in the past.

This allowed us to achieve two things: first, to support our approach in relation to existing research and, second, to build a comprehensive news database based on straightforward measures similar to those used in existing research, as presented in Table 1.2, in papers by Paul Tetlock and others who used non-proprietary methodology. Combining a number of methods and measuring a number of news characteristics, we have achieved a level of complexity in our database that is comparable to the proprietary methods used in existing research. We believe that this complexity, which combines simple and straightforward measurements and minute precision, is the main contribution of this database. It should allow researchers to utilize it for a number of news-related research projects.

Figure 1.3 illustrates the difference between what we retrieve, raw news stories, and the data we have in the database connected to the mentioned story and ready to be used in the research. Example shows an earnings announcement for the UnitedHealth Group. An article reads: 'UnitedHealth Group Inc (UNH.N), the largest U.S. health insurer, reported a better-than-expected profit in the third quarter...' and continues in a generally positive direction, mentioning the company's competitors Anthem and Aetna once. This is what an investor reading this news story would notice in the article. If we look at Figure 1.3 section (b), we can clearly see that our methodology correctly identified everything just mentioned. We can clearly see the date and time of report with minute precision and its topic identified as the earnings announcement. We can see that the article has a novelty level of 1 out of 1, meaning no similar article was published in the period prior to this article. Sentiment is identified as positive with 21 positive and 17 negative words identified and ambiguity is measured as 1 with the LOMC approach and 50 with the MAST approach (both approaches are described in detail in the sections below). We also correctly identified 3 companies mentioned in the article; in fact, the UnitedHealth Group was mentioned in the headline, as well, as it is the key company in this article, which is thus represented by a relevance score of 0.75 out of 1.

As presented, this is a comprehensive picture we can build from variables in our database. In order to understand variables in greater detail we describe them in subsections below: the variables themselves, as well as methods used for extracting them and a comparison with the existing research and the approaches it has taken to constructing similar variables. It is important to note that dates and times are extracted directly from news, while other variables are calculated.

[Insert Figure 1.3 about here]

A. Entity Identification

In order to be able to identify which news articles are relevant for which companies or industries and markets, we have to be able to identify named entities. This identification can be done relatively easily by applying the dictionary approach using n-grams. On the top of this, we identify whether a particular company was mentioned in the headline of the article or not. This means that we construct a dictionary of wanted equities, in this case listed companies, and we search for those entities in the text. This raises two issues, however. The first issue is: what if a company has a name that is more than one word long? This issue is mitigated by using n-gram identification, which translates to the ability to identify two or more word phrases in the raw text. The second issue is: what if an alternative name or ticker symbol is used instead of the full name? This issue is mitigated both by including alternative names in the dictionary as well as by making sure unnecessary parts of company name are not included. Examples of unnecessary parts in some cases include 'Inc' or 'Corp' at the end of a full company name.

We created a dictionary of company names using company names and descriptive variables from the Compustat database for both US and international companies. We followed the process mentioned above and stripped out unnecessary name parts, mostly company-type identifications like 'Inc', 'Ltd' or 'A/S' and then manually checked and fixed company names containing words or phrases used in common language. An example of this is the company name 'Global', which was changed back to 'Global Ltd' in order to avoid a false company identification every time the word 'global' appears in the text. Our final dictionary of companies contains more than 40 000 entries, including company names and their acronyms.

Using this approach, we were able to identify more than 12.5 million companies in our news articles. This means that the average number of companies per article is more than 1.5. Please bear in mind, however, that this number is deflated, as some articles clearly do not include any company in particular sport news, for example, or news about culture and cinema. This task was crucial and was among the first to be done as we use entity identification to calculate relevance score per company per news.

B. *Topic Identification*

We have decided to identify news topics, as well. This is useful as it allows us to distinguish between different types of news and allows us to focus on certain topics while excluding others. In the past, similar topic identification was done by (Antweiler and Frank, 2006); the drawback of this research, however, is the methodology used. The authors use the Naive Bayes statistical approach. In a nutshell, this method relies on a training dataset with correctly identified news articles and then uses probabilities to identify news articles that are close to them and assign the same topic. We believe that the drawbacks of this method are sufficient enough for us to introduce a more robust alternative. We have decided to create a rule engine and define rules for news topic identification. It is important to note that other authors, from computer science research fields, have used a combination of methods or a machine-learning approach. This research stream is represented by such studies as Allan, Harding, Fisher, Bolivar, Guzman-Lara, and Amstutz (2005), Ezzat, Ezzat, El-Beltagy, and Ghanem (2012), Tripathi, Oakes, and Wermter (2013) or Lau (2014). In financial research other than the research cited above, topic classification is present only in papers utilizing proprietary tools. These methods, despite being a black box for outsiders, are in fact methods similar to ones used in computer science research, relying heavily on a combination of different methods and machine learning. This stream of financial literature is represented by Groß-Klußmann and Hautsch (2011), Boudoukh, Feldman, Kogan, and Richardson (2013) and Smales (2014).

Another category is research focusing on a single news topic. Beaver (1968), Aharony and Swary (1980) and Kim and Verrecchia (1994), for example, and more recent papers, such as Landsman, Maydew, and Thornock (2012), Pozner, Naumovska, and Zajac (2014) and Chung, Kim, Lim, and Yang (2014) have studied the effect of earnings announcements on stock markets, which is probably the most studied topic of news. Other authors have focused on completely different news topics, Edmans, Garcia, and Norli (2007) focused on the international soccer results and their impact on the stock market, Sadorsky (1999) focused on the oil prices and their relation to the stock market, Hamilton (1995) revealed the effects of toxic release inventory on stock returns, and Niederhoffer, Gibbs, and Bullock (1970) studied the relation between news on US presidential election and Dow Jones Index. International evidence is represented by Wasserfallen (1989) who studied the impact of macroeconomic variables on international stock price indices.

We believe that the body of literature presented here illustrates both the need for a comprehensive topic identification as well as reasons why we decided to use a rule-based engine that avoids the unpredictable nature of machine learning and noise introduced by purely statistical methods. It also illustrates the only research which fully utilizes classification of news into topics and sheds a light on the impact of each topic on the stock market Antweiler and Frank (2006).

Before jumping to the methodology, we had to create a list of topics we will be identifying. Not focusing on studies of a single news topic, probably the most famous research working with many topics is Antweiler and Frank (2006). This study defined 67 basic news categories, but focused only on news directly related to an individual company. Another source of topics is the topic list of the Reuters RCV1 Corpus as presented by Lewis, Yang, Rose, and Li (2004)[1]. This list includes various topics, including politics, arts, health or sport, but the list is not granular enough, with some topics being very detailed and some not.

We decided to use an existing topics list to build our own topic categories where one can choose the granularity according to the level of detail required. We have decided to use the deductive approach and create the list of categories first, as opposed to inductive automatic topic generation in the spirit of Joachims (1996) or Fiaidhi, Mohammed, Islam, Fong, and Kim (2013).

We have combined a number of existing studies in order to derive our complex list of topics ranging from corporate news to news on disasters. We have split news into categories and created a 4-tier system enabling us to use as detailed a level of granularity as necessary. For business news, we used Antweiler and Frank (2006) for the company-specific news categories and Beber, Brandt, and Luisi (2015), Omrane and Hafner (2015) and Kurov, Sancetta, Strasser, and Wolfe (2015) for macroeconomic news categories. In order to create categories about markets, we used types of financial markets as defined by Pilbeam (2010) and Madura (2014). For political news categories, we used the policy agendas codebook as described by John (2006) and Bevan (2014). For science and technology, we defined our own categories, as this categorization was not defined before. For sport news categorization, we used a list of the major (most popular) sports in the world and the rest will be categorized into the Other sports category. For life and health, we used inspiration from the WSJ Magazine and for disasters, we used disasters categories as presented in the existing literature; for example by Guha-Sapir, Hargitt, and Hoyois (2004) or Eshghi and Larson (2008). A final list of top-level news topics is reported in Table 1.3; please note that this list excludes lower-level topics. For more details on how the list was compiled and what sources were used for which topic breakdown, please see the Appendix VI.

[Insert Table 1.3 about here]

We have decided to use a rule engine to extract topics from raw text. In order to be able to do it, we developed the engine itself, which is comprised of two parts. The first part is the user interface used for rule creating and editing. It allows for a simple rule visualization and stores rules in databases. This way, it is clear what the rule is, and when it is fulfilled, the text is classified as a topic set for given rule. The engine itself allows for the simple logical operations AND and OR and allows us to either include or completely exclude a given word. It also allows us to search for phrases. Second part of the engine is the analytic part that transforms rules to code and searches raw text to determine whether a rule is fulfilled or not.

The example or a simple rule with sample text is provided in the Table 1.4. In this case, the rule identifies the topic earnings announcements, it searches for the word 'earnings' and then a selection of words, with at least one of them required to be present in the text. The condition for the topic identification is fulfilled as the word 'rise' is present in the text as well. Please bear in mind that this is to illustrate the engine only and more advanced rules are defined. In theory, the engine is not limited to the number of phrases and logical operations.

[1] The topic list derived from the Reuters Corpus is available here: http://www.daviddlewis.com/resources/testcollections/rcv1/

The topic identification, besides helping to fit the puzzle of news characteristics, also allows us to control for topics and possibly evaluate the impact of news from each topic on the stock market. This allows us to exclude non-relevant topics from our sample or enables us to focus only on a handful of topics of interest.

[Insert Table 1.4 about here]

C. Relevance

The relevance of news stories is an important measure as it serves as a proxy for the extent to which a given news item has an impact on a company. Several existing studies measured the relevance, but most of them used proprietary software to do that, as in Groß-Klußmann and Hautsch (2011), Boudoukh, Feldman, Kogan, and Richardson (2013) or Smales (2014). Lau (2014) studied the relevance of Twitter topics. He used the logistic regression to evaluate which characteristics of tweets affect the relevance. We are constructing our measure of the relevance in a completely different manner.

Our measure takes into account the frequency of mentions of an identified entity as a relative fraction of all entities mentioned in the text. In order to calculate the relevance, we do not employ any additional analysis, but use identified entities and their frequencies in the text to calculate the relevance score.

$$relevance = \frac{mentioned_i}{\sum_{j=1}^{n} mentioned_j} \tag{1.1}$$

As equation (1.1) shows, the relevance score is calculated as the number of times a company is mentioned in news articles ($mentioned_i$) divided by the sum of occurrences of all mentioned companies in the news articles ($\sum_{j=1}^{n} mentioned_j$); this limits the relevance score in the interval between 0 and 1.

This approach, however, does not account for insightful news with many companies mentioned within. We therefore record the raw number of occurrences as well as the information whether the company was mentioned in the headline of news or not. This additional information can be later used to determine the relevance of news with greater precision using an alternative measure in the future.

D. Novelty

The novelty of news measures whether similar news was released in recent history. This measure is inspired by (Tetlock, 2011) who used 1 and 2 grams to evaluate the staleness of news. The so-called staleness is in fact novelty as it measures exactly same phenomena and that is if news with similar texts were published prior to the story of interest. We believe that the approach proposed by Paul Tetlock is the right proxy for novelty of news, yet is very straightforward to calculate and hence fits what we are trying to achieve a simple and straightforward methodology for variable construction perfectly.

$$similarity_1 = \frac{Words1_i \cap Words1_j}{Words1_i \cup Words1_j} \tag{1.2}$$

$$similarity_2 = \frac{Words2_i \cap Words2_j}{Words2_i \cup Words2_j} \tag{1.3}$$

Equations (1.2) and (1.3) describe the way we calculated the similarity between news of interest i and past news j. This is calculated as the intersect between the set of 1-gram fords from news i and past news j and divided by the union of two sets. In the case of (1.3), we use 2-grams instead of single words or 1-grams. Please note we have decided to exclude news with less than 50 words, as it might bias results if similar language is used even in a different context. This corresponds to (Tetlock, 2011), who excluded short news as well.

We calculate the similarity of all news j from $t-5$, until t, where t is the date and time of the release of news i. All similarities are stored in the database, allowing us for flexible calculation of novelty measurement. It allows for changes and alternative calculations of novelty. Not counting research utilizing proprietary black-box software, there is no other research focusing directly on novelty of news than (Tetlock, 2011) who used staleness measurement similar to our similarity and then calculate total staleness based on 10 prior news as a simple average of their staleness. This implies there might be more competing possibilities of how to calculate novelty. We have decided to build our novelty measurement around three pillars. First, we want only news mentioning the same company and prior to news of interest to be taken account for. Second, we want to reward news on same topic with higher similarity as well as we want to reward 2-gram similarity. Third, we want to punish news released a long time ago.

$$\begin{aligned} novelty_j = max\Bigg(0, 1 - \sum_{i=1}^{n} \Big(& max(1, dummy\_topic_i * 2) * max(1, dummy\_2gram_i * 1.5) \\ & * [1 - (T\_news_i - T\_news\_to_i) * 1/120] * similarity_{j,i} \Big) \Bigg) \end{aligned} \tag{1.4}$$

Novelty is calculated as presented in (1.4). It is evident that we limit the novelty to interval 0 to 1. This is done by making novelty equal to the maximum between 0 and 1 minus the sum of weighted similarities. As described above, we are using three weights to calculate a similarity multiplier and then we sum up those weighted similarities. We use $dummy\_topic_i$ to multiply the novelty impact of news 2 times if previously released news is on the same topic as analysed news, we use $dummy\_2gram_i$ to multiply novelty impact by 1.5 if previously released news similarity is a 2-gram similarity (similarity on a phrase level). The ratio behind these two multiplications is to reward news of the same topic and news similar on phrase level. Last multiplier is $[1 - (T\_news_i - T\_news\_to_i) * 1/120]$ and that is time multiplier. First, we take the difference in hours between the date and time when news of interest was released and the date and time previously released news was released. This difference is then multiplied by $1/120$ and then the final number is subtracted from 1 in order to get a final time multiplier. We are using $1/120$ as we go back in history to up to 5 days and we decided to use a linear decay of similarity, which means every hour difference between news releases is taking $1/120$ of similarity multiplier. All those weighted similarities for news j are summed and subtracted from 1.

Please note that the proposed novelty measurement is only one of potentially many novelty measurements. As one can change rules to filter news as well as to decide on different decay rates and reward for same topics or 2-gram similarity. We believe that the rationale for our measure is sound; however, more detailed research on alternative measures will have to be done in order to comment on which measure of novelty is the most suitable and when.

E. Sentiment

Previously described variables were all about identifying companies, topics of news and determining how new and relevant news articles are. Sentiment is used to measure what is inside news, what language is used and whether this news has positive or negative tone. It is arguably the most studied measure derived from news in existing literature. In the past, research utilized a wide range of methods to identify sentiment ranging from a simple dictionary approach to machine-learning methods. A useful survey of sentiment analysis methods was done by (Medhat, Hassan, and Korashy, 2014). In our research, we decided to use the dictionary approach. This approach has an advantage as it is very straightforward, easy to describe and understand and relatively accurate at the same time. We have used the Harvard 4 dictionary, similarly to (Tetlock, 2007). Then we developed our own algorithm for the word identification.

As sentiment has been described well by other authors in the past, we will only point out our approach of determining it. We have decided to use the dictionary approach as stated above. We first took raw textual news and cleaned it of all HTML tags as well as from generic things like menus or share and like buttons that might be left in the text. After this step, we split he text into words and then check it word-by-word against our dictionary. We created our dictionary from the Harvard Psychosociological Dictionary and its Harvard IV TagNeg file. Researchers in the past, for example Loughran and McDonald (2011), argued that the categorization scheme is discipline-specific, as the same word could have different translations across disciplines. They have evaluated the dictionary and found the misclassification for 10-Ks reports. They created their own dictionary with words having the sentiment meaning in the financial context. Based on our preliminary analysis, however, we found no significant difference between the usage of their dictionary or the Harvard Psychosociological Dictionary. Also, we believe, as our news is not limited to corporate or financial topics, the use of specialized dictionary may harm the performance of the analysis.

$$sentiment = \frac{POS - NEG}{POS + NEG} \tag{1.5}$$

Our methodology sums identified positive and negative words and calculates the 'raw' sentiment score as the difference between the number of positive words POS and the number of negative words NEG, divided by the sum of the two. This simple yet straightforward approach allows for a great transparency of used method; this is beneficial especially over machine-learning or statistical approaches who act more like black boxes in terms of the input-output relation.

F. Ambiguity

The last news characteristic we extract from the news is the ambiguity. We measure it in a similar vein as the sentiment described above. We take raw news, clean them of the rest of the tags and special characters and then split into single words. Then we apply the dictionary approach to identify ambiguous words. The main difference is the dictionary we used. In this case, we need a dictionary that will help us to identify ambiguous words, words that can be interpreted differently or those that introduce uncertainty into the text. We have acquired two dictionaries for this purpose. The first is based on the list of homonyms from the Clark's website. Homonyms are, in our case, used as a proxy for ambiguity of news. The alternative dictionary will measure uncertainty. The dictionary of uncertain words is obtained from Bill McDonald's website, which we use for words with an uncertainty category.

We proceed as in the case of sentiment and calculate raw ambiguity and uncertainty, calculating the sum of identified words for both dictionaries. We call the dictionary of homonyms the 'MAST' dictionary and dictionary of uncertain words 'LOMC'. Both measures are reported for all news as illustrated for example in the Figure 1.3 in section above. After some preliminary tests on earnings announcement news data, which showed similar behaviour if split by uncertainty as well as by ambiguity (LOMC and MAST dictionaries), along with the correlation between these two variables being 0.348, we believe that these measurements measure, in fact, one characteristic of news and are relatively strongly correlated. This implies there is no need to use both measures as their economic impacts will be relatively similar. It is interesting to note that, once weighted, the correlation between those variables drops to 0.196. Weighting is described in the separate section below and applies to both sentiment and ambiguity.

G. Weighted variables

In the previous two subsections, we defined sentiment and ambiguity. We defined them as raw calculations, simply summing up the number of occurrences in the text without regard for how frequent the word itself is, how many words there are in text, and so on. It is well known that some methods rely on the analysis of text with frequent words being discarded. This implies that not all words are made equal and we should account for this inequality, too.

In order to process 'raw' sentiment, we decided to weight all terms. We use the local weight, document weight and global weight, as proposed by Chisholm and Kolda (1999). Our weighting consists of three components: local weight ($L_{i,j}$), global weight (G_i) and normalization (N_j). We follow the square root local weight formula proposed by Chisholm and Kolda (1999), which calculates the local weight of term i for document j.

$$L_{i,j} = \sqrt{f_{i,j} - 0.5} + 1 \tag{1.6}$$

As visible in (1.6), $L_{i,j}$ is the local weight of term i for document j and $f_{i,j}$ is the local frequency of them i in document j. This formula applies if $f_{i,j} > 0$, if $f_{i,j} = 0$ then the $L_{i,j}$ is set to 0.

Global frequency is calculated using the global frequency IDF (inverted document frequency) approach. The IDF measure was derived by Sparck Jones (1972). Chisholm and Kolda (1999), however, proposes that the measurement be defined as $G_i = \frac{F_i}{n_i}$, where the global weight G_i for term i is calculated as the frequency of term i throughout the entire document collection F_i divided by the number of documents where term i appears n_i.

This is good approach if we want to analyse a static set of documents. However, we aim to create a measure which is stable and does not deteriorate over time, as an IDF measure would with an increasing number of documents n_i. One approach would be to limit the number of news articles in the set, this would however result in lower precision as older news would be excluded from the set. A second approach is to use relative values.

$$G_i = \frac{\frac{F_i}{\sum_{i=0}^{m} F_i}}{\frac{n_i}{N}} = \frac{F_i * N}{n_i * \sum_{i=0}^{m} F_i} \tag{1.7}$$

We propose to adjust the original global frequency IDF formula to include relative frequency and relative number of documents as described in (1.7). In this adjusted formula, we use, instead of the frequency of term i throughout the entire document collection F_i, the relative frequency defined as global frequency F_i divided by the sum of global frequencies for all identified terms $\sum_{i=0}^{m} F_i$. We adjust the number of documents where term i appears n_i to the relative number of documents where term i appears as the fraction of the total number of documents in the set N.

$$N_j = \frac{1}{\sqrt{\sum_{i=0}^{m} (G_i * L_{i,j})^2}} \tag{1.8}$$

The last term is the normalization. In this case, we follow the results of Chisholm and Kolda (1999), who proposed cosine normalization as the most familiar, yet very effective form of normalization. We defined the normalization formula in (1.8), where normalization weight N_j for document j is calculated as the square root of sums of squares of the local weight for term i and document j multiplied by the global weight of term i.

The term weighting formula for term i and document j is defined in the following equations (1.9) and (1.10).

$$weight_{i,j} = L_{i,j} * G_i * N_j \tag{1.9}$$

$$weight_{i,j} = \frac{(\sqrt{f_{i,j} - 0.5} + 1) * F_i * N}{n_i * \sum_{i=0}^{m} F_i * \sqrt{\sum_{i=0}^{m} (\frac{F_i * N * (\sqrt{f_{i,j} - 0.5} + 1)}{n_i * \sum_{i=0}^{m} F_i})^2}} \tag{1.10}$$

In order to calculate the weighted sentiment score, we take the equation (1.9) (as we have local, global and normalization weight calculated) and combine it with the not-weighted or 'raw' sentiment score calculation (1.5).

$$sentiment = \frac{POS_W - NEG_W}{POS_W + NEG_W} \tag{1.11}$$

Equation (1.11) is used to calculate weighted sentiment, with POS_W and NEG_W being sums of weighted sentiment scores. This way, we control for the length of document and number of word occurrences both locally in the file and globally in the whole set of news. The same weighting can be applied to ambiguity and uncertainty measurements with one difference: they are calculated simply as sum of identified ambiguous or uncertain words in text multiplied by appropriate weights, without the additional steps as described in (1.11).

Please note that this weighting is one of possible weights, and alternative weighting might be implemented. We based our weighting on Chisholm and Kolda (1999), who tested a number of alternative term weightings for information retrieval. Despite the fact that our application is slightly different, the context is rather similar as both Chisholm and Kolda (1999) and we want to retrieve text information from text.

IV. Sample Analysis

As presented above, our news database contains news of different topics and themes. There are 7 main news categories. Figure 1.4 show each main category and bubble size shows the total number of news identified in a given main category. It is clear that we have a wide range of news. While business news is the largest category, it also contains the largest amount of individual news topics. An overview of all topics is shown in Figure 1.5, where each bubble represents individual news topic, bubble size number of news identified in a given topic and bubble colour the main topic each topic belongs to.

[Insert Figure 1.4 & Figure 1.5 about here]

We have illustrated that our database contains a wide range of news covering a big universe of news topics. In order to relate the database to financial data, we have run three event studies. In all three cases, we used the value index of S&P 500 stocks, which we split around news events into two groups based on positive and negative sentiments. We have decided to do this for three news topics: earnings announcements, new technology news and bankruptcy news.

The first test was on earnings announcements news. This can be considered a 'standard' test as earnings announcements were studied extensively in the past and there is no doubt that positive earnings announcements yield positive returns, while negative earnings announcements yield negative returns. This story is confirmed in our data, as visible in Figure 1.6. We can see that the red line, representing negative sentiment news, yields negative returns in our time window. The story is consistent if we look at positive news on earnings announcements, which yield slightly positive returns and are represented by the blue line.

[Insert Figure 1.6 about here]

The second test is on new technology and science news. We can see that in Figure 1.7, there is a clear difference between positive sentiment and negative sentiment news on this topic. As expected, positive news yields higher returns than negative news in the time frame being studied. On the other hand, what is puzzling is that both returns are negative, with returns on positive news being less negative than on negative news. One explanation might be that news about new technology and scientific developments increase the uncertainty and investors price for this uncertainty. Exact causality and the reasoning for why this might be so are not, however, within the scope of this research and would need further investigation.

[Insert Figure 1.7 about here]

The third event study focuses on bankruptcy news. As visible in Figure 1.8, bankruptcy news means negative returns no matter how positive or negative they are. Consistent with the notion of news with a negative tone being more negative than news with a positive tone, the cumulative returns around the news event are more negative for negative sentiment news. This is observable in a period of one day before negative news is released to one day after it is released. What is puzzling is that negative news recover after that and reportedly yield less negative returns than does positive sentiment news.

[Insert Figure 1.8 about here]

As presented above, it is possible to relate a number of news topics and news characteristic variables to financial data. Despite the fact that we run only simple event studies, this illustrates to which extent news and their characteristic variables contain the information relevant for investors. We believe that these data could be used in further studies, not only limited to stock market or to finance in general.

V. Conclusion

The goal of this research was to build a news database containing news articles processed so that they are ready to be used in research. We developed a framework to identify news topics as well as other characteristic variables including sentiment, novelty, relevance and ambiguity. We also identify companies related to individual news in such a way that could enable more companies as well as more topics to be related to a single news story.

Our final database dates from the first of January 2007 and spans until 31st December 2015. We have more than 8 million raw news stories that are processed, with all characteristic variables as well as companies identified, as illustrated in Figure 1.3. We also illustrated, in three simple event studies, that our data could be connected with financial data in order to investigate stock markets around certain news releases. We understand that the analysis done in this research is limited and that more detailed analysis will be necessary to draw any conclusions.

On the other hand, we do prove the point that the news articles contained in our database carry information and could be potentially used in a variety of types of research. Beyond finance, where news and sentiment have been studied extensively, we see potential in political news as well as in news about disasters. We believe that these can be used for economic or political science studies. As we recognize the potential of our news database, we are happy to share processed news with other researchers if they are interested. We believe that this can be seen as a good conclusion for the work done in this research.

VI. Appendix: News Topics

A. Introduction

At the beginning, there was a task to classify raw textual news by topic in order to be able to investigate each news topic individually. Looking at the existing body of the literature on the topic categorization, we realized there is no comprehensive and general list of news topics. One part of existing lists is domain specific, for example Antweiler and Frank (2006) defined topics for financial and corporate news, John (2006) proposed a comprehensive topic list for political agendas. Other existing lists, despite covering a broad range of domains, are not comprehensive nor exhaustive. For example **?** created very general, but not very detailed list of topics.

The aim of this part is to propose a general and comprehensive list of news topics suitable for the topic categorization of textual news. This list is derived from lists of topics for particular domains from the literature combined with other sources for topics. It is structured in a way, one can use smaller or greater detail to achieve lower or higher granularity of topics. We use this list to classify Reuters News Web Archive news, but we believe this list can be utilized for the classification of any textual news or other texts.

B. News Topics

In this section, we discuss major sources of news topic categories we use for our topic list. These lists are different from automated topic categorization in the spirit of Joachims (1996) or Fiaidhi, Mohammed, Islam, Fong, and Kim (2013) in a way, we focused on retrieving detailed topic categorization covering a wide number of domains, rather than a list of individual topics generated automatically.

As the news data we aim to categorize are retrieved from Reuters, we fist look at the Reuters website and their categorization. We look at both, Reuters' news website today and one from the past stored at the web.archive.org. Reuters provides the categorization of topics to some extent, however not detailed enough for our purposes. For example, Reuters splits Politics to all political news, ”tales from the trail” commentaries and supreme court news, while our proposed categorization splits Politics to law, elections and political crises and then splits those sub-topics further. The assumption is, news on new legislature shouldn't be in the same topic category as news on elections as they are of different matter and content, having one political news topic category for both is not always sufficient and viable solution.

Following sections describe sources of topics sub-categories for main topic categories. Main (first tier) categories were defined based on the news categorization of Reuters and are Business news, Market news, Politics news, Science & Technology, Sport news, Life & Health and Disasters. The last category is Other for all news not specified in other categories.

Business news

The first topic category of news is Business news. This category is well defined by researchers in the finance and accounting who focus on news and announcements. We split this main category to two sub-categories both, corporate news and macroeconomic news, are well defined in the existing literature. For corporate news, we used the categorization proposed by Antweiler and Frank (2006) who used Naive Bayes method to categorize corporate news to 44 categories. Another approach was used by Boudoukh, Feldman, Kogan, and Richardson (2013) who used rule-based proprietary method to categorize news, however the number of categories was very limited.

In order to retrieve categories for macroeconomic news, we decided to combine a number of papers on macroeconomic announcements. Beber, Brandt, and Luisi (2015) combines macroeconomic announcements into daily factors, Omrane and Hafner (2015) studies economic fundamentals and the exchange rate volatility and Kurov, Sancetta, Strasser, and Wolfe (2015) studies the stock index and Treasury futures around US macroeconomic announcements. While first category consists of to some extent unpredictable and unscheduled news, macroeconomic announcement dates and times are known and disclosures are forecasted.

Market news

In order to categorize news about financial markets we looked at the categorization existing in the literature and online sources. We combined types of financial markets defined in Pilbeam (2010) and Madura (2014) with information on financial markets available from Investopedia.

Politics news

For the news on politics it is hard to define the categorization using existing literature. Existing literature usually focus on the Twitter posts and sentiment analysis. Twitter data are used for example to reveal political leaning Jiang and Argamon (2008), another example is Romero, Meeder, and Kleinberg (2011) studying the information diffusion. A simple, but not exhaustive list of politics news was created by Young and Soroka (2012), for more information about political texts classification see Albaugh, Sevenans, Soroka, and Loewen (2013).

Categories for the Politics news were inspired by a number of internet sources. First, for topics related to the law and policy agendas were inspired by the Policy agendas codebook available online http://www.policyagendas.org, for more information on the Political Agendas Project see John (2006) and for generalized (international) version see Bevan (2014) , the generalized version is available here. Topics on elections around the world are based on the IDEA.int and Wikipedia.

Science & Technology

Media reports on science was studied by Véiliverronen (1993). Another stream of the literature studies the communication between scientists and public, for example Bucchi (1996). Categorization of news on science and technology is however not very well defined in the existing literature, our categories were therefore defined as technology/discovery/invention announcements, other news and frauds for each of major categories.

Sport news

For the definition of sport categories we used lists of most popular sports in the world. The example of lists of the most popular sports are Sporteology and BiggestGlobalSports. According to the SportEncyclopedia, the total number of sports is more than 8000. This implies it is impossible to have a separate category for each individual sport, we have decided to select 10 to 15 major (most popular) sports and the rest will be categorized to ”Other sports” category.

Life & health

Life and health category is not well defined in existing literature on the news categorization as well. We have decided to split the major category to culture, lifestyle, health and environment. Sub-categories for those major categories were inspired by the number of internet sources. For lifestyle topics see WSJ Magazine.

Disasters

The last major category of news is disasters, this category was well defined in existing literature, for example by Guha-Sapir, Hargitt, and Hoyois (2004) and Eshghi and Larson (2008). Lists of disasters and their categorization to natural and man-made are available on the internet as well, for example at RestoreYourEconomy.

C. The news topics list

Our motivation was to create a comprehensive and general list of topics, usable for the textual news topic classification. We have combined topic lists available in existing literature and other sources in order to achieve our goal. Following table displays first and second tier topics, you can download full list of topics here: topic_list_v1.0.csv

VII. Appendix: Tables

Table 1.1. Following table
describes news characteristic variables as well as methods used to calculate them and a short description about the calculation.

News Characteristics	*Method*	*Calculation*
Entity	Dictionary / n-grams	Using own database of company names to identify companies in text
Topic	Rules / n-grams	Apply pre-defined rules to search for patterns in text
Relevance	Calculated	Calculated from the Entity information, based on occurrence frequency and positions
Novelty	n-grams	Following the approach introduced by (Tetlock, 2011), comparing 1 and 2-grams to past news
Sentiment	Dictionary / n-grams	Using HARVARD-4 dictionary to identify positive and negative words, weighting of words in the spirit of Chisholm and Kolda (1999) is introduced

Table 1.2. Overview table of existing research on the news. In this table we present key papers as well as details about them. We have decided to report the frequency research focused on, methodology they employed and what news characteristic variable they used in their research. We include this research for comparison as well. We have decided to note all black-box software as proprietary for simplicity.

Paper	*Data Frequency*	*Methodology*	*Entity*	*Topic*	*Relevance*	*Novelty*	*Sentiment*
(Antweiler and Frank, 2006)	Daily	Naive Bayes	Y/N	Yes	No	No	No
(von Beschwitz, Keim, and Massa, 2013)	Intraday	Proprietary	Yes	Yes	Yes	No	Yes
(Fang and Peress, 2009)	Daily	Proprietary[a]	Y/N	No	Yes	No	No
(Ferguson, Guo, Lam, and Philip, 2011a)	Daily	Dictionary	Y/N	No	No	No	Yes
(Garcia, 2013)	Daily	Dictionary	Y/N	No	No	No	Yes
(Groß-Klußmann and Hautsch, 2011)	Intraday	Proprietary	Yes	No	Yes	Yes	Yes
(Kelley and Tetlock, 2013)	Intraday	Proprietary	Yes	No	No	No	Yes
(Loughran and McDonald, 2011)	Daily	Dictionary	Y/N	No	No	No	Yes
(Smales, 2014)	Intraday	Proprietary	Yes	No	Yes	Binary[b]	Yes
(Tetlock, 2007)	Daily	Dictionary	Y/N	No	No	No	Yes
(Tetlock, SAAR-TSECHANSKY, and Macskassy, 2008)	Daily	Dictionary	Y/N	No	No	No	Yes
(Tetlock, 2011)	Daily	N-grams	Y/N	No	No	Yes	No

[a] They used LexisNexis entity identification and relevance data
[b] Takes values 1 if new and 0 otherwise

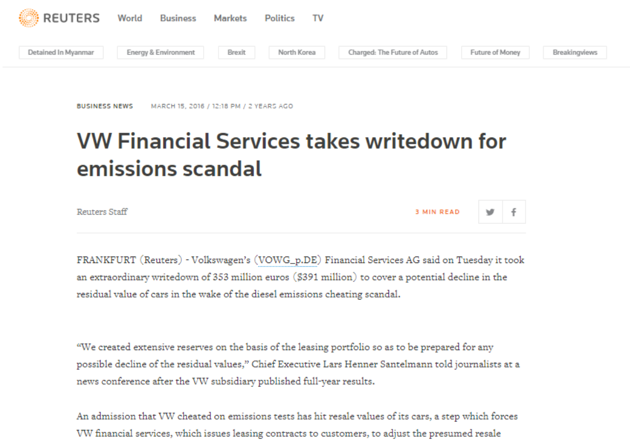
Table 1.3. Following table describes high level news topics. Topics are split to categories first by high level theme and then they are split to lower tiers up to 4th tier with last tier containing 150 news categories. As visible in the table we decided to identify topics for whole universe of news, not limiting ourselves to a certain topic or theme.

Topic	Identifiers	
	Id	Tier
Business	1000	1
Corporate	1100	2
Macroeconomics	1200	2
Markets	2000	1
Politics	3000	1
Law	3100	2
Elections	3200	2
Political crisis	3300	2
Policy announcements	3400	2
Government news	3500	2
International relations	3600	2
General news	3700	2
Legal system	3800	2
Science & Technologies	4000	1
Science	4100	2
Technology	4200	2
Innovations	4300	2
Millitary	4400	2
Sport	5000	1
Life & Health	6000	1
Culture	6100	2
Lifestyle	6200	2
Health	6300	2
Environment	6400	2
Crime	6500	2
Demonstrations	6600	2
Religion	6700	2
Disasters	7000	1
Natural Disasters	7100	2
Man-made Disasters	7200	2
Other	9000	1

Table 1.4. Topic Identification Example

Rule	**earnings** AND (fall OR up OR climb OR **rise** OR beat OR down OR report)
Text	American Express 4th-quarter **earnings rise**

VIII. Appendix: Figures


REUTERS World Business Markets Politics TV

Detained In Myanmar | Energy & Environment | Brexit | North Korea | Charged: The Future of Autos | Future of Money | Breakingviews

BUSINESS NEWS MARCH 15, 2016 / 12:18 PM / 2 YEARS AGO

VW Financial Services takes writedown for emissions scandal

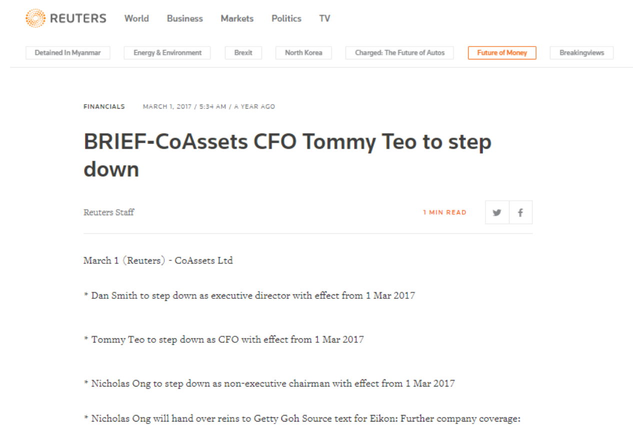
Reuters Staff 3 MIN READ

FRANKFURT (Reuters) - Volkswagen's (VOWG_p.DE) Financial Services AG said on Tuesday it took an extraordinary writedown of 353 million euros ($391 million) to cover a potential decline in the residual value of cars in the wake of the diesel emissions cheating scandal.

"We created extensive reserves on the basis of the leasing portfolio so as to be prepared for any possible decline of the residual values," Chief Executive Lars Henner Santelmann told journalists at a news conference after the VW subsidiary published full-year results.

An admission that VW cheated on emissions tests has hit resale values of its cars, a step which forces VW financial services, which issues leasing contracts to customers, to adjust the presumed resale

(a) Financial news


REUTERS World Business Markets Politics TV

Detained In Myanmar | Energy & Environment | Brexit | North Korea | Charged: The Future of Autos | Future of Money | Breakingviews

FINANCIALS MARCH 1, 2017 / 5:34 AM / A YEAR AGO

BRIEF-CoAssets CFO Tommy Teo to step down

Reuters Staff 1 MIN READ

March 1 (Reuters) - CoAssets Ltd

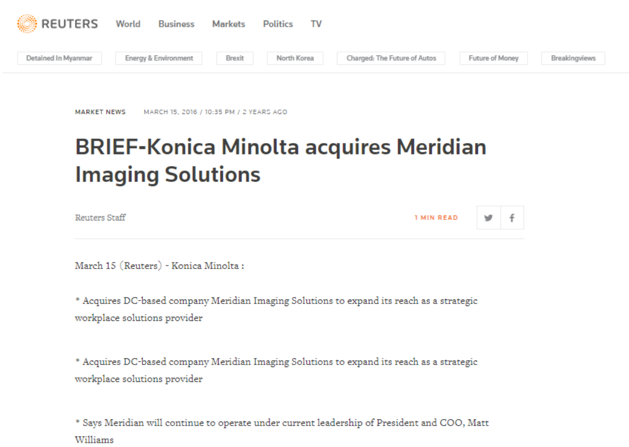
* Dan Smith to step down as executive director with effect from 1 Mar 2017

* Tommy Teo to step down as CFO with effect from 1 Mar 2017

* Nicholas Ong to step down as non-executive chairman with effect from 1 Mar 2017

* Nicholas Ong will hand over reins to Getty Goh Source text for Eikon: Further company coverage:

(b) Management change


REUTERS World Business Markets Politics TV

Detained In Myanmar | Energy & Environment | Brexit | North Korea | Charged: The Future of Autos | Future of Money | Breakingviews

MARKET NEWS MARCH 15, 2016 / 10:35 PM / 2 YEARS AGO

BRIEF-Konica Minolta acquires Meridian Imaging Solutions

Reuters Staff 1 MIN READ

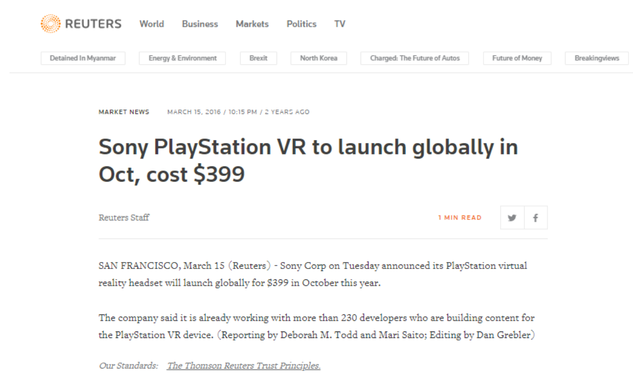
March 15 (Reuters) - Konica Minolta :

* Acquires DC-based company Meridian Imaging Solutions to expand its reach as a strategic workplace solutions provider

* Acquires DC-based company Meridian Imaging Solutions to expand its reach as a strategic workplace solutions provider

* Says Meridian will continue to operate under current leadership of President and COO, Matt Williams

(c) M&A


REUTERS World Business Markets Politics TV

Detained In Myanmar | Energy & Environment | Brexit | North Korea | Charged: The Future of Autos | Future of Money | Breakingviews

MARKET NEWS MARCH 15, 2016 / 10:15 PM / 2 YEARS AGO

Sony PlayStation VR to launch globally in Oct, cost $399

Reuters Staff 1 MIN READ

SAN FRANCISCO, March 15 (Reuters) - Sony Corp on Tuesday announced its PlayStation virtual reality headset will launch globally for $399 in October this year.

The company said it is already working with more than 230 developers who are building content for the PlayStation VR device. (Reporting by Deborah M. Todd and Mari Saito; Editing by Dan Grebler)

Our Standards: The Thomson Reuters Trust Principles.

(d) New Product announcement

Figure 1.1. Examples of raw news as displayed on Reuters website.

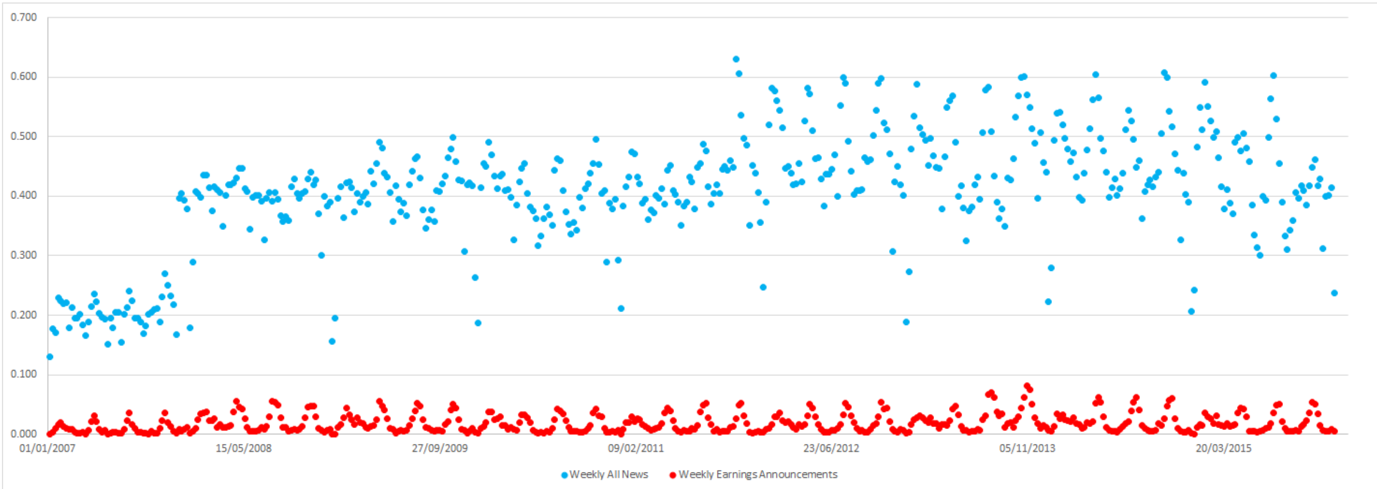
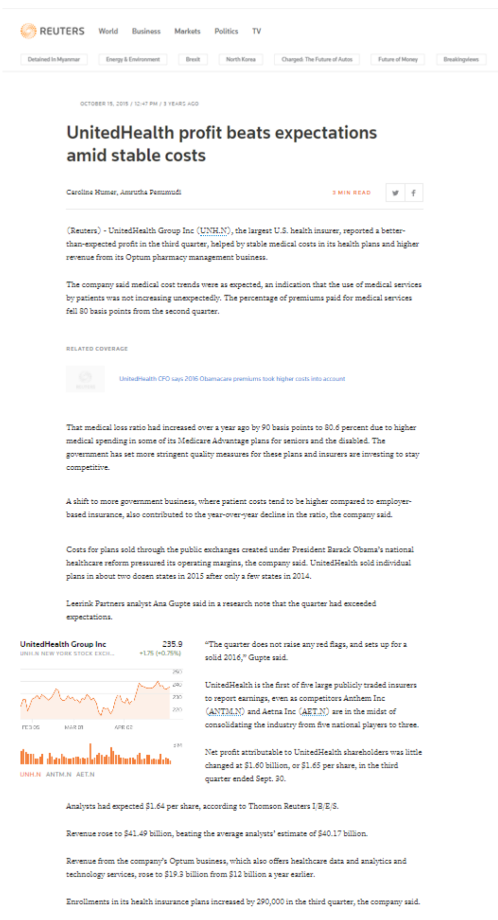


Figure 1.2. Average weekly fraction of S&P 500 firms mentioned in all news (blue) and mentioned in earnings announcement news (red).


REUTERS World Business Markets Politics TV

Detained In Myanmar | Energy & Environment | Brexit | North Korea | Charged: The Future of Autos | Future of Money | Breakingviews

OCTOBER 15, 2015 / 12:47 PM / 3 YEARS AGO

UnitedHealth profit beats expectations amid stable costs

Caroline Humer, Amrutha Penumudi 3 MIN READ

(Reuters) - UnitedHealth Group Inc (UNH.N), the largest U.S. health insurer, reported a better-than-expected profit in the third quarter, helped by stable medical costs in its health plans and higher revenue from its Optum pharmacy management business.

The company said medical cost trends were as expected, an indication that the use of medical services by patients was not increasing unexpectedly. The percentage of premiums paid for medical services fell 80 basis points from the second quarter.

RELATED COVERAGE

UnitedHealth CFO says 2016 Obamacare premiums took higher costs into account

That medical loss ratio had increased over a year ago by 90 basis points to 80.6 percent due to higher medical spending in some of its Medicare Advantage plans for seniors and the disabled. The government has set more stringent quality measures for these plans and insurers are investing to stay competitive.

A shift to more government business, where patient costs tend to be higher compared to employer-based insurance, also contributed to the year-over-year decline in the ratio, the company said.

Costs for plans sold through the public exchanges created under President Barack Obama's national healthcare reform pressured its operating margins, the company said. UnitedHealth sold individual plans in about two dozen states in 2015 after only a few states in 2014.

Leerink Partners analyst Ana Gupte said in a research note that the quarter had exceeded expectations.

UnitedHealth Group Inc 235.9
UNH.N NEW YORK STOCK EXCH... +1.75 (+0.75%)
UNH.N ANTM.N AET.N

"The quarter does not raise any red flags, and sets up for a solid 2016," Gupte said.

UnitedHealth is the first of five large publicly traded insurers to report earnings, even as competitors Anthem Inc (ANTM.N) and Aetna Inc (AET.N) are in the midst of consolidating the industry from five national players to three.

Net profit attributable to UnitedHealth shareholders was little changed at $1.60 billion, or $1.65 per share, in the third quarter ended Sept. 30.

Analysts had expected $1.64 per share, according to Thomson Reuters I/B/E/S.

Revenue rose to $41.49 billion, beating the average analysts' estimate of $40.17 billion.

Revenue from the company's Optum business, which also offers healthcare data and analytics and technology services, rose to $19.3 billion from $12 billion a year earlier.

Enrollments in its health insurance plans increased by 290,000 in the third quarter, the company said.

Our Standards: The Thomson Reuters Trust Principles.

(a) Raw News Story

General variables:

Variable	*Value*
Date & Time:	
	16-10-2015 9:16
Topic:	
	9001 - Earnings announcement
Novelty:	
	1
Sentiment:	
# positive words	21
# negative words	17
Total sentiment	4
Total weighted sentiment	1.85
Ambiguity:	
Ambiguity LOMC	1
Ambiguity MAST	50

Company variables:

	Value
Aetna Inc.	
In-headline:	0
Relevance score:	0.12
Unitedhealth Group Inc.	
In-headline:	1
Relevance score:	0.75
Anthem Inc.	
In-headline:	0
Relevance score:	0.12

(b) Identified News Variables

Figure 1.3. Example of raw news story (a) and the same news story processed (b).

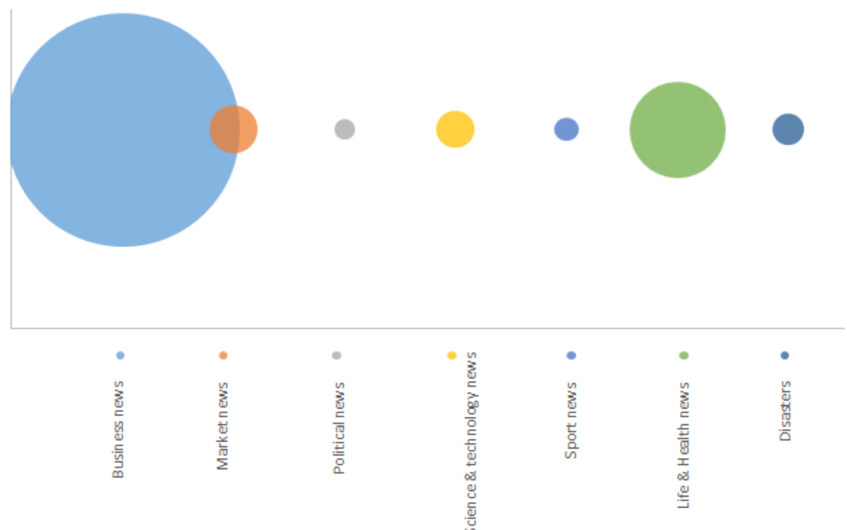


Figure 1.4. Individual bubbles show the total number of identified news stories per main news topics.

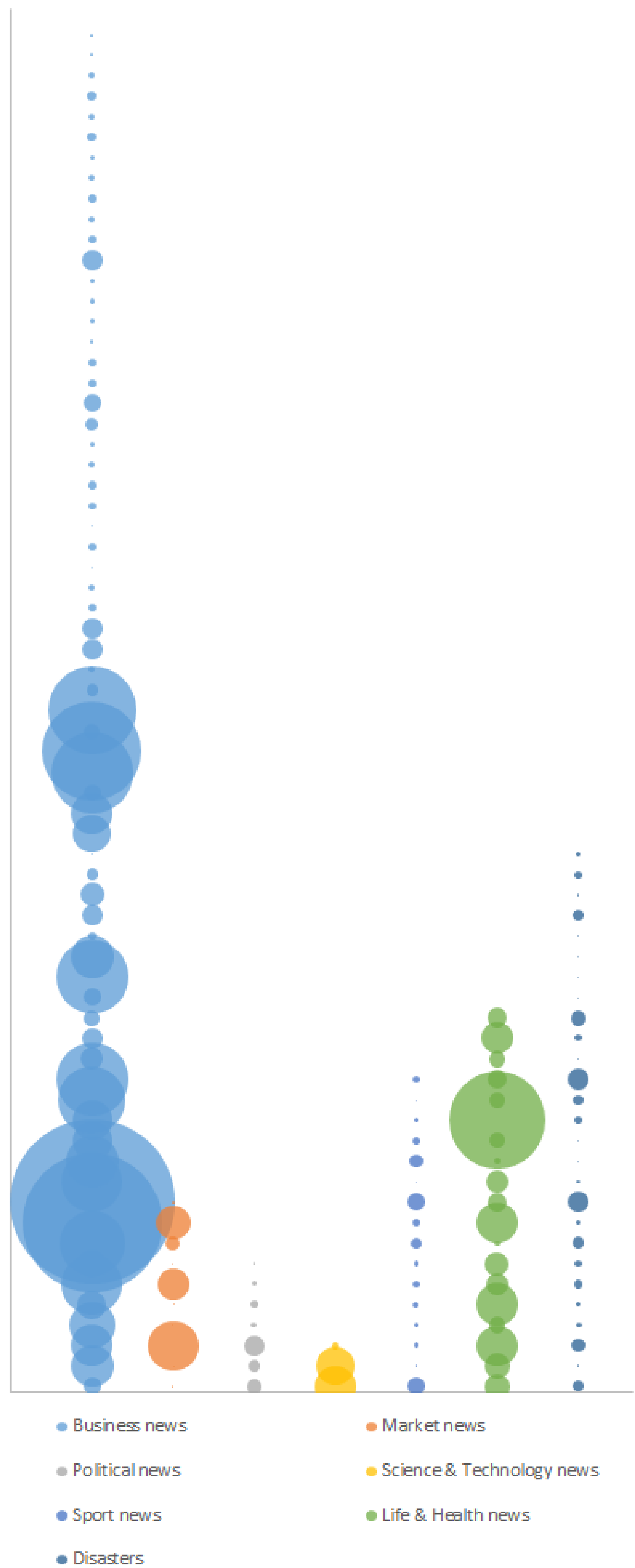


Figure 1.5. Individual bubbles show the number of identified news stories per sub-topics color coded by the main news topics.

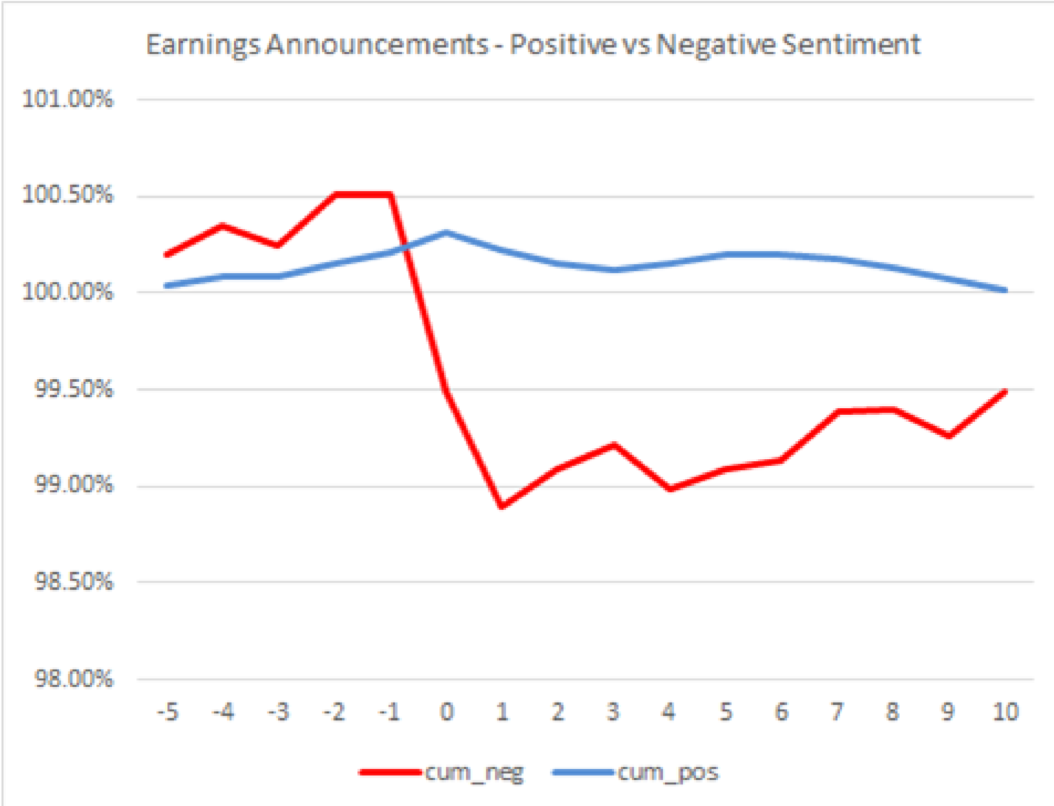


Figure 1.6. Red line shows the average value index of positive news on earnings announcements, while blue line shows the average value index of negative news on earnings announcements. Both lines are from 5 days before the news to 10 days after the news for S&P 500 companies.

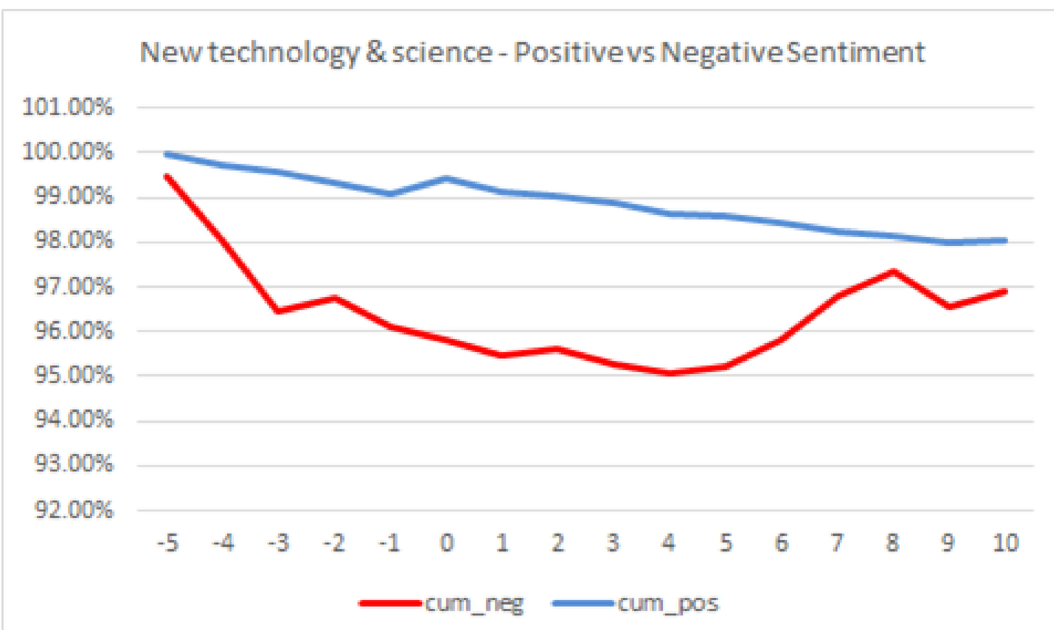


Figure 1.7. Red line shows the average value index of positive news on new technology, while blue line shows the average value index of negative news on new technology. Both lines are from 5 days before the news to 10 days after the news for S&P 500 companies.

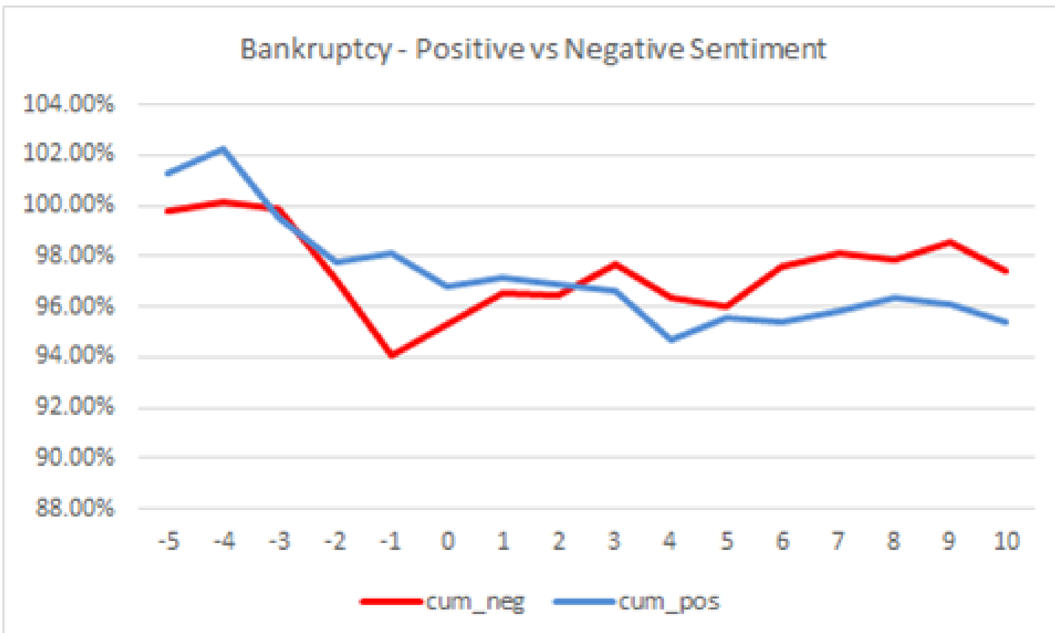


Figure 1.8. Red line shows the average value index of positive news on bankruptcy, while blue line shows the average value index of negative news on bankruptcy. Both lines are from 5 days before the news to 10 days after the news for S&P 500 companies.

Chapter 2

Trading Around EAs

This chapter on earnings announcements and post-earnings announcement drift is based on a draft written jointly with Pedro Saffi, Judge Business School, University of Cambridge.

I. Introduction

Earnings announcements have been studied for more than 50 years and researchers have focused on a number of aspects of the topic. We were able to identify the earnings announcement premium, which is one of the most important and significant asset pricing anomalies. Existing research shows that this premium cannot be simply explained by existing models. The reason why this premium exists is provided by a number of authors. Frazzini and Lamont (2007) argue that the premium is driven by small investors who trade when they see the announcements. On the other hand, Cohen, Dey, Lys, and Sunder (2007) point out that the premium is persistent due to limits to arbitrage.

Our goal is to contribute to this discussion by investigating whether the quality and nature of released information, or news, predict investors' decisions. We have decided to focus on earnings announcements as this area is already well described and hence it is clear what we expect. As it also provides quantitative description of announcements with earnings being quantitative in nature, we can easily distinguish whether investors were positively or negatively surprised. Our task is then to show whether additional information incorporated in textual news have an effect on investors' reactions to this surprise. We utilize our news database as it provides a wide range of news characteristic variables. In this research, we focus on news sentiment, relevance and ambiguity.

We do not aim to develop our own theoretical model, but rather focus on testing existing models on information asymmetry and heterogeneous beliefs around firm news announcements. Our main research question is:

RQ: How does the nature of textual news on earnings announcements predicts the reaction of stock markets to those announcements?

In a nutshell, we aim to investigate whether news-descriptive variables play a role in explaining the cumulative abnormal returns around earnings announcements, controlling for earnings surprise (the difference between what was expected and actual data). The theoretical background for our hypotheses is defined by Tetlock's model for asymmetric information (defined in Tetlock (2010)), models of heterogeneous interpretation of news (see for example Kandel and Pearson (1995)) and models combining both the information asymmetry and heterogenous beliefs represented by Saffi's model (defined in Saffi (2007)). What all models have in common is the conclusion that either heterogeneous beliefs or information asymmetry is responsible for the price drift and behaviour of stock prices before and after earnings announcements. Our goal is not to decide what causes the price change, but rather to investigate whether we can better explain the price drift by defining characteristic variables of textual news around announcements released by major news service companies. We have defined the following hypotheses in order to investigate whether the nature of news on earnings announcements has any effect on subsequent stock market reaction:

H1: News sentiment impacts stock price reaction on earnings announcements positively.

H2: News relevance impacts stock price reaction on earnings announcements positively.

H3: News ambiguity impacts stock price reaction on earnings announcements negatively.

A. Related Research

As proposed research relates to both (1) research on news and their impact on the stock market and (2) research on earnings announcements and market reactions around them, the related literature can be split into two main streams. The first stream represents the literature focused on news analysis and interpretation in relation to stock market data. Key studies in this field include Antweiler and Frank (2006), Tetlock (2007), Fang and Peress (2009), Groß-Klußmann and Hautsch (2011), Smales (2014) and Mamaysky and Glasserman (2015). Studies in this field are related to our research from the perspective of textual analysis. We are studying unstructured textual news on earnings announcements, which were obtained from the whole universe of news by news topic identification in the spirit of Antweiler and Frank (2006), who, however, used less precise methodology. We measure the sentiment of earnings announcements, which relates to studies on sentiment (e.g., Tetlock (2007), Tetlock, SAAR-TSECHANSKY, and Macskassy (2008), Ferguson, Guo, Lam, and Philip (2011a), Nagar and Hahsler (2012), Garcia (2013), Li (2006) and Sinha (2016)), we also measure the novelty of news in the spirit of Tetlock (2011). Two measures not defined in previous research are our proxies for the relevance of news and the ambiguity of news. The relevance was only measured by researchers using a black-box software to obtain their news characteristics (RavenPack used by (von Beschwitz, Keim, and Massa (2013), Kelley and Tetlock (2013), and Smales (2014)), Reuters NewsScope used by (Groß-Klußmann and Hautsch (2011)) or other similar software like The Stock Sonar used by Boudoukh, Feldman, Kogan, and Richardson (2013)).

The second stream is the literature on earnings announcements. Research on this this topic started more than 50 years ago with **?** with other researchers following (e.g., Aharony and Swary (1980), Bernard and Thomas (1990) and Krinsky and Lee (1996)). Since then, researchers have focused on a number of aspects of earnings announcements and how markets react around them. Some have focused on theoretical aspects, like Kim and Verrecchia (1994) or Barberis, Shleifer, and Vishny (1998), others on empirical aspects. Christophe, Ferri, and Angel (2004) studied short selling before earnings announcements. Landsman, Maydew, and Thornock (2012) studied how the adoption of IFRS affects earnings announcements. Barth and So (2014) focused on non-diversifiable risk and found that some announcements contain such risk. Savor and Wilson (2016) looked at systematic risk around announcements, while Leuz, Nanda, and Wysocki (2003) looked at their quality. Other authors focused on post-earnings announcement drift (e.g., Hung, Li, and Wang (2014) who focused on global markets or Bartov, Radhakrishnan, and Krinsky (2000) and Chen, Huang, and Jiang (2016) who studied the role of institutional investors in PEAD).

An interesting view is provided by Drake, Roulstone, and Thornock (2012); the authors used Google search data around announcements as a proxy for information demand around them.

More closely related to this research, DeFond, Hung, and Trezevant (2007) and Landsman and Maydew (2002) have analysed the information content of earnings announcement. However, they used financial data, rather than unstructured news. Baker and Wurgler (2006) studied investor sentiment constructed from financial data and its relation with abnormal return around announcements. Bird, Choi, and Yeung (2014) went even further and studied the role of market sentiment in PEAD; their study was conducted on the market as whole, rather than isolated sentiment of each individual announcement.

Despite an extensive existing body of literature, we are not able to explain the premium around earnings announcements. We believe that the combination of news and earnings announcement events is able to shed some light on this issue. Existing studies have examined the information content of earnings announcements (DeFond, Hung, and Trezevant (2007) and Landsman and Maydew (2002)), but they used financial data, rather than textual data. We argue that the use of textual data is a superior proxy in terms of which information is used and, more importantly, how it is shared. We proxy the quality of information by sentiment, ambiguity, novelty and relevance measures, which is in the line of existing theoretical models like Barberis, Shleifer, and Vishny (1998), who model investor sentiment around news events such as earnings announcements and the subsequent over- or under-reaction of stock prices. A clear contribution is not only the combination of unstructured news and earnings announcement data, which has not been done in the past, but the use of the news ambiguity measure, which is a completely novel news characteristic measure, not used in the previous research.

The rest of this chapter is organized as follows. In the next section, we describe the news database and variables and how we measure them. This is followed by Section II, where we describe the variables and the methods we use to calculate them. This is followed by Section III, which describes the data and contains some preliminary tests. The main section of this research is the Section IV, where we present the results of our event study. The last section draws conclusions about this research and provides avenues for future work.

II. Data

This study uses a news database based on the Reuters News data from January 2007 through December 2015. We have developed the above mentioned variables in order to describe news quantitatively. We utilize the Compustat and CRSP databases to identify companies in the news and assign them appropriate unique identifiers. We filter out news with the identified companies and earnings announcement topic assigned. Based on the numbers and release patterns of identified earnings announcements, we can clearly see two things. First, earnings announcement news is cyclical, with peaks every quarter as quarterly results are released. Second, their number, unlike all news, does not increase with time. This second observation is important as we can argue that there is no structural change in the number of earnings announcement news articles between the start of our period in 2007 and its end in 2015.

In order to put our data in perspective, the whole universe of news is more than 8 million news articles, and from those we were able to identify more than 5 million that mentioned at least one company at least once. We were able to identify topics in more than 2.5 million news stories and, out of those 2.5 million news stories, we identified 142,380 as earnings announcements. It is important to note here that although we have announcements for all of the companies that are included in the Compustat US and International file, in this study we focus only on US companies.

We combine news data with data retrieved from the I/B/E/S database. Our I/B/E/S sample covers the same period as our news sample. The database provides us with data on the date and time of earnings announcements, actual EPS as well as analysts' forecasts and dispersion right before the announcement. We merge this data with our news database in order to match analysts' forecasts and actual events as recorded in I/B/E/S to the news releases and news variables we calculated. Merging two databases (with I/B/E/S containing annual earnings only), we have a set of 9,232 observations (for the period from 2007 to 2015, or 1025 observations a year), while the original sample provides 142,380 news articles on earnings announcements from 2007 through 2015 (or 15,820 observations per year).

Once the data on news and earnings are merged, we combine those with appropriate CRSP data obtained from the WRDS. We use the WRDS Event Study toolkit to retrieve abnormal returns. We use the Fama and French 3-factor model with momentum for the abnormal return calculation. Using the combination of the aforementioned three data sets, we derive all variables needed for our analysis.

A. Variables

Earnings Surprise

Earnings surprise is the variable used in the past to measure the impact of earnings announcements on markets. Existing research introduced two methods for earnings surprise calculation. The first method is based on a time series model. This model uses past earnings data to calculate the surprise, subtracting the past year's earnings from recent earnings and scaling it to the stock price reported a few days before the earnings announcement. This method was used in the past by researchers who studied the post-earnings announcement drift, such as (Ball and Brown, 1968) and (Brown and Pope, 1996).

We will, however, use a method used in recent studies which used analysts' forecasts to calculate earnings surprise (e.g., (Mendenhall, 2004), (Truong, 2010) or (Truong, 2011)). Both methods were compared by (Livnat and Mendenhall, 2006), who found that the post-earnings announcement drift is larger when estimating the earnings surprise using analysts' forecasts. They also argue that both methods capture different phenomena (different forms of mispricing), leaving the decision on how to calculate earnings surprise inconclusive.

In our study we use both methods to test their performance and the results they yield. Based on the analysts' forecasts, we calculate the Standardized unexpected earnings or $SUEAF$. We define it as actual earnings less the median of analysts' forecasts of expected earnings divided by the standard deviation of analysts' forecasts. The stock price 10 days prior to the earnings announcement can be substituted for the standard deviation of forecasts without any change of results as demonstrated by (Truong, 2011). The equation for the earnings surprise calculation then becomes:

$$SUEAF_{j,t} = \frac{E_{j,t} - F_{j,t}}{P_{j,t}} \tag{2.1}$$

where $SUEAF_{j,t}$ is the earnings surprise measure for stock j in time t, $E_{j,t}$ is the earnings per share reported for stock j in time t, $F_{j,t}$ is the median of analysts' forecasts of earnings made by analysts prior to the earnings announcements for stock j in time t and the $P_{j,t}$ is the price of stock j 10 days before the earnings announcement in time t.

The second method to calculate earnings surprise is the time series model. In this case, the unexpected earnings based on the earnings a year ago are calculated. Following existing research, we calculate the unexpected earnings surprise or SUE as:

$$SUE_{j,t} = \frac{E_{j,t} - E_{j,t-1}}{P_{j,t}} \tag{2.2}$$

where $SUE_{j,t}$ is the earnings surprise measure for stock j in time t, $E_{j,t}$ is the earnings per share reported for stock j in time t, $E_{j,t-1}$ is the earnings per share measure reported for stock j in time $t-1$ (a year before the time t) and the $P_{j,t}$ is the price of stock j 10 days before the earnings announcement in time t.

Cumulative Abnormal Returns

Abnormal return, Cumulative abnormal return and Post-earnings announcement drift are variables based on stock returns indicating how much investors earn abnormally (or how high alpha investors get) if they invest in stock around its earnings announcement. These are important measurements when related to earnings surprise, news data and other controlling variables in a single model. In this way, we will test our news variables and whether they are able to explain the abnormal returns. The abnormal volume measures the abnormal trading activity around earnings announcements.

The abnormal return or alpha is calculated from realized returns and the difference between them and what is predicted by asset pricing models. In this paper, we use the abnormal return based on Fama and the French 3-factor model. Table 2.1 shows the calculations of this and alternative models. We have decided to go for a middle ground as the CAPM model is rather too simplistic and using more than a handful of asset pricing factors may be counterproductive (as some of them may already capture news flow to a certain extent).

[Insert Table 2.1 about here]

As seen in the Table 2.1, we need stock returns, the risk-free rate, market returns and, depending on the model, β , s, h and c as variables. Normal procedure would be to estimate needed variables using the regression with market return, risk free rate data and data on SMB, HML and CMA (these are available at French's website). In our case, we utilize the WRDS Event Study toolkit to retrieve abnormal returns for all considered models. WRDS ES also provides cumulative abnormal returns, but for greater flexibility on the time windows, we decided to calculate these manually from individual daily abnormal returns.

Cumulative abnormal return is the abnormal return over a period of time, it is hence based on realized returns less expected returns, as predicted by appropriate asset pricing models. We calculate the cumulative abnormal return as the sum of abnormal returns in the time window of interest:

$$CAR_{j,t} = \sum_{i=-n}^{m} R_{j,i} - E[R_{j,i}] \tag{2.3}$$

where $CAR_{j,t}$ is the cumulative abnormal return for stock j in time t, $R_{j,i}$ is the realized return for stock j in time i and $E[R_{j,i}]$ is expected return for stock j in time i calculated using one of asset pricing models as defined in Table 2.1.

III. Descriptive Statistics

In this section, we look at news around earnings announcements and their properties. We decided to split announcements into deciles and terciles based on the earnings surprise ($SUEAF$) followed by split to terciles based on news variables (sentiment, ambiguity and relevance). As a part of a standard exercise, we start by plotting the average cumulative abnormal returns split by earnings surprise both for highest and lowest SUEAF decile and high and low SUEAF tercile. These are available in Figure 2.1 in the case of SUEAF deciles and Figure 2.2 in the case of SUEAF terciles. We have to conclude that these figures are in line with our expectations, in which high (positive) earnings surprise yields positive cumulative abnormal return after earnings announcement, while low (negative) earnings surprise yields a loss.

[Insert Figure 2.1 about here]
[Insert Figure 2.2 about here]

In order to further illustrate the nature of our data, we have decided to report full descriptive statistics of decile and tercile splits by SUEAF. We are reporting average values for SUEAF, sentiment, relevance, ambiguity and novelty, as well as a number of average cumulative abnormal returns per each decile or tercile. We also report the standard error and T-stats for mean. Table 2.2 shows decile split by SUEAF. It is evident that in a number of variables these splits do not matter and average value stays constant. This is true for relevance and novelty, while the average ambiguity is also similar across all deciles. For SUEAF, we see a different story, with the lowest decile having an average of -2.8 and highest decile an average of 0.4. Deciles 2 to 9 have an average SUEAF of 0.0. Sentiment has lower averages for extreme deciles, with the lowest decile having an average sentiment of 1.9 and the highest decile an average of 1.7. All other deciles have higher average sentiment values ranging from 2.3 to 4.0, with middle deciles having the highest average. As for cumulative abnormal returns, we see relatively low returns for the lowest decile and high returns for the highest decile. Data presented in tables paint a picture consistent with our expectations that bad surprise or low SUEAF yields negative returns, while positive surprise or high SUEAF yields positive results. It is also important to note that CAR (-20,30) is not consistent with this as the lowest decile has higher average cumulative return than relatively higher deciles, we argue this is caused by an overly long period considered before the announcement date, in this case 20 days, simply cancelling out negative effects of bad surprise.

[Insert Table 2.2 about here]

A similar picture is painted by the tercile SUEAF split reported in Table 2.3. We see some variables being constant on average across all three terciles, this is true for relevance and novelty and to a certain extent to ambiguity, as well with very similar average values. Sentiment values are similar for low and high tercile and higher for middle tercile, this is consistent with data presented in the case of decile split. Unlike in the case of decile split, all cumulative abnormal return averages are in line with expectations, with low terciles yielding the lowest return and high terciles yielding the highest return. If we compare extreme deciles to low and high terciles, we see that the average earnings surprise (SUEAF) is much lower on average for low and high terciles than for the lowest and highest deciles. This is expected as we are splitting events into three equal groups rather than 10, meaning they converge to the middle.

[Insert Table 2.3 about here]

Tables and figures presented in this section are in line with expectations and only illustrate the earnings surprise captured by each decile or tercile of our data as well as the nature of news, which seems to be similar across all splits, especially for the lowest and highest deciles. Low and high terciles, as for those of all variables, including sentiment, are the same on average. It also shows the average cumulative abnormal return being lower for low splits and higher for high splits. Please note that the CAR in these and all following tables is reported in the form of a value index, where 1 means 0% cumulative abnormal returns and numbers other than 1 imply either negative (lower than 1) or positive (higher than 1) return.

IV. Results

In the previous section we described data from the perspective of earnings surprise. This showed expected picture of positive surprise yielding positive abnormal returns and negative surprise yielding negative abnormal returns, while news descriptive statistics being relatively similar. In the following section, we will be discussing figures and tables of further splits by news characteristic variables and reporting on average cumulative abnormal returns based on the Fama and French 3-factor model with momentum retrieved from the WRDS event study tool. Please note that in order to keep our results presentable and as clear as possible, we have decided to report only splits that are of interest to us.

A. SUEAF - sentiment split

Our first splits are based on earnings surprise and sentiment. Table 2.4 reports decile SUEAF split and tercile sentiment split. From 30 total groups, we decided to report on the lowest and highest SUEAF deciles and for each of them the low and high sentiment tercile. It is evident that some variables stay constant across all four reported groups. These are relevance, ambiguity and novelty. This is in line with what we would expect as our splits were on SUEAF and sentiments only. What is interesting is that while both sentiment terciles have a similar SUEAF in the case of the highest SUEAF decile, the lowest SUEAF decile shows a completely different story. In this case, the low sentiment tercile has an SUEAF of -0.3, while high sentiment tercile has an SUEAF of -7.8. This essentially implies that news about earnings announcements that missed their expectations by a large margin are much more positive on average, compared to news that missed their expectations only slightly. It is important to point out that our sentiment is adjusted for text size, how common the phrase in news on earnings announcements is, as well as how many times it is repeated in the text. This essentially means sentiment is not affected by anything other than news being positive or negative about the firm.

[Insert Table 2.4 about here]

When we look at the average cumulative abnormal returns as presented in Table 2.4 or in Figure 2.3. We can see a low sentiment tercile in the case of the highest SUEAF decile yielding much higher positive CAR compared to the high sentiment tercile. This is contrary to our expectations of higher positive sentiment, yielding higher returns compared to lower or negative sentiment. The same applies to the lowest SUEAF decile; in this case, the low sentiment tercile yields higher returns compared to the high sentiment tercile. This is again in the opposite direction to our expectations; we would expect low or negative sentiment to yield lower returns, while high or positive sentiment should yield higher returns.

[Insert Figure 2.3 about here]

A similar story is painted in Table 2.5 and in Figure 2.4. In this case, we split SUEAF into terciles and then did another split to terciles by sentiment. We see two things here. First, relevance, ambiguity as well as novelty is constant across all reported groups. SUEAF is lower, however, for the low sentiment group compared to the high sentiment group for the low SUEAF tercile. In the case of the high SUEAF tercile, we see a story similar to Table 2.4, where the low sentiment group has a higher SUEAF compared to the high sentiment group. Similarly, in the Figure 2.4, we can see that for the high SUEAF tercile, the low sentiment tercile yields on average higher cumulative abnormal returns compared to the high sentiment tercile. In the case of the low SUEAF tercile, we see almost exactly the same returns for both reported sentiment terciles. We argue that the tercile split is in line with results reported in the decile split and contrary to the expectations that higher sentiment would yield higher abnormal returns for a given earnings surprise.

[Insert Table 2.5 about here]
[Insert Figure 2.4 about here]

B. SUEAF - relevance split

The second split is based on earnings surprise and the relevance of news. In this case we look at SUEAF deciles and terciles and subsequent split by news Relevance. The expectation is that the more relevant news should strengthen the effect of earnings surprise. For simplicity we have decided in this case to report results only in figures as this split is informative and indicative. We report the SUEAF decile split and the relevance tercile split in Figure 2.9. As is visible in the figure, the results are consistent with our expectations of higher relevance in the case of the lowest SUEAF decile meaning lower average cumulative abnormal returns, while higher relevance in the case of the highest SUEAF decile means higher returns.

[Insert Figure 2.5 about here]

In the case of the SUEAF tercile split and the relevance tercile split, the story is different. In this case, as is visible in the Figure 2.10, the high relevance tercile of the high SUEAF tercile has a lower average cumulative abnormal return than the low relevance tercile of the same SUEAF tercile. In the case of the low SUEAF, both relevance terciles yield similar returns. This implies that other phenomena prevail in the case of the SUEAF tercile split. This may be due to fact that we are looking at less extreme surprises in terms of earnings compared to the decile splits.

[Insert Figure 2.6 about here]

C. SUEAF - ambiguity split

The last indicative results based on a two-way split is the split by SUEAF and ambiguity. In this case, we expect a reverse relationship: the lower the ambiguity, the higher the effect of earnings surprise. As we can see in Figure 2.7, our expectations are not met at all, while the highest earnings surprise decile yields higher returns for the high ambiguity tercile compared to the low ambiguity tercile. The opposite is true for the lowest SUEAF decile: it yields lower returns for the high ambiguity tercile and relatively higher returns for the low ambiguity tercile. This would support the argument that higher ambiguity multiplies the effect of earnings surprise and hence supports irrational behaviour of investors around ambiguous events. One way or another, the results are not in line with expectation nor with the traditional rationale that the weaker the news, lower its effect on the stock market.

[Insert Figure 2.7 about here]

The SUEAF tercile split shows a similar story. If we split it by earnings surprise to terciles and then by ambiguity to terciles, we see similar yields for low and high ambiguity terciles for both low and high SUEAF terciles. As presented in Figure 2.8, the results are once again not in line with our expectations; we would expect a difference in returns if we split the SUEAF further by ambiguity.

[Insert Figure 2.8 about here]

D. SUEAF - sentiment - relevance split

In addition to the two-way splits by SUEAF and news characteristic variables, we have decided to create a three-way split based on SUEAF, sentiment and relevance. Earnings surprises as well as news characteristic variables are in line with expectations, and novelty and ambiguity are constant through all reported groups for both SUEAF decile and SUEAF tercile splits. As visible in Table 2.6, the average novelty as well as ambiguity measurements are relatively stable across all groups. On the other hand, SUEAF, sentiment and relevance differ based on the group due to the split of these three variables. It is important to note that standard errors are relatively low and the T-statistics of the mean are significant for all three split variables. The same applies to Table 2.7 describing the SUEAF tercile split.

[Insert Table 2.6 about here]
[Insert Table 2.7 about here]

When we look at average cumulative abnormal returns, we see a story that is not in line with expectations. We would expect the lowest returns for the low SUEAF, low sentiment, high relevance group and highest returns for high SUEAF, high sentiment and high relevance. As is it visible both in Tables and Figures, the lowest returns are for the low SUEAF and high sentiment groups, while the highest returns are for the high SUEAF, low sentiment and low relevance groups. The same applies to long-term CAR (CAR(-20,30)) as presented in the Figure 2.9 and Figure 2.10 as for the shorter-term CARs presented in Table 2.6 and Table 2.7.

It is clear that both the SUEAF decile and SUEAF tercile splits do not meet our expectations. In the case of the SUEAF decile split, the lowest abnormal returns are reported for low SUEAF, high sentiment and high relevance, while the low SUEAF, low sentiment and high relevance, which are expected to have the lowest return, yield the highest return. In the case of the SUEAF tercile split, we see the lowest return in the case of low SUEAF, low sentiment and high relevance. This is in line with our expectations, but is not significantly different from other groups like low SUEAF, mid-sentiment and mid-relevance or low SUEAF, high sentiment and high relevance. A similarly counter-intuitive story is present in the case of the high SUEAF decile and tercile split. With the SUEAF decile split, the highest return is yielded by high SUEAF, low sentiment and high relevance group, while high SUEAF, high sentiment and high relevance yield among the lowest positive returns. In the case of SUEAF terciles, the highest return is yielded by high SUEAF terciles split only. This is followed by high SUEAF, high sentiment and high relevance, which is in line with expectations. This, however, is once again not significantly different from high SUEAF, mid-sentiment, mid-relevance or high SUEAF, low sentiment, and high relevance splits.

[Insert Figure 2.9 about here]
[Insert Figure 2.10 about here]

E. Big Picture

In this section, we would like to present the big picture of our results. In a previous part, we discussed the results for individual splits. As those results are relatively complex, we have decided to sum up all splits we did as well as expected compared to real lowest and highest average cumulative abnormal returns. These results are reported in Table 2.8. This table shows splits based on earnings surprise, both decile and tercile splits, as well as subsequent two-way splits based on SUEAF and sentiment, a three-way split by SUEAF, sentiment and relevance as well as a four-way split by SUEAF, sentiment, relevance and ambiguity. As is visible in the Table, the only time expectations that meet the real lowest and highest abnormal returns are for the SUEAF tercile split. In all other cases, groups that are not expected are yielding lower or higher returns compared to what is expected.

[Insert Table 2.8 about here]

We have decided to investigate the four-way split further. As presented in the Figure 2.11, we can see that in the case of low SUEAF, expectations are not met as low SUEAF, low sentiment, high relevance and low ambiguity group yields relatively average abnormal returns. A better story is provided by the high SUEAF tercile, where high SUEAF, high sentiment, high relevance and low ambiguity outperforms the rest. On the other hand, this is based on the CAR(-20,30) measure, and if we limit ourselves to shorter periods before and after the announcement, we can see a different situation, as presented in Table 2.8.

[Insert Figure 2.11 about here]

Based on the results presented in this section and in the previous section, we can see two important results. First, we can see that splits by earnings surprise work well as they provide results consistent with expectations across all tests we conducted. On the other hand, additional splits based on news characteristic variables are somehow limited in results, with SUEAF decile splits being completely the opposite of our expectations and SUEAF tercile splits showing consistent results to a certain limit, especially for positive earnings surprises.

V. Conclusion

In this chapter, we put the work done in Chapter I to a real-life test. We believe that the database might be useful for research, but our event study results presented in the tables and figures in this chapter yielded negative results. Based on our results, we can see that news characteristic variables are unable to provide additional insights for investors in the case of earnings announcements. This is documented by the fact that no significant abnormal return is provided if we control for news characteristic variables on top of earnings surprise variables.

This essentially means that our hypotheses about news characteristic variables as well as the research question are rejected. In order to challenge the hypotheses, we decided to do decile as well as tercile splits. Tercile splits supported our hypotheses only to a limited extent, while decile splits rejected them completely. Based on our results, no abnormal return can be earned on news analysis around earnings announcements. On the other hand, this does not mean that raw returns when looked at from the same perspective as this study looked at abnormal returns would not reveal an alternative story. We argue such research, despite potentially more positive results, would omit phenomena captured by factors included in standard asset pricing models. Looking at what is left after controlling for 'standard' asset pricing factors was our motivation for using abnormal, rather than raw returns. We simply wanted to focus on the unbiased effect of news variables.

Potential avenues for further research are available, and further tests could be done. One example would be regressions explaining cumulative abnormal returns with factors like earnings surprise and news characteristic variables included. We have tried to run a simplified version of those tests and they yielded completely insignificant results and zero R-squared. We also doubt the use of an alternative asset pricing model for abnormal return calculation would change these results. Another alternative would be the use of a different news database or news characteristic variables. Our tests and check performed on the news database itself indicate that this might yield only marginally different results, as all news stories were consistent with their characteristic variables. Ultimately, we believe that further investigation of the topic would not yield positive results and we would like to discourage researchers from doing it.

VI. Appendix: Tables

Table 2.1. Asset Pricing Models: Overview table of standard asset pricing models.

Model	*Equation*
CAPM	$\alpha_{CAPM} = R_{j,t} - (r_{f,t} + \beta_{j,t}(R_{m,t} - r_{f,t}))$
FF3	$\alpha_{FF3} = R_{j,t} - (r_{f,t} + \beta_{j,t}(R_{m,t} - r_{f,t}) + s_{j,t}SMB_t + h_{j,t}HML_t)$
FF3 + mom	$\alpha_{FF3+mom} = R_{j,t} - (r_{f,t} + \beta_{j,t}(R_{m,t} - r_{f,t}) + s_{j,t}SMB_t + h_{j,t}HML_t + c_{j,t}CMA_t)$

Table 2.2. Earnings Announcement news - SUEAF decile split (10):
Following table shows summary statistics for news related to earnings announcement events as well as average cumulative abnormal returns per earnings surprise (SUEAF) deciles. We report the average, the standard error as well as the T-stat of the mean. Please note standard errors and T-stats are reported for top and bottom deciles only, with exception of sentiment and ambiguity.

		Decile split by SUEAF									
Variable	Stat	Lowest	2	3	4	5	6	7	8	9	Highest
SUEAF	Avg	-2.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4
	S.E.	2.5									0.3
	T-stat	-1.0									2.4
Sentiment	Avg	1.9	2.5	2.7	3.0	4.0	3.7	3.6	2.3	2.5	1.7
	S.E.	0.7	0.6	0.3	0.3	0.3	0.3	0.3	0.3	0.5	0.3
	T-stat	-1.4	-0.4	-0.1	0.6	3.7	3.3	2.4	-1.8	-0.7	-4.1
Relevance	Avg	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
	S.E.	0.0									0.0
	T-stat	-2.0									-1.5
Ambiguity	Avg	8.6	8.3	7.3	7.3	7.5	7.1	8.2	7.3	8.3	7.9
	S.E.	0.8	0.8	0.3	0.3	0.3	0.3	0.3	0.2	0.7	0.2
	T-stat	1.0	0.7	-1.8	-1.8	-1.0	-2.2	1.3	-1.9	0.7	0.6
Novelty	Avg	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	S.E.	0.0									0.0
	T-stat	-0.1									-1.2
CAR(-20,30)	Avg	1.00	0.97	0.96	0.98	0.99	0.99	1.02	1.03	1.04	1.06
	S.E.	0.01									0.01
	T-stat	-0.24									4.11
CAR(-15,20)	Avg	0.98	0.97	0.96	0.98	0.99	1.00	1.02	1.03	1.04	1.04
	S.E.	0.01									0.01
	T-stat	-2.18									3.03
CAR(-5,10)	Avg	0.97	0.98	0.97	0.99	1.00	1.00	1.02	1.02	1.03	1.02
	S.E.	0.01									0.01
	T-stat	-4.37									2.77
CAR(0,10)	Avg	0.97	0.97	0.97	0.98	1.00	1.00	1.02	1.02	1.02	1.01
	S.E.	0.01									0.01
	T-stat	-4.51									2.54

Table 2.3. Earnings Announcement news - SUEAF tercile split (3): Following table shows summary statistics for news related to earnings announcement events as well as average cumulative abnormal returns per earnings surprise (SUEAF) terciles. We report the average, the standard error as well as the T-stat of the mean.

Variable	Stat	Tercile split by SUEAF Low	Mid	High
SUEAF	Avg	-0.8	0.0	0.1
	S.E.	0.8	0.0	0.1
	T-stat	-0.8	>99	4.3
Sentiment	Avg	2.4	3.6	2.3
	S.E.	0.3	0.2	0.2
	T-stat	-1.2	4.8	-2.6
Relevance	Avg	0.6	0.6	0.6
	S.E.	0.0	0.0	0.0
	T-stat	-2.0	1.3	0.7
Ambiguity	Avg	8.0	7.4	7.9
	S.E.	0.3	0.2	0.3
	T-stat	0.6	-2.0	0.5
Novelty	Avg	1.0	1.0	1.0
	S.E.	0.0	0.0	0.0
	T-stat	0.8	-0.5	-0.1
CAR(-20,30)	Avg	0.98	1.00	1.04
	S.E.	0.00	0.00	0.01
	T-stat	-5.21	-3.56	6.80
CAR(-15,20)	Avg	0.98	1.00	1.04
	S.E.	0.00	0.00	0.01
	T-stat	-7.14	-1.76	6.29
CAR(-5,10)	Avg	0.98	1.00	1.02
	S.E.	0.00	0.00	0.00
	T-stat	-9.31	1.37	7.54
CAR(0,10)	Avg	0.97	1.00	1.02
	S.E.	0.00	0.00	0.00
	T-stat	-10.14	1.99	7.70

Table 2.4. Earnings
Announcement news - SUEAF decile split, Sentiment terciles (10x3): Following table shows summary statistics for news related to earnings announcement events as well as average cumulative abnormal returns per earnings surprise (SUEAF) top and low deciles and high and low terciles of news Sentiment. We report the average, the standard error as well as the T-stat of the mean.

		Lowest SUEAF		**Highest SUEAF**	
Variable	Stat	Low Sentiment	High Sentiment	Low Sentiment	High Sentiment
SUEAF	Avg	-0.3	-7.8	0.2	0.1
	S.E.	0.1	7.6	0.2	0.1
	T-stat	-0.4	-1.0	7.7	13.3
Sentiment	Avg	-5.0	9.7	-5.3	9.1
	S.E.	0.3	1.8	0.4	0.5
	T-stat	-22.4	3.8	-22.3	13.9
Relevance	Avg	0.5	0.5	0.5	0.6
	S.E.	0.0	0.0	0.0	0.0
	T-stat	-3.1	-3.0	-3.5	-1.5
Ambiguity	Avg	9.1	12.1	9.4	9.8
	S.E.	0.4	2.3	0.4	0.5
	T-stat	3.5	1.9	4.2	4.4
Novelty	Avg	1.0	1.0	1.0	1.0
	S.E.	0.0	0.0	0.0	0.0
	T-stat	0.0	-0.4	1.4	-0.6
CAR(-20,30)	Avg	1.07	0.97	1.14	1.02
	S.E.	0.02	0.02	0.02	0.02
	T-stat	2.57	-1.84	5.33	0.50
CAR(-15,20)	Avg	1.02	0.96	1.09	1.02
	S.E.	0.01	0.01	0.02	0.02
	T-stat	1.19	-3.38	4.06	0.87
CAR(-5,10)	Avg	1.01	0.97	1.04	1.01
	S.E.	0.01	0.01	0.01	0.01
	T-stat	0.55	-3.33	2.76	0.83
CAR(0,10)	Avg	1.00	0.97	1.03	1.00
	S.E.	0.01	0.01	0.01	0.01
	T-stat	0.28	-3.49	2.17	0.51

Table 2.5. Earnings
Announcement news - SUEAF tercile split, Sentiment terciles (3x3): Following table shows summary statistics for news related to earnings announcement events as well as average cumulative abnormal returns per earnings surprise (SUEAF) high and low terciles and high and low terciles of news Sentiment. We report the average, the standard error as well as the T-stat of the mean.

		Low SUEAF		**High SUEAF**	
Variable	Stat	Low Sentiment	High Sentiment	Low Sentiment	High Sentiment
SUEAF	Avg	-0.1	0.0	0.1	0.0
	S.E.	0.0	0.0	0.0	0.0
	T-stat	8.7	17.2	16.6	33.0
Sentiment	Avg	-4.6	10.6	-5.1	10.3
	S.E.	0.2	0.7	0.2	0.4
	T-stat	-42.5	10.6	-5.1	10.3
Relevance	Avg	0.5	0.6	0.5	0.6
	S.E.	0.0	0.0	0.0	0.0
	T-stat	-7.9	0.3	-6.6	1.6
Ambiguity	Avg	8.6	10.9	9.0	10.2
	S.E.	0.2	1.0	0.2	0.7
	T-stat	3.5	3.2	5.1	3.6
Novelty	Avg	1.0	1.0	1.0	1.0
	S.E.	0.0	0.0	0.0	0.0
	T-stat	1.4	-0.5	3.2	-1.3
CAR(-20,30)	Avg	1.00	0.97	1.07	1.03
	S.E.	0.01	0.01	0.01	0.01
	T-stat	-1.10	-4.89	5.59	2.40
CAR(-15,20)	Avg	0.98	0.97	1.05	1.03
	S.E.	0.01	0.01	0.01	0.01
	T-stat	-2.98	-5.28	4.45	3.26
CAR(-5,10)	Avg	0.98	0.98	1.02	1.02
	S.E.	0.00	0.00	0.01	0.00
	T-stat	-3.54	-4.80	3.99	4.40
CAR(0,10)	Avg	0.98	0.98	1.02	1.02
	S.E.	0.00	0.00	0.01	0.00
	T-stat	-4.16	-5.46	3.95	4.58

Table 2.6. Earnings Announcement news -

SUEAF decile split, Sentiment terciles, Relevance terciles (10x3x3): Following table shows summary statistics for news related to earnings announcement events as well as average cumulative abnormal returns per earnings surprise (SUEAF) top and low deciles, high and low terciles of news Sentiment and low and high terciles of Relevance. In total, we report 4 terciles (Low Sentiment, Low Relevance - LS/LR, Low Sentiment, High Relevance - LS/HR, High Sentiment, Low Relevance - HS/LR and High Sentiment, High Relevance - HS/HR) for each decile. We report the average, the standard error as well as the T-stat of the mean.

		Lowest SUEAF				**Highest SUEAF**			
Variable	Stat	LS/LR	LS/HR	HS/LR	HS/HR	LS/LR	LS/HR	HS/LR	HS/HR
	Avg	-0.1	-0.4	-0.2	-30.3	0.2	0.4	0.1	0.1
SUEAF	S.E.	0.0	0.2	0.1	29.8	0.1	0.2	0.1	0.0
	T-stat	3.9	-1.2	1.0	-1.0	6.1	3.0	5.6	7.3
	Avg	-4.6	-5.1	8.7	6.2	-5.5	-4.6	8.9	7.6
Sentiment	S.E.	0.6	0.8	0.8	0.5	0.7	0.6	0.7	0.7
	T-stat	-13.3	-9.4	7.2	6.3	-12.6	-12.0	8.6	6.9
	Avg	0.2	0.9	0.2	0.9	0.2	0.9	0.2	0.9
Relevance	S.E.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	T-stat	-32.1	44.0	-34.0	41.0	-33.1	39.5	-33.5	40.5
	Avg	8.7	8.3	11.2	7.8	9.5	7.8	10.7	7.2
Ambiguity	S.E.	0.6	0.7	1.0	0.6	0.6	0.7	0.9	0.7
	T-stat	1.6	0.7	3.5	0.0	2.9	0.1	3.4	-0.9
	Avg	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Novelty	S.E.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	T-stat	9.6	-1.1	-0.6	-0.4	4.6	9.0	0.0	-0.2
	Avg	1.05	1.11	1.00	0.94	1.15	1.21	0.99	1.02
CAR(-20,30)	S.E.	0.04	0.05	0.03	0.04	0.04	0.05	0.04	0.03
	T-stat	1.10	2.07	-0.19	-1.66	3.70	3.84	-0.49	0.37
	Avg	1.01	1.03	0.98	0.94	1.13	1.12	0.99	1.01
CAR(-15,20)	S.E.	0.02	0.03	0.02	0.03	0.04	0.03	0.03	0.02
	T-stat	0.14	0.95	-1.05	-2.02	3.14	3.39	-0.31	0.24
	Avg	1.02	1.00	0.98	0.96	1.07	1.08	1.00	1.02
CAR(-5,10)	S.E.	0.02	0.02	0.01	0.02	0.02	0.02	0.02	0.02
	T-stat	1.26	-0.13	-2.05	-1.78	3.03	3.48	0.18	0.98
	Avg	1.02	0.97	0.98	0.95	1.06	1.06	0.99	1.01
CAR(0,10)	S.E.	0.02	0.02	0.01	0.02	0.02	0.02	0.02	0.02
	T-stat	1.63	-1.24	-1.81	-2.08	2.83	2.85	-0.25	0.82

Table 2.7. Earnings Announcement news

- SUEAF tercile split, Sentiment terciles, Relevance terciles (3x3x3): Following table shows summary statistics for news related to earnings announcement events as well as average cumulative abnormal returns per earnings surprise (SUEAF) high and low terciles, high and low terciles of news Sentiment and low and high terciles of Relevance. In total, we report 4 terciles (Low Sentiment, Low Relevance - LS/LR, Low Sentiment, High Relevance - LS/HR, High Sentiment, Low Relevance - HS/LR and High Sentiment, High Relevance - HS/HR) for each decile. We report the average, the standard error as well as the T-stat of the mean.

		Low SUEAF				**High SUEAF**			
Variable	Stat	LS/LR	LS/HR	HS/LR	HS/HR	LS/LR	LS/HR	HS/LR	HS/HR
SUEAF	Avg	-0.1	-0.1	-0.1	0.0	0.1	0.1	0.0	0.0
	S.E.	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0
	T-stat	6.8	1.8	8.1	32.9	15.3	15.8	12.9	17.6
Sentiment	Avg	-4.5	-3.9	13.0	7.4	-5.0	-3.7	12.1	8.8
	S.E.	0.2	0.3	1.8	0.3	0.3	0.3	1.2	0.4
	T-stat	-29.8	-19.2	5.6	17.6	-28.3	-20.7	7.9	15.1
Relevance	Avg	0.2	0.9	0.3	0.9	0.2	1.0	0.3	1.0
	S.E.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	T-stat	-60.8	84.4	-51.2	91.8	-57.1	102.0	-47.8	105.0
Ambiguity	Avg	8.9	6.7	14.2	6.5	9.2	6.3	13.7	7.2
	S.E.	0.4	0.4	2.3	0.4	0.3	0.4	2.2	0.4
	T-stat	2.6	-3.0	2.8	-3.3	4.1	-3.8	2.7	-1.3
Novelty	Avg	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	S.E.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	T-stat	8.4	-1.4	0.0	0.6	9.9	0.0	-0.5	0.6
CAR(-20,30)	Avg	1.02	0.98	0.96	0.97	1.06	1.03	1.01	1.03
	S.E.	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01
	T-stat	0.81	-1.08	-3.36	-2.35	3.63	1.44	0.57	1.55
CAR(-15,20)	Avg	1.00	0.97	0.97	0.97	1.06	1.01	1.02	1.02
	S.E.	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01
	T-stat	-0.17	-2.04	-3.44	-2.76	3.54	0.69	1.07	1.73
CAR(-5,10)	Avg	0.99	0.98	0.97	0.99	1.03	1.02	1.01	1.03
	S.E.	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	T-stat	-1.27	-2.14	-4.56	-1.75	3.14	1.72	1.12	2.99
CAR(0,10)	Avg	0.99	0.97	0.97	0.98	1.03	1.01	1.01	1.02
	S.E.	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01
	T-stat	-1.27	-3.33	-4.83	-2.33	3.67	1.70	1.01	3.08

Table 2.8. Lowest and Highest CAR - Expectations versus Reality: Following table shows group of announcements with lowest and highest cumulative abnormal return (CAR) per announcement split and compares expectations with reality. Split is defined by description and by number of deciles/terciles in the brackets. Returns are reportes as CAR values for period 5 days prior to announcement to 10 days after the announcement or CAR(-5,10), with standard errors in the bracket. Descriptions are abbreviated with first letter being L for low, H for high and M for middle (in the case of deciles is M followed by brackets with the number of decile) followed by split description SF for the SUEAF, S for the sentiment, R for the relevance, A for the ambiguity.

		Lowest CAR(-5,10)		**Highest CAR(-5,10)**	
Split	Stat	Expected	Real	Expected	Real
SUEAF (3)	Desc.	LSF	LSF	HSF	HSF
	Value & (S.E.)	0.98 (0.00)	0.98 (0.00)	1.02 (0.00)	1.02 (0.00)
SUEAF (10)	Desc.	LSF	MSF(3)	HSF	MSF(9)
	Value & (S.E.)	0.97 (0.01)	0.97 (0.00)	1.02 (0.01)	1.03 (0.00)
SUEAF-S (3x3)	Desc.	LSF-LS	LSF-MS	HSF-HS	HSF-MS
	Value & (S.E.)	0.98 (0.00)	0.96 (0.00)	1.02 (0.00)	1.03 (0.01)
SUEAF-S (10x3)	Desc.	LSF-LS	LSF-MS	HSF-HS	MSF(9)-MS
	Value & (S.E.)	1.01 (0.01)	0.95 (0.01)	1.01 (0.01)	1.04 (0.01)
SUEAF-S-R (3x3x3)	Desc.	LSF-LS-HR	LSF-MS-MR	HSF-HS-HR	HSF-MS-LR
	Value & (S.E.)	0.98 (0.01)	0.95 (0.01)	1.03 (0.01)	1.04 (0.01)
SUEAF-S-R (10x3x3)	Desc.	LSF-LS-HR	LSF-MS-MR	HSF-HS-HR	HSF-MS-MR
	Value & (S.E.)	1.00 (0.02)	0.94 (0.02)	1.02 (0.02)	1.09 (0.05)
SUEAF-S-R-A (3x3x3x3)	Desc.	LSF-LS-HR-LA	LSF-MS-MR-LA	HSF-HS-HR-LA	HSF-LS-LR-MA
	Value & (S.E.)	0.96 (0.02)	0.94 (0.01)	1.07 (0.01)	1.09 (0.01)
SUEAF-S-R-A (10x3x3x3)	Desc.	LSF-LS-HR-LA	LSF-MS-HR-HA	HSF-HS-HR-LA	MSF(9)-HS-LR-LA
	Value & (S.E.)	0.98 (0.04)	0.86 (0.05)	1.03 (0.02)	1.19 (0.01)

VII. Appendix: Figures

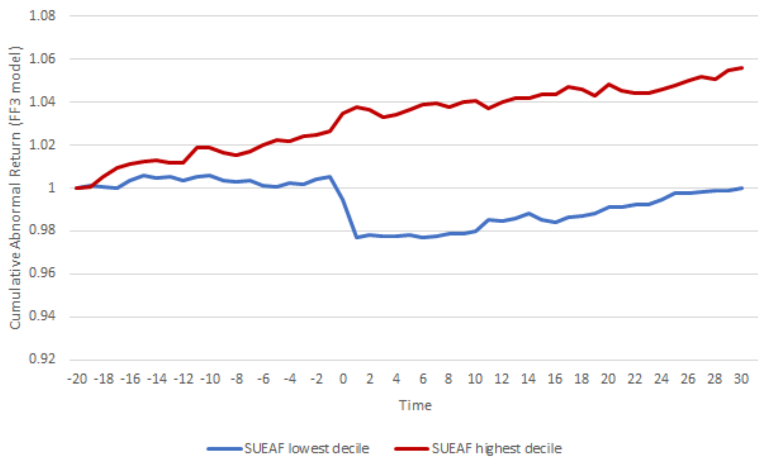


Figure 2.1. Average abnormal
returns around EAs - SUEAF deciles: Average cumulative abnormal return around earnings announcements split by deciles, figure shows top positive SUEAF decile (red) and bottom negative SUEAF decile (blue)

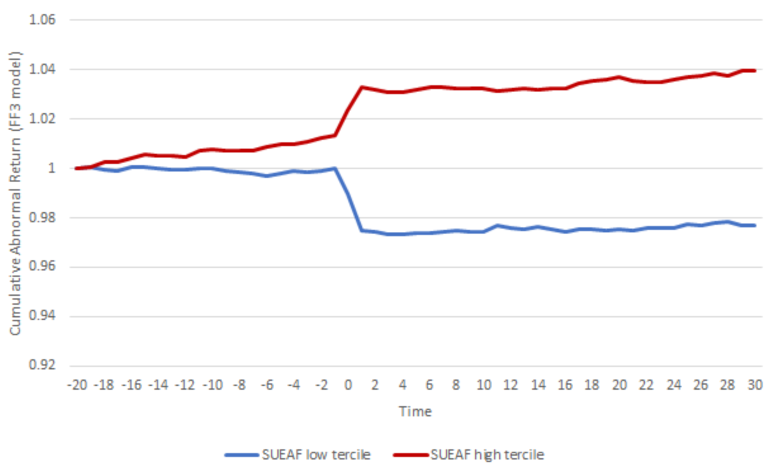


Figure 2.2. Average abnormal
returns around EAs - SUEAF terciles: Average cumulative abnormal return around earnings announcements split by terciles, figure shows top positive SUEAF decile (red) and bottom negative SUEAF decile (blue)

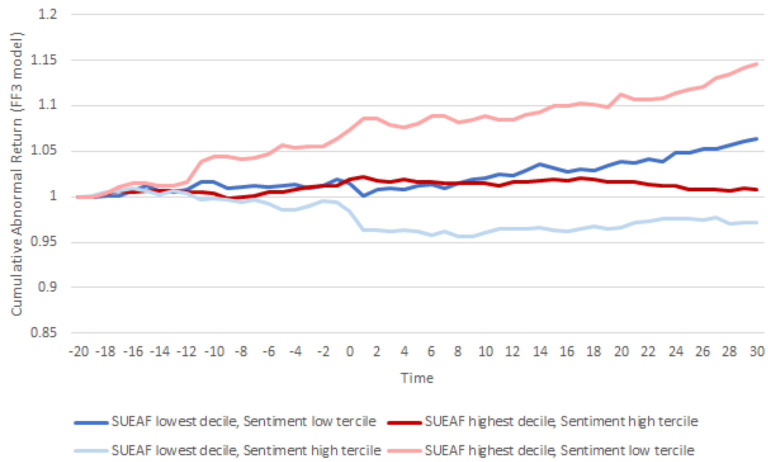


Figure 2.3. Average abnormal returns around EAs - SUEAF + Sentiment 10x3 split - Top/Bottom:
Average cumulative abnormal return around earnings announcements split by SUEAF and sentiment. Figure shows high and low sentiment tercile of top positive SUEAF decile (red lines) and high and low sentiment tercile of lowest SUEAF decile (blue lines)

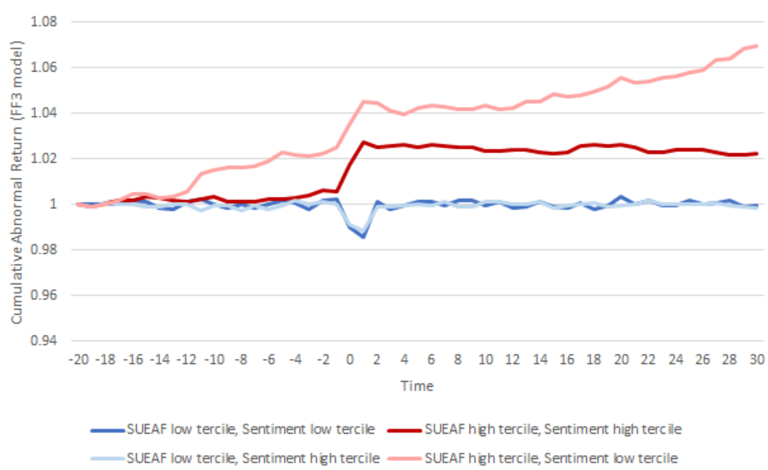


Figure 2.4. Average abnormal returns around EAs - SUEAF + Sentiment 3x3 split - Top/Bottom:
Average cumulative abnormal return around earnings announcements split by SUEAF and sentiment. Figure shows high positive sentiment tercile of high positive SUEAF tercile (red) and low negative sentiment tercile of low negative SUEAF tercile (blue)

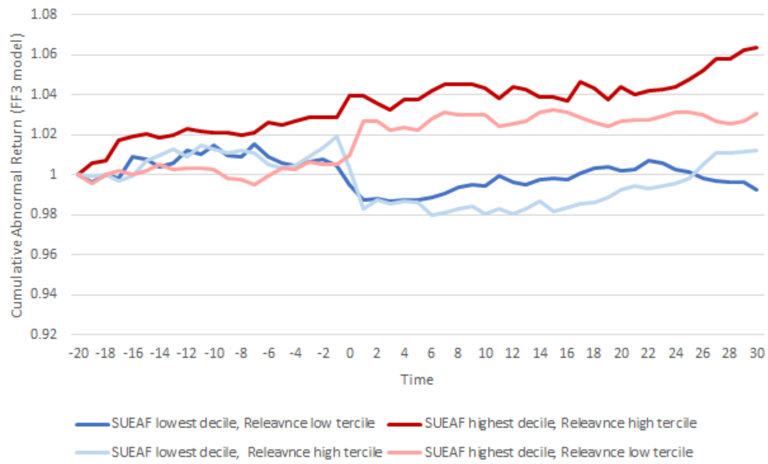


Figure 2.5. Average abnormal returns around EAs - SUEAF + Relevance 10x3 split - Top/Bottom: Average cumulative abnormal return around earnings announcements split by SUEAF and relevance. Figure plots in total four lines. It plots high and low relevance tercile for top SUEAF decile (red lines) and high and low relevance for bottom SUEAF decile (blue lines)

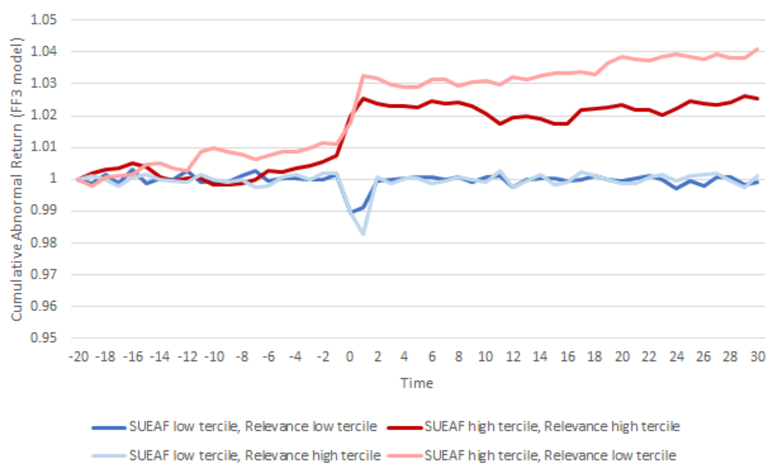


Figure 2.6. Average abnormal returns around EAs - SUEAF + Relevance 3x3 split - Top/Bottom: Average cumulative abnormal return around earnings announcements split by SUEAF and relevance. Figure plots in total four lines. It plots high and low relevance tercile for top SUEAF tercile (red lines) and high and low relevance for bottom SUEAF tercile (blue lines)

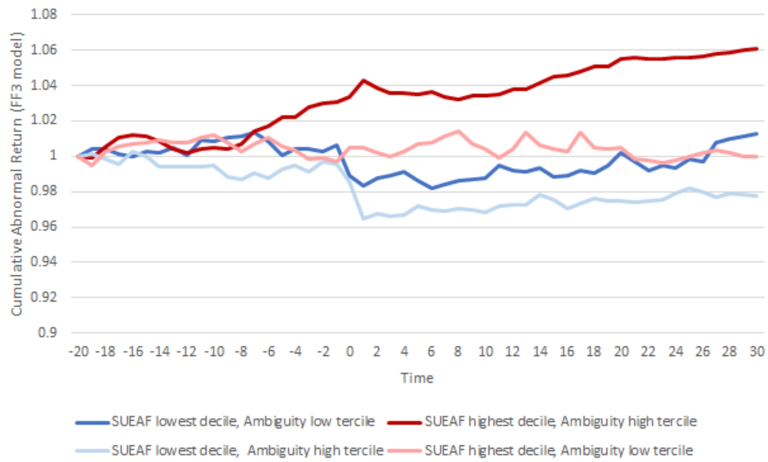


Figure 2.7. Average abnormal returns around EAs - SUEAF + Ambiguity 10x3 split - Top/Bottom: Average cumulative abnormal return around earnings announcements split by SUEAF and ambiguity. Figure plots in total four lines. It plots high and low relevance tercile for top SUEAF decile (red lines) and high and low relevance for bottom SUEAF decile (blue lines)

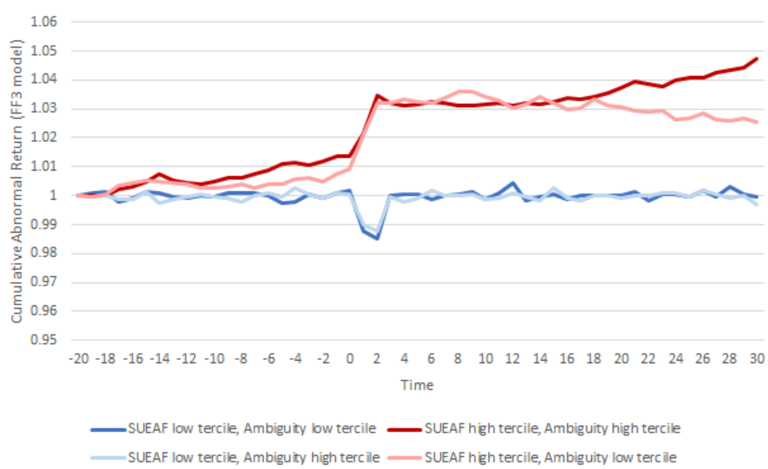
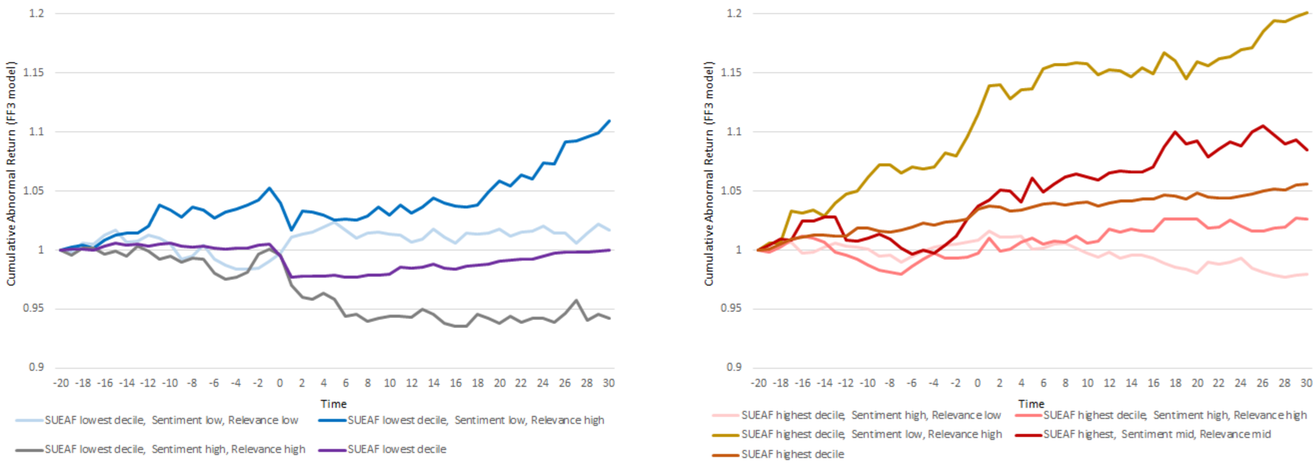
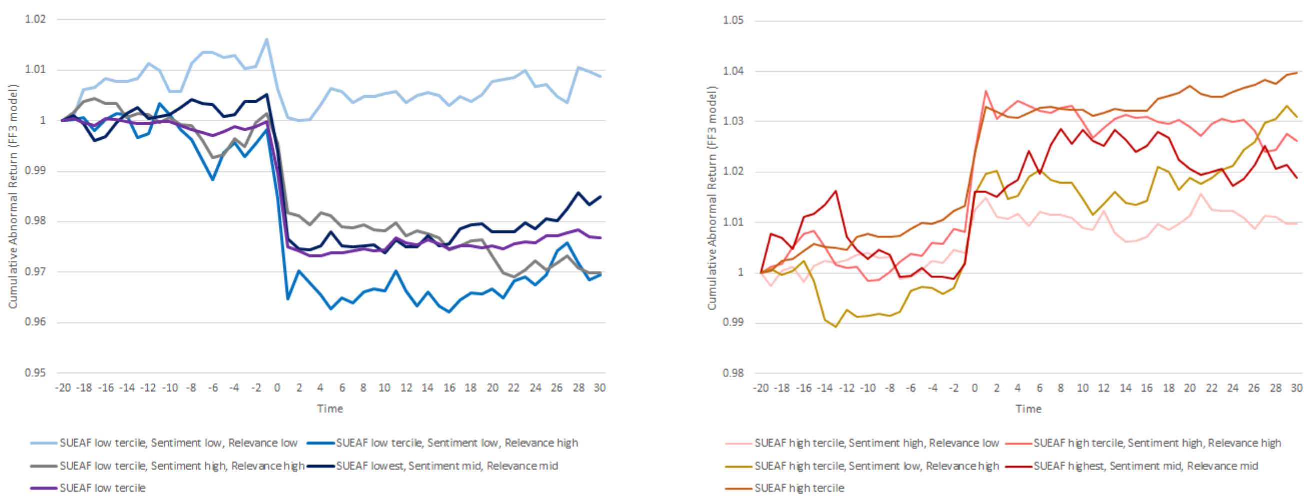


Figure 2.8. Average abnormal returns around EAs - SUEAF + Ambiguity 3x3 split - Top/Bottom: Average cumulative abnormal return around earnings announcements split by SUEAF and ambiguity. Figure plots in total four lines. It plots high and low relevance tercile for top SUEAF tercile (red lines) and high and low relevance for bottom SUEAF tercile (blue lines)



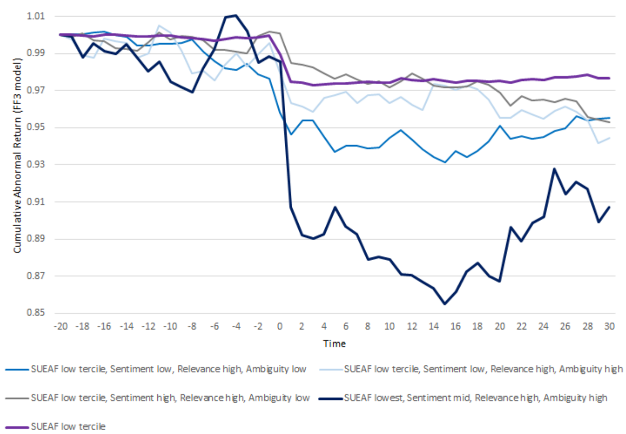
(a) Lowest SUEAF decile **(b)** Highest SUEAF decile

Figure 2.9. Graphs
for a sample of 10x3x3 splits of different Sentiment and Relevance terciles of highest and lowest SUEAF deciles.

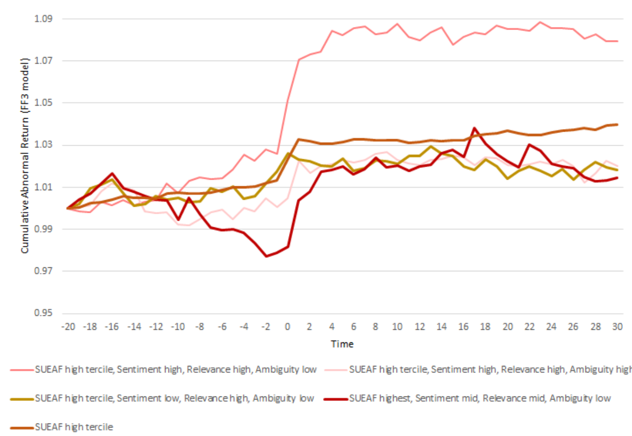


(a) Low SUEAF tercile **(b)** High SUEAF tercile

Figure 2.10.
Graphs for a sample of 3x3x3 splits of different Sentiment and Relevance terciles of high and low SUEAF terciles.



(a) Low SUEAF tercile



(b) High SUEAF tercile

Figure 2.11. Average abnormal returns around EAs - SUEAF + Sentiment + Relevance + Ambiguity 3x3x3x3 split: Graphs for a sample of 3x3x3x3 splits of different Sentiment,Relevance and Ambiguity terciles of high and low SUEAF terciles.

Bibliography

Aharony, Joseph, and Itzhak Swary, 1980, Quarterly dividend and earnings announcements and stockholders' returns: An empirical analysis, *The Journal of Finance* 35, 1–12.

Albaugh, Quinn, Julie Sevenans, Stuart Soroka, and Peter John Loewen, 2013, The automated coding of policy agendas: A dictionary-based approach, in *6th Annual Comparative Agendas Conference, Atnwerp, Beligum*.

Allan, James, Stephen Harding, David Fisher, Alvaro Bolivar, Sergio Guzman-Lara, and Peter Amstutz, 2005, Taking topic detection from evaluation to practice, in *System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on* pp. 101a–101a. IEEE.

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

——— , 2006, Do us stock markets typically overreact to corporate news stories?, *Available at SSRN 878091*.

Baker, Malcolm, and Jeffrey Wurgler, 2006, Investor sentiment and the cross-section of stock returns, *The Journal of Finance* 61, 1645–1680.

Ball, Ray, and Philip Brown, 1968, An empirical evaluation of accounting income numbers, *Journal of accounting research* pp. 159–178.

Bamber, Linda Smith, Orie E Barron, and Douglas E Stevens, 2011, Trading volume around earnings announcements and other financial reports: Theory, research design, empirical evidence, and directions for future research, *Contemporary Accounting Research* 28, 431–471.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny, 1998, A model of investor sentiment, *Journal of financial economics* 49, 307–343.

Barth, Mary E, and Eric C So, 2014, Non-diversifiable volatility risk and risk premiums at earnings announcements, *The Accounting Review* 89, 1579–1607.

Bartov, Eli, Suresh Radhakrishnan, and Itzhak Krinsky, 2000, Investor sophistication and patterns in stock returns after earnings announcements, *The Accounting Review* 75, 43–63.

Beaver, William H, 1968, The information content of annual earnings announcements, *Journal of accounting research* pp. 67–92.

Beber, Alessandro, Michael W Brandt, and Maurizio Luisi, 2015, Distilling the macroeconomic news flow, *Journal of Financial Economics* 117, 489–507.

Bernard, Victor L, and Jacob K Thomas, 1990, Evidence that stock prices do not fully reflect the implications of current earnings for future earnings, *Journal of Accounting and Economics* 13, 305–340.

Bevan, Shaun, 2014, Gone fishing: The creation of the comparative agendas project master codebook, .

Bird, Ron, Daniel FS Choi, and Danny Yeung, 2014, Market uncertainty, market sentiment, and the post-earnings announcement drift, *Review of Quantitative Finance and Accounting* 43, 45–73.

Bollen, Johan, Huina Mao, and Xiaojun Zeng, 2011, Twitter mood predicts the stock market, *Journal of Computational Science* 2, 1–8.

Boudoukh, Jacob, Ronen Feldman, Shimon Kogan, and Matthew Richardson, 2013, Which news moves stock prices? a textual analysis, Discussion paper, National Bureau of Economic Research.

Brown, Stephen J, and Peter F Pope, 1996, Post-earnings announcement drift?, .

Bucchi, Massimiano, 1996, When scientists turn to the public: Alternative routes in science communication, *Public understanding of science* 5, 375–394.

Carhart, Mark M, 1997, On persistence in mutual fund performance, *The Journal of finance* 52, 57–82.

Chan, Wesley S, 2003, Stock price reaction to news and no-news: drift and reversal after headlines, *Journal of Financial Economics* 70, 223–260.

Chen, Linda H, Wei Huang, and George J Jiang, 2016, Herding on earnings news: The role of institutional investors in post-earnings-announcement drift, .

Chisholm, Erica, and Tamara G Kolda, 1999, New term weighting formulas for the vector space method in information retrieval, *Computer Science and Mathematics Division, Oak Ridge National Laboratory*.

Christophe, Stephen E, Michael G Ferri, and James J Angel, 2004, Short-selling prior to earnings announcements, *The Journal of Finance* 59, 1845–1876.

Chung, Kee H, Oliver Kim, Steve C Lim, and Sean Yang, 2014, Information asymmetry, information content, and the measurement of underreaction to earnings announcements, *Robert H. Smith School Research Paper*.

Cohen, Daniel A, Aiyesha Dey, Thomas Z Lys, and Shyam V Sunder, 2007, Earnings announcement premia and the limits to arbitrage, *Journal of Accounting and Economics* 43, 153–180.

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

Davis, Angela K, Jeremy M Piger, and Lisa M Sedor, 2012, Beyond the numbers: Measuring the information content of earnings press release language, *Contemporary Accounting Research* 29, 845–868.

DeFond, Mark, Mingyi Hung, and Robert Trezevant, 2007, Investor protection and the information content of annual earnings announcements: International evidence, *Journal of Accounting and Economics* 43, 37–67.

Drake, Michael S, Darren T Roulstone, and Jacob R Thornock, 2012, Investor information demand: Evidence from google searches around earnings announcements, *Journal of Accounting Research* 50, 1001–1040.

Duarte, Jefferson, and Lance Young, 2009, Why is pin priced?, *Journal of Financial Economics* 91, 119–138.

Easley, David, Soeren Hvidkjaer, and Maureen Ohara, 2002, Is information risk a determinant of asset returns?, *The journal of finance* 57, 2185–2221.

Easley, David, Nicholas M Kiefer, Maureen O'hara, and Joseph B Paperman, 1996, Liquidity, information, and infrequently traded stocks, *The Journal of Finance* 51, 1405–1436.

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

Edmans, Alex, Diego Garcia, and Øyvind Norli, 2007, Sports sentiment and stock returns, *The Journal of Finance* 62, 1967–1998.

Eshghi, Kourosh, and Richard C Larson, 2008, Disasters: lessons from the past 105 years, *Disaster Prevention and Management: An International Journal* 17, 62–82.

Ezzat, Heba, Souraya Ezzat, Samhaa El-Beltagy, and Moustafa Ghanem, 2012, Topicanalyzer: A system for unsupervised multi-label arabic topic categorization, in *Innovations in Information Technology (IIT), 2012 International Conference on* pp. 220–225. IEEE.

Fama, EF, and KR French, 2014, A five-factor asset pricing model, fama-miller working paper, .

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

——— , Lawrence Fisher, Michael C Jensen, and Richard Roll, 1969, The adjustment of stock prices to new information, *International economic review* 10, 1–21.

Fama, Eugene F, and Kenneth R French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of financial economics* 33, 3–56.

Fang, Lily, and Joel Peress, 2009, Media coverage and the cross-section of stock returns, *The Journal of Finance* 64, 2023–2052.

Ferguson, Nicky, Jie Guo, Herbert Lam, and Dennis Philip, 2011a, Media content and stock returns: The predictive power of press, *Multinational Finance Journal*.

Ferguson, Nicky J, Jie Michael Guo, Herbert YT Lam, and Dennis Philip, 2011b, News media content and uk stock returns, Discussion paper, Durham Business School Working Paper.

Fiaidhi, Jinan, Sabah Mohammed, Aminul Islam, Simon Fong, and Tai-hoon Kim, 2013, Developing a hierarchical multi-label classifier for twitter trending topics, *International Journal of u-and e-Service, Science and Technology* 6, 1–12.

Frazzini, Andrea, and Owen A Lamont, 2007, The earnings announcement premium and trading volume, *NBER working paper*.

Garcia, Diego, 2013, Sentiment during recessions, *The Journal of Finance* 68, 1267–1300.

Groß-Klußmann, Axel, and Nikolaus Hautsch, 2011, When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions, *Journal of Empirical Finance* 18, 321–340.

Guha-Sapir, Debarati, David Hargitt, and Philippe Hoyois, 2004, *Thirty years of natural disasters 1974-2003: The numbers* (Presses univ. de Louvain).

Hamilton, James T, 1995, Pollution as news: media and stock market reactions to the toxics release inventory data, *Journal of environmental economics and management* 28, 98–113.

Hirshleifer, David, and Tyler Shumway, 2003, Good day sunshine: Stock returns and the weather, *The Journal of Finance* 58, 1009–1032.

Hung, Mingyi, Xi Li, and Shiheng Wang, 2014, Post-earnings-announcement drift in global markets: Evidence from an information shock, *Review of Financial Studies* p. hhu092.

Jiang, Maojin, and Shlomo Argamon, 2008, Exploiting subjectivity analysis in blogs to improve political leaning categorization, in *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval* pp. 725–726. ACM.

Joachims, Thorsten, 1996, A probabilistic analysis of the rocchio algorithm with tfidf for text categorization., Discussion paper, DTIC Document.

John, Peter, 2006, The policy agendas project: a review, *Journal of European Public Policy* 13, 975–986.

Kandel, Eugene, and Neil D Pearson, 1995, Differential interpretation of public signals and trade in speculative markets, *Journal of Political Economy* pp. 831–872.

Kelley, Eric K, and Paul C Tetlock, 2013, Retail short selling and stock prices, *Columbia Business School Research Paper*.

Kim, Oliver, and Robert E Verrecchia, 1994, Market liquidity and volume around earnings announcements, *Journal of accounting and economics* 17, 41–67.

Krinsky, Itzhak, and Jason Lee, 1996, Earnings announcements and the components of the bid-ask spread, *The Journal of Finance* 51, 1523–1535.

Kurov, Alexander, Alessio Sancetta, Georg Strasser, and Marketa Halova Wolfe, 2015, Price drift before us macroeconomic news: Private information about public announcements, *West Virginia University, University of London, Boston College, and Washington State University*.

Landsman, Wayne R, and Edward L Maydew, 2002, Has the information content of quarterly earnings announcements declined in the past three decades?, *Journal of Accounting Research* 40, 797–808.

——— , and Jacob R Thornock, 2012, The information content of annual earnings announcements and mandatory adoption of ifrs, *Journal of Accounting and Economics* 53, 34–54.

Lau, Cher Han, 2014, Detecting news topics from microblogs using sequential pattern mining, .

Leuz, Christian, Dhananjay Nanda, and Peter D Wysocki, 2003, Earnings management and investor protection: an international comparison, *Journal of financial economics* 69, 505–527.

Lewis, David D, Yiming Yang, Tony G Rose, and Fan Li, 2004, Rcv1: A new benchmark collection for text categorization research, *The Journal of Machine Learning Research* 5, 361–397.

Li, Feng, 2006, Do stock market investors understand the risk sentiment of corporate annual reports?, *Available at SSRN 898181*.

Livnat, Joshua, and Richard R Mendenhall, 2006, Comparing the post–earnings announcement drift for surprises calculated from analyst and time series forecasts, *Journal of Accounting Research* 44, 177–205.

Loughran, Tim, and Bill McDonald, 2011, When is a liability not a liability? textual analysis, dictionaries, and 10-ks, *The Journal of Finance* 66, 35–65.

Madura, Jeff, 2014, *Financial markets and institutions* (Cengage learning).

Mamaysky, Harry, and Paul Glasserman, 2015, Does unusual news forecast market stress?, *Available at SSRN 2632699*.

Medhat, Walaa, Ahmed Hassan, and Hoda Korashy, 2014, Sentiment analysis algorithms and applications: A survey, *Ain Shams Engineering Journal* 5, 1093–1113.

Mendenhall, Richard R, 2004, Arbitrage risk and post-earnings-announcement drift, *The Journal of Business* 77, 875–894.

Mian, G Mujtaba, and Srinivasan Sankaraguruswamy, 2012, Investor sentiment and stock market response to earnings news, *The Accounting Review* 87, 1357–1384.

Mitchell, Mark L, and J Harold Mulherin, 1994, The impact of public information on the stock market, *The Journal of Finance* 49, 923–950.

Molnar, Rastislav, 2016, News database, .

Nagar, Anurag, and Michael Hahsler, 2012, Using text and data mining techniques to extract stock market sentiment from live news streams, in *2012 International Conference on Computer Technology and Science* vol. 47 pp. 91–95.

Niederhoffer, Victor, Steven Gibbs, and Jim Bullock, 1970, Presidential elections and the stock market, *Financial Analysts Journal* pp. 111–113.

Omrane, Walid Ben, and Christian Hafner, 2015, Macroeconomic news surprises and volatility spillover in foreign exchange markets, *Empirical Economics* 48, 577–607.

Pilbeam, Keith, 2010, *Finance and financial markets* (Palgrave Macmillan).

Pozner, Jo-Ellen, Ivana Naumovska, and Edward Zajac, 2014, When do language wrappers matter? market reactions to language use in earnings announcements, in *Academy of Management Proceedings* vol. 2014 p. 16379. Academy of Management.

Romero, Daniel M, Brendan Meeder, and Jon Kleinberg, 2011, Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter, in *Proceedings of the 20th international conference on World wide web* pp. 695–704. ACM.

Sadorsky, Perry, 1999, Oil price shocks and stock market activity, *Energy Economics* 21, 449–469.

Saffi, Pedro AC, 2007, Differences of opinion, information and the timing of trades, *Information and the Timing of Trades (November 2007)*.

Savor, Pavel, and Mungo Wilson, 2016, Earnings announcements and systematic risk, *The Journal of Finance* 71, 83–138.

Sinha, Nitish Ranjan, 2016, Underreaction to news in the us stock market, *Quarterly Journal of Finance* 6, 1650005.

Smales, Lee A, 2014, Non-scheduled news arrival and high-frequency stock market dynamics: Evidence from the australian securities exchange, *Research in International Business and Finance* 32, 122–138.

Sparck Jones, Karen, 1972, A statistical interpretation of term specificity and its application in retrieval, *Journal of documentation* 28, 11–21.

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

———, 2010, Does public financial news resolve asymmetric information?, *Review of Financial Studies* 23, 3520–3557.

———, 2011, All the news that's fit to reprint: Do investors react to stale information?, *Review of Financial Studies* 24, 1481–1512.

———, MAYTAL SAAR-TSECHANSKY, and Sofus Macskassy, 2008, More than words: Quantifying language to measure firms' fundamentals, *The Journal of Finance* 63, 1437–1467.

Tripathi, Nandita, Michael Oakes, and Stefan Wermter, 2013, Hybrid classifiers based on semantic data subspaces for two-level text categorization, *International Journal of Hybrid Intelligent Systems* 10, 33–41.

Truong, Cameron, 2010, Post earnings announcement drift and the roles of drift-enhanced factors in new zealand, *Pacific-Basin Finance Journal* 18, 139–157.

———, 2011, Post-earnings announcement abnormal return in the chinese equity market, *Journal of International Financial Markets, Institutions and Money* 21, 637–661.

Véiliverronen, Esa, 1993, Science and the media: changing relations, *Science studies* 6, 23–34.

von Beschwitz, Bastian, Donald B Keim, and Massimo Massa, 2013, Media-driven high frequency trading: Implications of errors in news analytics, *Working Paper*.

Wasserfallen, Walter, 1989, Macroeconomics news and the stock market: evidence from europe, *Journal of Banking & Finance* 13, 613–626.

Young, Lori, and Stuart Soroka, 2012, Affective news: The automated coding of sentiment in political texts, *Political Communication* 29, 205–231.

Yu, Liang-Chih, Jheng-Long Wu, Pei-Chann Chang, and Hsuan-Shou Chu, 2013, Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news, *Knowledge-Based Systems* 41, 89–97.