CFS Working Paper Series

No. 583

Christina Bannier, Thomas Pauls, and Andreas Walter

CEO-speeches and stock returns

Center for Financial Studies | Goethe University | Theodor-W.-Adorno-Platz 3 | 60323 Frankfurt am Main | Germany
Phone +49 69 798 30050 | Fax +49 69 798 30077 | info@ifk-cfs.de | www.ifk-cfs.de

The CFS Working Paper Series

presents ongoing research on selected topics in the fields of money, banking and finance. The papers are circulated to encourage discussion and comment. Any opinions expressed in CFS Working Papers are those of the author(s) and not of the CFS.

The Center for Financial Studies, located in Goethe University Frankfurt's House of Finance, conducts independent and internationally oriented research in important areas of Finance. It serves as a forum for dialogue between academia, policy-making institutions and the financial industry. It offers a platform for top-level fundamental research as well as applied research relevant for the financial sector in Europe. CFS is funded by the non-profit-organization Gesellschaft für Kapitalmarktforschung e.V. (GfK). Established in 1967 and closely affiliated with the University of Frankfurt, it provides a strong link between the financial community and academia. GfK members comprise major players in Germany's financial industry. The funding institutions do not give prior review to CFS publications, nor do they necessarily share the views expressed therein.

Center for Financial Studies | Goethe University | Theodor-W.-Adorno-Platz 3 | 60323 Frankfurt am Main | Germany
Phone +49 69 798 30050 | Fax +49 69 798 30077 | info@ifk-cfs.de | www.ifk-cfs.de

CEO-speeches and stock returns

Christina Bannier[a] Thomas Pauls[b] Andreas Walter[c]

Abstract – We analyze the market reaction to the sentiment of the CEO speech at the Annual General Meeting (AGM). As the AGM is typically preceded by several information disclosures, the CEO speech may be expected to contribute only marginally to investors' decision-making. Surprisingly, however, we observe from the transcripts of 338 CEO speeches of German corporates between 2008 and 2016 that their sentiment is significantly related to abnormal stock returns and trading volumes following the AGM. Using a novel business-specific German dictionary based on Loughran and McDonald (2011), we find a negative association of the post-AGM returns with the speeches' negativity and a positive association with the speeches' relative positivity (i.e. positivity relative to negativity). Relative positivity moreover corresponds with a lower trading volume in a short time window surrounding the AGM. Investors hence seem to perceive the sentiment of CEO speeches at AGMs as a valuable indicator of future firm performance.

Keywords: Textual sentiment, CEO speeches, market efficiency, textual analysis, annual general meeting

JEL-Codes: G02, G12, G14

[a] Chair of Banking & Finance, University of Gießen, Licher Str. 66, 35394 Gießen, Germany. Christina.Bannier@wirtschaft.uni-giessen.de.

[b] *Corresponding author.* Chair of Financial Services, University of Gießen, Licher Str. 74, 35394 Giessen, Germany. Thomas.Pauls@wirtschaft.uni-giessen.de.

[c] Chair of Financial Services, University of Gießen, Licher Str. 74, 35394 Gießen, Germany. Andreas.Walter@wirtschaft.uni-giessen.de.

1. Introduction

Companies distribute information to relevant stakeholders by various means. Recent research has acknowledged the value not only of quantitative data disclosures but also of qualitative information, predominantly in the form of the textual sentiment of business communication. Sentiment is typically examined via content analyses which have been applied on several types of business communication such as annual reports (Feldman et al., 2008; Jegadeesh & Wu, 2013; Loughran & McDonald, 2011, 2015), earnings press releases (Davis et al., 2012; Davis & Tama-Sweet, 2012; Henry, 2008; Henry & Leone, 2016; Huang et al., 2014), IPO prospectuses (Demers & Vega, 2008; Ferris et al., 2013; Jegadeesh & Wu, 2013), CEO letters (Boudt & Thewissen, 2016), and earnings conference calls (Davis et al., 2015; Doran et al., 2012; Larcker & Zakolyukina, 2012; Price et al., 2012). In general, these studies find that qualitative information is indeed processed by investors and helps to predict future accounting returns, stock returns, stock volatility, and stock trading volume.[1]

Surprisingly, the Annual General Meeting (AGM) received only little attention so far and the CEO's speech held at the AGM hardly any. Only few studies investigate the market reaction to the AGM at all and those that do report inconclusive and partly diverging results. Firth (1981), for example, does not find a market reaction in terms of abnormal returns and trading volume. Brickley (1986) and Rippington and Taffler (1995) report only small price reactions around the AGM for US and UK firms, respectively. Olibe (2002) presents evidence of a minimal trading-volume response to UK companies' AGM. Martinez-Blasco et al. (2015) find no significant market reactions in Japan and Spain and only trading volume increases for US, UK, and French stocks. For German stocks, in contrast, they observe significant market responses to the AGM in terms of increased returns, return volatility and trading volume.

The generally weak market reaction to the AGM may be explained by the fact that the AGM is typically preceded by several information disclosures such as preliminary earnings announcements and the full release of the annual report. As a consequence, the AGM can hardly deliver any new quantitative information. However, to the best of our knowledge, no attempts have been made so far to investigate the qualitative content of the AGM and of the CEO's speech in particular. This is despite the fact that the AGM offers managers the rare opportunity to personally address the company's stockholders in order to share their views on the firm's prospects (Martinez-Blasco et al., 2015).

---

[1] See Kearney and Liu (2014) or Loughran and McDonald (2016) for a comprehensive overview.

The lack of studies on the qualitative content of CEO speeches is particularly surprising, since CEO communication in general has been shown to exhibit valuable qualitative information. For example, Arslan-Ayaydin et al. (2015) find that incentivized managers use positive words more aggressively in an attempt to influence share prices. Similarly, Boudt and Thewissen (2016) report that CEOs strategically present negative and positive words in CEO letters in order to prompt a more positive perception by the reader. Price et al. (2012) and Doran et al. (2012) show that the tone of earnings conference calls - which are typically conducted by the firm's top management team - is a significant predictor of subsequent returns and trading volume. We therefore hypothesize that CEO speeches held at AGMs contain valuable qualitative information that should influence the market reaction to the AGM. As Demers and Vega (2008) find that financial markets tend to incorporate qualitative information with delay, we furthermore presume investors to initially underreact to the speeches' sentiment so that the full market reaction will present itself only in a protracted time period after the AGM.

We test our hypothesis on the CEO speeches of publicly listed companies in Germany. We choose German firms as they regularly release the speeches' transcripts on their websites immediately after the AGM. US companies, in contrast, only rarely provide respective transcripts: While 72.50% of the German DAX and MDAX[2] companies offer transcripts, only 5.8% of all S&P 500 firms do so, rendering a meaningful empirical analysis on US data all but impossible.[3] We consider 338 CEO speeches of DAX and MDAX-listed corporations in Germany from 2008 to 2016. In a first step, we analyze whether AGMs systematically reveal new information per se and measure the financial market reaction subsequent to the AGM. Our univariate results show that AGMs do not seem to be followed by abnormal returns and we find a higher trading volume only in a short time window around the AGM.

In a second step, we examine whether the CEO speeches' sentiment at the AGM contains value-relevant information that is picked up by financial market participants. Sentiment is typically measured via a dictionary-based approach by assigning the words in a text or speech to different sentiment categories in accordance with a predefined dictionary (Manning & Schütze, 1999). Using a novel dictionary by Bannier et al. (2017), we gauge the sentiment of the CEO speeches

[2] The DAX and MDAX indices comprise the 80 largest German stock-listed companies in terms of order book volume and market capitalization. For more information on the indices, see http://www.dax-indices.com/EN/.

[3] Altogether, we were able to download only 54 speeches of US companies listed in the S&P 500.

and assess the financial market reaction to the AGM with respect to this sentiment. Our results show the sentiment to be significantly related to cumulative abnormal returns and trading volume. More precisely, we find the cumulative abnormal returns to decrease along with a speech's negativity and to increase with a speech's relative positivity, i.e., its positivity relative to negativity. Investigating the time structure of the sentiment's effect, we find that only a small part of the full market reaction occurs in the immediate vicinity around the AGM. Most of the market reaction, however, is observed in the time period between 2 and 30 days post AGM. This observation may be interpreted as an initial underreaction to the speeches' sentiment and could be an indication that qualitative information indeed needs more time to become fully incorporated in stock prices. Interestingly, the speeches' relative positivity is also significantly associated with a lower cumulative abnormal trading volume, but only in a short time window around the AGM. In summary, we find that a more positive relative to negative sentiment of a CEO speech goes along with higher cumulative abnormal returns and lower short-term cumulative abnormal trading volumes of the company's stock, whereas a lower positive relative to negative sentiment triggers lower returns and higher trading volumes.

Our paper's contribution to the existing literature is twofold. To begin with, we are the first to measure the sentiment of corporate texts in the German language using the business-specific dictionary introduced by Bannier et al. (2017). While the studies by Ammann and Schaub (2016) or Mengelkamp et al. (2015) also investigate the sentiment in German corporate texts, they either utilize only general German language dictionaries or ad-hoc dictionaries restricted to the respective sample of text documents at hand. The dictionary of Bannier et al. (2017), in contrast, is designed to capture the business-specific sentiment of any sample of German documents in a comprehensive way and follows the setup of the Loughran and McDonald (2011) dictionary for English documents.

As we are the first to employ this context-specific dictionary, we compare our results to those derived from using two general German language dictionaries. These are the "SentimentWortschatz" by Remus et al. (2010) and the German adaptation of the Linguistic Inquiry Word Count by Wolf et al. (2008). In line with content analyses on English documents (Henry & Leone, 2016; Loughran & McDonald, 2011, 2015; Price et al., 2012), we find the context-specific dictionary to be better suited for assessing the textual sentiment of business-related documents than general language dictionaries. Given the economic importance of firms in Germany and other German-speaking countries and the robust performance of the dictionary introduced by Bannier et al. (2017), the dictionary can hence be seen as a helpful tool to assess the qualitative information contained in these firms' communication. We also check the robustness of our results to different word weighting schemes, i.e., equal weighting vs. inverse document frequency weighting as proposed by Loughran and McDonald (2011). We find no improvement from using inverse document frequency weighting, similar to Henry and Leone (2016). Finally, we determine which measure of textual sentiment is most appropriate to gauge the qualitative information within German text documents. Consistent with Price et al. (2012) and Henry and Leone (2016) for English content analyses, we find the measure of relative positivity, which combines both positive and negative sentiment, to perform better than the positivity or negativity measure in isolation.

The second and main contribution of our study, however, is to show that there is valuable qualitative information hidden in the annual get-together of managers and shareholders. Our results suggest that financial market participants do indeed pick up the qualitative information contained in the CEO's speech for their investment decisions. However, both negativity and relative positivity - as the two most meaningful sentiment categories - are incorporated in the stock price only with a certain delay: While in the short time period around the AGM the association between the speeches' relative positivity and cumulative abnormal returns is only weak, the major part of the market reaction occurs in the time period between 2 and 30 days after the AGM. At the same time, however, we find a significant association between the relative positivity and the cumulative abnormal trading volume solely in the short time window immediately surrounding the AGM. The comparably long-lasting impact of the CEO speeches' sentiment on stock returns hence seems to be accompanied by an attention-capturing effect on the trading volume that is, however, quickly evaporating.

The remainder of the paper is structured as follows. The next section reviews the literature on the information provided in AGMs as well as on content analyses. Further, it introduces the dictionary developed by Bannier et al. (2017). Section 3 describes our data and the methodology employed. Section 4 presents the respective results. Finally, Section 5 concludes.

2. Literature

2.1. INFORMATIONAL CONTENT OF THE ANNUAL GENERAL MEETING

Companies typically release their annual results in three stages. First, a preliminary announcement is made including information about the company's profits, earnings per share, dividend per share, and sales turnover. A few weeks later, the company releases its annual report and finally, some weeks after that, the company's AGM takes place. Accordingly, Firth (1981) finds that the preliminary announcement and the release of the annual report induce significant abnormal returns and trading volume, while he finds no such market reaction following the AGM. Hence, he concludes that the AGM does not seem to provide new information to financial markets. This is supported by García-Blandón et al. (2012) who evaluate the AGM's information value in Spain and find no market reaction at all. Brickley (1986), Rippington and Taffler (1995) and Olibe (2002) observe only small price and trade volume reactions around the AGM. The most comprehensive study on the AGM's information value has been conducted by Martinez-Blasco et al. (2015) on a sample of common- and civil-law countries. The authors examine changes in abnormal returns, return volatility, and trading volume. Their analysis reveals no market reaction in Japan and Spain, and only small increases in trading volume in the US, the UK, and in France. In Germany, in contrast, the authors observe significant increases in abnormal returns, return volatility, and trading volume following the AGM, indicating that the AGMs of German companies exhibit substantial new information.

Despite the mixed results, none of the earlier studies - to the best of our knowledge - attempts to investigate the source or type of any potential information disclosure at the AGM. This is surprising since the AGM is a rare opportunity for a firm's management to get into direct contact with its shareholders (Martinez-Blasco et al., 2015) and since there is plenty of evidence on qualitative information inherent in the language of CEOs. Arslan-Ayaydin et al. (2015) find that managers adjust their language to specific situations at hand and inflate the use of positive language the higher their fraction of equity-based compensation. Doran et al. (2012) and Price et al. (2012) report that conference calls' positive sentiment is a significant predictor of subsequent returns and trading volume. Mayew and Venkatachalam (2012) and Hobson et al. (2012) analyze conference call audio files using vocal emotion analysis software. They come to the conclusion that positive and negative emotions expressed in the voice of managers can be informative about the firm's financial future and potential financial misreporting. It is hence reasonable to believe that qualitative information may be contained in the AGM even though substantial quantitative information has already been distributed to investors prior to the meeting. We therefore assess whether this qualitative information is inherent in the verbal communication of the CEO at the AGM.

2.2. DICTIONARY-BASED APPROACH

The dictionary-based approach has become a commonly used tool to measure the textual sentiment of various kinds of documents such as financial disclosures, analyst reports, earnings press releases, IPO prospectuses, internet board postings, or newspaper articles (Kearney & Liu, 2014). The individual dictionaries typically include various wordlists with respect to sentimental categories such as negativity or positivity. Text documents with a comparably high share of, for example, negative words are then considered to be more pessimistic compared to text documents with a comparably high share of positive words (Loughran & McDonald, 2015).

Early content analyses of financial texts (Davis et al., 2012; Davis & Tama-Sweet, 2012; Feldman et al., 2008; Ferris et al., 2013; Henry & Leone, 2016; Kothari et al., 2009; Larcker & Zakolyukina, 2012; Tetlock, 2007; Tetlock et al., 2008) utilized general English dictionaries such as the Harvard University's General Inquirer IV-4[4] dictionary, the dictionaries included in the Diction[5] software, or the Linguistic Inquiry Word Count[6] software. Henry (2008) is the first to compose a dictionary explicitly designed to examine the tone of financial documents. Despite the comparably small number of words in her positive and negative word lists, various studies comment on the superiority of the dictionary presented by Henry (2008) over the Diction and General Inquirer dictionaries (Doran et al., 2012; Henry & Leone, 2016; Price et al., 2012). Based on this finding, Loughran and McDonald (2011) create a more comprehensive dictionary (hereafter LM dictionary) by evaluating all words that appear in at least 5% of the entire 10-K disclosure universe. The LM dictionary contains 2,329 negative and 354 positive words. To assess the quality of their dictionary, the authors show that 73.8% of the General Inquirer dictionary's negative words do not have a negative meaning in financial documents and, in later work, demonstrate that none of the most frequently occurring negative words in the 10-K disclosures are included in the Henry (2008) dictionary (Loughran & McDonald, 2015). Due to its comprehensiveness and its appropriateness for financial documents, the LM dictionary has become the most widely used dictionary in business research and has been used to assess the textual sentiment of 10-K filings (Loughran & McDonald, 2011), earnings conference calls (Davis et al., 2015), news articles (García, 2013), or IPO prospectuses (Ferris et al., 2013; Jegadeesh & Wu, 2013).[7]

[4] See http://www.wjh.harvard.edu/~inquirer/.

[5] See http://www.dictionsoftware.com/.

[6] See http://www.liwc.net.

[7] For a comprehensive overview of dictionaries used in content analyses, see Kearney and Liu (2014) and Loughran and McDonald (2016).

2.3. GERMAN LANGUAGE DICTIONARIES

When it comes to the analysis of German text documents, two comprehensive general German language dictionaries but no business-specific dictionary exist: Remus et al. (2010) created the "SentimentWortschatz" (hereafter SENTIWS) dictionary, which is based on and extends the General Inquirer lexicon by Stone et al. (1966). SENTIWS has been used in studies of political communication (Haselmayer & Jenny, 2016), or art and literature (Zehe et al., 2016). The second general language dictionary was created by Wolf et al. (2008), who adapted the English version of the Linguistic Inquiry Word Count to the German language. Their dictionary (hereafter LIWC) puts special emphasis on analyzing essays in the context of expressive writing experiments, but has also been used in other research domains such as, for example, political analyses (Caton et al., 2015; Jacobi et al., 2016). However, with respect to business-related documents, there is no context-specific dictionary.

As many text documents containing relevant information on German companies are published exclusively in German, the absence of a context-specific dictionary in German is associated with very little research on German qualitative information. Rare exceptions are Ammann and Schaub (2016) and Mengelkamp et al. (2015), who investigate German corporate texts for their textual sentiment and utilize ad-hoc dictionaries that are constructed from - and thus restricted to – a given set of sample text documents. Similar to the studies conducted on English text documents, the authors also find that their ad-hoc dictionaries achieve more reliable results than the general German language dictionaries SENTIWS and LIWC.

In order to analyze German business-related texts comprehensively, Bannier et al. (2017) adapt the English business-specific dictionary by Loughran and McDonald (2011) to the German language. They follow the methodology of Wolf et al. (2008) and control for several linguistic issues such as inflections, compound words, or lexical morphology that are specific to the German language (Hawkins, 2015; König & Gast, 2012).[8] For a detailed explanation of the set up of word lists to measure sentiment in German corporate texts, see Bannier et al. (2017). The authors also test the equivalence of their adaptation (hereafter BPW dictionary)

[8] German speakers are forced to make certain inflectional distinctions which can regularly be left unspecified in English. Looking at verbs, for example, the German language distinguishes indicative and subjunctive forms whereas English employs a single form for both. Further, German verbs differ with respect to person and number, whereas the bare stem in English is used for all except the third person singular. As German nouns and adjectives need more inflections as well, a simple word-by-word translation of the LM dictionary will not fully cover the German inflectional morphology with the consequence of an underestimation of the German texts' sentiment.

using a broad sample of quarterly and annual reports of German companies that are available in German and English language.[9] The results show that all sentiment categories display high correlation and equivalence to their English counterparts, indicating the reliability of their adaptation. [10]

Table I presents a brief comparison of the LM dictionary, the BPW dictionary and the two general German dictionaries, SENTIWS and LIWC, that allows to put the specificities of the German language into perspective and helps to see the differences between general and context-specific wordlists.

<<< Insert Table I about here >>>

Table I shows that the German dictionaries' word lists contain far more individual words than the English LM dictionary. This is mainly due to the linguistic issues referred to above. However, even within the German language, there are strong differences between the dictionaries. Comparing the BPW to the SENTIWS dictionary reveals that SENTIWS includes about 50% more negative and about 700% more positive words than the BPW. Overall, SENTIWS contains as many negative as positive words. This stands in contrast to the other dictionaries, most obviously the LM dictionary, which contains a much smaller number of positive than negative words. Note that a direct comparison of the number of individual words between the BPW and the LIWC dictionaries is not feasible as LIWC includes word stems rather than inflections. However, both general German dictionaries are likely to include words that may misclassify sentiment in a business context. For example, "LEISTUNG(EN)" (service(s)), or "GEWINN" (profit), which are both classified as positive words by SENTIWS and LIWC, are regularly used in business documents without a necessarily positive connotation. Other examples such as "EIGENKAPITAL" (equity), "ANTEIL(E)" (share(s)), "INVESTITIONEN" (investments), "AKTIVITÄTEN (activities), and "WACHSTUM" (growth) are also counted as generally positive, while this may not be the case in business-related documents. As a consequence, both general language dictionaries and particularly the SENTIWS word lists might overestimate the positive sentiment of business-related text documents. While the higher fit of context-specific dictionaries has already been confirmed by English language studies (Henry & Leone, 2016; Loughran & McDonald, 2011, 2015; Price et al.,

[9] We estimate simple pairwise correlations, Spearman rank correlations, intra-class correlations Shrout and Fleiss (1979), and test the dictionaries' equivalence via two-sided equivalence testing following Blair and Cole (2002).

[10] For more information on the adaptation process and equivalence tests of the BPW dictionary, see Bannier et al. (2017).

2012), this issue is still unresolved in the German language. In the following analysis, we will therefore put some emphasis on evaluating the efficacy of the BPW dictionary relative to the two general language dictionaries, SENTIWS and LIWC, when employing the different dictionaries on the CEO speeches.

3. Data & Methodology

3.1. DATA AND VARIABLE MEASUREMENT

We attempt to capture the sentiment in CEO speeches held at German companies' AGMs and to assess whether this sentiment is associated with significant market reactions subsequent to the AGM. For that purpose, we gather the CEO speeches held at German DAX and MDAX companies' annual shareholder meetings from 2008 to 2016 by manually collecting transcripts from the companies' internet webpages. Our initial sample consists of 356 CEO speeches by 58 companies. We evaluate further documents, such as company charters, shareholder meeting invitations, and audio or video material from the companies' webpages, in order to confirm that the CEO speeches are indeed initially held in German. Based on this additional analysis, we exclude 18 speeches resulting in a final sample of 338 speeches.

Before we can segment the reports into vectors of word counts, we have to convert the documents, which are typically available in PDF file format, to TXT format. In this process, we also replace typographic ligatures and employ UTF-8 character encoding on all files in order to allow for German-specific characters such as 'Ä','Ü','Ö', or 'ß'. All characters are transformed into lower case and tokenized afterwards, whereby we define a token as any subsequent order of at least three alphabetic characters. In order to exclude potential spelling errors, we exclude tokens that do not occur in at least one percent of the speeches. After that, we apply a stop-word list on the reports to filter out words that might have important semantic functions, but rarely contribute information (Manning & Schütze, 1999). We use the stop-word list provided by Bannier et al. (2017) which includes common names, dates, numbers, geographic locations, currencies, the names of German DAX and MDAX companies, popular German pre- and surnames, and the names of the largest German and European cities. Hereafter, the documents are transformed to word count vectors using the Rapidminer software.[11] In a final step, the CEO speeches' numbers of negative and positive words

[11] The transformation to lower-case characters, the tokenization, the stop-word filtering and the generation of the word count vectors were conducted with the Rapidminer software. For more information, please see https://rapidminer.com/.

are counted with respect to the word lists of the BPW, SENTIWS and LIWC dictionaries.

Several measures to gauge textual sentiment have been utilized in the literature. Jegadeesh and Wu (2013) and García (2013) employ direct measures of positivity and find statistically significant market reactions. Loughran and McDonald (2011, 2016), however, point out that positive words are frequently used to frame negative words, whereas negative words are unambiguous in their usage. Consequently, Tetlock (2007) and Loughran and McDonald (2011) find little incremental information using only a positive wordlist and suggest using a documents' share of negative words to assess its textual sentiment. We therefore estimate the CEO speeches' share of negative words as follows:[12]

$$NEG\_BPW_j = \frac{NEGATIVE_j}{COUNT_j} * 100 \tag{1}$$

Here, $COUNT_j$ is the total number of words of CEO speech j and $NEGATIVE_j$ represents the number of negative words in CEO speech j with respect to the negative wordlist of the BPW dictionary. $NEG\_SENTIWS_j$ and NEG_LIWC_j are calculated analogously.

Recent studies point out, however, that recipients of financial documents might not consider positive and negative textual sentiment separately but rather in relation to each other. We therefore follow Henry (2008), Price et al. (2012), and Henry and Leone (2016) and estimate the CEO speeches' relative positivity (TONE) in the following way:

$$TONE\_BPW_j = \frac{POSITIVE_j - NEGATIVE_j}{POSITIVE_j + NEGATIVE_j} \tag{2}$$

Here, $POSITIVE_j$ is the number of positive words in CEO speech j with respect to the positive wordlist of the BPW dictionary. $TONE\_SENTIWS_j$ and $TONE\_LIWC_j$ are calculated analogously. The relative positivity measure - also referred to as tonality - hence combines the information of the negative and positive sentiment as it measures the positivity of speech j relative to its negativity. The TONE measures are scaled between -1 and 1, so that a purely positive CEO speech displays a score of 1, a purely negative speech a score of -1, and a neutral speech scores a 0.

[12] We re-estimate our main-analysis grasping the CEO speeches sentiment using a measure of positivity. The results are shown in Table VI.

In order to measure the stock price reaction subsequent to a CEO speech, we calculate Cumulative Abnormal Returns (CARs). For this, daily abnormal returns are calculated using the return of the CDAX[13] index as the expected return, which reflects the performance of the entire German equity market:

$$AR_{j,t} = R_{j,t} - R_{CDAX,t} \tag{3}$$

Here, $AR_{j,t}$ is the abnormal return on company j's stock at day t and $R_{j,t}$ is the actual return of company j's stock at day t. $R_{CDAX,t}$ is the return of the CDAX on day t. As Demers and Vega (2008) find that qualitative information is more difficult for market participants to process than quantitative information, we may expect any market reaction to the sentiment in CEO speeches to not be overly quick. We therefore examine the market reaction by cumulating the abnormal returns for each stock over a relatively long time period from day -1 to day 30, where 0 represents the day of the AGM at which speech j is held. To analyze the time structure of a potential market reaction in more detail, we then segregate this total time window into the three-day period around the AGM (-1,1) and the remaining period after the AGM (2,30). This approach should allow us to see whether the market reaction to the sentiment in CEO speeches operates in an immediate or a delayed fashion. We hence employ three CAR measures, estimated in the following way:

$$CAR(-1,30)_j = \sum_{t=-1}^{30} AR_{j,t} \tag{4}$$

$$CAR(-1,1)_j = \sum_{t=-1}^{1} AR_{j,t} \tag{5}$$

$$CAR(2,30)_j = \sum_{t=2}^{30} AR_{j,t} \tag{6}$$

In addition to analyzing the CEO speeches' sentiment effect on stock prices, we also measure the effect on actual trading. For this purpose, we estimate the Cumulative Abnormal Trading Volume (CAV) following Barber and Odean

[13] The CDAX comprises the price development of all 852 German stocks across the Deutsche Börse's prime and general standard. For more information on the CDAX, see http://www.dax-indices.com/EN/.

(2008) and Price et al. (2012), where the Abnormal Trading Volume (AV) is in a first step calculated as follows:

$$AV_{j,t} = \frac{VOLUME_{j,t}}{\overline{VOLUME}_{j,t}} - 1 \tag{7}$$

Here, $VOLUME_{j,t}$ is the trading volume for company j at day t, and $\overline{VOLUME}_{j,t}$ is the mean trading volume for company j from t-252 to t-1. Consequently, a value of zero for the abnormal trading volume $AV_{j,t}$ indicates that a company's stock j was not traded abnormally at day t compared to the previous 252 days, i.e., over the last year. A positive value indicates that the stock was traded more than usual and a negative value indicates that the stock was traded less than usual. Analogously to abnormal returns, $AV_{j,t}$ is accumulated over day -1 to 30, CAV(-1,30), day -1 to 1, CAV(-1,1), and day 2 to 30, CAV(2,30).

3.2. EMPIRICAL APPROACH

In a first univariate analysis, we sort the CEO speeches into quartiles with respect to the measures of textual sentiment and compare the mean and median CAR(-1,30), CAR(-1,1), and CAR(2,30) differences between the highest and lowest quartiles of textual sentiment. We then test the mean and median differences for statistical significance using t-tests and Wilcoxon rank sum tests, respectively. To check whether the univariate results of our sentiment measures hold in a multivariate setting, we then conduct cross-sectional OLS regressions with a comprehensive set of control variables of the following form:

$$CAR_j = \alpha_0 + \alpha_{1,i} * SENTIMENT_{i,j} + \alpha_{2,k} * CONTROLS_{k,j} + \varepsilon_j \tag{8}$$

Here, CAR_j is the measure of cumulative abnormal returns for CEO speech j, $SENTIMENT_{i,j}$ is a vector of the different sentiment measures i for speech j which are calculated as described above. $CONTROLS_{k,j}$ represents a vector of control variables for speech j which include the speech's length (COUNT), the speech's share of individual words (IND), the earnings surprise (EPS_SURP), the dividend surprise (DIV_SURP_POS and DIV_SURP_NEG), the market capitalization (SIZE), market to book ratio (M2B), leverage (LEVERAGE), return on assets (ROA), return volatility (VOLATILITY), and trading volume (VOLUME).[14]

[14] Note that we include COUNT, IND, SIZE and VOLUME in logarithmic format in the regressions.

COUNT represents the CEO speeches' length in terms of the total number of words. IND is the number of individual words in a CEO speech divided by the speech's total number of words. The earnings surprise (EPS_SURP) of CEO speech j is estimated in accordance with Price et al. (2012) as the difference between the last reported earnings per share for the company at time t minus the latest reported earnings per share in the year prior to date t, divided by the stock price one year before t:

$$EPS\_SURP_j = \frac{EPS_j - EPS_{j,t-1}}{STOCKPRICE_{j,t-1}} * 100 \qquad (10)$$

Here EPS_j is the most recent earnings-per-share release for the company at the time of speech j, $EPS_{j,t-1}$ is the most recent earnings-per-share release for the company one year before the day of speech j and $STOCKPRICE_{j,t-1}$ is the stock price of the company one year before the date of speech j. While the earnings surprise has been shown to affect returns and volatility following earnings announcements and earnings conference calls, we hypothesize that EPS_SURP should only have a limited effect on the CARs following the CEO speeches since the surprise is already known from the quarterly report und, thus, should already be incorporated in the stock price at the time the speech is held. We include the indicator variables DIV_SURP_POS and DIV_SURP_NEG to control for dividend surprises. Here, DIV_SURP_POS is equal to one if a company's dividend per share is increased compared to the previous year, zero otherwise, and DIV_SURP_NEG is equal to one if a company's dividend per share is decreased compared to the previous year, zero otherwise. In contrast to the earnings surprise, the dividend surprise might strongly influence the post AGM returns and trading volume, as the dividend is actually agreed on at the AGM. SIZE measures the company's equity market value at the day of the speech as the share price multiplied by the number of ordinary shares outstanding. It is displayed in Euro millions. We include the market to book ratio (M2B) to control for the company's growth opportunities. M2B is defined as the market value of the ordinary equity divided by the balance sheet value of the ordinary equity in the company. We include ROA, LEVERAGE and VOLATILITY to control for a potentially higher information demand by investors which might result from low profitability, financial distress or other forms of uncertainty, respectively. ROA is estimated as net income divided by total assets times one hundred. LEVERAGE is calculated as total liabilities divided by total assets and VOLATILITY is estimated as the daily returns' standard deviation in the time window of minus 90 days to minus 10 days prior to the AGM. Finally, VOLUME describes the number of shares traded of a stock on the day of the shareholder meeting and is expressed in thousands. While our sentiment measures, COUNT, and IND are collected directly from the CEO speeches, the data to estimate the remaining control variables are gathered from Thompson Reuters Datastream. We repeat all previously described analyses, substituting CAV for CAR. In the multivariate analyses we then utilize the same set of control variables except for VOLUME.

3.3. WEIGHTING SCHEME

The majority of studies employing the dictionary-based approach use equal weighting of individual words. This method values each individual word in a document equally and implies that a more frequent occurrence of a word indicates a higher importance.[15] However, as the impact of words might be diluted the more often they are used, Manning and Schütze (1999) propose a term-inverse document frequency measure (tf-idf) which weights each word inversely proportionally to its frequency in a document. Loughran and McDonald (2011) advocate the use of tf-idf weighting by arguing that a word's impact is likely to diminish with its frequency. Measuring the textual sentiment of annual 10-K reports with equal weights and with tf-idf weights and analyzing its impact on subsequent stock returns, they find that tf-idf weighting mitigates the impact of misclassified words in the measurement of textual sentiment. However, Henry and Leone (2016) point out that while tf-idf weighting might mitigate the impact of misclassification for frequent words, it concomitantly exacerbates the impact of misclassified words that are used only infrequently. They further argue that tf-idf weightings are sample-dependent and thus impede replication. In order to evaluate the efficacy of equal weighting versus tf-idf weighting, Henry and Leone (2016) gauge the textual sentiment in earnings announcements using both weighting schemes and analyze the subsequent capital market reaction. They find that using tf-idf weighting provides no improvement compared to equal weighting. As these issues have never been discussed for German language content analyses, we will not only measure the sentiment of CEO speeches using the context-specific BPW dictionary and compare the results to the general SENTIWS and LIWC dictionaries, but we will also evaluate the efficacy of equal weighting versus tf-idf weightings in measuring sentiment.

[15] For a comprehensive overview of studies using equal weighting, see Henry and Leone (2016).

4. Results

4.1. DESCRIPTIVE STATISTICS

Table II contains descriptive statistics for the CARs and CAVs (Panel A), for the CEO speeches' textual sentiment and other measures estimated from the CEO speeches (Panel B), as well as for the remaining control variables that we use in our multivariate regressions (Panel C).

<<< Insert Table II about here >>>

Panel A of Table II shows that, at the mean, all CARs under investigation are economically small and not statistically different from zero. This finding indicates that, on average, we do not observe a significant market reaction around the AGM. This is in contrast to Martinez-Blasco et al. (2015), who investigate companies from the German DAX30 index and report statistically significant positive cumulative abnormal returns around the AGM. With respect to cumulative abnormal trading volumes, we find statistically significant trading volumes for CAV(-1,1), indicating that German stocks are more frequently traded around the AGM. In contrast to our finding on CARs, our results on CAVs are in line with Martinez-Blasco et al. (2015), who also report an increase in trading volume around the day of the AGM.

Panel B of Table II presents summary statistics with regard to the CEO speeches and their sentiment and reveals that CEO speeches, on average, contain 1.15% negative words using the BPW dictionary and display a relative positivity, TONE_BPW, of 0.439. While the share of negative words is slightly larger using the SENTIWS dictionary (1.31%), it is much smaller employing the LIWC dictionary (0.36%). Both general dictionaries, however, also show a positive tonality. Altogether, this can be interpreted as a higher positivity than negativity of the average CEO speech. As CEOs should be expected to use public communication to present their company in a positive light, the higher positive word share does not come as a surprise. Boudt and Thewissen (2016), for instance, investigate CEO letters using the LM dictionary and find quite comparable values for negativity and relative positivity. On average, they report 1.03% of the letters' words to be negative and the relative positivity equals 0.485. Furthermore, Kim and Meschke (2014) investigate CEO interviews on CNBC using the Harvard University's General Inquirer IV-4 dictionary and find the share of negative words to be 1.38% and the relative positivity to equal 0.582. The results from the BPW word lists are hence well in line with the earlier studies.

Panel C of Table II presents the control variables that we use in our multivariate regressions. Surprisingly, in only 23.3% of our observations the dividend per share is unchanged compared to the previous year, while it is decreased in 17.8% of the cases and increased in 58.9%. This is quite high compared to the results by, for example, Andres et al. (2009), who investigate German companies from 1987 to 2005 and find the dividends for German companies to be stable in 46.4%, to increase in 33.7% and to decrease in 19.9% of all cases. However, as our sample period comprises the aftermath of the financial crises, the higher fraction of dividend increases is likely to reflect stepwise re-increases of the dividend after sharp dividend cuts due to the financial crises.

<<< Insert Table III about here >>>

Table III shows the Pearson and Spearman correlations among CARs, CAVs and the measures of textual sentiment for the BPW, SENTIWS and LIWC dictionaries. For all three dictionaries, the measures of negativity appear to be negatively correlated and the measures of positivity to be positively correlated to CARs of all three time windows. However, none of the measures' correlations to CAR(-1,1) are statistically significant, while they are statistically significant at the 1%-level to CAR(2,30) and CAR(-1,30). With respect to trading volumes, the picture is less clear: The BPW and SENTIWS measures of the speeches' negativity seem to be positively correlated to CAV(-1,1) and negatively to CAV(2,30) and CAV(-1,30). The BPW and SENTIWS measures of the speeches' relative positivity seem to be negatively correlated to CAV(-1,1) and positively to CAV(2,30) and CAV(-1,30). The LIWC measures, in contrast, show no significant correlation to trading volumes.

4.2. THE CEO SPEECHES' SENTIMENT EFFECT ON STOCK PRICES

4.2.1 THE CONTEXT-SPECIFIC BPW DICTIONARY

Before we proceed with the examination of the association between CEO speeches' sentiment and the stock price reaction in a multivariate analysis, we will consider the univariate dimension. In this respect, Figures I and II show the accumulation of abnormal returns from 5 days before to 30 days after the AGM for different levels of negativity and tonality. Figure I displays the accumulated abnormal returns of high and low negativity CEO speeches, where the sample is split at the median of NEG_BPW. As can be seen from the figure, at the day of the AGM, firms with less negative CEO speeches show a 0.55% higher accumulated abnormal return than firms with more negative speeches. Over the next days after the AGM, firms with less negative CEO speeches show positive and increasing CARs. Firms with more negative CEO speeches, in contrast, display CARs that are close to zero. While the spread in CARs between the two groups increases only slowly in the first days after the AGM, it accelerates drastically from day 15 on. This may be seen as a first indication that investors indeed process qualitative information only slowly, supporting the earlier findings by Demers and Vega (2008).

<<< Insert Figures I and II about here >>>

Figure II depicts the development of accumulated abnormal returns, differentiating between firms with high and low tonality speeches. The sample is split along the median TONE_BPW. Similarly to the results from Figure I, firms with high tonality speeches display positive and increasing CARs following the AGM, while firms with low tonality speeches show CARs that are close to zero in the first days after the AGM. From day 15 on, the difference between the CARs of the two groups increases strongly as firms with low tonality speeches then show strongly negative and decreasing CARs. Again, this might be interpreted as an initial underreaction of investors to the sentiment of the CEO speeches at the AGM.

<<< Insert Table IV about here >>>

Table IV gives further information on the univariate relation between sentiment and stock market reaction. The table sorts the CEO speeches into sentiment quartiles with respect to NEG_BPW and TONE_BPW and compares mean and median CARs of the highest and lowest sentiment quartiles for all time windows. Panel A of Table IV presents the results for the total time period. The CAR(-1,30) differences between the highest and lowest sentiment quartile are significantly different from zero with respect to both NEG_BPW and TONE_BPW. The CAR(-1,30) mean (median) difference between the highest and lowest NEG_BPW quartiles equals -5.7 (-6.1) percentage points, and 5.1 (5.2) percentage points between the highest and lowest TONE_BPW quartiles. Panel B of Table IV contains the univariate results for CAR(-1,1). In this short time window around the AGM, no statistically or economically significant differences can be found between the extreme quartiles, irrespective of the sentiment measure applied. Panel C of Table IV presents the results for CAR(2,30). In this longer time window, we observe economically and statistically significant differences Q4-Q1 both with respect to the speeches' negativity and relative positivity. More precisely, the CAR(2,30) mean (median) difference between the highest and lowest

NEG_BPW quartiles equals -5.2 (-5.6) percentage points. With respect to TONE_BPW, the CAR(2,30) difference between the highest and lowest quartile is positive and equals 4.7 (5.2) percentage points.

These first univariate results suggest that negative textual sentiment is negatively related to cumulative abnormal returns while relative positive textual sentiment shows a positive relation. Furthermore, our findings indicate that investors initially underreact to the CEO speeches' textual sentiment, as only little of the total effect is explained by an immediate reaction around the AMG.

<<< Insert Table V about here >>>

Table V finally presents multivariate regressions of CAR(-1,30), CAR(-1,1), and CAR(2,30) on NEG_BPW and TONE_BPW and a comprehensive set of control variables. Looking at the (-1,30) window, NEG_BPW and TONE_BPW have a strong statistically significant association with the CAR(-1,30). An increase in NEG_BPW by the interquartile change of 0.749 yields a 2.77 percentage points lower CAR(-1,30), while an increase in a CEO speech's TONE_BPW by the interquartile range of 0.353 induces a 3.14 percentage points higher CAR(-1,30).With regard to an immediate market reaction, i.e. the short-term event window (-1,1), NEG_BPW does not significantly affect the cumulative abnormal returns, thus confirming the univariate results. The relative positivity measure, TONE_BPW, in contrast, displays a statistically significant association with CAR(-1,1). However, this effect is only weakly significant and also quite small in economic terms. Nonetheless, this finding presents some first evidence that a combined positive and negative sentiment measure may capture qualitative information more effectively than a solely negative measure. In the more distant time period (2,30), NEG_BPW has a statistically significant negative effect on the cumulative abnormal returns. An increase in negativity by the interquartile change of 0.749 yields a 2.32 percentage points lower CAR(2,30). TONE_BPW also significantly affects CAR(2,30). An increase in a CEO speech's tonality by the interquartile range of 0.353 induces a 2.5 percentage points higher CAR(2,30).[16] In line with the univariate results, the mostly non-significant or only small effects of the sentiment measures in the immediate vicinity around the AGM indicate an initial investor underreaction to qualitative information as compared to the stronger reaction in the longer time period following the AGM. In

[16] Note that the CEO speeches' sentiment with respect to all our measures varies only little (Table II). As a consequence, interpreting the increase in terms of interquartile changes is more useful to illustrate our results.

this respect, our results are indeed consistent with Engelberg (2008), Demers and Vega (2008), and Price et al. (2012).

With respect to the control variables, neither the quantity of information as measured by the speeches' length (COUNT), nor the speeches' complexity as approximated by the share of individual words (IND) are significantly associated with cumulative abnormal returns. The same is true for EPS_SURP, supporting our conjecture that any EPS surprise is likely to be already processed by financial market participants after the earlier announcement in the annual report. In contrast, a change in dividends might have an effect on the CARs, as the dividend's payout is agreed upon at the AGM. Nevertheless, neither positive dividend surprises (DIV_SURP_POS), nor negative dividend surprises (DIV_SURP_NEG) seem to have a statistically significant effect on CARs.

To summarize, our analyses of cumulative abnormal returns highlight several interesting facts. Our measures of negative and relative positive sentiment show strong and statistically significant associations with CAR(-1,30) in univariate and multivariate analyses. When dissecting this time window into the period immediately surrounding the AGM (-1,1) and the subsequent period (2,30), we see that the market reaction occurs in a delayed fashion: Only a small fraction of the full effect is seen immediately and the larger part follows afterwards. It hence seems to be the case that the market indeed takes more time to process the qualitative information captured by the speeches' sentiment and to incorporate this in the stock price as compared to quantitative information.

Table VI re-estimates Table V, substituting NEG_BPW and TONE_BPW with the speeches' share of positive words (POS_BPW).[17] It reveals no significant relationship of POS_BPW with CAR(-1,1) or CAR(2,30). Looking at both time windows combined, we find a positive relation with CAR(-1,30) which is, however, statistically significant only at the 5% level. Compared to the speeches' negativity and relative positivity (Table V), the speeches' share of positive words hence seems to be less suited to capture the qualitative information of text-documents. Our results therefore support Tetlock (2007) and Loughran and McDonald (2011), who observe little incremental information using only a positive wordlist for the English language and suggest using a documents' share of negative words instead to assess its textual sentiment. We show that their observation holds for analyses on German text documents as well. Given the stronger statistical significance of the combined TONE_BPW measure, we furthermore underline the earlier suggestion that recipients tend to assess a text's positivity and negativity not in isolation but rather in relation to each other. As a consequence, tonality, i.e.,

[17] We calculate POS_BPW analogously to NEG_BPW, where the number of negative words is replaced by the number of positive words in the respective speech.

relative positivity, appears to be a superior measure for capturing the qualitative information in a text or speech in the German language as well.

<<< Insert Table VI about here >>>

4.2.2 THE BPW VS. GENERAL GERMAN LANGUAGE DICTIONARIES

In our analyses we so far applied the business-specific BPW dictionary. In order to evaluate its suitability for examining sentiment in business texts vis-à-vis more general word lists, we rerun our analyses using the general German language dictionaries instead. In this respect, Table VII re-estimates the earlier regression models using once the SENTIWS dictionary and once the LIWC dictionary to measure the sentiment of the CEO speeches. It should be noted that the following analyses employ standardized sentiment measures (with a mean of 0 and a standard deviation of 1) in order to facilitate comparisons between the results for each dictionary. We also include the (now standardized) regression coefficients for the sentiment measured via the BPW dictionary in the first line of Table VII.

<<< Insert Table VII about here >>>

Panel A considers the negative sentiment. As can be seen, irrespective of the dictionary used, none of the measures of negative textual sentiment has a statistically significant effect on CAR(-1,1). With respect to CAR(2,30) and CAR(-1,30), in contrast, all measures show a statistically significant negative relationship. However, NEG_BPW always delivers the highest and most strongly significant coefficient. Panel B of Table VII refers to the tonality measure, i.e. relative positivity. In the time window (-1,30), both TONE_BPW and TONE_SENTIWS are significantly related to CARs at the 1%-level while TONE_LIWC shows a significance only at the 10%-level. Still, the effect of TONE_BPW is of higher magnitude compared to TONE_SENTIWS and TONE_LIWC. In the time window (-1,1), TONE_BPW is significantly related to CARs, while the tonality measures based on the general language dictionaries are not. In time window (2,30), only TONE_BPW and TONE_SENTIWS show a significant association with CARs, with a stronger effect again for TONE_BPW.

Table VIII presents J-test (Davidson & MacKinnon, 1981) and Cox-Pesaran-Deaton (Pesaran & Deaton, 1978) test statistics for non-nested regressions in order to compare the in Table VII presented models' efficacy. The results show that none of the models using measures from the BPW dictionary can be rejected in favor of the respective models using measures from the SENTIWS or LIWC dictionaries according to both test statistics. Vice versa, the CAR(2,30) and CAR(-1,30) models including NEG_BPW (models (4) and (7)) and TONE_BPW (models (13) and (16)) are more favorable compared to the corresponding models using the general language SENTIWS and LIWC dictionaries according to both test statistics. Thus, our results indicate the superiority of context-specific dictionaries in capturing the textual sentiment of German business-related documents, underlining the earlier results from English text analyses (Henry & Leone, 2016; Loughran & McDonald, 2011, 2015; Price et al., 2012).

.

<<< Insert Table VIII about here >>>

4.2.3 WEIGHTING SCHEMES

The previous results have been estimated using equal weighting of words in calculating sentiment measures for the CEO speeches. In order to test whether the weighting scheme drives our results, Table IX re-estimates the regressions from Table V using equal weighting and tf-idf weighting for calculating the sentiment measures NEG_BPW and TONE_BPW in comparison.[18]

<<< Insert Table IX about here >>>

Panel A of Table IX shows that most of the coefficients estimated via tf-idf weighting are comparable in size and significance to those estimated via equal weighting. However, the coefficients of NEG_BPW in CAR(-1,30) and CAR(2,30) regressions equal -0.020 and -0.017 and are statistically significant at the 1%-level using equal weighting (models (1) and (9)), while they decrease to -0.012 and -0.011 and are only significant at the 10%-level (model (2)) and 5%-level (model (10)) with tf-idf weighting. In both cases, tf-idf weighting hence seems to unfavorably affect the results.

Panel B of Table IX reports the results from J-tests and Cox-Pesaran-Deaton tests. They show that none of the equally weighted NEG_BPW or TONE_BPW models can be rejected in favor of the tf-idf weighted models. Vice versa, all but two tf-idf weighted models cannot be rejected in favor of the respective equally weighted models. Only model (1) seems to be preferable compared to model (2). Model (9) seems to be preferable compared to model (10). Consequently, the results presented in Table IX indicate that, for our sample, tf-idf weighting seems to provide no improvement over equal weighting with respect to measures of

[18] We are still employing standardized sentiment measures.

relative positive textual sentiment. It may provide even less effective results with respect to measures of negative textual sentiment. With respect to the latter point, our results on NEG_BPW are in contrast to Loughran and McDonald (2011), who find tf-idf weighting to improve the effectiveness of their measure of negative textual sentiment. With respect to TONE_BPW, in contrast, our results are in line with Henry and Leone (2016), who find no improvement for measures of relative positivity using tf-idf weighting.

Loughran and McDonald (2011) argue that tf-idf weighting mitigates the impact of misclassified words (or noise) in the dictionaries, as words which appear more frequently are weighted less. To test this final aspect, we re-estimate Table VII using tf-idf weighting for all measures of textual sentiment, i.e., also those based on the SENTIWS and LIWC word lists, in Table X. Indeed, we find that some coefficients on general language sentiment SENTIWS and LIWC measures improve in magnitude and statistical significance. However, they still do not exceed the context-specific BPW measures. This finding is largely concordant with Henry and Leone (2016), who report that tf-idf weighting modestly increases statistical significance for general language measures of negative sentiment, but does not improve the results for measures of relative positivity.

<<< Insert Table X about here >>>

4.3. THE CEO SPEECHES' SENTIMENT EFFECT ON TRADING VOLUME

In addition to our analyses of stock prices, we also examine the relation between the CEO speeches' sentiment and the abnormal trading volume. For this examination, we employ the BPW dictionary and again start with a univariate analysis. Analogously to Table IV for CARs, Table XI shows the differences in CAV(-1,30), CAV(-1,1), and CAV(2,30) sorted for quartiles with respect to NEG_BPW and TONE_BPW.

<<< Insert Table XI about here >>>

As Table XI shows, we find statistically significant differences in the accumulated trading volume between the fourth and first sentiment quartiles only in the short time window, CAV(-1,1), and only with respect to tonality measure. Significance is given both with parametric and non-parametric test statistics. Firms with highest tonality speeches hence show a smaller abnormal trading in the time period immediately surrounding the AGM than firms with lowest tonality speeches. This may be taken as an indication that a higher "relative negativity" seems to draw investors' attention and leads to higher abnormal trading. As we find no significant Q4-Q1 differences with respect to the longer time windows CAV(2,30) and CAV(-1,30), the observed investor attention seems to be quickly evaporating.

<<< Insert Table XII about here >>>

Table XII is estimated analogously to Table V, substituting CAV for CAR. Table XII confirms the univariate findings from Table XI and shows that a higher tonality goes along with lower CAV(-1,1). Also in accordance with the univariate results, NEG_BPW does not seem to affect CAV(-1,1). With respect to the longer time horizons, we observe no statistically significant relationships between the measures of textual sentiment and CAV(-1,30) or CAV(2,30).

Table XIII tests whether results for CAVs are influenced by the word weighting scheme applied and Table XIV investigates the relationship among CAVs and the general language measures of textual sentiment. Similar to our results on CARs, Table XIII shows that tf-idf weighting does not seem to improve the results and Table XIV reports that general language SENTIWS and LIWC measures do not possess higher explanatory power compared to the context-specific BPW measures. In particular, measuring textual sentiment via the SENTIWS or LIWC dictionaries does not yield any statistically significant relationship between textual sentiment and CAVs.

<<< Insert Table XIII about here >>>

In sum, our findings on CAVs appear to some extent inverse to the results on cumulative abnormal stock returns: While the speeches' sentiment seems to be incorporated into returns rather slowly, it appears to draw investors' attention via trading volumes only during the short-term announcement period. For both returns and trading volumes, however, it is the relative positivity of the speeches that shows the predominant effect. In the longer time periods, (-1,30) and (2,30), none of the sentiment measures displays a significant association with the CAVs. The latter finding is in contrast to Price et al. (2012), who observe for US earnings conference calls that the sentiment's effect on abnormal trading volume is statistically significant only in longer time windows. Our findings are in accordance with Martinez-Blasco et al. (2015), however, who report that the trading volume of German stocks is economically and statistically significantly increased on the day of the AGM and the two days surrounding the AGM. According to our results, this observation may at least partly be explained by the sentiment of the CEO speeches at the AGM: Speeches with particularly low relative positivity, or high “relative negativity” respectively, should draw investors’ attention in the short term and go hand in hand with heightened trading volumes.

<<< Insert Table XIV about here >>>

5. Conclusion

CEOs’ language has been repeatedly shown to exhibit information that is relevant for financial market participants, for example, in analyses on earnings conference calls (Davis et al., 2015; Doran et al., 2012; Larcker & Zakolyukina, 2012; Price et al., 2012), or CEO letters (Boudt & Thewissen, 2016). Nevertheless, CEO speeches held at companies’ annual general meetings have received no attention in studies of qualitative content analysis yet. We try to fill this gap by analyzing the investor reaction to the textual sentiment in German CEO speeches held at the companies’ AGMs. We examine the speeches held by the CEOs of stock-listed German firms which regularly publish the speeches’ transcripts on their internet webpages. In order to be able to analyze German texts, we utilize a novel business-specific dictionary by Bannier et al. (2017) which converts the commonly used English dictionary by Loughran and McDonald (2011) to the German language. We gather the transcripts of 338 German CEO speeches, assess the speeches’ textual sentiment and measure the sentiment’s effect on both stock prices and trading volumes following the AGM.

We find that the CEO speeches’ textual sentiment is significantly related to abnormal stock returns and trading volume. In particular, the negativity of CEO speeches is negatively associated with abnormal returns, whereas the relative positivity of speeches is positively associated abnormal returns. With regard to the time structure of the information incorporation, we see a delayed reaction that may be interpreted as an initial underreaction to the speeches’ sentiment. With respect to cumulative abnormal trading volume, in contrast, sentiment seems to have only short term effects. CEO speeches with low relative positivity are followed by increased trading volume only in the three-day window surrounding the AGM. Further, similar to content analyses on English text documents, we find that context-specific measures of textual sentiment are better suited to capture the sentiment of business-related text documents compared to general language dictionaries. Moreover, and also in accordance with literature on English content analyses (Henry & Leone, 2016), we find using combined measures of a document’s positivity relative to its negativity to be advantageous compared to positive or negative measures of sentiment in isolation. Finally, our results also highlight that inverse term weighting does not yield improvements over equal weighting.

We are aware of some limitations of our analyses. First, our study is limited by the data availability of CEO speeches. As there is no compulsory register for CEO speeches, we are only able to gather CEO speeches whose transcripts are offered on the companies' homepages or sent to us on request. As most companies in our sample either offer transcripts of the speeches or do not, we can rule out the possibility that companies selectively publish only favorable speeches. However, the speeches are typically only offered a few years back, so that extending our sample poses difficulties and seems to be only possible using prospective CEO speeches. Further, the study at hand is limited to the examination of textual sentiment. Other channels of communication, for example the managers' voice, have been found to contain qualitative information as well (Hobson et al., 2012; Mayew & Venkatachalam, 2012). Future research might extend the analysis of textual sentiment by qualitative information communicated by the managers' voice, or other channels such as, for example, gestures.

6. References

Ammann, M., & Schaub, N. (2016). Social interaction and investing: Evidence from an online social trading network. *Working Paper.*

Andres, C., Betzer, A., Goergen, M., & Renneboog, L. (2009). Dividend policy of German firms. *Journal of Empirical Finance, 16*(2), 175-187. doi:10.1016/j.jempfin.2008.08.002

Arslan-Ayaydin, Ö., Boudt, K., & Thewissen, J. (2015). Managers set the tone: Equity incentives and the tone of earnings press releases. *Journal of Banking & Finance.* doi:10.1016/j.jbankfin.2015.10.007

Bannier, C. E., Pauls, T., & Walter, A. (2017). The BPW: Introducing a German dictionary for computer aided quantitative content analyses. *Working Paper.*

Barber, B. M., & Odean, T. (2008). All That Glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies, 21*(2), 785-818. doi:10.1093/rfs/hhm079

Blair, C., & Cole, S. R. (2002). Two-sided equivalence testing of the difference between two means. *Journal of Modern Applied Statistical Methods, 1*(1), 139-142. doi:10.22237/jmasm/1020255540

Boudt, K., & Thewissen, J. (2016). Jockeying for position in CEO letters: Impression management and sentiment analytics. *Working Paper.* doi:10.2139/ssrn.2326400

Brickley, J. A. (1986). Interpreting common stock returns around proxy statement disclosures and annual shareholder meetings. *Journal of Financial & Quantitative Analysis, 21*(3), 343-349. doi:10.2307/2331047

Caton, S., Hall, M., & Weinhardt, C. (2015). How do politicians use Facebook? An applied social observatory. *Big Data & Society.* doi:10.1177/2053951715612822

Caumanns, J. (1999). A fast and simple stemming algorithm for German words. *Freie Universität Berlin, Fachbereich Mathematik und Informatik Ser. B, Informatik: 99-16.* Berlin: Freie Univ. Fachbereich Mathematik und Informatik.

Davidson, R., & MacKinnon, J. G. (1981). Several tests for model specification in the presence of alternative hypotheses. *Econometrica, 49*(3), 781-793.

Davis, A. K., Ge, W., Matsumoto, D., & Zhang, J. L. (2015). The effect of manager-specific optimism on the tone of earnings conference calls. *Review of Accounting Studies, 20*(2), 639-673. doi:10.1007/s11142-014-9309-4

Davis, A. K., Piger, J. M., & Sedor, L. M. (2012). Beyond the numbers: Measuring the information content of earnings press release language. *Contemporary Accounting Research, 29*(3), 845‑868. doi:10.1111/j.1911-3846.2011.01130.x

Davis, A. K., & Tama-Sweet, I. (2012). Managers' use of language across alternative disclosure outlets: Earnings press releases versus MD&A. *Contemporary Accounting Research, 29*(3), 804‑837. doi:10.1111/j.1911-3846.2011.01125.x

Demers, E., & Vega, C. (2008). Soft information in earnings announcements: news or noise? *International Finance Discussion Papers.*

Doran, J. S., Peterson, D. R., & Price, S. M. (2012). Earnings conference call content and stock price: The case of REITs. *The Journal of Real Estate Finance and Economics, 45*(2), 402‑434. doi:10.1007/s11146-010-9266-z

Engelberg, J. (2008). Costly Information Processing: Evidence from Earnings Announcements. *Working Paper.* doi:10.2139/ssrn.1107998

Feldman, R., Govindaraj, S., Livnat, J., & Segal, B. (2008). The incremental information content of tone change in management discussion and analysis. *Working Paper.*

Ferris, S. P., Hao, Q., & Liao, M.-Y. (2013). The effect of issuer conservatism on IPO pricing and performance. *Review of Finance, 17*(3), 993‑1027. doi:10.1093/rof/rfs018

Firth, M. (1981). The relative information content of the release of financial results data by firms. *Journal of Accounting Research, 19*(2), 521‑529.

García, D. (2013). Sentiment during recessions. *The Journal of Finance, 68*(3), 1267‑1300. doi:10.1111/jofi.12027

García-Blandón, J., Martínez-Blasco, M., & González-Sabaté, L. (2012). Does the annual general meeting involve the release of relevant information in non-common law markets? Evidence from Spain. *Spanish Journal of Finance and Accounting / Revista Española de Financiación y Contabilidad, 41*(154), 209‑232. doi:10.1080/02102412.2012.10779723

Haselmayer, M., & Jenny, M. (2016). Sentiment analysis of political communication. Combining a dictionary approach with crowdcoding. *Quality & Quantity.* doi:10.1007/s11135-016-0412-4

Hawkins, J. A. (2015). A comparative typology of English and German: Unifying the contrasts (1st ed., Croom Helm, London, 1986). *Routledge library editions : english language: Vol. 10.* London [u.a.]: Routledge.

Henry, E. (2008). Are investors influenced by how earnings press releases are written? *Journal of Business Communication, 45*(4), 363-407. doi:10.1177/0021943608319388

Henry, E., & Leone, A. J. (2016). Measuring qualitative information in capital markets research: Comparison of alternative methodologies to measure disclosure tone. *The Accounting Review.* doi:10.2308/accr-51161

Hobson, J. L., Mayew, W. J., & Venkatachalam, M. (2012). Analyzing speech to detect financial misreporting. *Journal of Accounting Research, 50*(2), 349-392. doi:10.1111/j.1475-679X.2011.00433.x

Huang, X., Teoh, S. H., & Zhang, Y. (2014). Tone Management. *The Accounting Review, 89*(3), 1083-1113. doi:10.2308/accr-50684

Jacobi, C., Kleinen-von Königslöw, K., & Ruigrok, N. (2016). Political news in online and print newspapers. *Digital Journalism, 4*(6), 723-742. doi:10.1080/21670811.2015.1087810

Jegadeesh, N., & Wu, D. (2013). Word power: A new approach for content analysis. *Journal of Financial Economics, 110*(3), 712-729. doi:10.1016/j.jfineco.2013.08.018

Kearney, C., & Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis, 33,* 171-185. doi:10.1016/j.irfa.2014.02.006

Kim, Y. H., & Meschke, F. (2014). CEO Interviews on CNBC. *Working Paper.* doi:10.2139/ssrn.1745085

König, E., & Gast, V. (2012). Understanding English-German contrasts (3., neu bearb. und erw. Aufl.). *Grundlagen der Anglistik und Amerikanistik: Vol. 29.* Berlin: Schmidt.

Kothari, S. P., Li, X., & Short, J. E. (2009). The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: A study using content analysis. *The Accounting Review, 84*(5), 1639-1670. doi:10.2308/accr.2009.84.5.1639

Larcker, D. F., & Zakolyukina, A. A. (2012). Detecting deceptive discussions in conference calls. *Journal of Accounting Research, 50*(2), 495-540. doi:10.1111/j.1475-679X.2012.00450.x

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance, 66*(1), 35-65. doi:10.1111/j.1540-6261.2010.01625.x

Loughran, T., & McDonald, B. (2015). The use of word lists in textual analysis. *Journal of Behavioral Finance, 16*(1), 1-11. doi:10.1080/15427560.2015.1000335

Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research, 54*(4), 1187-1230. doi:10.1111/1475-679X.12123

Manning, C. D., & Schütze, H. (1999). Foundations of statistical natural language processing. Cambridge, Mass.: MIT Press.

Martinez-Blasco, M., Garcia-Blandon, J., & Argiles-Bosch, J. M. (2015). Does the informational role of the annual general meeting depend on a country’s legal tradition? *Journal of Management & Governance, 19*(4), 849-873. doi:10.1007/s10997-014-9294-9

Mayew, W. J., & Venkatachalam, M. (2012). The power of voice: Managerial affective states and future firm performance. *The Journal of Finance, 67*(1), 1-44. doi:10.1111/j.1540-6261.2011.01705.x

Mengelkamp, A., Hobert, S., & Schumann, M. (2015). Corporate credit risk analysis utilizing textual user generated content - A Twitter based feasibility study. *Working Paper.*

Olibe, K. O. (2002). The Information content of annual general meetings a price and trading volume analysis. *Journal of International Accounting, Auditing & Taxation, 11*(1), 19. doi:10.1016/S1061-9518(02)00052-6

Pesaran, M. H., & Deaton, A. S. (1978). Testing non-nested nonlinear regression models. *Econometrica, 46*(3), 677-694. doi:10.2307/1914240

Price, S. M., Doran, J. S., Peterson, D. R., & Bliss, B. A. (2012). Earnings conference calls and stock returns: The incremental informativeness of textual tone. *Journal of Banking & Finance, 36*(4), 992-1011. doi:10.1016/j.jbankfin.2011.10.013

Remus, R., Quasthoff, U., & Heyer, G. (2010). SentiWS-A publicly available German-language resource for sentiment analysis. *LREC. 2010.*

Rippington, F. A., & Taffler, R. J. (1995). The information content of firm financial disclosures. *Journal of Business Finance & Accounting, 22*(3), 345-362. doi:10.1111/j.1468-5957.1995.tb00878.x

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420-428. doi:10.1037/0033-2909.86.2.420

Stone, P. J., Dunphy, D. C., & Smith, M. S. (1966). The general inquirer: A computer approach to content analysis. *MIT Press.*

Tetlock, P. C. (2007). Giving content to investor sentiment. The role of media in the stock market. *The Journal of Finance, 62*(3), 1139-1168. doi:10.1111/j.1540-6261.2007.01232.x

Tetlock, P. C., Saar-Tsechansky, M., & MacsKassy, S. (2008). More than words. Quantifying language to measure firms' fundamentals. *The Journal of Finance, 63*(3), 1437-1467. doi:10.1111/j.1540-6261.2008.01362.x

Wolf, M., Horn, A. B., Mehl, M. R., Haug, S., Pennebaker, J. W., & Kordy, H. (2008). Computergestützte quantitative Textanalyse. *Diagnostica, 54*(2), 85-98. doi:10.1026/0012-1924.54.2.85

Zehe, A., Becker, M., Hettinger, L., Hotho, A., & Reger, I. (2016). Prediction of happy endings in German novels based on sentiment information. *Proceedings of the Workshop on Interactions between Data Mining and Natural Language Processing 2016,* 9-16.

7. Tables

Table I. Dictionaries for content analysis

This table shows the number of words contained in the positive and negative wordlists of existing English and German language dictionaries for content analysis. Note that the LIWC contains word stems rather than comprehensive sets of inflections as LM, BPW, and SENTIWS.

	English	German		
	LM	BPW	SENTIWS	LIWC
Negative	2,354	10,147	15,466	1,049
Positive	354	2,223	15,536	646

Table II. Descriptive statistics

This table provides descriptive statistics for the full sample of 338 CEO speeches. ***,**, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix I.

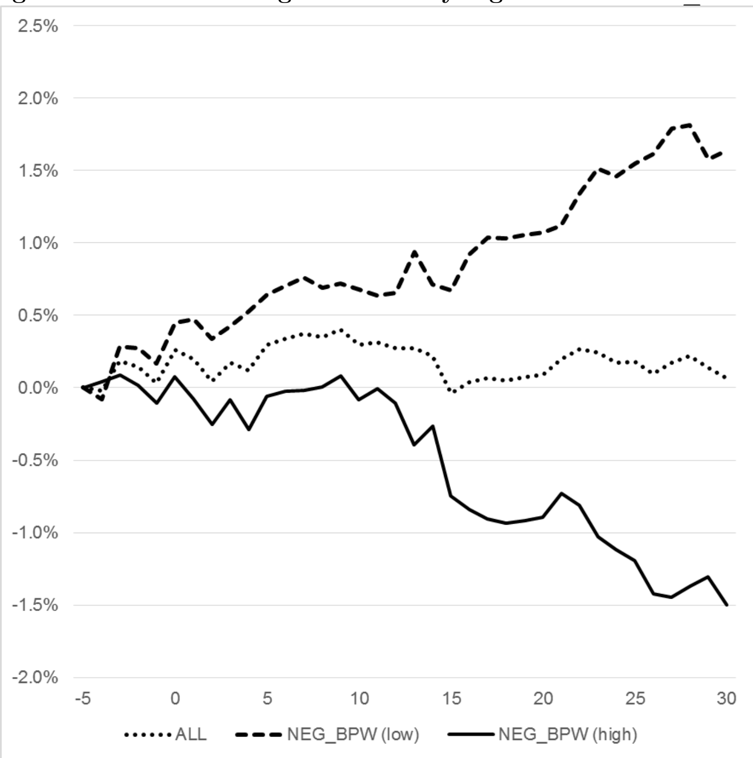
	Mean	Min	p25	p50	p75	Max	SD	N	T-Statistic
Panel A: CARs and CAVs									
CAR(-1,30)	-0.001	-0.277	-0.043	0.004	0.048	0.209	0.071	338	-0.202
CAR(-1,1)	0.001	-0.195	-0.017	0.000	0.017	0.095	0.029	338	0.317
CAR(2,30)	-0.001	-0.261	-0.042	-0.001	0.041	0.212	0.069	338	-0.346
CAV(-1,30)	0.999	-15.174	-4.403	-0.910	3.397	84.763	10.221	338	1.797*
CAV(-1,1)	1.502	-1.566	-0.147	0.626	1.954	19.424	3.089	338	8.942***
CAV(2,30)	-0.503	-13.948	-4.751	-1.975	1.310	83.692	8.839	338	-1.047
Panel B: CEO speeches and their sentiment									
COUNT	3,433	1,327	2,783	3,363	3,999	6,392	985	338	
IND	0.334	0.245	0.308	0.330	0.354	0.428	0.032	338	
NEG_BPW	1.154	0.235	0.759	1.057	1.508	3.237	0.549	338	
TONE_BPW	0.439	-0.207	0.268	0.459	0.621	0.894	0.237	338	
NEG_SENTIWS	1.309	0.293	0.917	1.231	1.637	2.832	0.521	338	
TONE_SENTIWS	0.740	0.420	0.670	0.754	0.824	0.947	0.105	338	
NEG_LIWC	0.359	0.000	0.233	0.337	0.460	0.962	0.182	338	
TONE_LIWC	0.717	0.213	0.649	0.741	0.815	1.000	0.140	338	
Panel C: Company-level controls variables									
EPS_SURP	0.030	-43.996	-1.567	0.374	2.055	57.060	7.933	330	
DIV_SURP_POS	0.589	0.000	0.000	1.000	1.000	1.000	0.493	338	
DIV_SURP_NEG	0.178	0.000	0.000	0.000	0.000	1.000	0.383	338	
SIZE	15,484	195	2,185	7,637	20,196	105,412	19,468	338	
M2B	2.08	0.16	1.05	1.75	2.70	10.33	1.56	338	
LV	0.07	-0.20	0.01	0.05	0.09	2.21	0.15	311	
ROA	3.68	-12.68	0.69	3.36	5.80	67.93	5.68	311	
VOLA	0.02	0.01	0.01	0.02	0.02	0.07	0.01	338	
VOLUME	28.65	0.00	2.10	6.00	31.70	406.60	53.63	337	

Table III. Correlations

This table shows pairwise correlations for the full sample of 338 CEO speeches. Note that the LIWC contains word stems rather than comprehensive sets of inflections as BPW and SENTIWS. Thus, we use a stemming algorithm by Caumanns (1999) on our sample of reports before gauging the textual sentiment using the LIWC. Pearson correlations are below the diagonal, Spearman correlations are above the diagonal. P-values are in parentheses. All variables are defined in Appendix I.

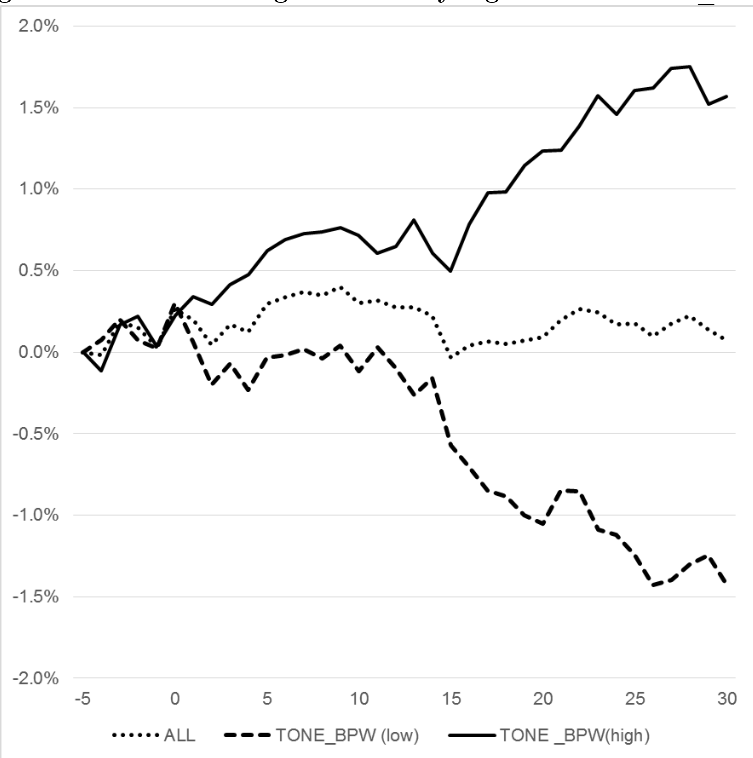
	CAR (-1,30)	CAR (-1,1)	CAR (2,30)	CAV (-1,30)	CAV (-1,1)	CAV (2,30)	NEG_ BPW	TONE_ BPW	NEG_ SENTIWS	TONE_ SENTIWS	NEG_ LIWC	TONE_ LIWC
CAR(-1,30)		0.258	0.904	-0.001	-0.008	0.023	-0.257	0.265	-0.215	0.241	-0.207	0.200
		(0.000)	(0.000)	(0.985)	(0.879)	(0.674)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CAR(-1,1)	0.296		-0.106	-0.090	0.031	-0.095	-0.073	0.075	-0.032	0.056	-0.038	0.059
	(0.000)		(0.051)	(0.100)	(0.573)	(0.082)	(0.179)	(0.168)	(0.561)	(0.303)	(0.489)	(0.282)
CAR(2,30)	0.912	-0.121		0.016	-0.039	0.048	-0.241	0.254	-0.211	0.230	-0.182	0.173
	(0.000)	(0.026)		(0.771)	(0.473)	(0.383)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
CAV(-1,30)	0.044	-0.077	0.078		0.713	0.949	-0.002	-0.013	0.004	-0.001	-0.058	0.005
	(0.423)	(0.159)	(0.151)		(0.000)	(0.000)	(0.975)	(0.810)	(0.946)	(0.985)	(0.291)	(0.921)
CAV(-1,1)	-0.037	0.008	-0.042	0.568		0.522	0.104	-0.129	0.091	-0.094	0.004	-0.048
	(0.504)	(0.879)	(0.447)	(0.000)		(0.000)	(0.057)	(0.018)	(0.095)	(0.084)	(0.938)	(0.383)
CAV(2,30)	0.063	-0.092	0.105	0.958	0.308		-0.044	0.038	-0.027	0.036	-0.068	0.019
	(0.246)	(0.093)	(0.054)	(0.000)	(0.000)		(0.417)	(0.484)	(0.615)	(0.513)	(0.211)	(0.724)
NEG_BPW	-0.265	-0.046	-0.255	-0.017	0.082	-0.048		-0.941	0.894	-0.880	0.692	-0.703
	(0.000)	(0.398)	(0.000)	(0.763)	(0.134)	(0.383)		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
TONE_BPW	0.274	0.060	0.259	-0.003	-0.114	0.037	-0.935		-0.857	0.904	-0.636	0.715
	(0.000)	(0.274)	(0.000)	(0.962)	(0.036)	(0.500)	(0.000)		(0.000)	(0.000)	(0.000)	(0.000)
NEG_SENTIWS	-0.221	-0.035	-0.214	-0.032	0.071	-0.062	0.901	-0.849		-0.959	0.657	-0.679
	(0.000)	(0.517)	(0.000)	(0.562)	(0.193)	(0.260)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
TONE_SENTIWS	0.239	0.052	0.226	0.029	-0.068	0.057	-0.886	0.908	-0.950		-0.630	0.709
	(0.000)	(0.337)	(0.000)	(0.602)	(0.213)	(0.299)	(0.000)	(0.000)	(0.000)		(0.000)	(0.000)
NEG_LIWC	-0.203	-0.048	-0.190	-0.054	-0.016	-0.056	0.673	-0.619	0.662	-0.638		-0.924
	(0.000)	(0.382)	(0.000)	(0.327)	(0.777)	(0.301)	(0.000)	(0.000)	(0.000)	(0.000)		(0.000)
TONE_LIWC	0.213	0.073	0.190	0.019	-0.007	0.025	-0.675	0.691	-0.652	0.700	-0.916	
	(0.000)	(0.182)	(0.000)	(0.727)	(0.896)	(0.653)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	

Figure I. CARs following the AGM by high vs. low NEG_BPW



This figure shows cumulative abnormal returns (CARs) across all CEO speeches as well as segregated by a median split on NEG_BPW. The speeches' negativity and abnormal returns are estimated as described in Appendix I. Abnormal returns are cumulated from 5 days before the annual general meeting (AGM) until 30 days after the AGM. CARs are shown in percent.

Figure II. CARs following the AGM by high vs. low TONE_BPW



This figure shows cumulative abnormal returns (CARs) across all CEO speeches as well as segregated by a median split on TONE_BPW. The speeches' negativity and abnormal returns are estimated as described in Appendix I. Abnormal returns are cumulated from 5 days before the annual general meeting (AGM) until 30 days after the AGM. CARs are shown in percent.

Table IV. Test of differences of cumulative abnormal returns

This table sorts the CARs following the annual general meeting into quartiles with respect to NEG_BPW and TONE_BPW and compares the differences in mean and median CARs between the highest and lowest quartiles of textual sentiment for all time windows under investigation. Statistical significance of the differences in CARs between the highest and the lowest quartile are assessed by t and z test statistics, respectively. ***,**, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix I.

		Q1	Q2	Q3	Q4	DIFF Q4-Q1	t-Statistic	Wilcoxon rank-sum z-Statistic
Panel A: CAR (-1,30)								
NEG_BPW	Mean	0.018	0.009	0.007	-0.038	-0.057	-5.117***	
	Median	0.019	0.008	0.008	-0.042	-0.061		-5.050***
TONE_BPW	Mean	-0.033	0.003	0.009	0.018	0.051	4.525***	
	Median	-0.034	0.006	0.008	0.017	0.052		4.441***
Panel B: CAR (-1,1)								
NEG_BPW	Mean	0.002	0.002	0.000	-0.002	-0.005	-0.980	
	Median	0.000	0.001	-0.001	-0.007	-0.008		-1.377
TONE_BPW	Mean	-0.003	0.002	0.001	0.001	0.004	0.868	
	Median	-0.005	0.000	0.000	0.000	0.005		1.077
Panel C: CAR (2,30)								
NEG_BPW	Mean	0.016	0.007	0.007	-0.036	-0.052	-4.920***	
	Median	0.016	0.012	0.009	-0.040	-0.056		-4.874***
TONE_BPW	Mean	-0.031	0.001	0.008	0.016	0.047	4.394***	
	Median	-0.036	0.007	0.010	0.016	0.052		4.451***

Table V. Determinants of cumulative abnormal returns

This table shows regression results of CAR(-1,30), CAR(-1,1), and CAR(2,30) on our measures of textual sentiment as well as on a comprehensive set of control variables. Robust standard errors are presented in parentheses. ***,**, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix I.

	CAR(-1,30)		CAR(-1,1)		CAR(2,30)	
	(1)	(2)	(3)	(4)	(5)	(6)
NEG_BPW	-0.037***		-0.006		-0.031***	
	(0.010)		(0.005)		(0.010)	
TONE_BPW		0.089***		0.018*		0.071***
		(0.022)		(0.010)		(0.021)
log(COUNT)	0.015	0.016	0.003	0.004	0.012	0.011
	(0.027)	(0.027)	(0.011)	(0.011)	(0.026)	(0.025)
log(IND)	0.100	0.090	0.020	0.021	0.080	0.069
	(0.084)	(0.082)	(0.034)	(0.034)	(0.084)	(0.082)
EPS_SURP	0.000	0.000	-0.000	0.000	0.000	0.000
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
DIV_SURP_POS	-0.005	-0.005	0.001	0.001	-0.006	-0.006
	(0.011)	(0.010)	(0.004)	(0.004)	(0.011)	(0.011)
DIV_SURP_NEG	0.003	0.005	-0.004	-0.003	0.006	0.008
	(0.013)	(0.014)	(0.006)	(0.006)	(0.013)	(0.013)
log(SIZE)	0.002	0.001	0.003	0.003	-0.001	-0.002
	(0.005)	(0.005)	(0.003)	(0.003)	(0.005)	(0.005)
M2B	-0.002	-0.003	-0.004**	-0.004**	0.001	0.001
	(0.004)	(0.004)	(0.002)	(0.002)	(0.003)	(0.003)
LEVERAGE	-0.129	-0.137	-0.009	-0.009	-0.119	-0.129
	(0.088)	(0.090)	(0.050)	(0.050)	(0.092)	(0.094)
ROA	0.004	0.004	0.001	0.001	0.003	0.004
	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)
VOLATILITY	-0.565	-0.689	0.040	0.025	-0.605	-0.713
	(0.835)	(0.807)	(0.614)	(0.601)	(0.850)	(0.840)
log(VOLUME)	-0.002	-0.002	-0.002	-0.002	0.000	0.000
	(0.004)	(0.004)	(0.002)	(0.002)	(0.004)	(0.004)
Constant	0.062	-0.024	-0.004	-0.011	0.023	0.026
	(0.143)	(0.148)	(0.063)	(0.061)	(0.143)	(0.139)
Year Dummies	yes	yes	yes	yes	yes	yes
Observations	304	304	304	304	304	304
R-squared	0.140	0.142	0.051	0.056	0.133	0.131

Table VI. Positive textual sentiment and cumulative abnormal returns

This table shows regression results of CAR(-1,30), CAR(-1,1), and CAR(2,30) on POS_BPW and on a comprehensive set of control variables. Robust standard errors are presented in parentheses. ***,**, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix I.

	CAR(-1,30)	CAR(-1,1)	CAR(2,30)
	(1)	(2)	(3)
POS_BPW	0.014**	0.004	0.010
	(0.006)	(0.003)	(0.006)
log(COUNT)	-0.008	0.000	-0.008
	(0.025)	(0.010)	(0.024)
log(IND)	0.024	0.008	0.016
	(0.078)	(0.034)	(0.078)
EPS_SURP	0.000	-0.000	0.001
	(0.001)	(0.000)	(0.001)
DIV_SURP_POS	0.001	0.002	-0.001
	(0.011)	(0.004)	(0.011)
DIV_SURP_NEG	0.004	-0.003	0.007
	(0.014)	(0.006)	(0.014)
log(SIZE)	0.002	0.003	-0.001
	(0.005)	(0.003)	(0.006)
M2B	-0.001	-0.004**	0.002
	(0.004)	(0.002)	(0.003)
LEVERAGE	-0.182*	-0.018	-0.164*
	(0.093)	(0.050)	(0.095)
ROA	0.006*	0.001	0.005
	(0.003)	(0.002)	(0.003)
VOLATILITY	-0.778	0.007	-0.785
	(0.821)	(0.593)	(0.868)
log(VOLUME)	-0.002	-0.002	-0.000
	(0.004)	(0.002)	(0.004)
Constant	0.075	-0.006	0.081
	(0.140)	(0.060)	(0.134)
Year Dummies	yes	yes	yes
Observations	304	304	304
R-squared	0.108	0.051	0.106

Table VII. Determinants of cumulative abnormal returns, by different word lists

This table shows regression results of CAR(-1,30), CAR(-1,1), and CAR(2,30) on our measures of textual sentiment as well as on a comprehensive set of control variables. Measures of textual sentiment are standardized to have a mean of 0 and a standard deviation of 1. Robust standard errors are presented in parentheses. ***,**, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix I.

Panel A: Regression Results for negative textual sentiment

	CAR(-1,30)			CAR(-1,1)			CAR(2,30)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
NEG_BPW	-0.020***			-0.003			-0.017***		
	(0.006)			(0.003)			(0.005)		
NEG_SENTIWS		-0.014**			-0.002			-0.011**	
		(0.006)			(0.003)			(0.005)	
NEG_LIWC			-0.01**			-0.002			-0.008*
			(0.005)			(0.002)			(0.005)
Constant	0.019	0.069	0.144	-0.004	0.003	0.016	0.023	0.066	0.128
	(0.147)	(0.146)	(0.137)	(0.063)	(0.063)	(0.058)	(0.143)	(0.143)	(0.134)
Year dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	304	304	304	304	304	304	304	304	304
R-squared	0.140	0.115	0.109	0.051	0.048	0.047	0.133	0.114	0.109

Panel B: Regression results for relative positive textual sentiment

	CAR(-1,30)			CAR(-1,1)			CAR(2,30)		
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
TONE_BPW	0.021***			0.004*			0.017***		
	(0.005)			(0.002)			(0.005)		
TONE_SENTIWS		0.015***			0.003			0.012**	
		(0.005)			(0.003)			(0.005)	
TONE_LIWC			0.010*			0.003			0.007
			(0.005)			(0.002)			(0.005)
Constant	0.015	0.08	0.128	-0.011	0.002	0.011	0.026	0.078	0.118
	(0.144)	(0.144)	(0.136)	(0.061)	(0.060)	(0.059)	(0.139)	(0.140)	(0.133)
Year dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	304	304	304	304	304	304	304	304	304
R-squared	0.142	0.119	0.108	0.056	0.051	0.051	0.131	0.115	0.106

Table VIII. Model comparison tests

This table present J-test and Cox-Pesaran Deaton test statistics for models presented in Table VII.

	J-test	Cox-Pesaran-Deaton test
Model (1) vs (2)	-1.08	1.00
Model (2) vs (1)	3.04***	-4.47***
Model (1) vs (3)	-0.10	0.11
Model (3) vs (1)	3.19***	-7.19***
Model (4) vs (5)	-0.24	0.23
Model (5) vs (4)	0.98	-1.36*
Model (4) vs (6)	0.07	-0.08
Model (6) vs (4)	1.10	-2.26**
Model (7) vs (8)	-1.01	0.93
Model (8) vs (7)	2.70***	-4.02***
Model (7) vs (9)	-0.14	0.14
Model (9) vs (7)	2.80***	-6.39***
Model (10) vs (11)	-0.90	0.86
Model (11) vs (10)	2.88***	-4.03***
Model (10) vs (12)	-0.16	0.16
Model (12) vs (10)	3.32***	-7.82***
Model (13) vs (14)	-0.36	0.34
Model (14) vs (13)	1.29	-1.77**
Model (13) vs (15)	0.36	-0.42
Model (15) vs (13)	1.31	-2.41***
Model (16) vs (17)	-0.77	0.73
Model (17) vs (16)	2.39**	-3.35***
Model (16) vs (18)	-0.33	0.31
Model (18) vs (16)	2.84***	-7.24***

Table IX. Determinants of cumulative abnormal returns, by weighting schemes employed

This table shows regression results of CAR(-1,30), CAR(-1,1), and CAR(2,30) on our measures of textual sentiment as well as on a comprehensive set of control variables. Measures of textual sentiment are standardized to have a mean of 0 and a standard deviation of 1. Robust standard errors are presented in parentheses. ***,**, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix I.

Panel A: Regression Results

	CAR(-1,30)				CAR(-1,1)				CAR(2,30)			
	equal	idf	equal	idf	equal	idf	equal	idf	equal	idf	equal	idf
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
NEG_ BPW	-0.020***	-0.012**			-0.003	-0.001			-0.017***	-0.011*		
	(0.006)	(0.006)			(0.003)	(0.003)			(0.005)	(0.006)		
TONE_ BPW			0.021***	0.020			0.004*	0.004			0.017***	0.016***
			(0.005)	(0.005)			(0.002)	(0.002)			(0.005)	(0.005)
Constant	0.019	-0.032	0.015	0.011	-0.004	0.003	-0.011	-0.009	0.023	-0.034	0.026	0.02
	(0.147)	(0.158)	(0.144)	(0.140)	(0.063)	(0.076)	(0.061)	(0.062)	(0.143)	(0.152)	(0.139)	(0.135)
Year dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
N	304	304	304	304	304	304	304	304	304	304	304	304
R^2	0.140	0.110	0.142	0.140	0.051	0.046	0.056	0.054	0.133	0.112	0.131	0.131

Panel B: Model comparison tests

	J-test	Cox-Pesaran-Deaton test
Model (1) vs (2)	-0.85	0.78
Model (2) vs (1)	3.25***	-5.85***
Model (3) vs (4)	0.69	-0.75
Model (4) vs (3)	0.91	-1.00
Model (5) vs (6)	-0.87	0.63
Model (6) vs (5)	1.59	-4.89***
Model (7) vs (8)	-0.11	0.11
Model (8) vs (7)	0.82	-0.95
Model (9) vs (10)	-0.48	0.46
Model (10) vs (9)	2.63***	-4.45***
Model (11) vs (12)	0.77	-0.84
Model (12) vs (11)	0.57	-0.61

Table X. IDF weighted CAR regressions with general language dictionaries

This table shows regression results of CAR(-1,30), CAR(-1,1), and CAR(2,30) on our measures of textual sentiment individually, as well as on a comprehensive set of control variables. Measures of textual sentiment are standardized to have a mean of 0 and a standard deviation of 1. Robust standard errors are presented in parentheses. ***,**, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix I.

Panel A: Regression results for negative textual sentiment

	CAR(-1,30)			CAR(-1,1)			CAR(2,30)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
NEG_BPW	-0.012**			-0.001			-0.011*		
	(0.006)			(0.003)			(0.006)		
NEG_SENTIWS		-0.008			-0.001			-0.008	
		(0.006)			(0.003)			(0.006)	
NEG_LIWC			-0.008*			0.000			-0.008*
			(0.004)			(0.002)			(0.004)
Constant	-0.032	0.030	0.050	0.003	0.009	0.018	-0.034	0.021	0.031
	(0.158)	(0.159)	(0.146)	(0.076)	(0.078)	(0.067)	(0.152)	(0.157)	(0.139)
Year dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	304	304	304	304	304	304	304	304	304
R-squared	0.110	0.103	0.103	0.046	0.045	0.045	0.112	0.106	0.107

Panel B: Regression results for relative positive textual sentiment

	CAR(-1,30)			CAR(-1,1)			CAR(2,30)		
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
TONE_BPW	0.020***			0.004			0.016***		
	(0.005)			(0.002)			(0.005)		
TONE_SENTIWS		0.016***			0.002			0.014***	
		(0.005)			(0.002)			(0.005)	
TONE_LIWC			0.010**			0.001			0.009**
			(0.004)			(0.002)			(0.004)
Constant	0.011	0.053	0.082	-0.009	0.004	0.009	0.020	0.049	0.074
	(0.140)	(0.140)	(0.137)	(0.062)	(0.061)	(0.060)	(0.135)	(0.138)	(0.132)
Year dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	304	304	304	304	304	304	304	304	304
R-squared	0.140	0.125	0.111	0.054	0.048	0.046	0.131	0.123	0.112

Table XI. Test of differences of cumulative abnormal trading volumes

This table sorts the CAVs following the annual general meeting into quartiles with respect NEG_BPW and TONE_BPW and compares the mean and median CAV differences between the highest and lowest quartiles of textual sentiment for all time windows under investigation. Statistical significance of the CAV differences between the highest and the lowest quartile are assessed by t and z test statistics, respectively. ***,**, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix I.

		Q1	Q2	Q3	Q4	DIFF Q4-Q1	t-Statistic	Wilcoxon rank-sum z-Statistic
Panel A: CAV (-1,30)								
NEG_BPW	Mean	0.663	1.586	1.423	0.323	-0.341	0.231	
	Median	-1.767	-0.182	-0.461	-1.274	0.493		-0.201
TONE_BPW	Mean	1.114	1.590	0.590	0.706	0.911	0.269	
	Median	-1.038	0.032	-0.999	-1.164	-0.126		-0.123
Panel B: CAV (-1,1)								
NEG_BPW	Mean	1.125	1.304	1.671	1.912	0.786	1.634	
	Median	0.377	0.556	0.853	0.735	0.358		1.481
TONE_BPW	Mean	1.853	2.250	0.877	1.033	-0.819	-1.857*	
	Median	1.076	0.924	0.320	0.327	-0.749		-1.984**
Panel C: CAV (2,30)								
NEG_BPW	Mean	-0.462	0.282	-0.248	-1.589	-1.127	-0.934	
	Median	-2.183	-1.178	-1.844	-2.982	-0.799		-0.871
TONE_BPW	Mean	-0.739	-0.660	-0.287	-0.327	0.411	0.310	
	Median	-2.362	-1.179	-2.049	-1.821	0.541		0.663

Table XII. Determinants of cumulative abnormal trading volume

This table shows regression results of CAV(-1,30), CAV(-1,1), and CAV(2,30) on our measures of textual sentiment as well as on a comprehensive set of control variables. Compared to the analyses of abnormal returns, we utilize the same set of control variables for our analyses on abnormal trading volume except for log(VOLUME), which is not included in the CAV regressions. Robust standard errors are presented in parentheses. ***,**, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix I.

	CAV(-1,30)		CAV(-1,1)		CAV(2,30)	
	(1)	(2)	(3)	(4)	(5)	(6)
NEG_BPW	-0.044		0.417		-0.461	
	(1.107)		(0.437)		(0.862)	
TONE_BPW		-2.413		-2.305**		-0.108
		(2.654)		(1.025)		(2.240)
log(COUNT)	0.917	0.056	1.520	1.063	-0.603	-1.007
	(3.033)	(3.068)	(0.942)	(0.908)	(2.573)	(2.641)
log(IND)	5.102	3.111	3.860	2.946	1.242	0.165
	(9.778)	(9.724)	(2.884)	(2.876)	(8.442)	(8.304)
EPS_SURP	0.144	0.147	0.036	0.037*	0.108	0.110
	(0.095)	(0.094)	(0.022)	(0.022)	(0.088)	(0.089)
DIV_SURP_POS	-1.429	-1.232	-0.068	0.037	-1.362	-1.269
	(1.597)	(1.573)	(0.438)	(0.432)	(1.416)	(1.386)
DIV_SURP_NEG	-0.272	-0.392	0.009	-0.077	-0.281	-0.314
	(1.674)	(1.683)	(0.647)	(0.653)	(1.280)	(1.293)
log(SIZE)	0.908***	0.944***	0.494***	0.527***	0.414	0.416
	(0.321)	(0.325)	(0.132)	(0.136)	(0.255)	(0.260)
M2B	0.273	0.375	0.250*	0.313**	0.023	0.061
	(0.402)	(0.418)	(0.138)	(0.142)	(0.330)	(0.344)
LEVERAGE	-4.190	-5.479	-2.037	-2.600	-2.153	-2.879
	(11.240)	(11.117)	(3.783)	(3.779)	(8.889)	(8.804)
ROA	-0.117	-0.075	-0.029	-0.010	-0.089	-0.065
	(0.358)	(0.351)	(0.108)	(0.108)	(0.287)	(0.281)
VOLATILITY	-240.315***	-243.972***	-99.783***	-100.202***	-140.531**	-143.771**
	(87.767)	(86.419)	(31.305)	(30.827)	(67.208)	(66.215)
Constant	-1.608	3.358	-7.746	-4.161	6.138	7.519
	(16.994)	(17.560)	(5.109)	(4.951)	(14.294)	(15.178)
Year Dummies	yes	yes	yes	yes	yes	yes
Observations	304	304	304	304	304	304
R-squared	0.168	0.170	0.255	0.267	0.104	0.103

Table XIII. CAV regressions and weighting

This table shows regression results of CAV(-1,30), CAV(-1,1), and CAV(2,30) on our measures of textual sentiment individually, as well as on a comprehensive set of control variables. Measures of textual sentiment are standardized to have a mean of 0 and a standard deviation of 1. Robust standard errors are presented in parentheses. ***,**, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix I.

Panel A: Regression results

	CAV(-1,30)				CAV(-1,1)				CAV(2,30)			
	equal	idf	equal	idf	equal	idf	equal	idf	equal	idf	equal	idf
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
NEG_BPW	-0.024	-0.346			0.229	0.106			-0.253	-0.452		
	(0.608)	(0.628)			(0.240)	(0.301)			(0.473)	(0.512)		
TONE_BPW			-0.571	-0.447			-0.546**	-0.450*			-0.025	0.003
			(0.628)	(0.607)			(0.243)	(0.242)			(0.530)	(0.510)
Constant	-1.659	-6.818	2.299	1.788	-7.265	-7.162	-5.173	-5.490	5.606	0.344	7.472	7.278
	(17.256)	(18.694)	(17.246)	(16.935)	(5.185)	(7.054)	(4.896)	(4.891)	(14.522)	(15.080)	(14.819)	(14.585)
Year dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	304	304	304	304	304	304	304	304	304	304	304	304
R-squared	0.168	0.168	0.17	0.169	0.255	0.253	0.267	0.263	0.104	0.105	0.103	0.103

Panel B: Model comparison tests

	J-test	Cox-Pesaran-Deaton test
Model (1) vs (2)	0.65	-6.65***
Model (2) vs (1)	-0.44	0.40
Model (3) vs (4)	-0.22	0.21
Model (4) vs (3)	0.53	-0.64
Model (5) vs (6)	-0.41	0.34
Model (6) vs (5)	1.00	-2.14**
Model (7) vs (8)	-0.38	0.38
Model (8) vs (7)	1.31	-1.55*
Model (9) vs (10)	0.60	-1.05
Model (10) vs (9)	-0.14	0.13
Model (11) vs (12)	0.11	0.05
Model (12) vs (11)	0.12	-0.43

Table XIV. CAV regressions and general language dictionaries

This table shows regression results of CAV(-1,30), CAV(-1,1), and CAV(2,30) on our measures of textual sentiment individually, as well as on a comprehensive set of control variables. Measures of textual sentiment are standardized to have a mean of 0 and a standard deviation of 1. Robust standard errors are presented in parentheses. ***,**, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Appendix I.

Panel A: Regression Results for negative textual sentiment

	CAV(-1,30)			CAV(-1,1)			CAV(2,30)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
NEG_BPW	-0.024			0.229			-0.253		
	(0.608)			(0.240)			(0.473)		
NEG_SENTIWS		-0.118			0.185			-0.303	
		(0.595)			(0.211)			(0.487)	
NEG_LIWC			-0.377			0.062			-0.440
			(0.485)			(0.160)			(0.401)
Constant	-1.659	-2.259	-2.077	-7.265	-7.609	-8.704*	5.606	5.350	6.627
	(17.256)	(17.157)	(16.845)	(5.185)	(5.203)	(5.169)	(14.522)	(14.442)	(14.130)
Year dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	304	304	304	304	304	304	304	304	304
R-squared	0.168	0.168	0.169	0.255	0.254	0.253	0.104	0.104	0.106

Panel B: Regression Results for TONE

	CAV(-1,30)			CAV(-1,1)			CAV(2,30)		
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
TONE_BPW	-0.571			-0.546**			-0.025		
	(0.628)			(0.243)			(0.530)		
TONE_SENTIWS		-0.329			-0.347			0.018	
		(0.597)			(0.233)			(0.474)	
TONE_LIWC			-0.097			-0.238			0.141
			(0.517)			(0.188)			(0.425)
Constant	2.299	0.168	-1.205	-5.173	-7.045	-8.083	7.472	7.213	6.878
	(17.246)	(16.948)	(16.959)	(4.896)	(5.084)	(5.146)	(14.819)	(14.385)	(14.272)
Year dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes
Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	304	304	304	304	304	304	304	304	304
R-squared	0.170	0.169	0.168	0.267	0.258	0.256	0.103	0.103	0.103

8. Appendix

Appendix I: Variable Descriptions

This table shows descriptions of the variables used in our analyses. COUNT, IND, and our sentiment measures are estimated directly from the CEO speeches. The data to estimate CARs, CAVs, and the remaining variables are gathered from Thompson Reuters Datastream.

Variable	Description
CAR(-1,30)	CAR(-1,30) is cumulative abnormal return from day -1 to day 30 where day 0 is the day of the AGM. Abnormal returns are estimated via a market return model as $AR_{j,t} = RI_{j,t} - RI_{CDAX,t}$ where $AR_{j,t}$ is the abnormal return for speech j at day t and $RI_{j,t}$ is the total return index for speech j at day t, which reflects the theoretical growth in value of a share over a specified period, assuming that dividends are re-invested to purchase additional units of an equity. $RI_{CDAX,t}$ *is the* mean total return index of the German CDAX index which 852 German stocks across the Deutsche Börse's prime and general standard.
CAR(2,30)	CAR(-1,1) is cumulative abnormal return from day -1 to day 1 where day 0 is the day of the AGM. Abnormal returns are estimated as described for CAR(-1,30).
CAR(2,30)	CAR(2,30) is cumulative abnormal return from day 2 to day 30 where day 0 is the day of the AGM. Abnormal returns are estimated as described for CAR(-1,30).
CAV(-1,30)	CAV(-1,30) is cumulative abnormal trading volume from day -1 to day 30 where day 0 is the day of the AGM. The abnormal trading volume is estimated as $AV_{j,t} = \frac{VOLUME_{j,t}}{\overline{VOLUME}_{j,t}} - 1$ where $VOLUME_{j,t}$ is the volume for firm j at day t, and $\overline{VOLUME}_{j,t}$ is the mean volume for firm j from day t=-252 to t=-1.
CAV(-1,1)	CAV(-1,1) is cumulative abnormal trading volume from day -1 to day 1 where day 0 is the day of the AGM. Abnormal trading volume are estimated as described for CAV(-1,30).
CAV(2,30)	CAV(2,30) is cumulative abnormal trading volume from day 2 to day 30 where day 0 is the day of the AGM. Abnormal trading volume are estimated as described for CAV(-1,30).
COUNT	COUNT represents the CEO speeches' length in terms of the total number of words.
IND	IND is the number if individual words in a CEO speech divided by the speech's total number of words.
POS_BPW	POS_BPW represents the CEO speeche's number positive words as classified by our BPW dictionary, divided by the speech's total number of words.
NEG_BPW	NEG_BPW represents the CEO speeche's number negative words as classified by our BPW dictionary, divided by the speech's total number of words. NEG_SENTIWS and NEG_LIWC are estimated analogously using the SENTIWS and LIWC dictionary, respectively.
TONE_BPW	TONE measures a speeches positivity relative to its negativity and is calculated as $TONE_{1,j} = \frac{POSITIVE_j - NEGATIVE_j}{POSITIVE_j + NEGATIVE_j}$ where POSITIVEj is the number of positive words, NEGATIVEj the number of negative words of speech j as classified by our BPW dictionary. TONE_SENTIWS and TONE_LIWC are estimated analogously using the SENTIWS and LIWC dictionary, respectively.
EPS_SURP	EPS_SURP is the earnings surprise and is calculated as $EPS\_SURP_j = \frac{EPS_j - EPS_{j,t-1YEAR}}{STOCKPRICE_{j,t-1YEAR}} * 100$ where EPS_j is the most recent earnings per share release for the CEO's company at the time of speech j, $EPS_{j,t\text{-}1YEAR}$ is the most recent earnings per share release for the CEO's company one year before the day of speech j and $STOCKPRICE_{j,t\text{-}1YEAR}$ is the stock price of the CEO's company one year before the date of speech j.
DIV_SURP_POS	DIV_SURP_POS is a dummy variable that equals one if the dividend was increased compared to the previous year. Zero otherwise.
DIV_SURP_NEG	DIV_SURP_NEG is a dummy variable that equals one if the dividend was decreased compared to the previous year. Zero otherwise.
SIZE	SIZE measures the companies' market value at the day of the speech as the share price multiplied by the number of ordinary shares in issue. It is displayed in Euro millions.
M2B	M2B reflect the market to book ratio and is defined as the market value of the ordinary equity divided by the balance sheet value of the ordinary equity in the company.
LEVERAGE	LEVERAGE describes the total liabilities by total assets ratio.
ROA	ROA describes the companies' return on assets and is estimated as net income divided by total assets times one hundred.
VOLATILITY	VOLATILITY is estimated as the daily returns' standard deviation for the time window of minus 90 days to minus 10 days prior the AGM.
VOLUME	VOLUME describes the number of shares traded for a stock on the day of shareholder meeting and is expressed in thousands.

CFS Working Paper Series

Recent Issues

All CFS Working Papers are available at *www.ifk-cfs.de*.

No.	Authors	Title
582	Nikolaus Hautsch and Stefan Voigt	*Large-Scale Portfolio Allocation Under Transaction Costs and Model Uncertainty*
581	Martin D. Gould, Nikolaus Hautsch, Sam D. Howison, and Mason A. Porter	*Counterparty Credit Limits: An Effective Tool for Mitigating Counterparty Risk?*
580	Nikolaus Hautsch, Michael Noé, and S. Sarah Zhang	*The Ambivalent Role of High-Frequency Trading in Turbulent Market Periods*
579	Klaus Adam and Henning Weber	*Optimal Trend Inflation*
578	Günter Coenen, Michael Ehrmann, Gaetano Gaballo, Peter Hoffmann, Anton Nakov, Stefano Nardelli, Eric Persson, and Georg Strasser	*Communication of monetary policy in unconventional times*
577	Francis X. Diebold and Frank Schorfheide	*Real-Time Forecast Evaluation of DSGE Models with Stochastic Volatility*
576	Mert Demirer, Francis X. Diebold, Laura Liu, and Kamil Yilmaz	*Estimating global bank network connectedness*
575	Francis X. Diebold, Laura Liu, and Kamil Yilmaz	*Commodity Connectedness*
574	Annamaria Lusardi, Olivia S. Mitchell, and Noemi Oggero	*Debt and Financial Vulnerability on the Verge of Retirement*