Bannier, Christina E.; Pauls, Thomas; Walter, Andreas

Conference Paper

CEO-speeches and stock returns

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2017: Alternative Geld- und Finanzarchitekturen - Session: Behavioral Finance, No. G16-V3

Provided in Cooperation with:
Verein für Socialpolitik / German Economic Association

Suggested Citation: Bannier, Christina E.; Pauls, Thomas; Walter, Andreas (2017) : CEO-speeches and stock returns, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2017: Alternative Geld- und Finanzarchitekturen - Session: Behavioral Finance, No. G16-V3, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel, Hamburg

This Version is available at:
https://hdl.handle.net/10419/168192

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 volume around the AGM. By adapting a finance-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 around the AGM. Investors hence seem to perceive the sentiment of CEO speeches at AGMs as a valuable indicator of future firm performance. Our results are robust against different weighting schemes and our dictionary appears to be better suited to grasp the sentiment of German business documents compared to general dictionaries.

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 use various means of distributing information to relevant stakeholders. 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 speech of the company's CEO held at the AGM hardly any. Only few studies investigate the market reaction to the AGM at all: For example, Firth (1981) 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 presents one of only few opportunities for managers 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 – that 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.

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 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 and measure the financial market reaction subsequent to the AGM in general. 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 this respect, we confirm only partly the results of the earlier study by Martinez-Blasco et al. (2015) who consider a slightly different and also shorter time period.

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).

Referring to Bannier et al. (2017), we gauge the sentiment of the CEO speeches and assess the financial market reaction to the AGM with respect to the sentiment in the CEO speeches. Our results show the sentiment to be significantly related to cumulative abnormal returns and trading volume. More precisely, we find that

[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.

the cumulative abnormal returns decrease along with the share of negative words in the speech over the interval from 2 to 30 days following the AGM. For the relative positivity of the speech, i.e., its positivity relative to negativity, we find a positive association with cumulative abnormal returns both immediately around the AGM and, even more strongly, over the longer time interval from 2 to 30 days post AGM. Interestingly, a speech's relative positivity is also related with a lower cumulative abnormal trading volume around the AGM. A more positive sentiment of the CEO speeches hence goes along with higher cumulative abnormal returns in the short and medium term and lower short-term cumulative abnormal 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 for the German language using a finance- and accounting-specific dictionary as 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 which are restricted to the respective sample of text documents. The dictionary provided in Bannier et al. (2017) and employed in this study, in contrast, is designed to capture the finance- and accounting-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 set up and employ this dictionary, we compare our results to existing general German language dictionaries. 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 in grasping the textual sentiment of business-related documents than general language dictionaries. We also check the robustness of our results to different weighting schemes, i.e., equal weighting vs. inverse document frequency weighting as proposed by Loughran and McDonald (2011). We find no improvement in 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 relative positivity, that combines both positive and negative tone, to perform better than other measures of sentiment. 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 be seen as a helpful tool to assess the qualitative information contained in these firms' communication in a replicable manner.

The second contribution of our study is to show that despite the limited novelty of the quantitative information provided in the AGM, there is valuable qualitative information hidden in the annual get-together of managers and shareholders. Our results suggest that financial market participants are indeed picking up the qualitative information contained in the CEO's speech for their investment decisions. As a consequence, the more positive the sentiment of a speech, the higher is the abnormal return around the AGM.

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

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 which contains 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 and 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 returns, return volatility, and trading volume and find 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 returns, return volatility and trading volume following the AGM, indicating that the AGMs of German companies are seen to exhibit material 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 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 the specific situation 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 find 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 distributed to (potential) investors prior to the meetings.. We therefore test whether this qualitative information is inherent in the verbal communication of the CEO at the AGM.

Surprisingly, US companies only rarely publish the transcripts of their CEOs' speeches. A search on the internet webpages of all S&P500 companies shows that only 5.8% of these firms offer any transcripts. Moreover, only 27.41% of those provide more than the most recent transcript. In contrast, 72.50% of the German DAX and MDAX companies offer transcripts and of those, 94.83% publish more than one transcript, allowing for a meaningful analysis. Some of the companies in our sample even offer English translated versions of the CEO's speech. However, even if such translations are provided, these translations are typically altered or shortened and published at a later date than the initial document. From the perspective of financial markets it would therefore be clearly preferable to use a tool that allows to evaluate the qualitative nature of information disclosed in the original German texts.

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 as 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 his 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 which 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 financial and accounting 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]

2.3. GERMAN LANGUAGE DICTIONARIES

When it comes to the analysis of German text documents, two comprehensive general German language dictionaries but no finance- and accounting-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).

[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 over dictionaries used in content analyses, see Kearney and Liu (2014) and Loughran and McDonald (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 analysis (Caton et al., 2015; Jacobi et al., 2016). However, with respect to business-related documents, no context-specific dictionary exists. As many text documents containing relevant information on German companies are published in German exclusively, 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 language corporate texts for their textual sentiment, utilize ad-hoc dictionaries that are constructed from - and thus restricted to – a given set of sample text documents. As the studies conducted for English, the authors also find their ad-hoc dictionaries to achieve more reliable results as compared to findings derived from the general German language dictionaries SENTIWS and LIWC.

In order to analyze German business-related texts comprehensively, Bannier et al. (2017) adapt the English dictionary by Loughran and McDonald (2011), which is most commonly used to analyze finance texts, to the German language, following the methodology of Wolf et al. (2008). As König and Gast (2012) and Hawkins (2015) highlight, 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. We therefore need to control for several linguistic issues such as inflections, compound words, or lexical morphology in setting up word lists to measure sentiment in corporate texts as explained in detail in Bannier et al. (2017). There, the authors also test the equivalence of their adaptation (hereafter BPW dictionary) using a broad sample of quarterly and annual reports of German companies that are available in German and English language.[8] The results show that all categories display high correlation and

[8] 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).

equivalence to their English counterparts, indicating the reliability of their adaptation. [9]

Table I shows an excerpt of a comparison between the LM dictionary, the BPW dictionary and the two general German dictionaries, SENTIWS and LIWC.

<<< Insert Table I about here >>>

Table I reveals, that the German dictionaries' word lists contain much larger numbers of individual words as compared to the English LM dictionary, due to the linguistic issues explained above. However, comparing the BPW to the SENTIWS dictionary reveals that SENTIWS includes 50% more negative and even 600% more positive words than BPW. SENTIWS contains as many negative as positive words, which stands in stark contrast to the finance- and accounting-specific LM dictionary. Note that a direct comparison of the number of individual words between the BPW and the LIWC word lists is not applicable as LIWC includes word stems rather than inflections. However, both general German dictionaries are likely to include words that may mispecify 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 example 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 (Bannier et al., 2017). 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 (Price et al. (2012); Loughran and McDonald (2011, 2015) Henry and Leone (2016), this issue still needs to be examined for 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.

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

3. Data & Methodology

3.1. DATA AND VARIABLE MEASUREMENT

We attempt to gauge the sentiment in CEO speeches held at German companies' AGMs and to assess whether this sentiment is associated with 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 and 90 speakers. 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. Thereby, we 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 quarterly reports. After that, we apply a stop-word list on the reports to filter words which might have important semantic functions, but rarely contribute information (Manning & Schütze, 1999). We use the stop-word lists 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.[10] In a final step, the CEO speeches' numbers of negative and positive words 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), for example, 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

[10] 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/.

usage. Consequently, Tetlock (2007) and Loughran and McDonald (2011) find little incremental information using only a positive wordlist and suggest to use a documents' share of negative words to assess its textual sentiment. We therefore estimate the CEO speeches' share of negative words in the following way:[11]

$$NEG\_BPW_j = \frac{NEGATIVE_j}{COUNT_j} * 100 \quad (1)$$

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:

$$TONE\_BPW_j = \frac{POSITIVE_j - NEGATIVE_j}{POSITIVE_j + NEGATIVE_j} \quad (2)$$

Here, $NEGATIVE_j$ is the number of negative words and $POSITIVE_j$ is the number of positive words in CEO speech j with respect to the negative and positive wordlist of the BPW dictionary. $TONE\_SENTIWS_j$ and $TONE\_LIWC_j$ are calculated analogously. The TONE measures represent the relative positivity of speech j and are scaled between -1 and 1, such 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.

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[12] index as the expected return, which reflects the performance of the entire German equity market:

$$AR_{j,t} = R_{j,t} - R_{CDAX,t} \quad (3)$$

[11] We re-estimate our main-analysis grasping the CEO speeches sentiment using a measure of positivity. We do not find a statistically significant relationship between cumulative abnormal returns and measures of positivity. The results are shown in Appendix II.

[12] 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/.

Here, $AR_{j,t}$ is the abnormal return for speech j at day t and $R_{j,t}$ is the actual return of the respective stock for speech j at day t. $R_{CDAX,t}$ is the return of the CDAX on day t. We analyze the immediate effect of the CEO speeches' sentiment by cumulating the abnormal returns for each stock over a three day period from day -1 to day 1, where day zero is the day of the AGM on which speech j is held. To allow also for a slightly longer market reaction, we also accumulate the abnormal returns for each stock from days 2 to 30 and from days 2 to 60. We thus follow Demers and Vega (2008) who emphasize that qualitative information such as textual sentiment is more difficult for market participants to process than quantitative information. We hence employ three CAR measures, estimated in the following way:

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

$$CAR(2,30)_j = \sum_{t=2}^{30} AR_{j,t} \tag{5}$$

$$CAR(2,60)_j = \sum_{t=2}^{60} AR_{j,t} \tag{6}$$

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 (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-2. 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 compared to the last year. Analogously to the abnormal returns, $AV_{j,t}$ is

accumulated over day -1 to +1, CAV(-1,1), day 2 to 30, CAV(2,30), and day 2 to 60, CAV(2,60).

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(-1,1), CAR(2,30), and CAR(2,60) 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:

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

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 company's size (MV), market to book ratio (M2B), leverage (LEVERAGE), return on assets (ROA), return volatility (VOLATILITY), and trading volume (VOLUME).

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:

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

Here EPS_j is the most recent earnings-per-share release for the company at the time of speech j, $EPS_{j,y-1}$ is the most recent earnings-per-share release for the company one year before the day of speech j and $STOCKPRICE_{j,y-1}$ 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 is increased compared to the previous year's dividend, zero otherwise, and DIV_SURP_NEG is equal to one if a company's dividend is decreased compared to the previous year's dividend, 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 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, TOTAL, and COUNT 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. Thereby, for the multivariate analyses, we utilize the same set of control variables except for log(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 higher occurrence of a word indicates a higher importance.[13] However, as words might diverge in their impact, Manning and Schütze (1999) propose a term-inverse document frequency measure (tf-idf)

[13] For a comprehensive overview over studies using equal weighting, cf. Henry and Leone (2016).

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. Further, 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, the authors find that tf-idf weighting mitigates the impact of misclassified words, or noise, 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 our BPW dictionary and compare the results to the SENTIWS and LIWC dictionaries, but we will also evaluate the efficacy of equal weighting versus tf-idf weightings in measuring sentiment.

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 find statistically significant abnormal returns around the AGM. With respect to cumulative abnormal trading volumes, we find statistically significant trading volumes only 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, show a positive tonality, which 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 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[14]. Furthermore, Kim and Meschke (2014) investigate CEO interviews on CNBC and find the share of negative words to be 1.38% and the relative positivity to equal 0.582[15]. The results from the BPW word lists are hence well in line with the earlier studies.

<<< 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 only occasionally statistically significant to CAR(2,60). 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(2,60). 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(2,60). The LIWC measures, in contrast, show basically no significant correlation to trading volumes.

[14] Calculated as (2.967-1.029)/(2.967+1.029).

[15] Calculated as (5.24-1.38)/(5.25+1.38).

4.2. THE CEO SPEECHES' SENTIMENT EFFECT ON STOCK PRICES

4.2.1 THE CONTEXT-SPECIFIC BPW DICTIONARY

Before we proceed to examining 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 the AGM to 60 days after for different levels of negativity and tonality, respectively. 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, a small spread develops around the day of the AGM: Less negative CEO speeches outperform more negative ones. Firms with less negative CEO speeches are moreover followed by positive and higher-than-average CARs. More negative CEO speeches, in contrast, are immediately followed by lower-than-average CARs that are close to zero. The spread between the two groups increases slowly until around day 15. After that, the spread accelerates drastically until day 37. Afterwards, the underperformance of the more negative speeches rebounds to some extent and the spread reaches 3.65% at day 60.

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

Figure II depicts the CARs from 5 days before the AGM to 60 days after the AGM for high and low tonality CEO speeches, where the sample is split at the median TONE_BPW. Similarly to the results from Figure I, above-median TONE_BPW speeches are followed by positive and higher-than-average CARs, while below-median TONE_BPW speeches are followed by lower-than-average CARs close to zero within the first days after the AGM. From day 14 on, where the spread equals 0.89%, it increases dramatically until day 37 to a level of 3.34%. Hereafter, the CARs following below-median TONE_BPW speeches rebound. At day 60, the spread between below and above-median TONE_BPW speeches equals 2.58%.

<<< 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 quartiles with respect to NEG_BPW and TONE_BPW and compares mean and median CARs between the highest and lowest quartiles of textual sentiment for all time windows. Panel A of Table IV contains the univariate results for CAR(-1,1) and shows no statistically or economically significant difference between the extreme

quartiles, no matter which sentiment measure is applied. Panel B of Table IV presents the results for CAR(2,30). In this longer time window, we observe economically and statistically significant Q4-Q1 CARs 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. All differences are statistically significant at the 1% level. Panel C of Table IV presents the results for CAR(2,60). The Q4-Q1 differences are significantly different from zero with respect to NEG_BPW and TONE_BPW, implying that the speeches' sentiment effect is still present 60 days after the AGM. The CAR(2,60) mean (median) difference between the highest and lowest NEG_BPW quartiles equals -4.2 (-2.7) percentage points, and 3.4 (3.0) percentage points between the highest and lowest TONE_BPW quartiles. The differences with respect to both sentiment measures are hence more pronounced in the medium-term time window but persist also for the longer 60-day period after the AGM.

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, the associations appear to become significant only after some days so that our results support the findings by Demers and Vega (2008), who argue that information grasped form textual sentiment needs some time to be incorporated by the financial markets.

<<< Insert Table V about here >>>

Table V finally presents multivariate regressions of CAR(-1,1), CAR(2,30), and CAR(2,60) on NEG_BPW and TONE_BPW and a comprehensive set of control variables. During the initial event window (-1,1), NEG_BPW does not significantly affect the cumulative abnormal returns, confirming the univariate results from Table IV. The combined positivity and negativity measure, TONE_BPW, in contrast, does display 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 compared to a solely negative measure.

For the time period (2,30), Table V shows that 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 TONE by the interquartile range of 0.353 induces 2.5 percentage points higher CAR(2,30). With respect to time period (2,60), NEG_BPW and TONE_BPW do not have a statistically significant association with the cumulative abnormal returns. Generally, the mostly non-significant or only small effects of our sentiment measures in the short term support the notion that qualitative information have greater explanatory power for abnormal stock returns over longer time horizons, consistent with Engelberg (2008), Demers and Vega (2008), or 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, a positive dividend surprise (DIV_SURP_POS) seems to have no statistically significant effect on the CARs and even a negative dividend surprise (DIV_SURP_NEG) shows a positive but only borderline statistically significant effect only over the longest time horizon, i.e., for CAR(2,60).

To summarize, our analyses of cumulative abnormal returns highlight several interesting facts. Our measure of a negative sentiment shows a strong and statistically significant negative relation to CAR(2,30) and CAR(2,60) in a univariate analysis. While the effect on CAR(2,60) does not hold in a multivariate regression, the negative sentiment's effect on CAR(2,30) remains robust and statistically strong. Appendix II re-estimates Table V, substituting NEG_BPW and TONE_BPW with the speeches' share of positive words (POS_BPW). There is no significant effect of the speeches positivity on CARs over any time window. Our results hence 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 underline furthermore the earlier suggestions that recipients tend to grasp a texts' positivity and negativity not in isolation but rather in relation to each other. As a consequence, tonality, i.e., relative positivity, appears to be a superior measure for assessing the qualitative information in a text or speech in the German language as well.

4.2.2 THE BPW VS. GENERAL GERMAN LANGUAGE DICTIONARIES

Table VI re-estimates the regression models presented in Table V, but uses the general German language SENTIWS and LIWC dictionaries to measure the sentiment of the CEO speeches. Note that in contrast to Table V, the textual sentiment measures are now standardized to have a mean of 0 and a standard deviation of 1 as this facilitates comparisons between the different sets of results. For ease of comparison, we also include the (now standardized) regression coefficients for the sentiment measured via our BPW dictionary in the first line of Table VI. 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) or CAR(2,60). With respect to CAR(2,30) in models (4)-(6), all measures show a statistically significant relationship. However, the magnitude and statistical significance of NEG_BPW is greater than those of NEG_SENTIWS or NEG_LIWC. Panel B of Table VI refers to the tonality measure, i.e. the relative positivity. As can be seen, TONE_BPW is significantly related to CAR(-1,1) while the tonality measures based on the general language dictionaries are not. With respect to the time windows (2,30), only TONE_BPW and TONE_SENTIWS are significantly related to CARs, but TONE_BPW is again of higher significance and magnitude. For CAR(2,60) in models (16)-(18), there is no significant association with neither of the tonality measures.

Appendix III presents J-test (Davidson & MacKinnon, 1981) and Cox-Pesaran-Deaton (Pesaran & Deaton, 1978) test statistics for non-nested regressions in order to compare the 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) models including NEG_BPW (model (4)) and TONE_BPW (model (13)) are both more favorable compared to the corresponding models including measures created using the general language SENTIWS and LIWC dictionaries according to both test statistics. Thus, our results indicate the superiority of context-specific dictionaries in grasping the textual sentiment of German business-related documents, underlining the earlier results from English text analyses (Price et al. (2012); Loughran and McDonald (2011, 2015); Henry and Leone (2016).

<<< Insert Table VI about here >>>

4.2.3 WEIGHTING SCHEMES

The previous results were estimated using equal weighting of the words in our dictionary. Table VII re-estimates the regressions from Table V, this time using equal weighting and tf-idf weighting of standardized measures NEG_BPW and TONE_BPW in comparison. Panel A of Table VII 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 coefficient of NEG_BPW in CAR(2,30) regressions equals -0.017 and is statistically significant at the 1%-level using equal weighting (model 5). Using tf-idf weighting, in contrast, the coefficient of NEG_BPW decreases to - 0.011 and is only significant at the 10%-level (model 6). In this case, tf-idf weighting seems to unfavourably affect the results.

Panel B of Table VII reports the results from J-tests and Cox-Pesaran-Deaton tests. They show that none of the equally weighted models can be rejected in favor of the tf-idf weighted models according to both tests. Vice versa, all but one tf-idf weighted models cannot be rejected in favor of the respective equally weighted models. Only model (5) seems to be preferable compared to model (6). Consequently, the results presented in Table VII indicate that, for our sample, tf-idf weighting seems to provide no improvement over equal weighting with respect to measures of relative positive textual sentiment and 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 & Leone, 2016), who find no improvement for measures of relative positivity using tf-idf weighting.

<<< Insert Table VII about here >>>

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 therefore 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 Appendix IV. Indeed, we find that 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.

4.3. THE CEO SPEECHES' SENTIMENT EFFECT ON TRADING VOLUME

In addition to our analyses on stock prices, we also examine the relation between the CEO speeches' sentiment and the abnormal trading volume. For this examination, we only employ the BPW dictionary and start again with a univariate analysis. Analogously to Table IV for CARs, Table VIII shows the differences in CAV(-1,1), CAV(2,30), and CAV(2,60) sorted for quartiles with respect to NEG_BPW and TONE_BPW.

<<< Insert Table VIII about here >>>

As Panel A shows, we find statistically significant differences between the fourth and first quartiles in the short time window, CAV(-1,1) only with respect to TONE_BPW. Significance is obtained both with parametric and non-parametric test statistics. Relative positivity seems to decrease the level of trading volume, such that a CEO speech with a higher share of positive relative to negative words induces a lower trading volume within the initial time window of minus one day to plus one day around the AGM. There are no significant Q4-Q1 differences in CAV(2,30), irrespective of the sentiment measure employed, while Panel C shows only weakly significant differences between the quartiles for the longest time window, CAV(2,60). These significances are only obtained via one type of test statistic, never with both, and are difficult to interpret.

<<< Insert Table IX about here >>>

Table IX is estimated analogously to Table V, substituting CAV for CAR. Table IX confirms the univariate findings from Table VIII 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(2,30) or CAV(2,60).

Appendix IV tests whether results for CAVs are influenced by the weighting scheme applied, and Appendix V investigates the relationship among CAVs and the general language measures of textual sentiment. Similar to our results on CARs, Appendix IV shows that tf.idf weighting does not seem to improve the results and Appendix V reports that general language SENTIWS and LIWC measures do not possess higher explanatory power compared to the context-spe-

cific 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.

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, i.e., only over the medium term period, it appears to drive 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 from two days until 30 or 60 days after the AGM, none of the sentiment measures displays a significant association with the CAVs. The latter finding is somewhat in contrast to Price et al. (2012), who observes 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), though, who observe 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 tonality of the CEO speeches at the AGM: Speeches with particularly low relative positivity should go hand in hand with heightened trading volumes.

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.

As a focus on the CEO speeches of US companies delivers an ineligibly small dataset, 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 adapt the most commonly used English dictionary by Loughran and McDonald (2011) to the German language (Bannier et al., 2017). Particularly, we adapt its positivity and negativity wordlists, controlling for several linguistic issues such as inflections, compound words, and lexical morphology. We gather the transcripts of 338 German CEO speeches and employ our BPW dictionary to 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, negative CEO speeches are followed by negative abnormal returns and relative positive CEO speeches are followed by positive abnormal returns. With respect to cumulative abnormal trading volume, sentiment in CEO speeches seems to have only short term effects. Thereby, relative positive CEO speeches are followed by reduced trading volume, while the speeches' share of negative words seems to possess no explanatory power for cumulative abnormal trading volume. 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, also in accordance with literature on English content analyses by Henry and Leone (2016), we find that using relative measures of a document's positivity to be advantageous compared to using positive or negative measures of sentiment, and inverse term weighting not to yield improved results compared to equal weighting. By that, we provide a framework for future adaptations in other languages and pave the way for further German content analyses using the dictionary-based approach.

We are aware of some limitations in our analyses. First, our study is certainly 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.*

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 English and German language dictionaries for content analysis.

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

Table II. Summary statistics

This table provides descriptive statistics for the full sample of 338 CEO speeches. All variables are defined in Appendix I.

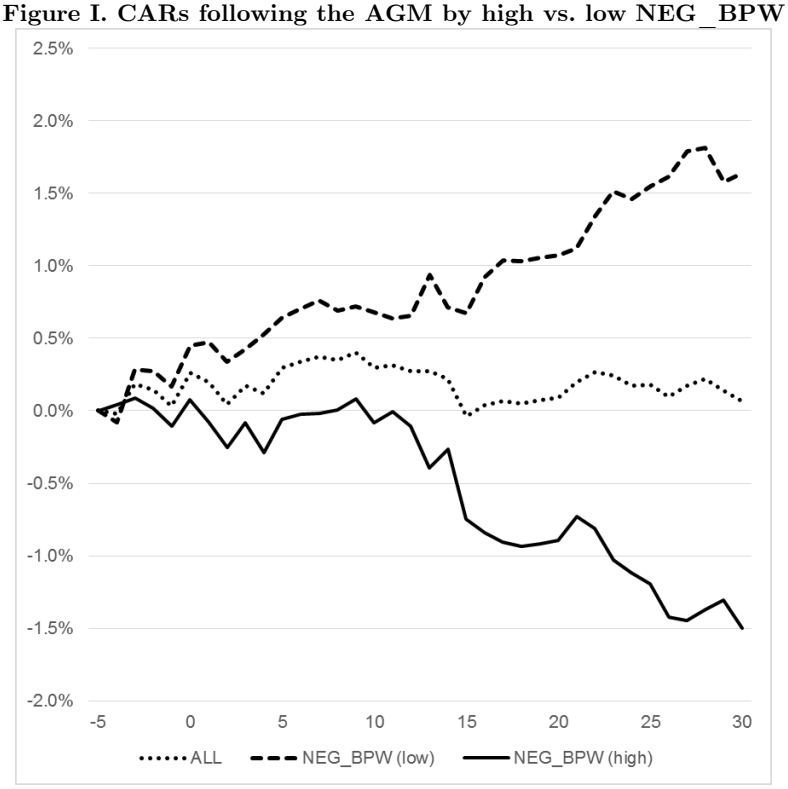
	Mean	Min	p25	p50	p75	Max	SD	N	T-Statistic
Panel A: CARs and CAVs									
CAR(-1,1)	0.001	-0.195	-0.017	0.000	0.017	0.095	0.029	338	0.3173
CAR(2,30)	-0.001	-0.261	-0.042	-0.001	0.041	0.212	0.069	338	-0.3457
CAR(-1,30)	-0.001	-0.277	-0.043	0.004	0.048	0.209	0.071	338	-0.2015
CAV(-1,1)	1.502	-1.566	-0.147	0.626	1.954	19.424	3.089	338	8.9419***
CAV(2,30)	-0.503	-13.948	-4.751	-1.975	1.310	83.692	8.839	338	-1.0470
CAV(-1,30)	0.999	-15.174	-4.403	-0.910	3.397	84.763	10.221	338	1.7970*
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.

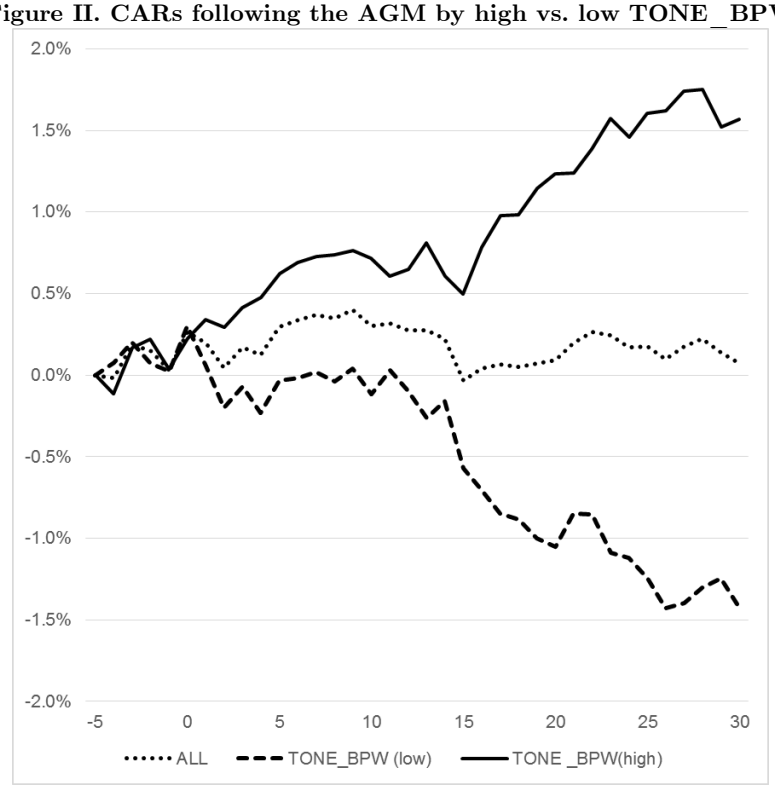
	CAR (-1,1)	CAR (2,30)	CAR (2,60)	CAV (-1,1)	CAV (2,30)	CAV (2,60)	NEG_ BPW	TONE_ BPW	NEG_ SENTIWS	TONE_ SENTIWS	NEG_ LIWC	TONE_ LIWC
CAR(-1,1)		-0.106	-0.047	0.040	-0.086	-0.016	-0.073	0.075	-0.032	0.056	-0.038	0.059
		(0.051)	(0.392)	(0.466)	(0.115)	(0.766)	(0.179)	(0.168)	(0.561)	(0.303)	(0.489)	(0.282)
CAR(2,30)	-0.121		0.617	-0.031	0.046	-0.041	-0.241	0.254	-0.211	0.230	-0.182	0.173
	(0.026)		(0.000)	(0.571)	(0.403)	(0.458)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
CAR(2,60)	0.296	0.912		-0.051	-0.018	-0.095	-0.096	0.104	-0.081	0.116	-0.093	0.095
	(0.000)	(0.000)		(0.351)	(0.742)	(0.080)	(0.079)	(0.057)	(0.139)	(0.034)	(0.087)	(0.082)
CAV(-1,1)	0.008	-0.042	-0.037		0.531	0.441	0.103	-0.126	0.096	-0.095	-0.006	-0.036
	(0.879)	(0.447)	(0.504)		(0.000)	(0.000)	(0.060)	(0.021)	(0.079)	(0.083)	(0.914)	(0.506)
CAV(2,30)	-0.092	0.105	0.063	0.308		0.813	-0.040	0.034	-0.022	0.031	-0.068	0.018
	(0.093)	(0.054)	(0.246)	(0.000)		(0.000)	(0.466)	(0.536)	(0.690)	(0.574)	(0.216)	(0.739)
CAV(2,60)	-0.077	0.078	0.044	0.568	0.958		-0.112	0.113	-0.097	0.101	-0.118	0.086
	(0.159)	(0.151)	(0.423)	(0.000)	(0.000)		(0.039)	(0.038)	(0.076)	(0.064)	(0.031)	(0.115)
NEG_BPW	-0.046	-0.255	-0.265	0.082	-0.048	-0.017		-0.941	0.894	-0.880	0.692	-0.703
	(0.398)	(0.000)	(0.000)	(0.134)	(0.383)	(0.763)		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
TONE_BPW	0.060	0.259	0.274	-0.114	0.037	-0.003	-0.935		-0.857	0.904	-0.636	0.715
	(0.274)	(0.000)	(0.000)	(0.036)	(0.500)	(0.962)	(0.000)		(0.000)	(0.000)	(0.000)	(0.000)
NEG_SENTIWS	-0.035	-0.214	-0.221	0.071	-0.062	-0.032	0.901	-0.849		-0.959	0.657	-0.679
	(0.517)	(0.000)	(0.000)	(0.193)	(0.260)	(0.562)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
TONE_SENTIWS	0.052	0.226	0.239	-0.068	0.057	0.029	-0.886	0.908	-0.950		-0.630	0.709
	(0.337)	(0.000)	(0.000)	(0.213)	(0.299)	(0.602)	(0.000)	(0.000)	(0.000)		(0.000)	(0.000)
NEG_LIWC	-0.048	-0.190	-0.203	-0.016	-0.056	-0.054	0.673	-0.619	0.662	-0.638		-0.924
	(0.382)	(0.000)	(0.000)	(0.777)	(0.301)	(0.327)	(0.000)	(0.000)	(0.000)	(0.000)		(0.000)
TONE_LIWC	0.073	0.190	0.213	-0.007	0.025	0.019	-0.675	0.691	-0.652	0.700	-0.916	
	(0.182)	(0.000)	(0.000)	(0.896)	(0.653)	(0.727)	(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 the 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 and by below and above median TONE_BPW CEO speeches. The speeches' TONE and the abnormal returns are estimated as described in Appendix I. Abnormal returns are cumulated from 10 days before the annual general meeting (AGM) until 60 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 (-1,1)								
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 B: 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***
Panel C: 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***

Table V. Determinants of cumulative abnormal returns

This table shows regression results of CAR(-1.1), CAR(2,30), and CAR(2,60) 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(-1,1)		CAR(2,30)		CAR(2,60)	
	(1)	(2)	(3)	(4)	(5)	(6)
NEG_BPW	-0.006		-0.031***		-0.037 ***	
	(0.005)		(0.010)		(0.010)	
TONE_BPW		0.018*		0.071***		0.089 ***
		(0.010)		(0.021)		(0.022)
log(COUNT)	0.003	0.004	0.012	0.011	0.015	0.016
	(0.011)	(0.011)	(0.026)	(0.025)	(0.027)	(0.027)
log(IND)	0.020	0.021	0.080	0.069	0.100	0.090
	(0.034)	(0.034)	(0.084)	(0.082)	(0.084)	(0.082)
EPS_SURP	-0.000	-0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
DIV_SURP_POS	0.001	0.001	-0.006	-0.006	-0.005	-0.005
	(0.004)	(0.004)	(0.011)	(0.011)	(0.011)	(0.010)
DIV_SURP_NEG	-0.004	-0.003	0.006	0.008	0.003	0.005
	(0.006)	(0.006)	(0.013)	(0.013)	(0.013)	(0.014)
log(SIZE)	0.003	0.003	-0.001	-0.002	0.002	0.001
	(0.003)	(0.003)	(0.005)	(0.005)	(0.005)	(0.005)
M2B	-0.004**	-0.004**	0.001	0.001	-0.002	-0.003
	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)
LEVERAGE	-0.009	-0.009	-0.119	-0.129	-0.129	-0.137
	(0.050)	(0.050)	(0.092)	(0.094)	(0.088)	(0.090)
ROA	0.001	0.001	0.003	0.004	0.004	0.004
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
VOLATILITY	0.040	0.025	-0.605	-0.713	-0.565	-0.689
	(0.614)	(0.601)	(0.850)	(0.840)	(0.835)	(0.807)
log(VOLUME)	-0.002	-0.002	0.000	-0.000	-0.002	-0.002
	(0.002)	(0.002)	(0.004)	(0.004)	(0.004)	(0.004)
Constant	-0.004	-0.011	0.023	0.026	0.062	-0.024
	(0.063)	(0.061)	(0.143)	(0.139)	(0.143)	(0.148)
Year Dummies	yes	yes	yes	yes	yes	yes
Observations	304	304	304	304	304	304
R-squared	0.051	0.056	0.133	0.131	0.140	0.142
Adj. R-squared	-0.012	-0.007	0.075	0.072	0.082	0.084

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

This table shows regression results of CAR(-1.1), CAR(2,30), and CAR(2,60) on our measures of textual sentimentas 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(-1,1)			CAR(2,30)			CAR(-1,30)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
NEG_BPW	-0.003			-0.017***			-0.020***		
	(0.003)			(0.005)			(0.006)		
NEG_SENTIWS		-0.002			-0.011**			-0.014 **	
		(0.003)			(0.005)			(0.006)	
NEG_LIWC			-0.002			-0.008*			-0.010**
			(0.002)			(0.005)			(0.005)
Constant	-0.004	0.003	0.016	0.023	0.066	0.128	0.019	0.069	0.144
	(0.063)	(0.063)	(0.058)	(0.143)	(0.143)	(0.134)	(0.147)	(0.146)	(0.137)
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	0.051	0.048	0.047	0.133	0.114	0.109	0.140	0.115	0.109
Adj. R2	-0.012	-0.015	-0.016	0.075	0.054	0.049	0.082	0.056	0.049

Panel B: Regression results for relative positive textual sentiment

	CAR(-1,1)			CAR(2,30)			CAR(-1,30)		
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
TONE_BPW	0.004*			0.017***			0.021***		
	(0.002)			(0.005)			(0.005)		
TONE_SENTIWS		0.003			0.012**			0.015***	
		(0.003)			(0.005)			(0.005)	
TONE_LIWC			0.003			0.007			0.010 *
			(0.002)			(0.005)			(0.005)
Constant	-0.011	0.002	0.011	0.026	0.078	0.118	0.015	0.080	0.128
	(0.061)	(0.060)	(0.059)	(0.139)	(0.140)	(0.133)	(0.144)	(0.144)	(0.136)
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	0.056	0.051	0.051	0.131	0.115	0.106	0.142	0.119	0.108
Adj. R2	-0.007	-0.013	-0.013	0.072	0.056	0.046	0.084	0.060	0.049

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

This table shows regression results of CAR(-1.1), CAR(2,30), and CAR(2,60) 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(-1,1)				CAR(2,30)				CAR(-1,30)			
	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	-0.003	-0.001			-0.017***	-0.011*			-0.020***	-0.012**		
	(0.003)	(0.003)			(0.005)	(0.006)			(0.006)	(0.006)		
TONE_BPW			0.004*	0.004			0.017***	0.016***			0.021***	0.020***
			(0.002)	(0.002)			(0.005)	(0.005)			(0.005)	(0.005)
Constant	-0.004	0.003	-0.011	-0.009	0.023	-0.034	0.026	0.020	0.019	-0.032	0.015	0.011
	(0.063)	(0.076)	(0.061)	(0.062)	(0.143)	(0.152)	(0.139)	(0.135)	(0.147)	(0.158)	(0.144)	(0.140)
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	0.051	0.046	0.056	0.054	0.133	0.112	0.131	0.131	0.140	0.110	0.142	0.140
Adj. R^2	-0.012	-0.018	-0.007	-0.010	0.075	0.053	0.072	0.073	0.082	0.050	0.084	0.083

Panel B: Model comparison tests

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

Table VIII. Test of differences of cumulative abnormal trading volumes

This table sorts the CAVs following the annual general meeting into quartiles with respect POS_BPW, NEG_BPW, TONE, and UNC_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 (-1,1)								
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**
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
Panel C: 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

Table IX. Determinants of cumulative abnormal trading volume

This table shows regression results of CAV(-1.1), CAV(2,30), and CAV(2,60) 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(-1,1)		CAV(2,30)		CAV(2,60)	
	(1)	(2)	(3)	(4)	(5)	(6)
NEG_BPW	0.417		-0.461		-0.044	
	(0.437)		(0.862)		(1.107)	
TONE_BPW		-2.305**		-0.108		-2.413
		(1.025)		(2.240)		(2.654)
log(COUNT)	1.520	1.063	-0.603	-1.007	0.917	0.056
	(0.942)	(0.908)	(2.573)	(2.641)	(3.033)	(3.068)
log(IND)	3.860	2.946	1.242	0.165	5.102	3.111
	(2.884)	(2.876)	(8.442)	(8.304)	(9.778)	(9.724)
EPS_SURP	0.036	0.037*	0.108	0.110	0.144	0.147
	(0.022)	(0.022)	(0.088)	(0.089)	(0.095)	(0.094)
DIV_SURP_POS	-0.068	0.037	-1.362	-1.269	-1.429	-1.232
	(0.438)	(0.432)	(1.416)	(1.386)	(1.597)	(1.573)
DIV_SURP_NEG	0.009	-0.077	-0.281	-0.314	-0.272	-0.392
	(0.647)	(0.653)	(1.280)	(1.293)	(1.674)	(1.683)
log(SIZE)	0.494***	0.527***	0.414	0.416	0.908***	0.944***
	(0.132)	(0.136)	(0.255)	(0.260)	(0.321)	(0.325)
M2B	0.250*	0.313**	0.023	0.061	0.273	0.375
	(0.138)	(0.142)	(0.330)	(0.344)	(0.402)	(0.418)
LEVERAGE	-2.037	-2.600	-2.153	-2.879	-4.190	-5.479
	(3.783)	(3.779)	(8.889)	(8.804)	(11.240)	(11.117)
ROA	-0.029	-0.010	-0.089	-0.065	-0.117	-0.075
	(0.108)	(0.108)	(0.287)	(0.281)	(0.358)	(0.351)
VOLATILITY	-99.783***	-100.202***	-140.531**	-143.771**	-240.315 ***	-243.972 ***
	(31.305)	(30.827)	(67.208)	(66.215)	(87.767)	(86.419)
Constant	-7.746	-4.161	6.138	7.519	-1.608	3.358
	(5.109)	(4.951)	(14.294)	(15.178)	(16.994)	(17.560)
Year Dummies	yes	yes	yes	yes	yes	yes
Observations	304	304	304	304	304	304
R-squared	0.255	0.267	0.104	0.103	0.168	0.170
Adj. R-squared	0.208	0.221	0.047	0.046	0.115	0.117

8. Appendix

Appendix I: Variable Descriptions

This table shows descriptions of the variables used in our analyses. TOTAL, 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(-1,1)	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 via a market return model as $AR_{j,t} = RI_{j,t} - RI_{CDAX,t}$ where $AR_{j,t}$ ist 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(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(-1,1).
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 as described for CAR(-1,1).
CAV(-1,1)	CAV(-1,1) is cumulative abnormal trading volume from day -1 to day 1where 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 company j at day t, and $\overline{VOLUME}_{j,t}$ is the mean volume for firm j from day t=-252 to t=-1.
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(-1,1).
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. Abnormal trading volume are estimated as described for CAV(-1,1).
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.

Appendix II. Positive textual sentiment and cumulative abnormal returns

This table shows regression results of CAR(-1.1), CAR(2,30), and CAR(-1,30) 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(-1,1)	CAR(2,30)	CAR(-1,30)
	(1)	(2)	(3)
POS_BPW	0.004	0.010	0.014**
	(0.003)	(0.006)	(0.006)
log(COUNT)	0.000	-0.008	-0.008
	(0.010)	(0.024)	(0.025)
log(IND)	0.008	0.016	0.024
	(0.034)	(0.078)	(0.078)
EPS_SURP	-0.000	0.001	0.000
	(0.000)	(0.001)	(0.001)
DIV_SURP_POS	0.002	-0.001	0.001
	(0.004)	(0.011)	(0.011)
DIV_SURP_NEG	-0.003	0.007	0.004
	(0.006)	(0.014)	(0.014)
log(SIZE)	0.003	-0.001	0.002
	(0.003)	(0.006)	(0.005)
M2B	-0.004**	0.002	-0.001
	(0.002)	(0.003)	(0.004)
LEVERAGE	-0.018	-0.164*	-0.182*
	(0.050)	(0.095)	(0.093)
ROA	0.001	0.005	0.006*
	(0.002)	(0.003)	(0.003)
VOLATILITY	0.007	-0.785	-0.778
	(0.593)	(0.868)	(0.821)
log(VOLUME)	-0.002	-0.000	-0.002
	(0.002)	(0.004)	(0.004)
Constant	-0.006	0.081	0.075
	(0.060)	(0.134)	(0.140)
Year Dummies	yes	yes	yes
Observations	304	304	304
R-squared	0.051	0.106	0.108
Adj. R-squared	-0.012	0.046	0.049

Appendix III

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

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

Appendix IV. IDF weighted CAR regressions with general language dictionaries

This table shows regression results of CAR(-1.1), CAR(2,30), and CAR(-1,30) 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(-1,1)			CAR(2,30)			CAR(-1,30)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
NEG_BPW	-0.001			-0.011*			-0.012 **		
	(0.003)			(0.006)			(0.006)		
NEG_SENTIWS		-0.001			-0.008			-0.008	
		(0.003)			(0.006)			(0.006)	
NEG_LIWC			0.000			-0.008*			-0.008 *
			(0.002)			(0.004)			(0.004)
Constant	0.003	0.009	0.018	-0.034	0.021	0.031	-0.032	0.030	0.050
	(0.076)	(0.078)	(0.067)	(0.152)	(0.157)	(0.139)	(0.158)	(0.159)	(0.146)
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	0.046	0.045	0.045	0.112	0.106	0.107	0.110	0.103	0.103
Adj. R2	-0.018	-0.019	-0.019	0.053	0.046	0.048	0.050	0.042	0.043

Panel B: Regression results for relative positive textual sentiment

	CAR(-1,1)			CAR(2,30)			CAR(-1,30)		
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
TONE_BPW	0.004			0.016***			0.020***		
	(0.002)			(0.005)			(0.005)		
TONE_SENTIWS		0.002			0.014***			0.016***	
		(0.002)			(0.005)			(0.005)	
TONE_LIWC			0.001			0.009**			0.010**
			(0.002)			(0.004)			(0.004)
Constant	-0.009	0.004	0.009	0.020	0.049	0.074	0.011	0.053	0.082
	(0.062)	(0.061)	(0.060)	(0.135)	(0.138)	(0.132)	(0.140)	(0.140)	(0.137)
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	0.054	0.048	0.046	0.131	0.123	0.112	0.140	0.125	0.111
Adj. R2	-0.010	-0.016	-0.017	0.073	0.064	0.052	0.083	0.066	0.051

Appendix V. CAV regressions and weighting

This table shows regression results of CAV(-1.1), CAV(2,30), and CAV(2,60) 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(-1,1)				CAV(2,30)				CAV(2,60)			
	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	0.229	0.106			-0.253	-0.452			-0.024	-0.346		
	(0.240)	(0.301)			(0.473)	(0.512)			(0.608)	(0.628)		
TONE_BPW			-0.546**	-0.450*			-0.025	0.003			-0.571	-0.447
			(0.243)	(0.242)			(0.530)	(0.510)			(0.628)	(0.607)
Constant	-7.265	-7.162	-5.173	-5.490	5.606	0.344	7.472	7.278	-1.659	-6.818	2.299	1.788
	(5.185)	(7.054)	(4.896)	(4.891)	(14.522)	(15.080)	(14.819)	(14.585)	(17.256)	(18.694)	(17.246)	(16.935)
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	0.255	0.253	0.267	0.263	0.104	0.105	0.103	0.103	0.168	0.168	0.17	0.169
Adj. R2	0.208	0.206	0.221	0.216	0.047	0.048	0.046	0.046	0.115	0.116	0.117	0.117

Panel B: Model comparison tests

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

Appendix VI. CAV regressions and general language dictionaries

This table shows regression results of CAV(-1.1), CAV(2,30), and CAV(-1,30) 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(-1,1)			CAV(2,30)			CAV(-1,30)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
NEG_BPW	0.229			-0.253			-0.024		
	(0.240)			(0.473)			(0.608)		
NEG_SENTIWS		0.185			-0.303			-0.118	
		(0.211)			(0.487)			(0.595)	
NEG_LIWC			0.062			-0.440			-0.377
			(0.160)			(0.401)			(0.485)
Constant	-7.265	-7.609	-8.704*	5.606	5.350	6.627	-1.659	-2.259	-2.077
	(5.185)	(5.203)	(5.169)	(14.522)	(14.442)	(14.130)	(17.256)	(17.157)	(16.845)
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	0.255	0.254	0.253	0.104	0.104	0.106	0.168	0.168	0.169
Adj. R2	0.208	0.207	0.205	0.047	0.047	0.049	0.115	0.115	0.117

Panel B: Regression Results for TONE

	CAV(-1,1)			CAV(2,30)			CAV(-1,30)		
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
TONE_BPW	-0.546**			-0.025			-0.571		
	(0.243)			(0.530)			(0.628)		
TONE_SENTIWS		-0.347			0.018			-0.329	
		(0.233)			(0.474)			(0.597)	
TONE_LIWC			-0.238			0.141			-0.097
			(0.188)			(0.425)			(0.517)
Constant	-5.173	-7.045	-8.083	7.472	7.213	6.878	2.299	0.168	-1.205
	(4.896)	(5.084)	(5.146)	(14.819)	(14.385)	(14.272)	(17.246)	(16.948)	(16.959)
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	0.267	0.258	0.256	0.103	0.103	0.103	0.170	0.169	0.168
Adj. R2	0.221	0.212	0.209	0.046	0.046	0.047	0.117	0.116	0.115