Jockeying for position in CEO letters: Impression management and sentiment analytics

March 9, 2016

ABSTRACT

This paper evidences the strategic positioning of positive and negative news within a CEO letter as a subtle form of impression management. Based on a unique sample of CEO letters published by DJIA firms between 2000 and 2011, we provide empirical support for the hypothesis that managers exploit the serial position effect by presenting the information in such an order that the reader is likely to have a more positive perception of the underlying message. We find that there is a smile in the frequency of positive words within the letter, and a half-smile in the intratextual distribution of negative words, with a prevalence of negative words at the beginning of the letter. It follows that the difference (net sentiment) shows a right-sided smirk with more positive than negative words overall. We propose sentiment analytics that can compensate for the strategic management of narrative structure by using a novel weighting scheme to aggregate the within-text net sentiment dynamics into a single proxy for the CEO's sentiment. Consistent with the presence of CEO incentives to inflate sentiment, we find that the proposed position weighted sentiment is more pessimistic than the traditional equally-weighted sentiment measure and has more predictive power for the firm performance over the next year.

KEYWORDS: CEO sentiment, Firm profitability, Impression management, Intratextual analysis, Sentiment dynamics

AUTHORS INFO

PAPER INFO

JEL CLASSIFICATION: *G14, G17, M40*

1 Introduction

Prior research in accounting and finance suggests there is an ambiguity in the sentiment expressed in the accounting narratives of financial disclosures by firms [Arslan-Ayaydin et al., 2015; Huang et al., 2014]. Most authors agree that the sentiment of the qualitative sections in corporate communications reduces the information asymmetry between the firm management and the firm stakeholders, and has information value to predict future performance [see, e.g., Davis et al., 2012; Patelli and Pedrini, 2013]. This also explains why the market reaction around the release of the accounting narrative tends to be positively associated with the expressed managerial sentiment. There is however increasing empirical evidence that this potentially valuable information channel is misused by managers, using these qualitative disclosures to influence the perceptions of third parties for their own benefit through various impression management techniques [Arslan-Ayaydin et al., 2015; Huang et al., 2014]. These practices not only make the signal provided by financial disclosures biased, but they also reduce the investors' confidence in the information disclosed by managers [Arslan-Ayaydin et al., 2015; Clatworthy and Jones, 2003; Heaton, 2002; Huang et al., 2014; Patelli and Pedrini, 2013].[1] Understanding the use of sentiment in accounting narratives is thus of vital importance for improving the efficiency of financial markets.

The recent literature on tone inflation focuses mainly on an aggregate analysis, studying how

[1]The central importance of financial disclosures to the efficiency of securities markets is frequently mentioned in speeches given by Securities and Exchange Commission (SEC) commissioners. For instance, "Audited financial statements provide the foundation for our securities markets. Audited financial statements allow investors to make decisions on whether to buy, hold, or sell a particular security" [SEC, 2002a]. "Accurate information also improves the quality of markets by allowing markets to discover the true price at which specific securities trade" [SEC, 2002b].

an equally-weighted average of intratextual sentiment is influenced by managerial incentives and how such biases affect the information signal in the corporate disclosure. This paper innovates by investigating in detail the management of the narrative structure in accounting narratives as a subtle form of impression management technique used by managers to influence investors' perception of a firm's future performance. Based on the serial position effect introduced by Ebbinghaus [1885], we hypothesize that managers set the intratextual frequency of sentiment within their financial disclosures in a way that increases the likelihood of leaving a positive perception on the reader. The serial position effect argues that the order in which the information is disclosed is a central factor influencing the sentiment perceived by the reader. That is, the reader tends to recall the first and last items of a series best, and the middle items worst. For this reason, we argue that managers concentrate most of the discussion of positive news at the beginning and end of the disclosure, leading to a U-shape form (also referred to as a smile) of positive sentiment. Likewise, we expect managers to concentrate most of the negative information at the beginning of the letter, leading to a left-sided half-smile (also called a smirk).

We examine the intratextual dynamics of sentiment within CEO letters to shareholders of the DJIA constituents between 2000 and 2011.[2] CEO letters are widely used accounting narratives and considered important in the investment decisions of private and institutional investors [Abrahamson and Amir, 1996; Kohut and Segars, 1992; Patelli and Pedrini, 2013]. In fact, CEO letters are unaudited and, unlike disclosures to the SEC, the message in these letters can, to a substantial extent, be shaped as the CEO sees fit. This gives management opportunities to select, discuss

[2] Compared to other studies on textual sentiment, our sample stands out in terms of its time series length. This is needed to ensure a separation between the sample used for the estimation of the position weighted sentiment and the performance measure to predict, as well as guaranteeing a long enough panel for accurate estimation of the panel regressions. In order to make the homogeneity assumption of the slope parameters in the panel regression plausible and to keep the data collection feasible, we focused on the DJIA firms.

and explain corporate financial performance largely untroubled by mandatory constraints. CEO letters therefore offer a natural test case for studying the effects of impression management on tone structure, as their textual characteristics reflect the self-serving goals of CEOs rather than the actual performance results being communicated [Clatworthy and Jones, 2006].

Consistent with the hypothesis that the narrative structure is used as a vehicle for impression management, we find strong evidence of a U-shape in the intratextual dynamic of CEOs' positive sentiment, with a significantly larger peak at the end of the letter. The intratextual number of negative words peaks at the start of the text, to fall to an almost constant low level towards the middle and the end of the text. The combination of a smile in positive sentiment and a left-sided half-smile (or smirk) in negative sentiment, together with the overall average of positive words being higher then the number of negative words, leads to a right-sided smirk in the difference of positive and negative sentiment (called net sentiment). The sharply increasing prevalence of positive words towards the end of the letter is obviously no coincidence, but a reflection of careful writing strategies in which CEOs jockey positive words for position, giving them the best exposure within the CEO letter.

We next investigate how intratextual dynamics influence automated approaches for the analysis of sentiment. It is generally expected that the sentiment expressed in the CEO letter is informative of future firm performance. The caveat is that, because of impression management, it may also sketch a sugar-coated view on the firm's future performance. Today's workhorse in estimating sentiment consists of a simple spread between the percentage of words that can be classified as positive and those that can be classified as negative. Statistical theory predicts that, when the intratextual number of positive and negative words is not uniformly distributed, the standard "bag of words" approach to proxy the author's sentiment on future firm performance is likely to be inefficient. In the case of impression management, the equally weighted average of intratextual sentiment may even be upward biased and thus provide a too optimistic view on the future corporate achievements.

We introduce a new generation of sentiment analytics that yields estimates of the factual sentiment in accounting narratives that are more robust to impression management. The proposed method weights the information value of words depending on their position within the text. The weights are optimized to maximize the sentiment measure's predictive power for the firm's return on assets (ROA) over the year following the publication of the CEO letter. To avoid overfitting, the weights are parsimoniously specified as a linear combination of third-order Almon polynomials that are smooth functions of the word's position in the text [Almon, 1965]. Consistent with the hypothesis that sentiment at the beginning and end of a text is overstated, we find that the optimized weights attach a relatively higher information value to the net sentiment in the middle of the text than at the beginning and end of the text. We also find that the position weighted sentiment is on average more pessimistic than the equally weighted sentiment measure, which is expected when position weighting corrects for sentiment inflation due to impression management.

We then evaluate the gains of optimizing the intratextual sentiment weights for forecasting future performance of the DJIA constituents. We find a significant increase in the R-square of the prediction model relative to the classical approach used in the prior literature. This result indicates that the structure of the sentiment within CEO letters provides a signal to investors concerning future performance and that an intratextual analysis is required to accurately measure

CEO sentiment within the CEO letter.

This study makes several contributions. First, together with Allee and DeAngelis [2015], this paper is the first to study the intratextual dynamics of sentiment of financial disclosures and to show the existence of impression management through the structure of sentiment within accounting narratives.[3] While prior literature reports the existence of impression management based on the manipulation of the tone level [see e.g., Arslan-Ayaydin et al., 2015], our paper uncovers a more subtle form of impression management, where managers recurrently structure sentiment within their accounting narratives in such a way that it positively influences investors' expectations.

Second, we develop a more efficient sentiment aggregation method to predict future firm performance, as compared to the usual spread of positive and negative words used in prior literature. Our approach is in line with Jegadeesh and Wu [2013], who suggest to estimate the information value of each word by regressing the market impact of the disclosure on dummy variables indicating the use of a particular word in the textual communication. Instead of exploiting the market impact of particular words, we introduce the dynamics of sentiment and the importance of the position of words within narrative disclosures.

Third, our study improves upon our understanding of how managers conceive and shape their qualitative financial disclosures. Although prior research has increasingly shown that the qualitative information in accounting narratives is complementary to the information in quantitative disclosures for predicting a firm's future performance [Arslan-Ayaydin et al., 2015; Davis et al.,

[3] Allee and DeAngelis [2015] summarize the complexity of narrative structure by means of the intratextual dispersion of sentiment and find that managers tend to concentrate the discussion of bad news in a text, while good news is more spread out.

2012; Davis and Tama-Sweet, 2012], few focus on how sentiment should be accurately measured. We add to this research by showing that the traditional equally-weighted measure of sentiment defined thus-far in the literature is likely to be an inefficient sentiment aggregation method for predicting future performance and that users of accounting narratives can be better off by utilizing position weighted sentiment measures that exploit the recurrent patterns in the underlying dynamics of sentiment.

We proceed as follows. Section 2 first sets our motivation and develops our hypotheses. Section 3 describes our sample as well as the word libraries used. Section 4 shows the presence of a common pattern in CEO sentiment dynamics within annual letters. Section 5 explains our new weighted measure of sentiment. Section 6 contains the main analysis of the forecast performance of the sentiment measure for future firm performance. Section 7 presents the conclusions and sketches directions for further research.

2 Literature Review and Hypothesis Development

Our main hypothesis is that the narrative structure in CEO letters is used for impression management and that, as a consequence, a weighted measure of CEO sentiment in which the weights are defined as a function of the position of the (positive or negative) word in the text is more accurate in predicting future performance than its equally-weighted counterpart. To test this hypothesis, we proceed in two steps. First, we investigate the intratextual distribution of net sentiment in CEO letters, and show that this distribution is far from uniform and that it has an explainable periodic shape. Second, based on these stylized facts, we investigate whether allocating weights to words as a function of their position in their text increases the prediction accuracy of future firm performance relative to the equally-weighted metrics used in prior literature.

In this section, we first motivate our choice for the analysis of impression management in the context of CEO letters. We then review the most important results from the narratology and computational linguistic literature in terms of the potential impact of impression management on the strategic positioning of positive and negative words within the CEO letter. Finally, we present our hypotheses on how position weighting improves sentiment aggregation with respect to the traditional "bag-of-words" approach.

2.1 Hypotheses on impression management in CEO letters

Because of information asymmetries between the firm management and the firm stakeholders, users of financial information have to rely their evaluation of management effort and future performance, at least partly, on reports that are prepared by managers themselves. Early research on the qualitative information of accounting narratives mainly interprets sentiment as an unbiased signal of a manager's private information about future corporate performance and generally ignores the managerial incentives to manage investors' expectations about the firm's future performance [see e.g. Davis et al., 2012; Henry, 2008]. It is only recently that increasing evidence shows that managers can intentionally affect the optimistic language in their accounting narratives through impression management techniques [see e.g. Arslan-Ayaydin et al., 2015; Huang et al., 2014].

Hooghiemstra [2000] describes impression management as "a field of study within social psychology studying how individuals present themselves to others to be perceived favourably by others." In a corporate reporting context, impression management is regarded as attempts to control and manipulate the impression conveyed to users of accounting information [Clatworthy and Jones, 2001]. Managers can distort the expectations of third parties by selecting only positive information to discuss in their communications, by choosing which quantitative information to highlight, distorting graphical presentation of data or withholding bad news information. For instance, Clatworthy and Jones [2001] find that profitable companies are more inclined to discuss their results and acquisitions and disposals, while unprofitable companies include more discussion of board changes. Other forms of impression management in corporate disclosures include the transitivity structure (active and passive verb choice). For instance, Thomas [1997] finds that in the CEO letter of a company, active voices are associated with success, while passive voices distance writers from the message. She also finds that the use of the pronoun "we" declines with profitability. Similarly, Sydserff and Weetman [2002] argue that the use of passive constructions gives the text a veneer of objectivity or neutrality, and can be used by writers as a linguistic mechanism to disassociate themselves from the text.

More recently, Huang et al. [2014] provide evidence that managers manipulate investors' perceptions to hype a stock before important events. They find that sentiment in earnings press releases is, on average, more positive when firms are issuing new equity or undertaking mergers and acquisitions, and more negative when granting stock options. Similarly, Davis and Tama-Sweet [2012] argue that managers act strategically in choosing the narrative outlets to describe firm performance. Schleicher and Walker [2010] study the sentiment in the outlook section of the annual reports of UK firms and find evidence that firms with an impending performance decline tend to bias sentiment in the outlook section upwards. Finally, Arslan-Ayaydin et al. [2015] show that equity-based incentives induce managers to inflate the sentiment of earnings press releases to increase the value of their stock and option portfolios.

CEO letters are ideal for our analysis on the strategic positioning of positive and negative words by CEOs. The main reason is that CEOs have a significant freedom to choose the content and the layout of the information reported. The auditor's role remains limited to verifying that the information in it is consistent with the numbers presented in the financial statements [see, e.g., Clatworthy and Jones, 2003]. This stands in contrast with the MD&A section of the firm's annual 10-K filling which is heavily influenced by corporate lawyers. This lack of control provides the management with an excellent opportunity to manage outsiders' impressions on the company without regulatory repercussions.

Notwithstanding the greater opportunities for impression management, the information value of CEO letters for predicting future performance is generally recognized [Abrahamson and Amir, 1996; Patelli and Pedrini, 2013]. CEOs tend to include in their letters the (non-financial) explanations and interpretations, which cannot be included in the audited financial statements [Abrahamson and Amir, 1996]. Their importance as a complementary means of communicating with shareholders could explain why the length of CEO letters has substantially increased over the last 20 years. From an average of 1,230 words per letter between 1987 and 1988 [Abrahamson and Amir, 1996], the average number of words in CEO letters in our sample shows an increase to approximately 1,900 as of 2012.

Our baseline hypothesis is that CEO letters are subject to impression management. Because impression management is inherently unobservable, we cannot test this hypothesis directly. In the next subsections, we build on this hypothesis to formulate testable predictions on the shape of the intratextual distribution of positive and negative sentiment in a CEO letter. We also present hypotheses on how the intratextual dynamics of sentiment can be exploited in the aggregation of intratextual sentiment into a single sentiment estimate for the overall CEO letter.

2.2 Hypotheses on the effect of impression management on the positioning of sentiment in CEO letters

The value of the position of a word within a text has been thoroughly investigated in the narratology and computational linguistics literature. We first review two generally accepted theories (the serial position effect and peak-end-rule theory) and then discuss how they apply to the effects of impression management on the narrative structure of CEO letters.

According to the serial position effect, readers recall information better when it is presented first (primacy) or last (recency) in a vector of words, rather than in the middle [Baddeley and Hitch, 1977; Glanzer and Cunitz, 1966; Roediger and Crowder, 1976]. Some studies have examined this issue in prose. One finds that recall of propositions in the text was higher for the first propositions, followed by the last propositions and finally the middle propositions in two of eight passages [Freebody and Anderson, 1986], while another study finds only primacy effects [Frase, 1969]. Furthermore, Deese and Kaufman [1957] find both primacy and recency effects and Meyer and McConkie [1973] and Kieras [1980] find that information was recalled better if it appeared early in the text and that this information was more important than other information in the text. This evidence thus suggests that position and information value interact in some way.

This pattern in readers' recall is usually referred to as the U-shaped free-recall curve and is consistent with the position method defined by Edmundson [1969] in computational linguistics. Edmundson [1969] develops automated text summarization techniques to aid readers in accessing information at a faster pace and defines a weight-based method that computes the weight of each sentence based on certain features, such as cue phrase, keyword (i.e., term-frequency-based), title and location. He evaluates each of the criteria by comparison against manually created extracts. He finds that the combination of cue phrases/title/location dominates word frequency measures in the creation of better extracts, with keywords alone being the worst performing algorithms and location being the best individual feature. This research suggests that the relationship between the position of a word in the text and its information value should be considered to optimally measure sentiment in corporate disclosures.

The peak-end-rule theory developed in Varey and Kahneman [1992] predicts that the peak and final event of an experience influences the evaluation more than all other events in the experience, which contradicts a simple hedonic calculus in which years of please and pain are summed or averaged. Experiences that end very well or with a large positive moment are rated as more pleasurable than longer, more moderately pleasant experiences despite the total happiness experienced ostensibly being greater in the longer case [Diener et al., 2001; Do et al., 2008; Fredrickson and Kahneman, 1993]. As a consequence, following the peak-end rule theory, investors reading two sentiment-neutral CEO letters (both with the same number of positive and negative words) have a more positive (negative) assessment of the firm's future performance depending on whether positive words are at the end (beginning) and negative words at the beginning (end) of the letter and whether a large positive (negative) peak occurred in the letter.

In order to inflate the perceived sentiment, we therefore expect the firm management to release CEO letters that are logically organized discourses in which the most salient elements of the text are discussed at the beginning and end of the text, while the more neutral elements are discussed in the middle. For the positive impression to dominate, we expect a higher incidence of positive words than negative words and that the number of words classified as positive will be higher at the beginning and at the end. Therefore, following the serial position effect and for a given total number of positive words, we expect CEOs to be disproportionately more positive at the beginning and end of the letter than in the middle, where the firm's operations and developments are discussed. This leads to our first testable hypothesis:

H1a: Textual positive sentiment within CEO letters to shareholders is U-shaped on average, with a peak in positive sentiment at the end of the text.

Because the end of the letter is recalled best, we expect the end of the letter to contain a larger number of positive words than the beginning. The U-shape of CEOs' positive sentiment within their letters can also be understood in the context of the peak-end-rule, which predicts that framing financial performance in positive terms with a peak at the end will cause investors to think about the results in terms of increases relative to the reference point (average) [Kahneman et al., 1993].

The pattern of the intratextual frequency of CEOs' negative sentiment is more difficult to predict. Based on the peak-end-rule theory and the U-shaped free-recall curve, one may expect an inverse U-shape (a "frown") in negative sentiment. In practice, the objective of avoiding negative sentiment in the key positions of a text has to be balanced off with the constraint of providing a realistic view on the firm's achievement. In fact, since the Sarbanes-Oxley Act of 2002 and the establishment of the Public Company Accounting Oversight Board in the USA (and similar bodies in other countries), CEOs have to be more conscious of the words they choose in discharging their accountability to stakeholders. In the wake of recent accounting and corporate governance scandals, audit committees, regulatory authorities and others involved in the oversight of CEOs are now more alert to their obligations implicit in the narratives signed by CEOs, especially concerning past or prospective negative events.

We therefore expect the use of negative words to be a trade-off for the CEO. On the one hand, it is important for the CEO letter to be in agreement with prior knowledge to assist the reader's comprehension [Pearson et al., 1979]. Readers will stop reading the CEO letter if it is unrealistic. Because our sample includes economically volatile periods with two important crises (the dot-com bubble in 2001 and the great recession in 2007-2009), realistic CEO letters cannot avoid the use of negative words. On the other hand, the CEO wants to maximize the firm value and communicate positively to investors. We expect that the CEO optimally achieves these objectives by placing the majority of the negative words at the beginning of the text. Because the introduction is thereby realistic, the CEO will avoid losing the reader, as the text is in agreement with his understanding of the economic situation. Because of the recency theory and peak-end rule, investors will remember these negative words less after having read the entire text. The concentration of negative sentiment at the beginning of the text, as opposed to the hypothesized repetition of positive sentiment words at the beginning and end of the text, is also consistent with the recent results in Allee and DeAngelis [2015] showing that the discussion of negative news is more concentrated in a text than the discussion of positive news. In visual terms, the U-shape in positive words can be seen as a smile, while the shape of negative sentiment is only a left-sided half-smile, to which we refer henceforth using the term left-sided smirk.

H1b: Textual negative sentiment within CEO letters to shareholders is characterized by a *left-sided smirk on average.*

As a result and consistent with the peak-end rule, we expect CEOs' net sentiment, measured as the spread between positive and negative words, to show a right-sided smirk.

H1c: Textual net sentiment within CEO letters to shareholders is characterized by a right-sided smirk on average.

2.3 Hypotheses on the value of position weighting in the aggregation of intratextual sentiment

We now turn to the question of the information value of the textual sentiment expressed in CEO letters for predicting the one-year-ahead firm performance, as measured by the firm's return on assets (ROA) over the year following the publication of the CEO letter. Linguistic communication is a potentially important source of information about firms' fundamental values. Because very few stock market investors directly observe firms' production activities, they get most of their information secondhand. Their main sources are analysts' predictions, quantifiable publicly disclosed accounting variables, and accounting narratives of firms' current and future profit-generating activities, such as the CEO letter to shareholders. Abrahamson and Amir [1996] and Patelli and Pedrini [2013] evidence the effect of sentiment of CEO letters on the perception of investors about its future performance. Although they find evidence of sugar-coating in CEO letters to shareholders, they show that future return on asset increases with the sentiment of financial disclosures.[4] This evidence supports the fact that the qualitative information contained

[4]Similarly, Henry [2008], Davis et al. [2012], Demers and Vega [2010] and Price et al. [2012], among others, conclude that the sentiment of earnings press releases is significantly positively correlated with future firm per-

in earnings releases provides a signal regarding managers' future earnings expectations to the market that is incremental to quantitative information. It remains however an open question on how to capture best that signal in a data-driven manner.

The automated analysis of sentiment requires to aggregate the numbers of positive and negative words into a manageable metric for further analysis. Typically, the total textual sentiment is measured as the spread in the proportion of positive and negative words in the document [Davis et al., 2012; Davis and Tama-Sweet, 2012; Demers and Vega, 2010; Huang et al., 2014; Patelli and Pedrini, 2013] or simply as the proportion of negative words [Abrahamson and Amir, 1996; Tetlock et al., 2008]. However, these approaches implicitly assume that all words in the negative (resp. positive) word list are equally negative (resp. positive). Jegadeesh and Wu [2013] propose to estimate the information value of each word in the list by regressing the market impact on dummy variables indicating the use of a particular word in the textual communication. Given the large number of possible words, this so-called "word power approach" is only feasible for the analysis of a high-dimensional set of communications. In addition, when the focus is on forecasting, typically rolling window estimations are used, reducing further the number of degrees of freedom in the data to estimate the word power. For similar reasons, the approach based on setting weights inversely proportional to the frequency of documents in which the word is used (see e.g. Loughran and McDonald 2011) is likely to lead to noisy weights.[5]

One of the major consequences of the non-uniform distribution of the intratextual number

formance and short window contemporaneous returns around the date that the disclosures are made even after controlling for a firm's financial information and earnings surprises.

[5]Compared to the word power and inverse document frequency approaches, the position weighting approach that we develop in this paper is fundamentally different in terms of the type of information it exploits (the position in the text). It is also more parsimonious by modelling the weight of a word as a semiparametric function of the position in the text (instead of having as many degrees of freedom as the number of words in the corpus analyzed).

of positive and negative words is that total sentiment measures that aggregate the intratextual sentiment without considering the position in the text may be suboptimal. As we show in the Appendix, one notable exception is when the observed sentiment is a noisy (but unbiased) proxy of the true underlying sentiment and the noise satisfies the condition of being independently and identically normally distributed with zero mean. Whenever impression management leads to managing the narrative structure, these assumptions will be violated and the equally-weighted sentiment measure will tend to be biased. In particular, whenever the beginning and end of the letter are dominated by impression management and overconfidence biases, it implies that these parts of the text contain less information value and should be underweighted when measuring sentiment.

H2a: The information content of textual sentiment estimates to predict firm performance can be improved by underweighting the intratextual sentiment of the words located at textual positions that are systematically inflated by impression management.

In the position weighted sentiment measure, the intratextual weights need to sum to unity. This implies that underweighting the intratextual sentiment of the words located at textual positions that are systematically inflated by impression management will lead to position weighted sentiment measures that are on average more pessimistic than the approaches based on equal weighting of intratextual sentiment.

H2b: The position weighted measure of sentiment is on average more pessimistic than the equal weighted sentiment measure.

Ultimately, the position weighted sentiment measure can be used to forecast future performance. The estimation of intratextual weights is of course only useful if sentiment is not uniformly distributed within a text. When some parts of the text are systematically more informative than others, then it may be more efficient to measure sentiment as a weighted average of the intratextual net sentiment such that the words with a higher (lower) information value are overweighted (resp. underweighted). This leads us to formulate the following hypothesis:

H2c: When the intratextual sentiment is not uniformly distributed, a weighted measure of CEO sentiment with weights that are a function of the position of a word in a text is more informative of future firm performance than their equally-weighted counterparts.

3 Data Collection and Financial Dictionaries

In the next sections, we analyze the intratextual dynamics of sentiment in CEO letters over the period 2000 and 2011. This section first describes our data set of CEO letters and then introduces the libraries used to extract the sentiment from the observed words.

3.1 Collection of DJIA CEO letters

We hand-collect the CEO letters of the firms included in the Dow Jones Industrial Average Index (DJIA) for the twelve consecutive fiscal years 2000 to 2011. We choose the DJIA constituents for reasons of importance and tractability. The DJIA encompasses 30 of the largest firms in the United States and is considered a leading indicator of the stock market. We obtain the letters from each firm's respective website. If the annual report is not directly available, we contact the firm's public relations department. Because firms typically file their annual reports in the next calendar year, our sample mostly covers fiscal years 2001 to 2012. To avoid double-counting, we only select the text portion and delete any table, graph or figure included in the letter. Each letter is then saved as a text file for compatibility with our content analyzer.[6]

In terms of comparison, Abrahamson and Amir [1996] and Patelli and Pedrini [2013] cover two-year periods between 1987–1988 and 2008–2009, respectively. Although both papers cover a larger cross-section of firms, the 12-year period of this paper stands in clear contrast to the short time-series adopted in their research and covers different market regimes, while the 1987–1988 and 2008–2009 periods adopted by Abrahamson and Amir [1996] and Patelli and Pedrini [2013] correspond to a high-market-uncertainty regime.

Our second hypothesis focuses on the relationship between CEO sentiment and future firm performance, which requires a date on which the letter was made publicly available. Thus, we manually collect each firm's annual report SEC filing date on the Edgar system. The firm is required to have at least one filing date at the SEC over the 2000-2012 period. If we obtain no SEC filing date for some years, we extrapolate the missing date(s) based on the latest date available for that firm.

As described below, our analyses require stock price and accounting data. Market prices and returns data are taken from the Center for Research in Security Prices (CRSP) database, while the COMPUSTAT database is our source for accounting data. Our final sample consists of 342 CEO letters with a total of 1,002,054 words.

[6] It is the CEO who signs the shareholder letter. Prior literature confirms that shareholder letters reflect the CEO more than other legally vetted communications such as the commonly studied MD&A section of form 10-K and regulatory filings [Abrahamson and Amir, 1996; Amernic et al., 2010]. Dombalagian [2014] notes that in financial filings, "narrative disclosures are typically prepared by teams of attorneys who are versed in the relevant disclosure standards as well as the associated civil and criminal standards." Dikolli et al. [2014] indicate that while firms' legal teams are heavily involved in writing sections of the annual report that are regulated by the SEC (such as the MD&A), attorneys "almost never even comment on the shareholder letter."

3.2 Financial dictionaries

Several lists of words, called dictionaries, exist but there is no consensus in the literature regarding which wordlist is more appropriate for the analysis of language in corporate financial disclosures. As argued by Rogers and Van Buskirk [2009], among others, the tone obtained by using a single wordlist should be seen as a (noisy) proxy for the true, but unknown, tone of the text. To avoid the model risk of choosing an inappropriate library, we average over the (standardized) tone obtained by using three established lists of words, namely the positive and negative wordlists defined by Henry [2008], the positive and negative wordlists defined by Loughran and McDonald [2011] and the so-called "optimism-increasing" and "optimism-decreasing" word lists in the DICTION 7.0 software.[7] All three of them are already popular choices in practice. The optimism-in/decreasing wordlists in DICTION were used by Davis et al. [2012] and Davis and Tama-Sweet [2012] to analyze earnings press releases. A limitation of general word lists such as DICTION is that they do not analyze language in the context of financial disclosures.

The building block of a sentiment measure is the qualification of words as positive, negative or neutral. This is usually performed by a content analysis that verifies whether the words belong to a pre-specified list of positive and negative words, called a dictionary. Most of the early research uses general lists of words, such as the Diction software program, that automatically generates a score of a document's optimism. These libraries were built for the study of sociological and psychological text and may not be suitable for the content analysis of corporate disclosures. The

[7] More precisely, the"optimism-increasing" wordlists in DICTION are the lists labeled "Praise", "Satisfaction" and "Inspiration", while the "optimism-decreasing" word list is the union of the words in the word lists "Blame", "Hardship" and "Denial". As in Davis et al. [2012], the tone measure $Tone^{DIC}_{j,q,t}$ is then defined as the difference in the percentage of words in the press release that are "optimism increasing" and the percentage of words in the press release that are "optimism decreasing", all relatively to the total number of words in the press release.

current trend in text analysis research is to refer to domain-specific dictionaries. For the analysis of CEO letters, this implies the use of specialized financial dictionaries, such as those developed by Loughran and McDonald [2011] and Abrahamson and Amir [1996].[8]

We will use three different libraries of words, each of which has been used to study firms' financial disclosures. The first is obtained from Loughran and McDonald [2011], who provide finance-oriented (positive and negative) lists of words. [9] The second library of positive and negative words, that we use, consists of the so-called "optimism-increasing" and "optimism-decreasing" word lists in the DICTION 7.0 software.[10] Finally, the library of Abrahamson and Amir [1996] only consists of a list of negative words, which was specifically designed for the study of CEO letters. The Abrahamson and Amir [1996] library is available in their paper.

We report in Table 1 the number of words found in our sample of CEO letters between 2000 and 2011 for each library. We find that 40,034 words out of the total 1,002,054 words in our sample can be matched with words in the Loughran and McDonald [2011] library, 47,429 with words in Diction and 2,835 with words in the Abrahamson and Amir [1996] library. A comparison between the words found by Diction and Loughran and McDonald [2011] demonstrates the broader scope of the Diction library, especially for negative words. The top five words in Diction include 'not', 'needs', 'no' and 'hard', those of the Loughran and McDonald [2011] contain more financially oriented words such as 'crisis', 'critical', 'challenges'.

[8]Although domain-specific libraries are progressively being defined, a consensus has yet to be reached as to which library to use. This explains why research usually refers to various multiple lists of words to evidence the robustness of their results.

[9]The lists are publicly available on the authors' website: http://www3.nd.edu/~mcdonald/Word_Lists.html.

[10]More precisely, the"optimism-increasing" wordlists in DICTION are the lists labeled "Praise", "Satisfaction" and "Inspiration", while the "optimism-decreasing" word list is the union of the words in the word lists "Blame", "Hardship" and "Denial".

[Insert Table 1 here.]

We expect sentiment estimates using finance-oriented dictionaries to be more powerful in predicting future performance than generic dictionaries, such as the Diction library. This is consistent with prior studies suggesting that generic linguistic algorithms may yield noisy measures of "positive" and "negative" linguistic sentiment in the context of financially oriented text passages. For instance, Loughran and McDonald [2011] show that each discipline has its own dialect in which words take on specific meanings in specific contexts that may not translate effectively in other disciplines. For these reasons, we use in the following sections the word lists provided by Loughran and McDonald [2011] as our main library and report the results of Diction and Abrahamson and Amir [1996] for comparison purposes.[11]

4 Intratextual Dynamics of CEO Sentiment

In this section, we first introduce the definition of the intratextual sentiment proxies. We then provide strong empirical evidence in favour of the hypotheses H1a, b and c on the shape of the intratextual dynamics of CEO sentiment within the letters to shareholders.

4.1 Notation

The length of each text is standardized to correspond to the $[0, 1]$ interval, which we divide in B bins such that each bin contains the same number of total words. In the remainder of the

[11] The choice of Loughran and McDonald [2011] as our main library is further substantiated in Subsection 6.3 where we find that the sentiment measured obtained using the word lists of Loughran and McDonald [2011] predict better future firm performance than the alternative word lists from Diction and Abrahamson and Amir [1996].

paper we use $B = 20$ bins.[12] For each bin, we then compute the percentage number of positive (resp. negative) words out of the total number of words in each bin. As such, the positive CEO sentiment in bin b ($b = 1, \dots, B$) for firm j for the CEO letter of fiscal year t (expressed in percentage points) is

$$PosSent_{b,j,t} = 100 \cdot \frac{PW_{b,j,t}}{TW_{b,j,t}}, \tag{4.1}$$

where $PW_{b,j,t}$ and $TW_{b,j,t}$ are the number of positive words and the total number of words for firm j in bin B for fiscal year t, respectively. Similarly, the negative CEO sentiment for bin b for firm j for the CEO letter of fiscal year t is given by

$$NegSent_{b,j,t} = 100 \cdot \frac{NW_{b,j,t}}{TW_{b,j,t}}, \tag{4.2}$$

where $NW_{b,j,t}$ is the number of negative words for firm j in bin b for fiscal year t.

Finally, for each bin, we also compute the difference between the positive and negative sentiment, and call this the net sentiment of that bin:

$$NetSent_{b,j,t} = PosSent_{b,j,t} - NegSent_{b,j,t}. \tag{4.3}$$

Our interest is in the dynamics of textual sentiment within the bin and how these dynamics affect the estimation of total sentiment. The traditional aggregation of the B estimates of

[12] Results are qualitatively similar for 10 bins or when considering random bin selections as in Allee and DeAngelis [2015].

intratextual sentiment is based on simple averaging:

$$NetSent_{j,t}^{EW} = \frac{1}{B} \sum_{b=1}^{B} NetSent_{b,j,t}. \tag{4.4}$$

The subscript EW refers to the fact that the sentiment of each bin is equally weighted. The definition of $PosSent_{j,t}^{EW}$ and $NegSent_{j,t}^{EW}$ is analogous.

To summarize the intratextual dynamics into univariate statistics, we recommend to use the intratextual slope, curvature and Herfindahl Index statistics. The intratextual slope shows the direction of the intratextual sentiment evolution. An increasing value of sentiment corresponds to a positive slope, and vice versa for a negative slope. We implement the slope statistic as the average spread between the sentiment of the last two bins minus the sentiment of the first two bins. As such, we obtain the net sentiment slope statistic as:

$$Slope_{j,t}^{net} = \frac{1}{2}\Big((NetSent_{B-1,j,t} + NetSent_{B,j,t}) - (NetSent_{1,j,t} + NetSent_{2,j,t}) \Big), \tag{4.5}$$

and similarly for $Slope_{j,t}^{pos}$ and $Slope_{j,t}^{neg}$.

The slope statistic can be seen as the average change of sentiment in a text. As mentioned above, this change is not expected to be constant. For positive sentiment, e.g., we expect a U-shape implying a convex-shaped curvature of intratextual sentiment. We measure the curvature of sentiment as the spread between the average net sentiment of the first and last two bins and the average sentiment of the three most central bins. This curvature statistic will be positive in case of a U-shaped sentiment. The net sentiment curvature statistic is defined as follows:

$$Curvature^{net}_{j,t} = \frac{1}{4}\Big(NetSent_{1,j,t} + NetSent_{2,j,t} + NetSent_{B-1,j,t} + NetSent_{B,j,t}\Big)$$
$$-\frac{1}{3}\Big(NetSent_{\lfloor B/2\rfloor-1,j,t} + NetSent_{\lfloor B/2\rfloor,j,t} + NetSent_{\lfloor B/2\rfloor+1,j,t}\Big), \qquad (4.6)$$

where $\lfloor\cdot\rfloor$ denotes the floor operator. $Curvature^{pos}_{j,t}$ and $Curvature^{neg}_{j,t}$ are similarly defined.

Finally, in order to verify the result of Allee and DeAngelis [2015] that managers tend to concentrate the discussion of bad news in a text, while good news is more spread out, we compute the Herfindahl index of intratextual sentiment. For net sentiment, the Herfindahl Index is given by:

$$HI^{net}_{j,t} = \sum_{b=1}^{B}\left(\frac{NetSent_{b,j,t}}{\sum_{a=1}^{B} NetSent_{a,j,t}}\right)^2. \qquad (4.7)$$

The definition of the Herfindahl Index for positive and negative sentiment ($HI^{pos}_{j,t}$ and $HI^{neg}_{j,t}$) is analogous.[13] The Herfindahl Index ranges from $1/B$ (maximum dispersion) to one (maximum concentration), where B is the number of intratextual bins. Based on Allee and DeAngelis [2015], we hypothesize that $HI^{neg}_{j,t}$ exceeds the $HI^{pos}_{j,t}$.

4.2 Findings

We discuss in this section the results on the slope, curvature and Herfindahl Index statistics presented in Table 2 and the intratextual positive, negative and net sentiment frequency plots shown in Figures 1-2.

[13]The calculation of the Herfindahl Index requires the sentiment measures for each bin to be positive. For $PosSent_{b,j,t}$ and $NegSent_{b,j,t}$, this is by definition the case. For our sample, $NetSent_{b,j,t}$ is always positive and thus does not require any truncation.

[Insert Table 2 here.]

Consider first of all the aggregated sentiment measures in Panel A of Table 2. We see that, the net sentiment estimated from both the Loughran and McDonald [2011] and DICTION word lists, is on average positive. The average expressed positive sentiment by CEOs in their letter is as expected and in line with previous results in the literature. It confirms the general consensus that CEOs tend to be optimistic, overconfident and have an intrinsic interest of portraying a positive image about their firm (see e.g. Arslan-Ayaydin et al. 2015; Heaton 2002; Malmendier and Tate 2005).

Panel D reports the Herfindahl Index for positive, negative and net sentiment. The results confirm the hypothesis of Allee and DeAngelis [2015] that firm managers tend to concentrate the discussion of negative news and spread more the discussion of positive news, since we find that the Herfindahl Index of negative news is more than double the Herfindahl Index of positive news.

The novelty of our research is regarding the narrative structure of where sentiment is positioned, which, as we show next, is primarily in the beginning of the text for negative sentiment, and U-shaped with a peak at the end of the text for positive sentiment. We hypothesize that the intratextual analysis of CEO sentiment reveals a new and subtler form of impression management that occurs within the CEO letter: CEOs first refer to negative past events and, whenever they introduce bad events, they swamp them with many positive words. They then progressively talk about the future in positive terms while clearly reducing the negative tone within their letter. Such a swamping strategy is compatible with the serial position effect. Because investors will recall the end of the text best (recency effect), CEOs will increase the number of positive words towards the last bins. Similarly, the first bins are expected to report an above-average positive sentiment, as these are the bins that are recalled more frequently than the middle items (the primacy effect). This interpretation of impression management by means of strategic positioning of positive and negative sentiment in a text is consistent with prior research on CEO behavior, which shows that CEOs tend to conceal bad news by not reporting it to the same extent as good news [see, e.g., Clatworthy and Jones, 2003].[14]

These predictions are confirmed by both the summary slope and curvature statistics in Table 2, and the more detailed plots in Figures 1-2 showing the average $PosSent_{b,j,t}$, $NegSent_{b,j,t}$ and $NetSent_{b,j,t}$, over all CEO letters in our sample, as function of the bin b, with $b = 1 \ldots, 20$.

As can be seen from Panel B, positive and net intratextual sentiment are upward sloping, while the negative slope for negative sentiment indicates a concentration of negative sentiment at the beginning of a text. The curvature statistics confirm that sentiment is especially expressed at the beginning and the end of a text, while the middle of the text is more neutral. These results are consistent with the hypotheses H1a, H1b and H1c predicting a U-shaped pattern in sentiment, with a peak at the end of the text for positive and net sentiment, and a peak at the beginning of the text for negative sentiment. We interpret this as *prima facie* evidence that CEOs carefully position positive and negative words in their CEO letter in order to transmit a positive sentiment to the stakeholder reading the letter and thereby influencing expectation of future firm performance compared to the rational prediction based on the objective firm sentiment and

[14]CEOs are consistently more optimistic than pessimistic, even during financial crises: only 8 out of 342 (2%) CEO letters have a negative value for their net sentiment between 2000 and 2012, among which five are for JPMorgan between 2007 and 2011.

reported performance numbers.

The predicted intratextual patterns of sentiment also appear clearly in Figure 1 showing the $PosSent_{b,j,t}$, $NegSent_{b,j,t}$ and $NetSent_{b,j,t}$, as function of the bin b (with $b = 1\ldots,20$), as an average over all CEO letters in our sample and estimated using the word lists of Loughran and McDonald [2011]. Consistent with Hypothesis 1a, we find in Figure 1a a U-shaped frequency plot for positive tone, where the average tone decreases from 3.4% in the first bin to approximately 3% between bins 4 and 18. The positive tone then increases again to 4% in the last two bins. In contrast with the smile in positive sentiment, the negative sentiment frequency plot in the top right of Figure 1 shows a left-sided half-smile. The left-sided smirk in negative words starts with an average value of 1.4% in bin 1. Then, the average number of negative words declines sharply to 0.8% and becomes relatively flat for the last two thirds of the letter. A single-sided t-test shows that the negative tone at the end of the letter is significantly lower than that at beginning at a 99% confidence level. This result is consistent with Hypothesis 1b.

Similar patterns are found for the sentiment frequency plots based on the positive and negative word lists of Diction and the negative word list of Abrahamson and Amir [1996], which are shown in Figure 2. One exception is in terms of the negative words based on Diction, where the frequency is more U-shaped than a left-sided smirk. We believe this is due to the general nature of the Diction word list (see e.g. the top 10 negative words according to Diction, as reported in Table 1), which makes it less suitable for the analysis of the sentiment expressed by CEOs in their letter to shareholder. All other plots confirm the smile in intratextual positive sentiment, the left-sided smirk in the intratextual frequency of negative sentiment and the right-sided smirk in net sentiment. These plots are consistent with the peak-end-rule and the recency effect, thus confirming Hypotheses H1c.

[Insert Figure 1 here.]

The bottomline of the analysis is that there is a strong commonality in the intratextual dynamics in the positive, negative and net sentiment expressed in CEO letters and that they are consistent with the research hypotheses H1a, H1b and H1c. CEO letters are strategically crafted corporate discourses in which positive and negative words are distributed throughout the text in such a way that readers are left with a positive impression about the firm. In the next Section, we test whether we can improve the prediction of future firm performance, by weighting sentiment in function of the position of words within the text. If words at the beginning and end of the letter manage investors' expectations, they should contain on average less incremental value and, therefore, be underweighted in predicting future performance.

5 A Position-Weighted Measure of Sentiment

The possible existence of impression management in terms of strategic positioning of positive and negative words within a text raises questions on the validity of measuring textual sentiment as a simple average of the sentiment expressed in a text. The next question that we address is how to aggregate the intratextual net sentiment measures $NetSent_{b,j,t}$ into a single overall sentiment per text that has predictive power for future firm performance. In this section, we first outline our estimation methodology and then discuss the resulting differences between the standard equally-weighted measure of textual sentiment versus the position-weighted sentiment estimates obtained for our panel of CEO letters over the period 2000-2011.

5.1 Methodology

As mentioned in our hypotheses and literature review, the workhorse aggregation technique in the literature has hitherto been to use simple averaging. The underlying assumptions are twofold. First of all, it specifies that the total sentiment measure of the CEO letter j in year t is given by the linear mapping of the B intratextual sentiment measures $NetSent_{b,j,t}$ on $NetSent_{j,t}$ with weights $w = (w_1, \ldots, w_B)'$:

$$NetSent_{j,t}(w) = \sum_{b=1}^{B} w_b NetSent_{b,j,t}, \tag{5.1}$$

and where all weights sum to unity. Secondly, it assumes equal importance of each part of the text. Since the bins have the same text length, this implies an equal weighting, i.e. w is set to

$$w^{\text{EW}} = (1/B, 1/B, \ldots, 1/B)'. \tag{5.2}$$

Specification of position weighted sentiment Under the approach of position weighting, we use a flexible parametric structure that maps the intratextual position b to its position weight w_b, based on a parameter vector θ:

$$w_b = f_\theta(b). \tag{5.3}$$

The formal definition of the function $f_\theta(b)$ is given in Appendix. Important features are that it is linear in θ and uses Almon polynomials to describe in a flexible and smooth way the potentially complex intratextual dynamics with a small number of regressors. To simplify notation, we use henceforth $NetSent^{\text{EW}}_{j,t}$ to the equally-weighted measure of net sentiment and $NetSent^{\text{PW}}_{j,t}$ is the position-weighted measure of net sentiment.

It is important to note that the position-weighted sentiment measure is only a potential improvement for the CEO letters for which there are substantial intratextual dynamics. If there are no dynamics, e.g. when $NetSent_{b,j,t} = NetSent^{\text{EW}}_{j,t}$, for all b, any combination of weights will lead to (almost) the same sentiment measure. The previous analysis has shown that there are *on average* significant and intuitively appealing dynamics in the DJIA firms. In order to distinguish the CEO letters for which it is potentially relevant to model the intratextual weights from those for which the equal-weighting approach is expected to work best (see in particular the case described in Appendix), we will follow the approach based on observed state variables and least squares estimation of the threshold. In particular, the state variable, that we use as a signal of possible dynamics in the intratextual net sentiment of the CEO letter of firm j in year t is the standard deviation of the intratextual net sentiment ($SD^{net}_{j,t}$), which is given by:

$$SD^{net}_{j,t} = \left[\frac{1}{B-1} \sum_{b=1}^{B} \left(NetSent_{b,j,t} - NetSent^{EW}_{j,t} \right)^2 \right]^{\frac{1}{2}}. \tag{5.4}$$

When the $SD^{net}_{j,t}$ is below a threshold parameter κ, the equally-weighted sentiment measure will be used. Otherwise, the alternative position-weighted approach is to be used. The optimal value of κ is of course application-specific and will be determined by least squares estimation.[15] For each CEO letter and for a given value of κ, we can thus define the indicator for high intratextual

[15] The least squares estimator of the threshold parameter κ is known to be super-fast convergent (see e.g. Chan [1993] and Hansen [2000] for the asymptotic distribution, and e.g. Boudt et al. [2015], for a recent application to the modeling of time-varying parameters).

volatility in net sentiment,

$$\mathbf{1}_{j,t}^{net}(\kappa) = \begin{cases} 1 \text{ if } SD_{j,t}^{net} > \kappa, \\ 0 \text{ otherwise.} \end{cases} \tag{5.5}$$

Based on these assumptions, the two unknowns needed to define the optimized weights are the threshold κ determining the use of equal–weights versus optimized intratextual weights, and the value of the parameter vector θ. These values will be application–specific and we will next describe a data–based procedure to determine the weights' particular functional form.

Panel threshold least squares estimation of intratextual weights The optimized weights are application-specific. Our focus is to use the textual sentiment for predicting future firm performance. Based on Engelberg [2008] and Davis et al. [2012], we proxy future firm performance using the firm's return on assets (ROA) over the year following the publication of the CEO letter.[16] Because of the yearly frequency of our data, there are insufficient observations to do the estimations for each firm separately and we therefore resort to panel estimation in order to estimate the weights on the intratextual sentiment.

[16]To avoid any look-ahead bias, we start measuring ROA in the quarter following the quarter in which the annual report has been filed at the SEC. Specifically, firm future performance $ROA_{j,t+1}$ is measured as the sum of quarterly earnings before extraordinary items $Y_{j,q+i,t+1}$ (Compustat data item #18) over the four quarters after the SEC filing quarter q, scaled by total assets (#6) at the end of quarter q.

This leads us to specify the following fixed effects regression model:

$$ROA_{j,t+1} = \alpha_j + \beta_1 \cdot NetSent^{\text{EW}}_{j,t} \cdot \left(1 - \mathbf{1}^{net}_{j,t}(\kappa)\right) \tag{5.6}$$

$$+\beta_2 \cdot NetSent^{\text{PW}}_{j,t} \cdot \mathbf{1}^{net}_{j,t}(\kappa) + \epsilon_{j,t+1},$$

where $\mathbf{1}^{net}_{j,t}(\kappa)$ is a dummy variable that equals one if the intratextual dispersion in sentiment is higher than a threshold value of κ and is zero otherwise. $NetSent^{\text{EW}}_{j,t}$ refers to the equally-weighted measure of net sentiment and $NetSent^{\text{PW}}_{j,t}$ is the position-weighted measure of net sentiment.

In the panel, we distinguish between texts for which there is a sufficiently high intratextual dispersion in sentiment ($SD^{net}_{j,t} > \kappa$) from those for which the intratextual sentiment dynamics are less important. We estimate the parameters (α_j, β_1, β_2, θ, κ) by non-linear least squares. This can be performed in a computationally convenient way by a loop over κ and by noting that, for a given value of κ, the model is linear in the parameters α, β_1 and the parameter product $\tilde{\theta} = \beta_2 \cdot \theta$.[17] Because of the bound constraint that $\theta \geq 0$, $\tilde{\theta}$ is also bound constrained, but the problem of least squares estimation is still simple to solve by reformulating the estimation

[17] The linearity in $\tilde{\theta} = \beta_2 \cdot \theta$ becomes obvious by using (5.1) and (8.2) to rewrite the regressor in (5.6) as:

$$\beta_2 \cdot NetSent^{\text{PW}}_{j,t} = \beta_2 \sum_{b=1}^{B} w_b^{\text{Almon}}(\theta) NetSent_{b,j,t}$$

$$= (\beta_2\theta_1) \sum_{b=1}^{B} NetSent_{b,j,t} + \sum_{c=1}^{3} (\beta_2\theta_{1+c}) \sum_{b=1}^{B} P_c(b/B) NetSent_{b,j,t} + (\beta_2\theta_{4+c}) P_c((B-b)/B) NetSent_{b,j,t}.$$

problem conditional on κ as a bound constrained quadratic optimization problem that can be easily solved numerically.[18] The constraint that all intratextual weights need to sum up to unity in Equation (8.3) implies that the estimated θ is the normalized version of the estimated $\tilde{\theta}$.

Estimation window We consider two types of estimation of the parameter θ determining the weights. First, for the descriptive analysis of optimized weights in Subsection 5.2, we will consider the full sample of CEO letters over the period 2000-2011. Second, for the economic validation of the forecast performance of the position weighted sentiment measure versus the equally weighted sentiment measure in Section 6, we will use rolling estimation samples of three years such that the estimated weights are independent of the firm performance that is predicted using the position weighted sentiment measure.

5.2 Equally- and Position-weighted sentiment in CEO letters

Let us now investigate the resulting optimized weights for our 2000-2011 sample of CEO letters using the Loughran and McDonald [2011] library to decode the intratextual sentiment. The first step is to estimate the threshold value κ that distinguishes the CEO letters for which the equally–weighted measure is used versus those for which the weights are heterogeneous and depend on the position in the text. As shown in Figure 4a, where we report the sum of squared residuals as a function of the threshold parameter κ for the complete 2000-2011 sample, the least squares estimate for κ is 1.9%, resulting in 81.579% of the letters for which position–specific weights are to be used.

[18] Given a value of κ, the least squares regression estimator minimizes a sum of squared residuals that can be rewritten as $(y-Xb)'(y-Xb)$, with y the vector of future ROA, X the matrix of explanatory variables and b the parameter estimates. This is equivalent to minimizing $-2b'X'y + b'X'Xb$, which is a quadratic objective function of b.

[Insert Figure 4 here.]

The corresponding pattern of estimated weights is reported in Figure 4b. It is immediate to see that the optimized weights are bell-shaped. In line with the expectation formulated in the hypothesis section, the position-weighted measure attaches lower weights to the beginning and end of the text compared to the middle parts of the text that are overweighted. In Figure 4b, the optimized weights are around 1.9% for the words used in the beginning of the text, then rise to their maximum of 7% in the middle of the text and fall to almost 0% at the end of the text. This is consistent with Hypothesis 2 and shows that the sentiment at the beginning and end of the letter are partly biased by impression management motives and therefore should be underweighted when predicting future firm performance.

Such a strong intratextual variation in the weights does not necessarily imply a large difference between the position– and equally–weighted measures. Indeed, if the measure of sentiment were the same for each bin, then the weighting has no impact on the sentiment measure. Figure 5 investigates the impact of the weighting on the estimated sentiment. It shows the scatter plot of the position–weighted versus the equally–weighted sentiment measure for our panel of DJIA CEO letters over the period 2003–2011. We see that there is a strong agreement across the measures, but that differences exist. To inspect these differences, consider as the reference line, the 45° line corresponding to equality of the two approaches. We find that 14% of the observations are on the line (perfect agreement), 54% below the line (the position-weighted measure is less than the equally-weighted measure and thus more pessimistic) and 32% above the line. The outcome that position-weighted leads on average to a less optimistic view on sentiment is as expected, since, as explained in the hypothesis section, managers have a tendency to be overly optimistic and convey this optimism by overloading the beginning and end of the text with positive words.

[Insert Figure 5 here.]

In the next section, we evaluate the forecasting performance of the position-weighted sentiment measure. To avoid look ahead bias, we use, at each point in time, the sample of the three most recent years, in order to estimate the optimal intratextual weights. Denote these estimates based on years $t-2$, $t-1$ and t as $\hat{\theta}_t$ and $\hat{\kappa}_t$. The estimated threshold varies between 1.068 % and 2.003% of the sample with an average value of 1.572%, corresponding to 124 firms. Figure 6 plots the kappa values for each rolling sample, as well as the percentage of firms with an intratextual standard deviation of sentiment that is higher than kappa. We see that over time, the threshold κ increases gradually and the percentage of firms for which position weighting is applied tends to reduce. This is weak evidence in favor of the hypothesis that the qualitative information in CEO letters has become more reliable and that the percentage number of firms engaging in impression management by strategically positioning sentiment in their CEO letter decreases.

[Insert Figure 6 here.]

6 CEO Sentiment and Future Firm Performance

We now turn to the question of the economic relevance of using the position-weighted sentiment measure rather than the equally weighted sentiment measure for predicting future firm performance. We do this analysis for the out-of-sample period 2003-2011, where, as described in the previous section, the weights underlying the position-weighted sentiment measure are estimated on the three preceding years.

In Subsection 6.1, we first do the rat race of comparing models that predict future ROA, using only the information in the intratextual net sentiment. Then in Subsection 6.2 we control for other influences that may have an impact on ROA.

6.1 Comparison of pure sentiment-based prediction models for future ROA

The benchmark model is the traditional approach consisting of a linear prediction model in which the equally-weighted measure of sentiment is used to forecast future ROA:

$$ROA_{j,t+1} \quad = \quad \varrho_j + \delta_{t+1} + \beta \cdot NetSent^{\mathrm{EW}}_{j,t} + \epsilon_{j,t+1}, \tag{6.1}$$

and where ϱ_j and δ_{t+1} correspond to the industry and year fixed effects. Under this "EW model" approach, we thus regress future $ROA_{j,t+1}$ on $NetSent^{\mathrm{EW}}_{j,t}$, as in Abrahamson and Amir [1996] and Patelli and Pedrini [2013].[19]

Our leading hypothesis is that, for CEO letters with marked intratextual sentiment dynamics,

[19] Our primary interest is in predicting future performance. We are not interested in the behavioral interpretation of the coefficient of the impact of sentiment on future performance, which would require to deal with the endogeneity of the sentiment variable and either require instrumental variable type of estimation or, at least mitigate the endogeneity issue by taking the performance and sentiment variables in first differences, as recommended by Li [2010] and Kravet and Muslu [2013].

a more accurate forecast can be obtained by using the proposed position-weighted sentiment measure through the "PW model":

$$
\begin{aligned}
ROA_{j,t+1} \quad &= \quad \varrho_j + \delta_{t+1} + \beta_1 \cdot NetSent^{\text{EW}}_{j,t} \cdot \big(1 - \mathbf{1}^{net}_{j,t}(\kappa)\big) \qquad (6.2) \\
&\quad + \beta_2 \cdot NetSent^{\text{PW}}_{j,t} \cdot \mathbf{1}^{net}_{j,t}(\kappa) + \epsilon_{j,t+1}.
\end{aligned}
$$

We will compare the two test regression based on their goodness of fit as measured by the adjusted (within) R^2, but also through an F-test comparing their fit with the one of the generalized unrestricted model (GUM):

$$
\begin{aligned}
ROA_{j,t+1} \quad &= \quad \varrho_j + \delta_{t+1} + \beta \cdot NetSent^{\text{EW}}_{j,t} + \beta_1 \cdot NetSent^{\text{EW}}_{j,t} \cdot \big(1 - \mathbf{1}^{net}_{j,t}(\kappa)\big) \\
&\quad + \beta_2 \cdot NetSent^{\text{PW}}_{j,t} \cdot \mathbf{1}^{net}_{j,t}(\kappa) + \epsilon_{j,t+1}. \qquad (6.3)
\end{aligned}
$$

Note that the GUM in (6.3) nests the EW model in (6.1) and the PW model in (6.2) as special cases and is therefore also a useful reference model to test for the significance of the model simplifications using F-tests by omitting the different types of sentiment measures.

The results of these regressions are reported in Panel A (without firm and year fixed effects) and Panel B (with fixed effects) of Table 4. We test for the significance of the coefficients using standard errors clustered by firm and year. The first result shown in Panel A of Table 4 is that the sentiment in CEO letters contains information to predict future firm performance. This can be seen through the positive and significant coefficient of sentiment in the EW model and its large explanatory power in predicting future firm performance (as measured by the $Adj.\ within\ R^2$, comparing the fit of the proposed model with the fit of the model including only the firm and year fixed effects). This result implies that, despite their strategic approach to communicate with shareholders, managers use language in their annual letters to communicate relevant information about the firm's future performance.

[Insert Table 4 here.]

The second finding in Table 4 is that the proposed position-weighted measure has a significantly higher power to predict future firm performance than the traditional position-weighted measure. The $Adj.\ within\ R^2$ increases from 14.357% (EW Model) to 18.891% (PW Model) once the sentiment measure considers the position of a word in the document. The increase in $Adj.R^2$ is statistically significant at a 99% confidence level. From the F-tests comparing the GUM with its restricted versions, we observe that, ignoring the position of a word in a document, decreases significantly the fit of the model, while omitting the equally–weighted sentiment variable has no significant effect on the fit of the model.

Consistent with the presence of impression management in terms of strategic positioning of sentiment within a text, our main conclusion is that the structure of the sentiment within CEO letters provides a signal to investors concerning future performance. However, the presence of intratextual analysis requires to adjust the measure of sentiment within the CEO letter that was used so far in the literature. Indeed, the total aggregated textual sentiment are better predictors of future performance, if an appropriate weighting is assigned based on the position of words within a text. This result indicates that the structure of the sentiment within CEO letters provides a signal to investors concerning future performance and that an intratextual analysis is required to accurately measure CEO sentiment within the CEO letter

6.2 Multivariate prediction model: Controlling for hard information

In this subsection, we expand the number of regressors in the EW Model (Equation (6.1)) and the PW Model (Equation (6.2)) in order to test whether the explanatory power of CEOs' net sentiment in predicting future firm performance survives after controlling for "hard information". We define hard information as quantitative information easily processed from annual reports and select three sets of variables that have been shown in prior literature to predict future firm performance. The first set of variable relates to the firm's past profitability:

- *Return on assets* – Return on assets ($ROA_{j,t}$) is measured as the earnings before extraordinary items at the end of fiscal year t, scaled by the total assets at the beginning of the year. The $ROA_{j,t}$ coefficient is predicted to be positive and lower than one, consistent with prior research documenting mean reversion in performance metrics [Barber and Lyon, 1997].

- *Past stock returns* – We define $Ret_{j,t}$ as the firm's past stock returns between the end of fiscal year $t-1$ and the filing of the annual report of fiscal year t.[20]. Based on Fama and French [2006], we expect current stock market performance to be positively related to future firm performance.

- *Size* – Firm size $MC_{j,t}$ is measured as the natural logarithm of market value of equity (Compustat item #25 · #199) at the end of the fiscal year. We expect smaller firms to be less

[20]The annual report for fiscal year t is usually filed at the SEC in fiscal year $t+1$.

profitable [Fama and French, 1995].

The second set of factors that we consider proxy for the risks that firms face:

- *Book-to-market* – Firms with smaller book-to-market ratios $BTM_{j,t}$ are growth firms that are valued more for their growth opportunities and, hence, are likely to be more profitable. Book-to-market is defined as the book value of equity (#6-#18), divided by $MC_{j,t}$.
- *Volatility* – Based on Core et al. [1999], we also introduce $\sigma_{ROA,j,t}$ to capture firm risk, which is defined as the standard deviation of $ROA_{j,t}$ over the preceding five years.

There is also evidence that dividends and accruals forecast profitability [Fairfield et al., 2003a,b; Fama and French, 2001; Sloan, 1996]. We include the ratio of dividends to book equity ($D_{j,t}$), as Fama and French [2001] show that dividend-paying firms tend to be more profitable. Initiated by Sloan [1996], we also consider accruals, as they have been shown to be negatively related to future profitability: investors overestimate the future earnings of firms with high accruals in current earnings and underestimate the future earnings of firms with low current accruals. As in Sloan [1996], we compute the accrual as the one-year change in current assets excluding cash minus change in current liabilities excluding long-term debt in current liabilities and taxes payables minus depreciation, standardized by the book value of equity. We distinguish between positive accruals ($+AC_{j,t}$) and negative accruals ($-AC_{j,t}$).

Table 3 reports the summary statistics for these variables and their Pearson correlation with the next period's ROA. Thes sign of the correlations are as predicted.

[Insert Table 3 here.]

The estimates are reported in Panel B of Table 4. Consistent with Hypothesis 2, the main evidence of this section is that our conclusions persist after we control for the earnings-predicting variables defined by Fama and French [2006]. Indeed, the coefficient of $NetSent_{j,t}^{\text{PW}} \cdot \mathbf{1}_{j,t}^{net}(\kappa)$ in the multivariate PW Model is positive and significant at a 95% confidence level, which indicates that there is forward-looking information in the sentiment of CEO letters that is incremental to more quantitative financial and accounting information. From the F-tests comparing the GUM with its restricted versions, we see that, ignoring the position–weighting of a word in a text, decreases significantly the fit of the model, while omitting the equally–weighted sentiment variable has no significant effect on the fit of the model. The F-test that compares the $Adj.\ within$ R^2 of the EW Model and the GUM Model has a value of 6.371, which is significant at a 99% confidence level.

The bottom line of the regression results in Panel A and B of Table 4 is that, when there is intratextual dispersion in sentiment, a weighted measure of CEO sentiment with weights that are a function of the position of a word in the text is more informative to predict future firm performance than the equally-weighted metrics used in prior literature. This result holds after we control for hard, financial information.

6.3 Sensitivity of forecast performance to the choice of library and sentiment type

The analysis on the forecast performance of position-weighted sentiment for firm performance is clearly conditional on the utilization of an appropriate method for classifying the sentiment of the words in the CEO letter. Throughout the paper, our main results use the word lists proposed by Loughran and McDonald [2011] to capture the sentiment in terms of firm performance based on financial corporate disclosures. For the sake of comparison, we also report additional results for the general purpose positive and negative word lists embedded in Diction 7.0 and the list of negative words proposed by Abrahamson and Amir [1996].

We already showed in Section 4 that, for all three word lists (Loughran and McDonald [2011]; Diction 7.0; Abrahamson and Amir [1996]), there is an agreement on the general shape of the average intratextual distribution of sentiment for the sample of DJIA firms over the period 2000–2011: a smile in positive sentiment, a left-sided smirk in negative sentiment and a right-sided smile in positive sentiment. In Table 5, we investigate the impact of the choice of library on the accuracy of predicting firm performance using the equally-weighted and position-weighted sentiment measures. We also explore the sole use of the positive and negative libraries to forecast firm performance.

There are three main take-away points from Table 5. The first important result is that the domain-specific libraries, i.e. the ones of Loughran and McDonald [2011] and Abrahamson and Amir [1996] , always lead to substantially better forecasts of future firm performance. Secondly, for the domain-specific libraries, the forecast performance is always improved by using position-weighted sentiment rather than equally-weighted sentiment. Thirdly, the highest accuracy in predicting future performance is obtained using our baseline approach: the position-weighted net sentiment using the word lists of Loughran and McDonald [2011] in combination with the control variables, which yields an $Adj.\ within\ R^2$ of 60.818%.

[Insert Table 5 here.]

7 Conclusions

Investors routinely use summary statistics to avoid being overwhelmed by the massive amount of information available to them. The linguistic sentiment of a corporate disclosure, such as the CEO letter to shareholders, is one such statistic that recently has become popular. The sentiment of letters to shareholders is related to both current and future firm profitability, which is consistent with the notion that the linguistic sentiment of a CEO letter conveys the manager's private information about the expected performance of the firm. Since the letters to shareholders are unaudited, an important caveat is that CEOs may have incentives to engage in impression management and that the CEO letter to shareholders is an ideal outlet to do so. It is thus likely that the signal in the managerial sentiment expressed in CEO letters is biased when managers engage in impression management to influence investors' expectations about the firm's future prospects.

Based on DJIA firms between 2000 and 2011, this paper is the first to uncover the presence of a more subtle form of impression management within financial disclosures. We find the presence of a U-shape (resp. decreasing smirk) in the intratextual frequency of positive (resp. negative) sentiment in corporate disclosures. The intratextual net sentiment is characterized by an increasing smirk. This result is consistent with impression management in CEO letters, since according to the serial position effect, readers recall information better when it is presented first (primacy) or last (recency) in a vector of words, rather than in the middle.

CEOs thus jockey positive words for position, giving them the best exposure within the CEO letter. CEO letters are thus documents crafted in such a way that they give a positive impression to the reader. The undeniable intratextual patterns are relevant not only as evidence of impression management, but also as the data feature predicting the inefficiency and potential bias in the classical summary statistics for linguistic sentiment, which assign equal weights to the intratextual sentiment. We test this second hypothesis by first proposing a methodology for weighted sentiment measurement and then applying it to forecasting future firm performance for the panel of 2000-2011 DJIA CEO letters.

The proposed position-sentiment measurement framework consists of replacing the equal weight design in linguistic sentiment with a flexible weighting scheme that is optimized to predict future firm performance. A convenient threshold least squares estimator is proposed to optimize the weights linking the intratextual sentiment to future firm performance. The threshold is needed to distinguish the letters with negligible sentiment dynamics from the majority of the texts showing significant intratextual dynamics in sentiment. The optimization is done on rolling estimation samples to avoid look-ahead bias.

When modeling the weights as a function of the position in the text, we find that for our sample of 342 CEO letters the optimized sentiment measure significantly outperforms the standard equally-weighted measure in terms of explaining future firm performance. This result is robust to the inclusion of all types of control variables.

We have applied our framework and based our conclusions on the sample of CEO letters of the Dow Jones Industrial Average constituents over the period of 2000 to 2011. An important direction for future research is to test our two hypotheses and apply the proposed optimized sentiment measurement framework on other types of corporate communication tools, such as earnings press releases or forward-looking statements in corporate filings.

References

Abrahamson, E. and Amir, E. (1996). The information content of the president's letter to shareholders. *Journal of Business Finance & Accounting*, 23:1157–1182.

Allee, K. D. and DeAngelis, M. (2015). The structure of voluntary disclosure from narratives: Evidence from tone dispersion. *Journal of Accounting Research*, 53:241–274.

Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica*, 33:178–196.

Amernic, J., Craig, R., and Tourish, D. (2010). *Measuring and Assessing Tone at the Top Using Annual Report CEO Letters*. Institute of Chartered Accountants of Scotland, Edinburgh.

Andersen, T., Bollerslev, T., and Cai, J. (2000). Intraday and interday volatility in the japanese stock market. *Journal of International Financial Markets, Institutions and Money*, 10:107–130.

Arslan-Ayaydin, Ö., Boudt, K., and Thewissen, J. (2015). Managers set the tone: Equity incentives and the tone of earnings press releases. *Journal of Banking and Finance*, page forthcoming.

Baddeley, A. and Hitch, G. (1977). *Recency reexamined*. In S. Dornic (Eds.) Attention and performance vol. VI, pp. 647–667. Hillsdale, NJ: Lawrence Erlbaum and Associates Inc.

Barber, B. and Lyon, J. (1997). Detecting long-run abnormal stock reutnrs: The empirical power and specification of test statistics. *Journal of Financial Economics*, 43:341–371.

Boudt, K., Croux, C., and Laurent, S. (2011). Robust estimation of intraweek periodicity in volatility and jump detection. *Journal of Empirical Finance*, 18:353–369.

Boudt, K., Liu, F., and Sercu, P. (2015). Exporters' exposures to currencies: Beyond the loglinear model. *Review of Finance*, page forthcoming.

Chan, K.-S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *The Annals of Statistics*, 21:520–533.

Clatworthy, M. and Jones, M. (2001). The effect of thematic structure on the variability of annual report readability. *Accounting, Auditing and Accountability Journal*, 14:311–326.

Clatworthy, M. and Jones, M. (2003). Financial reporting of good news and bad news: Evidence from accounting narratives. *Accounting and Business Research*, 33:171–185.

Clatworthy, M. and Jones, M. (2006). Differential reporting patterns of textual characteristics and company performance in the chairman's statement. *Accounting and Business Research*, 19:493–511.

Core, J., Holthausen, R., and Larcker, D. (1999). Corporate governance, chief executive officer compensation and firm performance. *Journal of Financial Economics*, 51:371–406.

Davis, A., Piger, J., and Sedor, L. (2012). Beyond the Numbers: Measuring the Information Content of Earnings Press Release Language. *Contemporary Accounting Research*, 29:845–868.

Davis, A. and Tama-Sweet, I. (2012). Managers' use of language across alternative disclosure outlets: Earnings press releases versus MD&A. *Contemporary Accounting Research*, 29:804–837.

Deese, J. and Kaufman, R. (1957). Serial effects in recall of unorganized and sequentially organized verbal material. *Journal of experimental psychology*, 54:180–187.

Demers, E. and Vega, C. (2010). Soft information in earnings announcements: News or noise? *Working paper, INSEAD*.

Diener, E., Wirtz, D., and Oishi, S. (2001). End effects of rated life quality: The James Dean effect. *Psychological Science*, 12:124–128.

Dikolli, S., Keusch, T., Mayew, W., and Steffen, T. (2014). Using shareholder letters to measure ceo integrity. *Working paper, Duke University*.

Do, A., Rupert, A., and Wolford, G. (2008). Evaluations of pleasurable experiences: The peak-end rule. *Psychonomic Bulletin & Review*, 15:96–98.

Dombalagian, O. (2014). *Chasing The Tape: Information Law and Policy in U.S. Capital Markets*. MIT Press.

Ebbinghaus, H. (1885). *Über das gedächtnis: untersuchungen zur experimentellen psychologie*. Duncker & Humblot.

Edmundson, H. (1969). New methods in automatic extracting. *Journal of the ACM*, 16:264–285.

Engelberg, J. (2008). Costly Information Processing: Evidence from Earnings Announcements. *Working paper, Northwestern University*.

Fairfield, P., Whisenant, S., and Yohn, T. (2003a). Accrued earnings and growth: Implications for future profitability and market mispricing. *Accounting Review*, 78:353–371.

Fairfield, P., Whisenant, S., and Yohn, T. (2003b). The differential persistence of accruals and cash flows for future operating income versus future profitability. *Review of Accounting Studies*, 8:221–243.

Fama, E. and French, K. (1995). Size and book-to-market factors in earnings and returns. *Journal of Finance*, 50:131–155.

Fama, E. and French, K. (2001). Disappearing dividends: Changing firm characteristics or lower propensity to pay. *Journal of Financial Economics*, 60:3–43.

Fama, E. and French, K. (2006). Profitability, investment and average returns. *Journal of Financial Economics*, 82:491–518.

Frase, L. (1969). Paragraph organization of written materials: The influence of conceptual clustering upon the level and organization of recall. *Journal of Educational Psychology*, 60:394–401.

Fredrickson, B. and Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology*, 65:45–55.

Freebody, P. and Anderson, R. (1986). Serial position and rated importance in the recall of text. *Discourse Processes*, 9:31–36.

Glanzer, M. and Cunitz, A. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning And Verbal Behavior*, 5:351–360.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica*, pages 575–603.

Heaton, J. (2002). Managerial optimism and corporate finance. *Managerial Optimism and Corporate Finance*, 31:33–45.

Henry, E. (2008). Are investors influenced by how earnings press releases are written? *Journal of Business Communications*, 45:363–407.

Hooghiemstra, R. (2000). *Corporate communication and impression management - New perspectives why companies engage in corporate social reporting*. Springer.

Huang, X., Teoh, S. H., and Zhang, Y. (2014). Tone management. *The Accounting Review*, 89:1083–1113.

Jegadeesh, N. and Wu, D. (2013). Word power: A new approach for content analysis. *Journal of Financial Economics*, 110(3):712–729.

Kahneman, D., Fredrickson, D., Schreiber, C., and Redelmeier, D. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science*, 4:401–405.

Kieras, D. (1980). Initial mention as a signal to thematic content in technical passages. *Memory And Cognition*, 8:345–353.

Kohut, G. and Segars, A. (1992). The president's letter to stockholders: An examination of corporate communication strategy. *Journal of Business Communication*, 29:7–21.

Kravet, T. and Muslu, V. (2013). Textual risk disclosures and investors' risk perceptions. *Review of Accounting Studies*, 18:1088–1122.

Li, F. (2010). The Information content of forward looking statements in corporate filings – A naive bayesian machine learning approach. *Journal of Accounting Research*, 48:1049–1102.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-ks. *Journal of Finance*, 66:35–65.

Malmendier, U. and Tate, G. (2005). CEO overconfidence and corporate investment. *Journal of Finance*, 60:2661–2700.

Meyer, B. and McConkie, G. (1973). What is recalled after hearing a passage? *Journal of Educational Psychology*, 65:106–117.

Patelli, L. and Pedrini, M. (2013). Is the optimism in CEO's letters to shareholders sincere? Impression management versus communicative action during the economic crisis. *Journal of Business Ethics*, pages 1–16.

Pearson, P., Hansen, J., and Gordon, C. (1979). The effect of background knowledge on young children's comprehension and implicit information. *Journal of Reading Behaviour*, 11:201–209.

Price, S., Doran, J., Peterson, D., and Bliss, B. (2012). Earnings conference calls and stock returns: The incremental informativeness of textual tone. *Journal of Banking & Finance*, 36:992 – 1011.

Roediger, H. and Crowder, R. (1976). A serial position effect in recall of United States presidents. *Bulletin of The Psychonomic Society*, 8:275–278.

Rogers, J. L. and Van Buskirk, A. (2009). Bundled forecasts and selective disclosure of good news. *Working paper, University of Chicago*.

Schleicher, T. and Walker, M. (2010). Bias in the tone of forward-looking narratives. *Accounting and business research*, 40(4):371–390.

SEC (2002a). Speech by SEC Commissioner Hunt Jr Isaac C Remarks before the ABA Committee on Federal Regulation of Securities. Available at http://www.sec.gov/new/speech/spch550.htm.

SEC (2002b). Speech SEC Chairman Harvey L. Pitt. Remarks before the annual meeting of the Bond Market Association. Available at http://www.sec.gov/new/speech/spch553.htm.

Sloan, R. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *Accounting Review*, 71:289–315.

Sydserff, R. and Weetman, P. (2002). Developments in content analysis: a transitivity index and diction scores. *Accounting, Auditing & Accountability Journal*, 15(4):523–545.

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

Thomas, J. (1997). Discourse in the marketplace: The making of meaning in annual reports. *Journal of Business Communication*, 34(1):47–66.

Varey, C. and Kahneman, D. (1992). Experiences extended across time: Evaluation of moments and episodes. *Journal of Behavioral Decision Making*, 5:169–195.

8 Appendix

Optimality of the equally-weighted sentiment measures under the Gauss-Markov assumptions.

In this section, a sufficient set of conditions are derived under which the equally-weighted sentiment measure is optimal in terms of mean squared error. Denote the true net sentiment underlying the CEO letter of firm j in year t as $NetSent^*_{j,t}$. The $NetSent^*_{j,t}$ is a latent variable and thus can only be observed through proxies. Imagine that we can split the text in B parts of equal text length and that, for the letter of firm j in year t, the observed proxy for net sentiment of text part b is given by $NetSent_{b,j,t}$. Since $NetSent_{b,j,t}$ proxies $NetSent^*_{j,t}$, it is intuitive to rewrite the relationship as a linear regression model:

$$NetSent_{b,j,t} = \alpha_b + \beta_b NetSent^*_{j,t} + \varepsilon_{b,j,t}, \tag{8.1}$$

where α_b and β_b are the position-based intercept and slope parameters, and $\varepsilon_{b,j,t}$ is the error term. If, for all b, $\alpha_b = 0$ and $\beta_b = 1$, i.e. all intratextual sentiment measures are unbiased, then the model simplifies to $NetSent_{b,j,t} = NetSent^*_{j,t} + \varepsilon_{b,j,t}$ and the equally–weighted sentiment measure is the ordinary least squares estimate of $NetSent^*_{j,t}$, i.e., the OLS estimate, for fixed t, over $b = 1 \ldots, B$). In case of homoskedastic errors (i.e. all intatextual sentiment measures are not only unbiased but also equally efficient proxies), then $NetSent^{\text{EW}}_{j,t}$ is the best linear unbiased (BLUE) estimator of $NetSent^*_{j,t}$. If we can additionally assume that the $\varepsilon_{b,j,t}$ **are normally distributed**, then the equally–weighting approach is even efficient.

But, when there is heterogeneity in the regression parameter and/or the variance of the error terms, i.e. when some parts of the text are biased and/or systematically more informative than others, then it may be more efficient to measure sentiment as a weighted average of the intratextual net sentiment such that the words with a higher (lower) information value are overweighted (resp. underweighted).

The Almon approach to specifying flexible and smooth intratextual weights

The most straightforward way to specify the functional form of the position-based weights $f_\theta(b)$ in (5.3) is to use dummy variables. The disadvantage is that, because of estimation error, the weights may be erratic and show jumps, whilst a smooth intratextual weight pattern is expected. We therefore recommend to impose smoothness and parsimony on the optimized weights by specifying the weights as a combination of Almon polynomials of the normalized bin index (b/B): the B weights can be represented as a linear combination of first, second and third–order Almon polynomials of the normalized bin index (b/B):

$$w_b^{\text{Almon}}(\theta) \quad = \quad \theta_1 + \sum_{c=1}^{3} \theta_{1+c} P_c(b/B) + \theta_{4+c} P_c((B-b)/B), \tag{8.2}$$

where $P_c(u) = (1 - u^c)u^{3-c}$ is a c^{th}-order Almon polynomial of u. These polynomials are positively valued functions such that the positivity of the weights is guaranteed by requiring $\theta \geq 0$. We further require that all weights add up to unity:

$$\sum_{b=1}^{B} w_b^{\text{Almon}}(\theta) = 1. \tag{8.3}$$

The flexibility of the third-order Almon polynomials is illustrated in Figure 3. The linear combination of those six curves can closely fit almost every smooth periodic pattern and has been used to fit periodic patterns by Andersen et al. [2000] and Boudt et al. [2011], among others.

[Insert Figure 3 here.]

Main tables

Table 1: **Frequency of (top 5) words in DJIA CEO letters between 2000-2011**

	Loughran and McDonald [2011]		Diction 7.0		Abrahamson and Amir [1996]	
Words found	Positive	Negative	Positive	Negative	Positive	Negative
No.	29,726	10,308	37,248	10,181	-	2,835
%	2.967	1.029	3.717	1.016	-	0.283
Top 5 words found	Word	No.	Word	No.	Word	No.
Positive	Strong	1,763	Growth	4,470	-	-
	Leadership	1,080	Strong	1,763	-	-
	Opportunities	1,077	Important	1,116	-	-
	Innovation	1,006	Health	1,090	-	-
	Better	1,006	Leadership	1,080	-	-
Negative	Challenges	536	Not	1953	Difficult	328
	Difficult	328	No	673	Crisis	281
	Critical	324	Needs	628	Tough	235
	Crisis	281	Risk	598	Losses	212
	Challenging	275	Hard	305	Loss	175

Note: This table reports the number and frequency of (positive & negative) words in DJIA CEO letters between 2000 and 2011. The Loughran and McDonald [2011], Diction 7.0 and Abrahamson and Amir [1996] list of words are used to identify the positive and negative words. This table also reports the top 5 words found in the CEO letters and their associated frequency.

Table 2: **Equally weighted average, slope, curvature, and Herfindahl Index statistics of positive, negative and net sentiment. Statistics are averaged across CEO letters, grouped by subperiod indicated in the column header.**

Library		2000-2002	2003-2005	2006-2008	2009-2011	2000-2011
Panel A – Equally-weighted average sentiment						
Pos	LM	3.129	3.267	3.258	3.193	3.212
	Diction	3.818	3.931	3.888	3.903	3.885
Neg	LM	0.950	0.723	0.889	0.846	0.851
	Diction	0.824	0.723	0.815	0.871	0.808
	AA	0.257	0.148	0.248	0.221	0.218
Net	LM	2.179	2.544	2.368	2.347	2.360
	Diction	2.994	3.208	3.073	3.032	3.078
Panel B – Slope of intratextual sentiment						
Pos	LM	0.442	0.116	0.422	0.212	0.296
	Diction	1.458	1.429	1.553	1.382	1.456
Neg	LM	-0.449	-0.106	-0.187	-0.606	-0.339
	Diction	-0.080	-0.045	0.207	-0.313	-0.057
	AA	-0.225	-0.084	-0.095	-0.351	-0.189
Net	LM	0.892	0.222	0.609	0.818	0.636
	Diction	1.538	1.474	1.346	1.695	1.513
Panel C – Curvature of intratextual sentiment						
Pos	LM	0.103	0.539	0.667	0.746	0.515
	Diction	0.862	1.093	1.220	1.756	1.232
Neg	LM	0.430	0.176	0.151	0.445	0.302
	Diction	0.359	0.029	0.016	0.098	0.125
	AA	0.209	0.056	0.095	0.157	0.129
Net	LM	-0.327	0.363	0.517	0.301	0.214
	Diction	0.503	1.064	1.204	1.658	1.107
Panel D – Herfindahl Index						
Pos	LM	0.074	0.072	0.073	0.070	0.073
	Diction	0.071	0.072	0.072	0.069	0.071
Neg	LM	0.162	0.249	0.175	0.194	0.195
	Diction	0.175	0.218	0.201	0.196	0.198
	AA	0.395	0.419	0.375	0.403	0.398
Net	LM	0.271	0.260	0.317	0.209	0.264
	Diction	0.131	0.095	0.128	0.166	0.130

Note: This table reports the average sentiment by year, the average curvature, the average slope for positive, negative and net sentiment measures for CEO letters to shareholders between 2000 and 2011. Sentiment is measured as the spread between the percentage of positive and negative words in the letter. Curvature of sentiment is measured as the spread between the average net sentiment of the first and last two bins and the average sentiment of the three most central bins. This curvature statistic is positive in case of a U-shaped sentiment. The slope statistic is measured as the average spread between the sentiment of the last two bins minus the sentiment of the first two bins. An increasing value of sentiment corresponds to a positive slope, and vice versa for a negative slope. The Herfindahl Index is defined as as the sum of the squares of the net, positive or negative sentiment by bin for a CEO letter.

Table 3: **Summary statistics of control variables for predicting future firm performance**

Variable	Mean	Median	Std.	Q1	Q3	Correlation with $ROA_{j,t+1}$
$ROA_{j,t}$ (in %)	8.543	7.895	6.030	3.560	12.977	0.871
$\sigma_{ROA,j,t}$	0.022	0.018	0.018	0.009	0.028	0.455
$+AC_{j,t}$	0.048	0.000	0.151	0.000	0.028	-0.097
$-AC_{j,t}$	-0.069	-0.004	0.246	-0.036	0.000	0.144
$BTM_{j,t}$	-1.155	-1.187	0.602	-1.556	-0.777	-0.594
$MC_{j,t}$	11.464	11.564	0.773	10.912	12.048	0.254
$D_{j,t}$	0.091	0.076	0.079	0.049	0.116	0.336
$Ret_{j,t}$	1.051	1.063	0.243	0.930	1.184	0.072

Note: This table reports the average, standard deviation, the 1^{st} and rd quartile for the control variables. The last column reports the Spearman correlation factor between future firm performance and the different variables. Return on assets ($ROA_{j,t}$) is measured as the earnings before extraordinary items at the end of fiscal year t, scaled by the total assets at the beginning of the year. $Ret_{j,t}$ is defined as the firm's past stock returns between the end of fiscal year $t-1$ and the filing of the annual report of fiscal year t. Firm size $MC_{j,t}$ is measured as the natural logarithm of market value of equity at the end of the fiscal year. Book-to-market ($BTM_{j,t}$) is defined as the book value of equity, divided by $MC_{j,t}$. $\sigma_{ROA,j,t}$ is defined as the standard deviation of $ROA_{j,t}$ over the preceding five years. $D_{j,t}$ is defined as the ratio of dividends to book equity. $AC_{j,t}$ are the accruals, defined as the one-year change in current assets excluding cash minus change in current liabilities excluding long-term debt in current liabilities and taxes payables minus depreciation, standardized by the book value of equity. We distinguish between positive accruals ($+AC_{j,t}$) and negative accruals ($-AC_{j,t}$).

Table 4: **Equally-weighted versus position-weighted CEO sentiment and future firm performance**

	Panel A: Univariate models			*Panel B: Multivariate models*		
	EW model	PW model	GUM model	EW model	PW model	GUM model
Sentiment measures						
$NetSent^{\text{EW}}_{j,t}$	$\mathbf{2.032}^{***}$		0.265	$\mathbf{0.476}^{*}$		-0.524
	(0.309)		(1.130)	(0.256)		(0.658)
$NetSent^{\text{PW}}_{j,t} \cdot \mathbf{1}^{net}_{j,t}(\kappa)$		$\mathbf{2.039}^{***}$	1.789		$\mathbf{0.569}^{**}$	$\mathbf{1.062}^{*}$
		(0.281)	(1.101)		(0.247)	(0.636)
$NetSent^{\text{EW}}_{j,t} \cdot \left(1 - \mathbf{1}^{net}_{j,t}(\kappa)\right)$		0.542	0.300		-0.451	0.028
		(0.431)	(1.133)		(0.480)	(0.726)
Control variables						
$ROA_{j,t}$				$\mathbf{0.571}^{***}$	$\mathbf{0.560}^{***}$	$\mathbf{0.561}^{***}$
				(0.107)	(0.099)	(0.099)
$Ret_{j,t}$				$\mathbf{0.022}^{*}$	$\mathbf{0.024}^{*}$	$\mathbf{0.025}^{**}$
				(0.013)	(0.012)	(0.013)
$BTM_{j,t}$				$\mathbf{-0.020}^{**}$	$\mathbf{-0.019}^{**}$	$\mathbf{-0.020}^{**}$
				(0.009)	(0.008)	(0.008)
$\sigma_{ROA,j,t}$				0.230	0.160	0.163
				(0.146)	(0.147)	(0.147)
$MC_{j,t}$				0.005	0.005	0.004
				(0.003)	(0.003)	(0.003)
$+AC_{j,t}$				-0.017	-0.018	-0.018
				(0.012)	(0.013)	(0.013)
$-AC_{j,t}$				-0.003	-0.001	-0.002
				(0.009)	(0.010)	(0.010)
$D_{j,t}$				0.032	0.030	0.030
				(0.031)	(0.025)	(0.025)
Goodness of fit statistics – F-test of equal fit between GUM and its restrictions (EW model, PW model)						
Within R^2	14.733	19.603	19.620	60.408	62.537	62.604
Adj. within R^2	14.357	18.891	18.549	58.781	60.818	60.709
RSS	0.526	0.496	0.496	0.244	0.231	0.231
F-test EW/PW vs. GUM	6.840	0.0498	-	6.371	0.391	-
$pvalue$ EW/PW vs GUM	0.001	0.8237	-	0.002	0.533	-

Note: This table presents the estimation results for the EW, PW and GUM models with industry and year fixed effects. Panel A and Panel B report the results for the EW (Equation (6.1)), PW (Equation (6.2)) and GUM (Equation (6.3)) models, where Panel B includes the control variables defined in Subsection 6.2. The equally- and position-weighted measures of CEO sentiment are defined by Equation (5.2) and Equation (8.2), respectively. The word lists used to estimate sentiment is from the Loughran and McDonald [2011] library. The within R^2 compares the fit of the model with the fit obtained using only the firm and year fixed effects. The significance of coefficients is tested using standard errors clustered by firm and year. *, **, and *** denote statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively, based on a two-sided t-test.

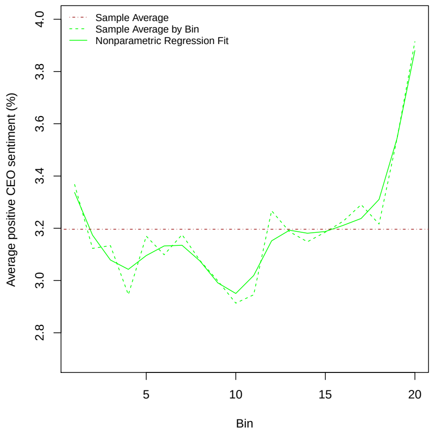
Table 5: **Sensitivity of forecast performance (measured by the $Adj.\ within\ R^2$ (in %) of the ROA forecasting regression) to the choice of library and the type of sentiment**

Sentiment:	Net		Positive		Negative	
Hard Information:	No	Yes	No	Yes	No	Yes
Dictionary						
Panel A – Adj. Within R^2 of the equally-weighted sentiment measure						
Loughran and McDonald [2011]	14.357	58.781	14.716	59.748	3.892	58.167
Diction 7.0	1.980	58.106	0.516	58.137	2.396	58.116
Abrahamson and Amir [1996]					6.232	58.096
Panel B – Adj. Within R^2 of the position-weighted sentiment measure						
Loughran and McDonald [2011]	18.891	60.818	16.222	60.115	4.923	58.174
Diction 7.0	2.688	58.844	0.982	57.957	1.768	57.984
Abrahamson and Amir [1996]					7.755	58.232

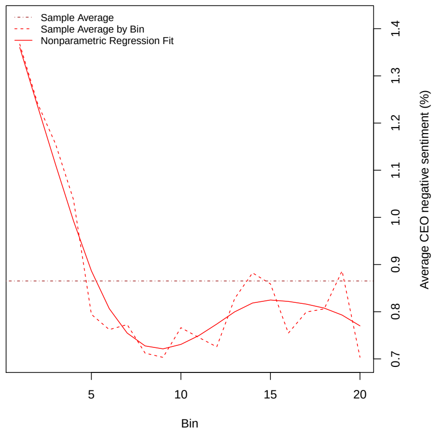
Note: This table reports the robustness of our results to the choice of library, by comparing the $Adj.R^2$ of the PW model (Equation (6.2)) using the Diction 7.0, Loughran and McDonald [2011] and Abrahamson and Amir [1996] word lists. It also tests the inclusion of hard financial information. Positive, negative and net sentiment measures are also distinguished.

Main figures

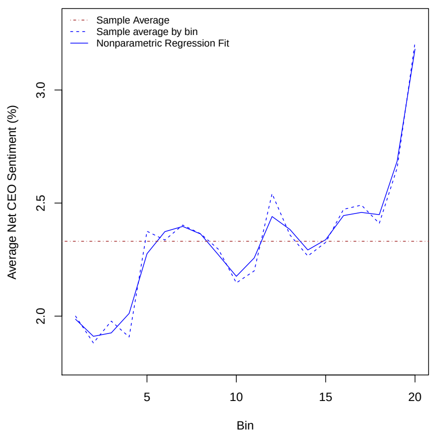
Figure 1: **Frequency plots of the intratextual distribution of CEO positive, negative and sentiment based on the library of Loughran and McDonald [2011] to identify positive and negative words in the CEO letters by DJIA firms**



(a) Frequency plot of average positive sentiment by bin - A U-shape and smile in positive sentiment



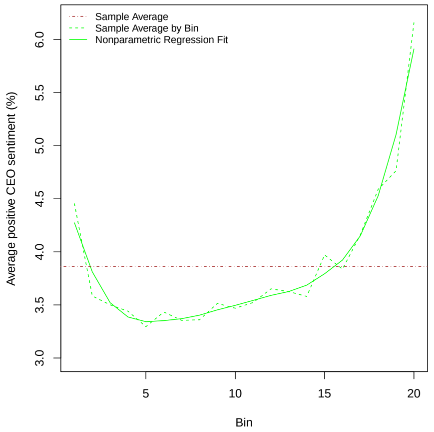
(b) Frequency plot of average negative sentiment by bin - A left-sided smirk in negative sentiment



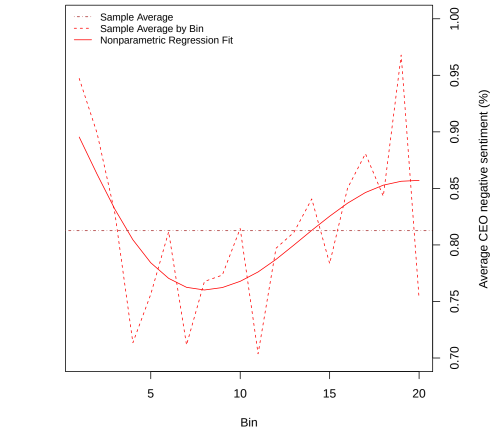
(c) Frequency plot of average net sentiment by bin - A right-sided smirk in net sentiment

Note: This figure depicts the dynamics of CEO sentiment within letters to shareholders. The length of each text is standardized to correspond to a $[0, 1]$ interval, which is divided in B bins such that each bin contains the same number of total words. For each bin, the percentage number of positive words out of the total number of words in each bin is reported (Figure 1a). Similarly, for each bin, the percentage number of negative words out of the total number of words in each bin is computed (Figure 1b). For each bin, the net sentiment is then measured as the spread between the positive and negative tone (Figure 1c). Positive and negative tones are measured based on the Loughran and McDonald [2011] word lists.

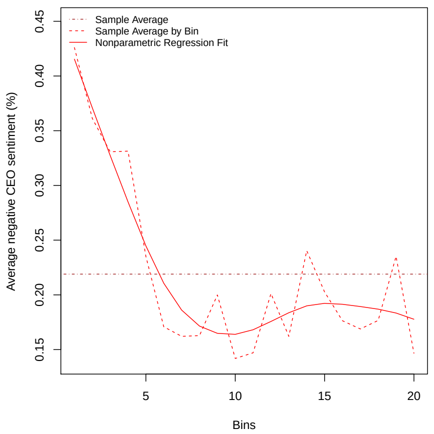
Figure 2: **Frequency plots of the intratextual distribution of CEO positive, negative and sentiment based on the positive and negative word lists of Diction and the negative word lists of Abrahamson and Amir [1996] to identify the positive and negative words in CEO letters by DJIA firms**



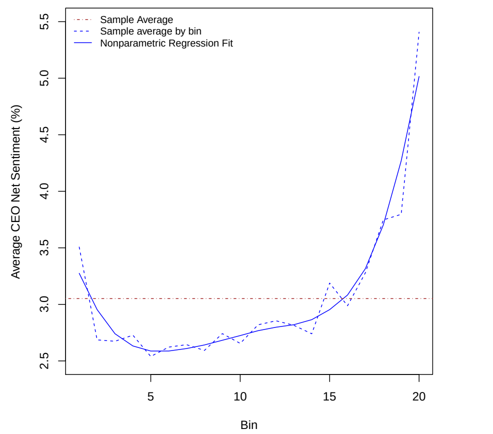
(a) Frequency plot of average positive sentiment by bin (Diction) - A U-shape and smile in positive sentiment



(b) Frequency plot of average negative sentiment by bin (Diction)



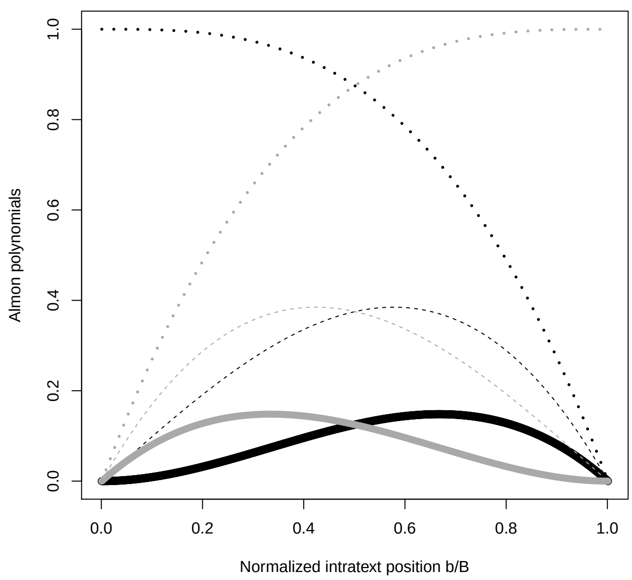
(c) Frequency plot of average negative sentiment by bin (Abrahamson-Amir) - A left-sided smirk in negative sentiment



(d) Frequency plot of average net sentiment by bin - A right-sided smirk in net sentiment

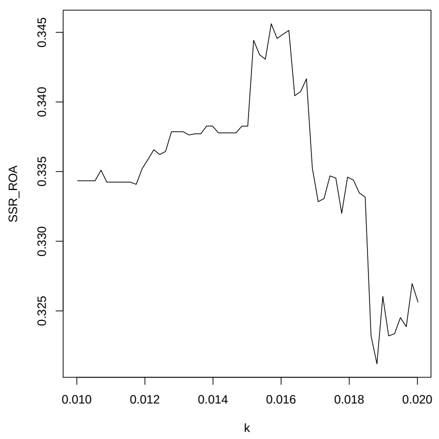
Note: This figure depicts the dynamics of CEO sentiment within letters to shareholders. The length of each text is standardized to correspond to a $[0, 1]$ interval, which is divided in B bins such that each bin contains the same number of total words. For each bin, the percentage number of positive words out of the total number of words in each bin is reported (Figure 2a). Similarly, for each bin, the percentage number of negative words out of the total number of words in each bin is computed (Figures 2b-2c). For each bin, the net sentiment is then measured as the spread between the positive and negative tone (Figure 2d).

Figure 3: **Left- (grey) and right-centered (black) Almon polynomials of order one (full), two (dashed) and three (dotted) used to model the intratextual weights**

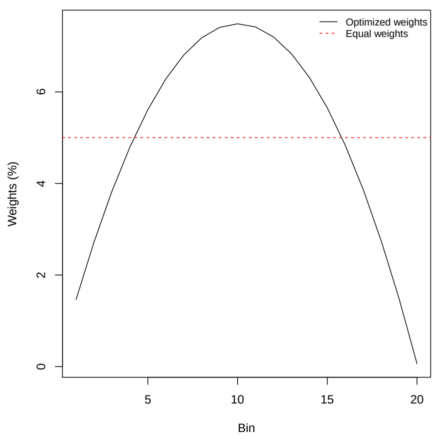


Note: This graph reports the third-order Almon polynomials denoted as $P_c(u)$ in Equation (8.2), for $u = b/B$ (left-centered) and $u = (B - b)/B$ (right-centered).

Figure 4: **The sum of squared residuals as a function of the threshold parameter κ (left figure) and the bell shape in optimized weights of intratextual net sentiment as a function of position of a word in the text (right figure)**

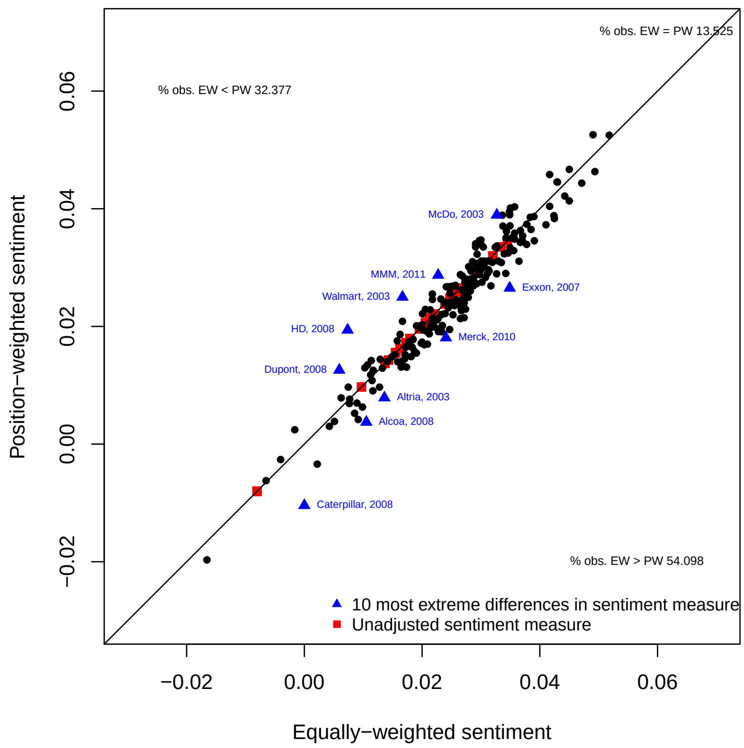


(a) Sum of squared residuals as a function of the threshold parameter kappa (κ) using Equation 5.6.



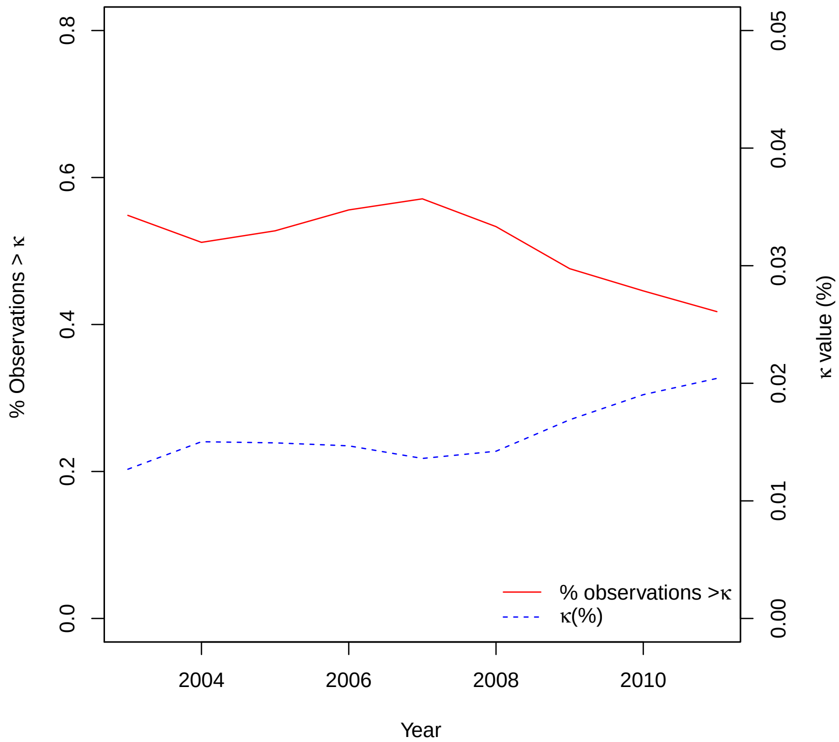
(b) Optimized weights of intratextual net sentiment as a function of position of a word in the text.

Figure 5: **Scatter plot of position-weighted versus equally-weighted sentiment measures for CEO letters of DJIA firms between 2000-2011**



Note: This figure presents the scatter plot of the position-weighted sentiment measures versus the equally-weigted measures of sentiment expressed in the CEO letters of DJIA firms between 2000-2011, and computed following Equation (5.2) and Equation (8.2), respectively. The word lists used to estimate sentiment is from the Loughran and McDonald [2011] library. The ten most extreme differences in sentiment measures are depicted with blue triangles.

Figure 6: **Time series plot of rolling three-year sample estimates of the threshold values κ and the percentage firms with intratextual standard deviation higher than the estimated value of the threshold parameter κ**



Note: This figure presents the plot of the kappa values for each rolling sample of Equation (5.6) (dashed line) and the percentage of firms with intratextual standard deviation higher than kappa (full line).