# A new measure of financial texts' readability

Abhinav Anand*, Sankarshan Basu†, Jalaj Pathak‡

**Abstract**

We introduce a new measure of financial texts' readability: 'tone shift', which captures the incremental contribution of multi-clausal phrases (e.g., 'slowdown in business activity'), and adjectives and adverbs (e.g., 'although', 'faintly') which alter the connotation and tone of financial text. All else equal, high values of tone shift denote high complexity in text-parsing, which in turn implies increased ambiguity and more investor uncertainty. We show that during 1994–2018, yearly tone shift of US firms' 10-K filings display significantly positive associations with their subsequent idiosyncratic volatility and standardized unexpected earnings (SUE); and significantly negative associations with their subsequent earnings, above and beyond that attributable to several prevalent readability measures. We also show that the SEC Plain English Rule (October 1998) has improved the readability of US firms' 10-Ks, and has led to consistently falling levels of tone shift post-1999 for a large majority of US firms—observations at odds with other popular measures of readability.

*Keywords: Financial Disclosures, Financial Text Analysis, Plain English, Readability*

*JEL Classification: G12, G14, G18, M41*

*IIM Bangalore, Finance & Accounting

†IIM Bangalore, Finance & Accounting

‡Great Lakes Institute of Management, Finance & Accounting

# 1 Introduction

Verbosity and complexity of financial disclosures can be effective tools when managers of firms wish to hide unpleasant news or disagreeable future possibilities from shareholders. Similar concerns have been echoed by the SEC [Cox, 2007] and prominent investors such as Warren Buffett [Blanco and Dhole, 2017]. Poor readability of financial texts such as 10-K filings have also been found to be associated with poor financial performance [Li, 2008], earnings management [Lo et al., 2017] and higher stock price crash risk [Kim et al., 2019].

We introduce a new measure of financial texts' readability, 'tone shift', which isolates and quantifies the tone due to the incremental contribution of complex, multi-clausal phrases (e.g., 'buoyancy in animal spirits') and the impact of 'valence shifters': adjectives and adverbs which alter the meaning, and hence the tone of sentences (e.g., 'barely', 'however') [Anand et al., 2021a,b]. All else equal, higher prevalence of multi-clausal phrases and valence shifters tends to make the text-parsing process more complex which leads to more ambiguity regarding the text's connotation; and leads to more uncertainty for investors and analysts alike.

There is a rich collection of prior studies which investigates the readability of financial text and its putative impact on a wide variety of financial outcomes. Some early pioneering studies are Li [2008], Biddle et al. [2009] and Miller [2010], all of which use the Fog Index [Gunning, 1952]. The Fog Index is a popular text analysis technique which comprises two components: 'complex words' which are words with more than two syllables, and 'average words per sentence'. In a well-known paper, Loughran and McDonald [2014b] criticize the usage of the Fog Index since polysyllabic words such as 'telecommunication', 'corporations' etc. are readily understood by readers of financial documents, and hence their 'complexity' is suspect; and the second component: 'average words per sentence' is prone to measurement error since in financial documents such as the 10-Ks, it is not clear what the definition

of a sentence ought to be.[1] Similar criticism applies to other readability indices such as the Flesh-Kincaid index and the SMOG index which are highly correlated with the Fog Index.

Early criticisms of readability formulas dates back to Ojemann [1934] and Dolch [1939] who pointed out that such formulas are often used out-of-context. Redish and Selzer [1985] further clarified how readability formulas were meant for children and were not intended to be used for assessing readability for adults or for technical documents. Further, with respect to the usage of complex words as a measure of readability, Entin [1981] states that when reader interest is high—as is the case with analysts/investors parsing financial documents—comprehension does not increase by writing below grade level. On similar lines DuBay [2007] specifies more than 200 formula based readability metrics by 1980s, all of them being the subject of criticism in a variety of studies [Manzo, 1970, Maxwell, 1978, Bruce et al., 1981, Duffy, 1985, Connatser, 1999].

Apart from formula based readability metrics, two other prominent categories of financial texts' readability are: i) vocabulary-based, and ii) size-based. In our study, we consider both types for the purpose of comparison with our proposed measure 'tone shift'. We employ two vocabulary based measures: i) 'Vocab' which is defined as the number of unique words in a 10-K divided by the total number of words in the Loughran and McDonald [2011] dictionary; and ii) 'Financial_Term' defined as the number of unique words in a 10-K that also appear in Campbell Harvey's hypertextual finance glossary divided by the number of unique 10-K words. Similarly, we use two measures based on size: i) log(file size) advocated by Loughran and McDonald [2014b];[2] and ii) log(total number of words) in the MD&A section as well

[1] For example, as noted in Loughran and McDonald [2014b] financial documents contain several abbreviations, bullet points, numbered lists, tables, figures, nonstandard headings etc. which make the identification of a sentence a much more nontrivial task than that for a conventional piece of text, such as a news report or a novel.

[2] We note however, that DuBay [2007] states that file size may depend on the typeface and layout of the document and hence is more a measure of legibility than readability.

as in the whole 10-K document.

Our approach, however, is different from that of the techniques outlined above. We rely on accurate quantification of that part of the text which i) contributes to complexity in parsing, and ii) alters the connotation of sentences by the use of nuanced writing. The tone shift is calculated as the absolute value of the difference between the tone computed according to Anand et al. [2021a,b] and that according to Loughran and McDonald [2011]. The well-known approach in Loughran and McDonald [2011] stipulates that the tone of financial texts be computed according to the LM dictionary and one-word-at-a-time (bag-of-words) approach. In contrast, the modified tone extraction process described in Anand et al. [2021a,b] considers $n$-words-at-a-time where $n$ varies, and is derived from the length of sentences; and uses adjectives and adverbs ('valence shifters') which can alter the connotation and tone of the financial text. The tone shift is the absolute value of the difference in tones calculated as described. Hence, high levels of tone shift imply high contribution of valence shifters and multi-clausal phrases, which, all else equal, contribute towards more complexity in text-parsing, leading to more ambiguity and uncertainty.

Pennebaker et al. [2003] specify how a text can be analyzed within the context of previously defined psychological content dimensions or by analyzing the word count and/or word pattern strategies. Hart [2001] compares the two approaches by drawing upon a metaphor of two people trying to understand a city by driving on the streets versus viewing it from a helicopter. The word count based approaches provide linguistic information of the text content from an 'aerial distance' (using a helicopter) which could, in principle, lead to missing information on the details around specific 'corners of the street'. The new readability measure introduced in this study improves upon the 'corners of the street' details by the use of 'valence shifters' and by using the whole sentence as a unit of tone quantification which is akin to providing binoculars to the person in the helicopter, thus ensuring he/she gets a more detailed view of the corners while also receiving an 'aerial' per-

spective. Our proposed readability metric (tone shift) is also compatible with Pennebaker et al. [2003] which argue that the entire corpus of text and individual sentences within it, must be considered while assessing the meaning of the text. DuBay [2007] also specifies how cognitive theorists and linguists in the 1970s elaborated that the meaning of a text is not in the independent words but is rather constructed by making inferences and interpretations on the whole. Our readability metric ensures that this dictum is obeyed since it is able to assign proper weights to complex, multi-clausal phrases, as well as to adjectives and adverbs—both of which can completely alter the connotation of the text. Similarly, Kintsch and Vipond [2014] mention that readability metrics should accommodate the interaction between the reader and the text. This aspect cannot be captured by simple readability formulas but can be accounted for by tone shift since it explicitly quantifies the effect of adjectives and adverbs (valence shifters) which, according to Hull [1979] is necessary for assessing the readability of technical writing. On a similar note, Larcker and Zakolyukina [2012] also state that pure word counting does not categorize combination of words that might imply different meanings from the constituent words. While other readability metrics share this weakness, the usage of valence shifters when calculating tone shift overcomes this challenge. Moreover, since tone shift is not based on either complex words or the average number of words, it remains immune to the criticisms which afflict readability metrics such as the Fog, Flesh-Kincaid or the SMOG indices. Further, since tone shift is built upon the tone of the financial text at its core, it is able to capture the essence of the MD&A section of the 10-K reports in particular.

Prior studies on readability of annual reports have examined both the 10-K and the management discussion and analysis (MD&A henceforth) section, and there seems to be no consensus as to which document is more desirable from the perspective of analyzing the effects of readability [Xu et al., 2018]. For example, Li [2008], Lehavy et al. [2011] and Loughran and McDonald [2014b] are some prominent studies which examine the impact of readability

of the 10-K, whereas Feldman et al. [2010], Li [2010] and Lo et al. [2017] analyze the impact of readability of the MD&A section. According to Lo et al. [2017] the management has substantial leeway in the content and layout of the MD&A section, and its inclusion is mandated by law. Additionally, the MD&A section provides investors with new and important supplementary information in addition to the financial statement numbers in the 10-K report [Feldman et al., 2010, Loughran and McDonald, 2011, Jegadeesh and Wu, 2013]. In light of this, we conduct our preliminary analysis on the MD&A section and ensure robustness by repeating the entire exercise with 10-K specific readability measures as additional controls.

In line with Loughran and McDonald [2014b] we use post-10-K-filing stock return volatility (market model RMSE) as a proxy of firms' information environment. We implicitly assume that more readable financial documents produce less ambiguity in valuation, which should be reflected in lower price volatility of the stock in the period immediately following the 10-K filing even after controlling for other relevant variables, including the historical level of volatility.

Our main results are as follows. We analyze a comprehensive sample of US firms' 10-K filings during 1994–2018 and test whether their subsequent idiosyncratic voltility—measured as the market model's RMSE [Loughran and McDonald, 2014b]—is significantly associated with tone shift, over and above the impact of other popular measures of readability and relevant controls. We find that firms' MD&A tone shift has a significant association with their subsequent volatility and that the coefficient is uniformly positive in all regression specifications, even after accounting for other prevalent readability measures and controls. We also find that firms' standardized unexpected earnings (SUE) are significantly positively associated with their MD&A tone shift. Both these findings are in line with our hypothesis that higher tone shift—or equivalently, poorer MD&A readability—leads to more ambiguity and investor uncertainty which leads to more subsequent volatility and increases uncertainty regarding unexpected earnings. Further, we test if firms'

tone shift has any predictive association with their subsequent earnings. We find that the MD&A tone shift is significantly associated with firms' future earnings, up to two years in advance, and renders all other popular readability metrics insignificant in its presence. The signs of the coefficients are uniformly negative which suggests that all else equal, an increase in firms' MD&A tone shift—equivalently an increase in unreadability which entails higher ambiguity and uncertainty—is associated with a fall in future earnings for both a one-year and two-year horizon.

We run a battery of robustness exercises to ensure that our results can be relied upon under varying circumstances. To this end, we additionally control for business complexity, to allay concerns that complex, hard-to-parse text may be a necessary outcome for firms operating in a complex business environment, and find that business complexity has no impact on our results. We isolate only the least frequently used valence shifters and recalculate the tone shift to account for possible over-representation of the most numerous valence shifters and find that the results remain the same. We augment the LM dictionary [Loughran and McDonald, 2011] with verb-noun combinations which assign weights to phrases (such as 'increasing instability', 'decreasing returns' etc.) which are ignored in traditional dictionary-based text analysis approaches and find no changes in our benchmark results. Finally, we isolate two special classes of valence shifters—negators (e.g., 'never') and adversative conjunctions (e.g., 'however', 'but')—which can flip the sign of the text tone, recompute the tone shift, and rerun our regressions and find that the results remain unaffected.

We also study trends in firms' MD&A and 10-K readability over the years and evaluate whether the SEC Plain English Rule, imposed in October 1998 had the desired impact in making firms' financial disclosures more readable. We document that for a large majority of US firms, tone shift has been steadily declining after 1999 suggesting that the rule had the intended effect, in agreement with prior findings reported in Loughran and McDonald [2014a]. Moreover, we show that firms which did not use overly complex language in

their MD&A section are not impacted by the SEC Plain English Rule, but those which featured nuanced, hard-to-parse language prior to the imposition of the rule exhibit the maximum impact by changing their financial disclosure behavior so as to conform to the SEC's initiative. However, other readability measures such as 'average words per sentence', '% of complex words', 'log of total words', 'gross/net file size of 10-K' etc. do not show this behavior and are at odds with our results that demonstrate improved readability of firms' 10-K filings over the years.

The remainder of the paper proceeds as follows. Section 2 reviews the relevant literature, section 3 outlines the data gathering process, section 4 describes our paper's methodology, section 5 discusses our benchmark results, section 6 outlines trends in readability, and finally section 7 offers concluding remarks.

## 2 Literature Review

Among the first studies (in finance and accounting) to address the issue of readability (or lack thereof) of financial texts is Li [2008] in which he examines the impact of readability—proxied by the Fog Index—on earnings persistence and finds that annual reports of firms with lower earnings are hard to read. Biddle et al. [2009] also examine the impact of financial reports' readability on investment efficiency and report significant results. On a related note, Miller [2010] finds that more complex financial reports are associated with lower trading activity due to reduced activity by small investors. Lehavy et al. [2011] also use Fog Index as measure of readability and find that a higher Fog Index (lower readability) is significantly associated with higher analyst following. Lawrence [2013] also uses the Fog Index, as well as financial disclosures' log of word count, and finds that individuals invest more with firms which have clear and concise disclosures. On the other hand, in a well-known study, Loughran and McDonald [2014b] argue that readability measures based on average words per sentence and percentage of complex

words as constituents (i.e., the Fog Index, SMOG Index and Flesch–Kincaid Index) are misleading for the purposes of financial reports and disclosures. Instead, they advocate the usage of the file size of the financial report as a proxy of readability. Lo et al. [2017] analyze the association between readability of the MD&A section of the 10-K reports and earnings management using the Fog Index as a measure of readability. Ertugrul et al. [2017] and Kim et al. [2019] further examine the impact of readability using file size and a modified Fog Index respectively as proxies and report that firms with more complex reports have higher risk of future stock price crashes. They also note how file size has a severe measurement error problem in gauging information obfuscation, since graphics, XBRL and HTML significantly enlarge the file sizes of 10-K reports but actually improve the information gathering process.

# 3 Data

The 10-K documents are downloaded from 1994 to 2020 from the EGDAR website. The Loughran and McDonald word list is downloaded from the website https://sraf.nd.edu/textual-analysis/resources/ for constructing the LM based vocabulary measure. Similarly, the Harvey Campbell word list is downloaded from http://people.duke.edu/~charvey/Classes/wpg/glossary.htm. This word list is used to specify the "Financial Term" measure of readability as specified in the section on methodology. The control variables are downloaded from CRSP and COMPUSTAT and are discussed further in the methodology section. The analyst data are downloaded from Thomson Reuters.

# 4 Methodology

## 4.1 Measures based on the complexity of words

We follow Loughran and McDonald [2011] in parsing text from 10-K reports and remove tables and exhibits during the parsing process. However, there is one main difference between our procedure and theirs in that we additionally classify sentences as a collection of words between: i) two full stops, ii) a full stop and a question mark, and iii) two question marks. Further, sentences with fewer than 10 characters are excluded from the sample to ensure that instances where a decimal point, say, is incorrectly identified as a full stop, are omitted. The complex words are identified as words with 3 or more syllables. This leads to the formation of three measures based on the percentage of complex words in the text and the average number of words: the Fog Index, the Flesh-Kincaid index and the SMOG Index [Li, 2008, Lehavy et al., 2011].[3]

## 4.2 Measures based on vocabulary and size

The LM 'vocab' measure is calculated as the number of unique words in a MD&A report divided by the the number of entries in the LM dictionary [Loughran and McDonald, 2014b]. 'Financial_Term' is defined by the number of unique words in the 10-K report which appear in Campbell Harvey's hypertextual finance glossary (http://people.duke.edu/~charvey/Classes/wpg/glossary.htm) divided by the total number of unique words in the MD&A [Loughran and McDonald, 2014b]. Size-based measures include the log of the total number of words in the 10-K; and the log of net, as well as the gross file size of the whole 10-K. [Loughran and McDonald, 2014b].

[3]Fog Index is defined as $0.4\times$(average words per sentence + percentage of complex words). Flesh-Kincaid index is defined as $206.835 - 1.015\times$ (average words per sentence) $-$ $84.6 \times$ (percentage of complex words). SMOG Index is defined as $1.043 \times$ sqrt(percentage of complex words) $\times$ 30/number of sentences.)

### 4.3 The new measure of readability: 'Tone shift'

We refer to our new, proposed measure of financial texts' readability as 'tone shift'. We define and calculate tone shift as the absolute value of the difference between the tone of the financial text calculated according to Anand et al. [2021a] and that obtained from the usage of the LM dictionary and a 'bag-of-words' (one word at a time) approach. The tone quantification technique in Anand et al. [2021a] stipulates that i) the sentence be used a unit of analysis, which solves the problem of how many words to include in the ngram analysis, and ii) proper weights be assigned to 'valence shifters', which are adjectives and adverbs such as 'but', 'despite', 'faintly', 'very' etc. which affect the connotation of sentences but have been ignored in the traditional tone extraction process for financial texts.

These two innovations help us to capture financial texts' connotations derived from multi-clausal phrases (e.g., 'slowdown in business activity'); as well as the effect of adjectives and adverbs which can amplify/de-amplify (e.g., 'severely', 'faintly' etc.), negate (e.g., 'not/nor' etc.) or provide additional nuance (e.g., 'despite', 'although' etc.); but have been ignored so far by current methods [Anand et al., 2021a,b].[4] The standard LM method of financial texts' tone quantification relies on a unigram analysis which considers the whole text as a 'bag of words'. Thus tone shift—constructed as the absolute value of the difference of tones—is, in fact, exactly the incremental value of the connotation contributed by valence shifters and multi-clausal phrases. The central idea of the paper is that large (small) values of tone shift—implying high (low) levels of parsing complexity in financial texts—lead to low (high) levels of readability. All else equal, low readability entails more complexity in parsing financial text which leads to more ambiguity and more investor uncertainty.

For example, the sentence below is taken from the MD&A section of the

[4]The valence shifters can be divided into four categories: adversative conjunction (e.g. 'although', 'however'), negator (e.g. 'never', 'not'), amplifier (e.g. 'very') and de-amplifier (e.g. 'few').

10-K of AAC Holdings Inc. on 2015-03-11.

> *"the gross profit margin percentage declined slightly from the prior year primarily due to start up activities at the indianapolis airframe maintenance facility."*

The tone of this sentence using the "bag-of-words" approach and LM dictionary is:

$$\frac{(-1)[=\text{declined}]}{16} = -0.0625$$

However, the sentence has one valence shifter: "slightly" which is a de-amplifier. Thus, the value of the texts' tone using the sentence as a unit and valence shifters is:

$$\frac{(-1)[=\text{declined}] + (0.8)[=\text{slightly}]}{16} = -0.0125$$

Hence, the new readability score, or 'tone shift' for this sentence is:

$$\text{tone shift} = |-0.0625 - (-0.0125)| = 0.05$$

Table 1 illustrates tone shift calculations based on different types of valence shifters. Tables A.1 and A.2 in the appendix, list the full collection of valence shifters encountered in this study.

## 4.4 CRSP and COMPUSTAT control variables

Root mean square error (RMSE) is calculated using the market model for trading days [6,28] with firm-return downloaded from CRSP, and market return from Kenneth French's website `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`. On a similar note, pre-filing RMSE is calculated for trading days [-257,-6] [Loughran and McDonald, 2014b]. Pre-filing alpha is calculated for trading days [-252,-6] using data from CRSP. Book-to-market is calculated using book value from most recent year prior to filing date and market value of equity from CRSP [Fama and

Table 1: MD&A Valence Shifters

| Company (CIK) &Date | Type | Word | Sentence | Tone New | Tone LM | Tone Shift | Comment |
|---|---|---|---|---|---|---|---|
| Aaon Inc. (824142) 25-02-2016 | Adversative Conjunction | "but" | "the price levels of our raw materials have remained relatively consistent the past few years, but the market continues to be volatile and unpredictable as a result of the uncertainty related to the u.s. economy and global economy." | -0.264 | -0.091 | 0.173 | The **but** preceding the phrase "market continues to be volatile and unpredictable" accentuates the negative tone of the sentence. |
| Aar Corp. (1750) 17-07-2006 | De-amplifier | "slightly" | "the gross profit margin percentage declined slightly from the prior year primarily due to start up activities at the indianapolis airframe maintenance facility." | -0.012 | -0.062 | 0.05 | The presence of the word **slightly** before the phrase "gross profit margin percentage declined" reduces the severity of negative connotation. |
| Avx Corp. (859163) 21-06-2002 | De-amplifier | "few" | "during fiscal 2002, the company recognized a $3.0 million special charge for doubtful customers receivables, as the economic environment caused the receivables due from a few smaller customers to be significantly past due and determined to be uncollectable." | -0.090 | -0.121 | 0.031 | The de-amplifier **few** reduces the coefficient for tone of the negative word uncollectable. |
| Avx Corp. (859163) 30-05-2006 | Amplifier | "very" | "the company and our industry have gone through a very difficult cycle over the past five years." | -0.257 | -0.111 | 0.146 | Usage of **very** amplifies the coefficient for tone of the negative phrase "difficult cycle". |

Note: This table presents illustrations for various types of valence shifters encountered in firms' MD&A section in our sample. The column 'Tone Shift' is the new readability measure introduced in this study and is calculated as the absolute value of the difference between Tone LM: the tone calculated using the Loughran and McDonland dictionary and "bag-of-words approach"; and Tone New: the tone using the sentence as a base unit and valence shifters.

French, 2001]. Size is proxied by log of market value of equity. The business segment index is calculated using the measure specified in Jennings et al. [2014] by taking the sum of squared business segment proportions from the COMPUSTAT segment database.

# 5 Results and analysis

## 5.1 Sample creation and correlation

Table 2 presents the sample creation process for our study. We start with all 10-K files from 1994 to 2020 (10-K, 10-KSB, 10-K405 and 10-KSB40) and extract the MD&A section from these files leading to an intial sample size of 165,616 observations. In line with Loughran and McDonald [2014b] we remove duplicate filings with respect to CIK and year combination; and also if the filing date is less than 180 days from prior filing which reduces our sample size to 162,859 observations. Next we drop files for which relevant control variables are not available from CRSP and COMPUSTAT and if the MD&A section has fewer than 250 words, which narrows down our sample size to 60,112 observations. Finally we drop 10-K's for which the RMSE value is missing. This bring the final sample to 57,518 firm-year observations.

Table 2: Sample Creation

| | Dropped | Sample Size |
|---|---|---|
| MD&A extracted from SEC 10-K files 1994:2018 | | 165,616 |
| Remove duplicates within year/CIK | 2,409 | 163,207 |
| Drop if file date is $<$ 180 days from prior filing | 349 | 162,859 |
| Drop if corresponding data unavailable | 101,949 | 60,910 |
| Drop if MD&A has fewer than 250 words | 198 | 60,112 |
| Drop if RMSE value is missing | 2,594 | 57,518 |

Note: This table presents the details of sample construction and the number of observations dropped in each filtering step. The unavailability of data referred to in row 4 refers to the CRSP and COMPUSTAT databases.

Table 3 presents the summary statistics of readability variables for all

three categories: based on complex words and average words per sentence (Fog Index, Flesch-Kincaid Index, SMOG Index); vocabulary based (LM vocab and Campbell Harvey Vocab); and size based (file size and number of words). The table also contains summary statistics for tone shift, the readability measure introduced in this study. The values of tone shift are the smallest when compared to other measures on account of its construction as the incremental contribution of text employing multi-clausal phrases and valence shifters.

Table 3: Summary Statistics

| Readability Measures | Mean | Median | SD | IQR |
|---|---|---|---|---|
| AWPS | 29.82 | 29.46 | 5.39 | 5.63 |
| Per_CW | 23.73 | 23.74 | 0.021 | 0.027 |
| Fog_Index | 21.42 | 21.38 | 2.30 | 2.38 |
| FK_Index | 30.78 | 30.76 | 1.45 | 1.97 |
| SMOG_Index | 18.26 | 18.27 | 1.36 | 1.60 |
| Log(Words) | 8.93 | 9.04 | 0.79 | 0.91 |
| Vocab | 0.57 | 0.55 | 0.26 | 0.31 |
| Financial_Term | 0.01 | 0.01 | 0.0001 | 0.00001 |
| Tone Shift | 0.0060 | 0.0042 | 0.0073 | 0.0061 |

Note: This table presents the summary statistics for all readability measures (MD&A) used in this study. 'AWPS' denotes 'average words per sentence' and 'Per_CW' denotes 'percentage of complex words'.

Table 4 presents correlations among readability measures. Tone shift has a quite low positive correlation with the three formula-based readability measures: Fog Index, Flesch Kincaid (FK) and SMOG Index. There is almost no correlation with the metrics 'average words per sentence' and 'Financial_Term', and it displays a low negative correlation with LM based variable 'Vocab' and the log of total number of words. Among the existing readability measures, the Fog index, the SMOG index and average words per sentence are extremely positively correlated ($> 0.90$).

We repeat the central idea behind the new variable: by the usage of multi-clausal sentences, and adverbs and adjectives, more nuance and/or

Table 4: Correlation among Readability Measures

| Variable | Fog_Index | FK_Index | SMOG_Index | AWPS | per_CW | log_words | Vocab | Fin_Term | Tone Shift |
|---|---|---|---|---|---|---|---|---|---|
| Fog_Index | 1 | | | | | | | | |
| FK_Index | 0.143 | 1 | | | | | | | |
| SMOG_Index | 0.944 | 0.246 | 1 | | | | | | |
| AWPS | 0.924 | -0.098 | 0.806 | 1 | | | | | |
| per_CW | 0.354 | 0.619 | 0.498 | -0.028 | 1 | | | | |
| log_words | 0.210 | -0.021 | 0.238 | 0.254 | -0.072 | 1 | | | |
| Vocab | 0.249 | -0.070 | 0.267 | 0.297 | -0.076 | 0.896 | 1 | | |
| Fin_Term | 0.004 | 0.003 | 0.002 | 0.011 | -0.015 | 0.032 | 0.042 | 1 | |
| Tone Shift | 0.069 | 0.106 | 0.086 | 0.014 | 0.146 | -0.274 | -0.194 | -0.024 | 1 |

Note: This table presents the correlations of various meausres of readability along with the new measure 'tone shift'.

complexity can be introduced into a sentence. All else equal, therefore, it is likely that text with more complex writing leads to more ambiguity and hence higher uncertainty among readers of the financial text.

This hypothesis may be tested by the following regression specification in which the root-mean-squared error of the market model is attributed to controls, prior measures of readability; and the new readability measure: tone shift.

$$RMSE_{tk} = a_0 + a_1 Tone\_Shift_{tk} + a_2 Readability\_Measures_{tk} + a_3 Controls_{tk} + u_{tk} \quad (1)$$

The dependent variable is the RMSE for trading days [6,28] (post-filing date market model root mean square error). The regression includes an intercept, calender year dummies, and Fama and French 48-Industry dummies.

## 5.2 Impact of MD&A tone shift on RMSE

Table 5 presents the results of the regression which evaluates the hypothesis that the market model's residual size (RMSE) has a significant association with tone shift over and above that for other readability scores. The control variables include pre-filing alpha $[-252, -6]$, pre-filing RMSE $[-257, -6]$,

holding period return $[0, +1]$, size measured as log(market equity), log(BM) and NASDAQ dummy in line with Loughran and McDonald [2014b]. The regression also contains complex word and average words per sentence, as well as other popular readability measures: Fog Index, FK Index and SMOG Index.

Table 5: Tone shift impact on RMSE [in presence of complex-word based measures]

| Readability Variables | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Fog_Index | | | 0.0001 | | |
| | | | (0.0001) | | |
| FK_Index | | | | 0.0001 | |
| | | | | (0.0001) | |
| SMOG_Index | | | | | 0.0001 |
| | | | | | (0.0001) |
| awps | | 0.0000 | | | |
| | | (0.0000) | | | |
| per_CW | | 0.010 | | | |
| | | (0.008) | | | |
| Tone Shift | 0.039** | 0.034** | 0.038** | 0.036** | 0.037** |
| | (0.014) | (0.011) | (0.013) | (0.012) | (0.013) |
| Control Variables | | | | | |
| Pre-filing alpha | −1.566*** | −1.565*** | −1.565*** | −1.565*** | −1.595*** |
| | (0.310) | (0.310) | (0.310) | (0.310) | (0.310) |
| Pre-filing RMSE | 0.581*** | 0.580*** | 0.580*** | 0.580*** | 0.580*** |
| | (0.048) | (0.049) | (0.049) | (0.048) | (0.049) |
| Filing Period Return | −0.016*** | −0.016*** | −0.016*** | −0.016*** | −0.016** |
| | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) |
| Size | −0.001*** | −0.001*** | −0.001*** | −0.001*** | −0.001*** |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| BM | −0.001*** | −0.001*** | −0.001*** | −0.001*** | −0.001*** |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| NASDAQ Dummy | 0.002** | 0.002** | 0.002*** | 0.002** | 0.002*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Adjusted $R^2$ | 43.96% | 44% | 43.96% | 44% | 43.96% |
| $N$ | 57,510 | 57,510 | 57,510 | 57,510 | 57,510 |

Note: This table presents the results from the regression of RMSE on various complex-words based readability measures. The dependent variable is the RMSE for trading days [6,28] (post-filing date market model root mean square error). The regression includes an intercept, calender year dummies, and Fama and French 48-Industry dummies. The results are reported in line with equation 1. The standard errors (reported in parentheses) are clustered by industry and year. ***, ** and * indicate that the coefficient estimate are significantly different from zero at the 1 percent, 5 percent and 10 percent levels respectively.

The new readability score, tone shift—the tone spread between the benchmark LM based bag-of-words approach and the modified approach based on valence shifters—shows persistent significance in all specifications. In the presence of controls, the tone shift displays significant impact on the market residual RMSE above and beyond that attributable to prior complex-words

based readability measures. In particular, in the presence of tone shift, the FOG index, the Flesch-Kincaid index, the SMOG index; and 'average words per sentence' and '% of complex words' are all rendered insignificant. Further, the positive sign of the tone shift coefficient in all regression specifications suggests that all else equal, a rise (fall) in tone shift—an increase (decrease) in parsing complexity in the form of the tone spread attributable to valence shifters and multi-clausal phrases—leads to a rise (fall) in RMSE.

Table 6 presents the results of the regression which evaluates the hypothesis that the market model's residual size (RMSE) has a significant association with tone shift over and above that for other vocab and size-based readability scores.

Table 6: Tone shift impact on RMSE [in presence of vocab and size-based measures]

| Readability Variables | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Vocab | 0.003*** | | 0.003*** | |
| | (0.001) | | (0.001) | |
| Financial_Term | | −0.421 | −0.462 | |
| | | (0.590) | (0.586) | |
| Log Words | | | | 0.0004 |
| | | | | (0.0003) |
| Tone Shift | 0.051*** | 0.051*** | 0.049*** | 0.037*** |
| | (0.016) | (0.014) | (0.016) | (0.016) |
| Control Variables | | | | |
| Pre-filing alpha | −1.559*** | −1.566*** | −1.559*** | −1.562*** |
| | (0.310) | (0.310) | (0.310) | (0.310) |
| Pre-filing RMSE | 0.578*** | 0.581*** | 0.578*** | 0.580*** |
| | (0.048) | (0.048) | (0.048) | (0.048) |
| Filing Period Return | −0.016*** | −0.016*** | −0.016*** | −0.016*** |
| | (0.004) | (0.004) | (0.004) | (0.004) |
| Size | −0.001*** | −0.001*** | −0.001*** | −0.001*** |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| BM | −0.001*** | −0.001*** | −0.001*** | −0.001*** |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| NASDAQ Dummy | 0.002*** | 0.002** | 0.002** | 0.002** |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| Adjusted $R^2$ | 44.1% | 44% | 44.1% | 44% |
| $N$ | 57,510 | 57,510 | 57,510 | 57,510 |

Note: This table presents the results from the regression of RMSE on vocab and size-based readability measures. The dependent variable is the RMSE for trading days [6,28] (post-filing date market model root mean square error). The regression includes an intercept, calender year dummies, and Fama and French 48-Industry dummies. The results are reported in line with equation 1. The standard errors (reported in parentheses) are clustered by industry and year. ***, ** and * indicate that the coefficient estimate are significantly different from zero at the 1 percent, 5 percent and 10 percent levels respectively.

Tone shift shows persistent significance in all specifications. In the pres-

ence of controls, the tone shift displays significant impact on the market residual RMSE above and beyond that attributable to prior vocab and size-based readability measures. The readability measures 'log words' and 'financial term' are rendered insignificant in the presence of tone shift; and the LM based 'vocab' and tone shift are both significant when tested together. The positive sign of the tone shift coefficient in all regression specifications suggests that all else equal, a rise (fall) in tone shift—an increase (decrease) in the marginal contribution of valence shifters and multi-clausal phrases to the tone of firms' MD&A—leads to a rise (fall) in RMSE.

In table 7 we present the results of regressions in which all readability scores—from both the complex words and vocab and size-based methods—are included as controls to test whether there is any incremental significance for the tone shift measure. We test the impact of tone shift in the presence of all combinations of complex-words based, and vocab and size-based readability scores and find that for each model, the tone shift continues to display significance over and above other readability measures. In its presence other prior readability measures, such as 'average words per sentence', '% of complex words', the FOG, Flesch-Kincaid and SMOG indices; as well vocab-based readability measures such as 'Financial_Term' are all rendered insignificant. Only the LM-based 'vocab' and the size-based log of words retain their significance in the presence of tone shift. The coefficient of tone shift is positive in all specifications suggesting that all else equal, a rise in the tone spread on account of increased parsing complexity leads to an increase in RMSE.

## 5.3 File size of 10-K as an additional Control

We further examine if tone shift is significantly associated with RMSE in the presence of firms' 10-K file size. Thus, we add log of gross file size [Loughran and McDonald, 2014b] to tables 5 and 6 and report the results in table 8.

Tone shift shows positive significance in all specifications and renders all other readability scores (except vocab) insignificant in its presence. The table

Table 7: Impact of tone shift on RMSE [in presence of all readability measures]

| Readability Variables | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Vocab | | | | | $0.005^{***}$ | $0.004^{***}$ |
| | | | | | (0.001) | (0.001) |
| Financial_Term | | | | | | $-0.474$ |
| | | | | | | (0.584) |
| Log Words | $0.0004^{*}$ | 0.0004 | 0.0004 | 0.0004 | $-0.0008^{*}$ | $-0.0008^{*}$ |
| | (0.0002) | (0.0002) | (0.0002) | (0.0002) | (0.0004) | (0.0004) |
| Fog_Index | | 0.0001 | | | | |
| | | (0.0001) | | | | |
| FK_Index | | | 0.0001 | | | |
| | | | (0.0001) | | | |
| SMOG_Index | | | | 0.0001 | | |
| | | | | (0.0001) | | |
| awps | 0.0000 | | | | $-0.0000$ | 0.0000 |
| | (0.0000) | | | | (0.0000) | (0.0000) |
| per_CW | 0.012 | | | | 0.012 | 0.010 |
| | (0.008) | | | | (0.008) | (0.008) |
| Tone Shift | $0.044^{***}$ | $0.047^{***}$ | $0.046^{***}$ | $0.046^{***}$ | $0.037^{***}$ | $0.037^{***}$ |
| | (0.014) | (0.015) | (0.014) | (0.015) | (0.014) | (0.013) |
| Control Variables | | | | | | |
| Pre-filing alpha | $-1.560^{***}$ | $-1.561^{***}$ | $-1.560^{***}$ | $-1.561^{***}$ | $-1.559^{***}$ | $-1.559^{***}$ |
| | (0.310) | (0.310) | (0.310) | (0.310) | (0.309) | (0.309) |
| Pre-filing RMSE | $0.579^{***}$ | $0.579^{***}$ | $0.579^{***}$ | $0.579^{***}$ | $0.577^{***}$ | $0.577^{***}$ |
| | (0.048) | (0.048) | (0.048) | (0.048) | (0.048) | (0.048) |
| Filing Period Return | $-0.015^{***}$ | $-0.015^{***}$ | $-0.015^{***}$ | $-0.015^{***}$ | $-0.015^{**}$ | $-0.015^{**}$ |
| | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) |
| Size | $-0.001^{***}$ | $-0.0014^{***}$ | $-0.0014^{***}$ | $-0.0014^{***}$ | $-0.0014^{***}$ | $-0.0015^{***}$ |
| | (0.0003) | (0.0002) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| BM | $-0.001^{***}$ | $-0.0013^{***}$ | $-0.0013^{***}$ | $-0.0013^{***}$ | $-0.0014^{***}$ | $-0.0014^{***}$ |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| NASDAQ Dummy | $0.001^{**}$ | $0.001^{***}$ | $0.0017^{**}$ | $0.001^{***}$ | $0.001^{***}$ | $0.001^{***}$ |
| | (0.0006) | (0.0006) | (0.0006) | (0.0006) | (0.0006) | (0.0006) |
| Adjusted $R^2$ | 43.96% | 43.96% | 43.96% | 43.96% | 43.99% | 43.99% |
| $N$ | 57,427 | 57,428 | 57,428 | 57,428 | 57,426 | 57,425 |

Note: This table presents the results from the regression of RMSE on various complex-words as well as vocab and size based readability measures. The dependent variable is the RMSE for trading days [6,28] (post-filing date market model root mean square error). The regression includes an intercept, calender year dummies, and Fama and French 48-Industry dummies. The results are reported in line with equation 1. The standard errors (reported in parentheses) are clustered by industry and year. ***, ** and * indicate that the coefficient estimate are significantly different from zero at the 1 percent, 5 percent and 10 percent levels respectively.

suggests that all else equal, a rise in tone shift contributes to a corresponding rise in the level of RMSE.

## 5.4 Impact of tone shift on SUE

The central thesis of this paper is that the incremental effect of the tone of multi-clausal phrases and valence shifters on financial texts' readability is to make the text more complex to parse, leading to higher ambiguity and more investor uncertainty. To this end, we examine the impact of tone shift on 'standardized unexpected earnings' (SUE), with the measure calculated according to Chung and Hrazdil [2011].

Table 8: Tone shift impact on RMSE controlling for file size

| Readability Variables | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| awps | | 0.00000 | | | | | | | |
| | | (0.00003) | | | | | | | |
| per_CW | | 0.010 | | | | | | | |
| | | (0.007) | | | | | | | |
| Fog_Index | | | 0.00004 | | | | | | |
| | | | (0.0001) | | | | | | |
| FK_Index | | | | 0.0001 | | | | | |
| | | | | (0.0001) | | | | | |
| SMOG_Index | | | | | 0.0001 | | | | |
| | | | | | (0.0001) | | | | |
| vocab | | | | | | 0.002** | | 0.002** | |
| | | | | | | (0.001) | | (0.001) | |
| Financial_Term | | | | | | | −0.420 | −0.457 | |
| | | | | | | | (0.583) | (0.583) | |
| log_wrds | | | | | | | | | 0.0003 |
| | | | | | | | | | (0.0003) |
| log(GrossFileSize) | 0.001** | 0.0005* | 0.0005* | 0.0005* | 0.0005* | 0.0003 | 0.001* | 0.0003 | 0.0004 |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| Tone Shift | 0.041*** | 0.036*** | 0.039*** | 0.037*** | 0.039*** | 0.050*** | 0.041*** | 0.050*** | 0.048*** |
| | (0.014) | (0.012) | (0.014) | (0.013) | (0.014) | (0.016) | (0.014) | (0.016) | (0.016) |
| Control variables | | | | | | | | | |
| Pre-filing RMSE | 0.580*** | 0.579*** | 0.580*** | 0.579*** | 0.580*** | 0.578*** | 0.580*** | 0.578*** | 0.579*** |
| | (0.048) | (0.049) | (0.049) | (0.049) | (0.049) | (0.048) | (0.048) | (0.048) | (0.048) |
| BM | −0.001*** | −0.001*** | −0.001*** | −0.001*** | −0.001*** | −0.001*** | −0.001*** | −0.001*** | −0.001*** |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| Size | −0.001*** | −0.001*** | −0.001*** | −0.001*** | −0.001*** | −0.002*** | −0.001*** | −0.002*** | −0.001*** |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| Filing Period Return | −0.016*** | −0.016*** | −0.016*** | −0.016*** | −0.016*** | −0.016*** | −0.016*** | −0.016*** | −0.016*** |
| | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) |
| Pre-filing alpha | −1.563*** | −1.562*** | −1.563*** | −1.562*** | −1.563*** | −1.558*** | −1.563*** | −1.558*** | −1.560*** |
| | (0.310) | (0.310) | (0.310) | (0.309) | (0.310) | (0.309) | (0.310) | (0.309) | (0.310) |
| NASDAQ_Dummy | 0.002** | 0.002** | 0.002*** | 0.002** | 0.002*** | 0.002*** | 0.002*** | 0.002*** | 0.002** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Adjusted $R^2$ | 44.0% | 44.0% | 44.0% | 44.0% | 44.0% | 44.0% | 44.0% | 44.0% | 44.0% |
| $N$ | 57,510 | 57,510 | 57,510 | 57,510 | 57,510 | 57,510 | 57,510 | 57,510 | 57,510 |

Note: This table presents the results from the regression of RMSE on all readability measures along with gross file size of the 10-K as an additional control. The dependent variable is the RMSE for trading days [6,28] (post-filing date market model root mean square error). The regression includes an intercept, calender year dummies, and Fama and French 48-Industry dummies. The results are reported in line with equation 1. The standard errors (reported in parentheses) are clustered by industry and year. ***, ** and * indicate that the coefficient estimate are significantly different from zero at the 1 percent, 5 percent and 10 percent levels respectively.

Table 9 displays the results of the regression in which the putative impact of tone shift is examined on the variable 'standardized unexpected earnings' (SUE) in the presence of other measures of readability—both from the complex words based, and size and vocab based methods. We retain the same

Table 9: Tone shift impact on standardized unexpected earnings (SUE)

| Readability | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| awps | −0.011 | | | | | |
| | (0.011) | | | | | |
| per_CW | 2.418 | | | | | |
| | (2.388) | | | | | |
| Fog_Index | | −0.013 | | | | |
| | | (0.025) | | | | |
| FK_Index | | | −0.002 | | | |
| | | | (0.027) | | | |
| SMOG_Index | | | | −0.019 | | |
| | | | | (0.039) | | |
| vocab | | | | | −0.542* | |
| | | | | | (0.284) | |
| Financial_Term | | | | | | 92.767 |
| | | | | | | (136.213) |
| Tone Shift | 10.054** | 10.901** | 10.594** | 10.891** | 8.330 | 10.573** |
| | (4.763) | (4.849) | (5.115) | (4.825) | (5.630) | (5.109) |
| Control Variables | | | | | | |
| Pre-filing RMSE | −114.841*** | −114.634*** | −114.712*** | −114.633*** | −113.851*** | −114.724*** |
| | (21.941) | (21.932) | (21.942) | (21.930) | (21.879) | (21.913) |
| BM | −0.508*** | −0.510*** | −0.511*** | −0.510*** | −0.492*** | −0.511*** |
| | (0.144) | (0.145) | (0.143) | (0.145) | (0.145) | (0.150) |
| Size | 0.071 | 0.071 | 0.070 | 0.071 | 0.097 | 0.070 |
| | (0.068) | (0.070) | (0.071) | (0.072) | (0.064) | (0.065) |
| Filing Period Return | 3.479** | 3.464** | 3.466** | 3.464** | 3.473** | 3.464** |
| | (1.396) | (1.391) | (1.397) | (1.386) | (1.408) | (1.517) |
| Pre-filing alpha | 356.058*** | 355.862*** | 355.943*** | 355.853*** | 354.623*** | 355.964*** |
| | (46.967) | (47.013) | (47.031) | (47.017) | (47.215) | (47.003) |
| NASDAQ_Dummy | 0.574*** | 0.575*** | 0.576*** | 0.575*** | 0.571*** | 0.576*** |
| | (0.199) | (0.201) | (0.202) | (0.204) | (0.202) | (0.193) |
| Number_of_Analyst | 0.015 | 0.016 | 0.016 | 0.016 | 0.015 | 0.016 |
| | (0.015) | (0.016) | (0.017) | (0.016) | (0.015) | (0.016) |
| Adjusted $R^2$ | 16.30% | 16.30% | 16.30% | 16.30% | 16.30% | 16.30% |
| *N* | 37,787 | 37,787 | 37,787 | 37,787 | 37,787 | 37,787 |

Note: This table presents the results from the regression of SUE on complex-words based as well as vocab and size based readability measures. The dependent variable is the Standardized Unexpected Earnings (SUE). The regression includes an intercept, calender year dummies, and Fama and French 48-Industry dummies. The results are reported in line with equation 1. The standard errors (reported in parentheses) are clustered by industry and year. ***, ** and * indicate that the coefficient estimate are significantly different from zero at the 1 percent, 5 percent and 10 percent levels respectively.

regression specification as in equation (1) but with the addition of one more control: the number of analysts following the firm in line with Loughran and McDonald [2014b].

Again, it is striking to note that similar to benchmark results in table 7 the variable 'tone shift' displays significance over and above other readability measures in all specifications. In other words, in the presence of tone shift, all other prior readability measures, such as 'average words per sentence', '% of complex words', Fog, Flesch-Kincaid and SMOG indices; as well size-based

readability measures such as 'Financial_Term' are all rendered insignificant. Only the LM-based 'vocab' retains its significance in the presence of tone shift and renders it insignificant in one specification.

We report that tone shift is positively associated with standardized unexpected earnings and this result is analogous to those found in Loughran and McDonald [2014b]. This implies that all else equal, an increase in tone shift—i.e., an increase parsing complexity—is associate with higher ambiguity and analysts' uncertainty, leading to higher earnings surprises.

## 5.5 Earnings Predictability with Tone Shift

We also examine if there is any impact of tone shift on future earnings of the firm. Li [2008] suggests that poor readability of 10-Ks lead to lower earnings in future periods. Thus, all else equal, higher levels of tone shift—which imply poorer levels of readability—should depress future earnings. We test this hypothesis for future earnings one and two years ahead in tables 10 and 11 respectively.

$$Earnings_{t+1,k} = a_0 + a_1 Tone\_Shift_{tk} + a_2 Readability\_Measures_{tk} + a_3 Earnings_{tk} + a_4 Controls_{tk} + u_{tk} \quad (2)$$

$$Earnings_{t+2,k} = a_0 + a_1 Tone\_Shift_{tk} + a_2 Readability\_Measures_{tk} + a_3 Earnings_{t+1,k} + a_4 Controls_{tk} + u_{tk} \quad (3)$$

For one year ahead future earnings (table 10), tone shift displays significantly negative coefficients in all model specifications and in its presence, the readability variables 'average words per sentence', Flesh-Kincaid index, the LM-based vocab measure and the 'Financial Term' measure are all rendered insignificant. Among prior readability scores, only the 'percentage of complex words', Fog index, and the SMOG index show significance in the presence of

Table 10: Tone shift impact on next year's earnings (t+1)

| Readability Variables | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| awps | | −0.016 | | | | | |
| | | (0.012) | | | | | |
| per_CW | | −3.786** | | | | | |
| | | (1.785) | | | | | |
| Fog_Index | | | −0.049* | | | | |
| | | | (0.028) | | | | |
| FK_Index | | | | −0.027 | | | |
| | | | | (0.026) | | | |
| SMOG_Index | | | | | −0.085** | | |
| | | | | | (0.043) | | |
| vocab | | | | | | −0.122 | |
| | | | | | | (0.285) | |
| Financial_Term | | | | | | | 95.932 |
| | | | | | | | (110.225) |
| Tone Shift | −9.662** | −7.925** | −8.407** | −9.146** | −8.175** | −10.169** | −9.644** |
| | (3.968) | (3.777) | (3.778) | (3.887) | (3.789) | (4.259) | (3.975) |
| Control variables | | | | | | | |
| Earnings (t) | 0.047*** | 0.047*** | 0.047*** | 0.047*** | 0.047*** | 0.047*** | 0.047*** |
| | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) | (0.006) | (0.005) |
| Pre-filing RMSE | −25.148*** | −24.602*** | −24.766*** | −24.979*** | −24.671*** | −24.919*** | −25.135*** |
| | (2.691) | (2.744) | (2.706) | (2.707) | (2.714) | (2.597) | (2.685) |
| BM | 0.044 | 0.047 | 0.048 | 0.044 | 0.048 | 0.049 | 0.044 |
| | (0.053) | (0.052) | (0.053) | (0.053) | (0.053) | (0.046) | (0.053) |
| Size | 0.419*** | 0.426*** | 0.426*** | 0.422*** | 0.427*** | 0.426*** | 0.420*** |
| | (0.060) | (0.063) | (0.065) | (0.062) | (0.066) | (0.061) | (0.060) |
| Filing Period Return | 0.703** | 0.693** | 0.701** | 0.702** | 0.698** | 0.707** | 0.702** |
| | (0.278) | (0.289) | (0.277) | (0.279) | (0.279) | (0.280) | (0.279) |
| Pre-filing alpha | 79.490*** | 78.951*** | 79.088*** | 79.241*** | 78.956*** | 79.309*** | 79.476*** |
| | (16.562) | (16.617) | (16.564) | (16.574) | (16.568) | (16.446) | (16.556) |
| NASDAQ_Dummy | −0.146* | −0.148* | −0.149* | −0.147* | −0.149* | −0.147* | −0.146* |
| | (0.081) | (0.082) | (0.081) | (0.083) | (0.080) | (0.081) | (0.082) |
| Number_of_Analyst | −0.001 | −0.0002 | −0.0004 | −0.001 | −0.0003 | −0.001 | −0.001 |
| | (0.013) | (0.013) | (0.013) | (0.014) | (0.013) | (0.013) | (0.013) |
| Adjusted $R^2$ | 27.90% | 27.90% | 27.90% | 27.90% | 27.90% | 27.90% | 27.90% |
| $N$ | 30,216 | 30,216 | 30,216 | 30,216 | 30,216 | 30,216 | 30,216 |

Note: This table presents the results from the regression of one-year ahead earnings on complex-words based as well as vocab and size based readability measures. The dependent variable is the earnings for the year t+1. The regression includes an intercept, calender year dummies, and Fama and French 48-Industry dummies. The results are reported in line with equation 1. The standard errors (reported in parentheses) are clustered by industry and year. ***, ** and * indicate that the coefficient estimate are significantly different from zero at the 1 percent, 5 percent and 10 percent levels respectively.

tone shift. The uniformly negative coefficient of tone shift suggests that all else equal, an increase in tone shift, which leads to poorer MD&A readability,

Table 11: Tone shift impact on earnings two years ahead (t+2)

| Readability Variables | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| awps | | −0.008 | | | | | |
| | | (0.014) | | | | | |
| per_CW | | −3.453 | | | | | |
| | | (2.549) | | | | | |
| Fog_Index | | | −0.030 | | | | |
| | | | (0.036) | | | | |
| FK_Index | | | | −0.033 | | | |
| | | | | (0.031) | | | |
| SMOG_Index | | | | | −0.056 | | |
| | | | | | (0.058) | | |
| vocab | | | | | | 0.024 | |
| | | | | | | (0.260) | |
| Financial_Term | | | | | | | 191.804 |
| | | | | | | | (128.074) |
| Tone Shift | −9.577** | −8.178** | −8.801** | −8.895** | −8.597** | −9.474** | −9.550** |
| | (4.237) | (3.626) | (3.831) | (4.040) | (3.819) | (4.212) | (4.243) |
| Control variables | | | | | | | |
| Earnings (t+1) | 0.062 | 0.061 | 0.062 | 0.062 | 0.061 | 0.062 | 0.062 |
| | (0.614) | (0.614) | (0.614) | (0.614) | (0.614) | (0.614) | (0.614) |
| Pre-filing RMSE | −18.926 | −18.504 | −18.717 | −18.738 | −18.647 | −18.972 | −18.909 |
| | (13.119) | (12.826) | (12.955) | (13.036) | (12.911) | (13.118) | (13.112) |
| BM | 0.134* | 0.136* | 0.137* | 0.134* | 0.137* | 0.133** | 0.134* |
| | (0.070) | (0.070) | (0.071) | (0.070) | (0.071) | (0.067) | (0.070) |
| Size | 0.434* | 0.439* | 0.439* | 0.437* | 0.439* | 0.433* | 0.435* |
| | (0.230) | (0.234) | (0.234) | (0.232) | (0.235) | (0.229) | (0.230) |
| Filing Period Return | 0.669 | 0.661 | 0.669 | 0.669 | 0.667 | 0.668 | 0.668 |
| | (0.489) | (0.483) | (0.488) | (0.488) | (0.487) | (0.490) | (0.492) |
| Pre-filing alpha | 66.455 | 66.183 | 66.267 | 66.198 | 66.198 | 66.489 | 66.432 |
| | (41.306) | (41.089) | (41.131) | (41.240) | (41.070) | (41.295) | (41.313) |
| NASDAQ_Dummy | −0.189 | −0.191 | −0.191 | −0.191 | −0.192 | −0.189 | −0.190 |
| | (0.155) | (0.156) | (0.156) | (0.156) | (0.155) | (0.156) | (0.157) |
| Number_of_Analyst | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.0005 |
| | (0.017) | (0.017) | (0.017) | (0.018) | (0.017) | (0.017) | (0.018) |
| Adjusted $R^2$ | 10.0% | 10.0% | 10.0% | 10.0% | 10.0% | 10.0% | 10.0% |
| *N* | 24,647 | 24,647 | 24,647 | 24,647 | 24,647 | 24,647 | 24,647 |

Note: This table presents the results from the regression of two-year ahead earnings on complex-words based as well as vocab and size based readability measures. The dependent variable is the earnings for the year t+2. The regression includes an intercept, calender year dummies, and Fama and French 48-Industry dummies. The results are reported in line with equation 1. The standard errors (reported in parentheses) are clustered by industry and year. ***, ** and * indicate that the coefficient estimate are significantly different from zero at the 1 percent, 5 percent and 10 percent levels respectively.

leads to a fall in earnings one year in the future.

For two year ahead future earnings (table 11), tone shift displays significantly negative coefficients in all model specifications and in its presence, all readability variables are uniformly rendered insignificant i.e., except tone shift, no other prior readability measure has any significant predictive association with future earnings two years ahead. The persistently negative coefficient of tone shift suggests that all else equal, an increase in tone shift,

which leads to poorer readability, leads to a fall in earnings two years into the future.

## 5.6 Robustness

We conduct a variety of auxiliary tests to ensure robustness of our results. We add business complexity of firms as an additional control, test for the significance of tone shift using a smaller set of valence shifters, use modified dictionaries for financial texts' tone calculations; and finally, use only adversative conjunctions and negators to test the impact of tone shift on RMSE.

### 5.6.1 Business Complexity

A financial document's readability (or lack thereof) can be influenced by two factors: i) operational complexity (ontological explanation); and ii) deliberate obfuscation on part of the firm's executives (opportunistic explanation) [Bloomfield, 2008]. For example, Loughran and McDonald [2014b] discuss how readability of 10-Ks is intertwined with firms' complexity. Perhaps high-complexity firms necessarily need to use more complex language in their 10-K filings, and hence their financial disclosures' unreadability may not be motivated by obfuscation. Hence to account for this aspect, we introduce 'business complexity' as an additional control variable [Jennings et al., 2014, Loughran and McDonald, 2014b]. The measure is calculated as the sum of the squared business segment proportions as reported for the firm in the COMPUSTAT Segment database. For our sample, the value for business complexity ranges from 0.11 to 1.00 with lower values implying more firm-specific complexity.

Tone shift shows a significantly positive coefficient for all model specifications and in its presence other readability measures such as Financial Term, Fog, Flesh-Kincaid and SMOG indices, average words per sentence and percentage of complex words are rendered insignificant. Only the LM based vocab measure and log of words retain their significance in the presence

Table 12: Impact of tone shift on RMSE controlling for business complexity

| Readability | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Vocab | | | | | $0.005^{***}$ | $0.005^{***}$ |
| | | | | | (0.002) | (0.002) |
| Financial_Term | | | | | | $-0.651$ |
| | | | | | | (0.725) |
| Log Words | $0.0007^{*}$ | 0.0006 | 0.0006 | 0.0006 | $-0.0007^{*}$ | $-0.0007^{*}$ |
| | (0.0004) | (0.0004) | (0.0004) | (0.0004) | (0.0006) | (0.0006) |
| Fog_Index | | 0.0000 | | | | |
| | | (0.0001) | | | | |
| FK_Index | | | 0.0001 | | | |
| | | | (0.0001) | | | |
| SMOG_Index | | | | 0.0000 | | |
| | | | | (0.0001) | | |
| awps | 0.0000 | | | | $-0.0000$ | 0.0000 |
| | (0.0000) | | | | (0.0000) | (0.0000) |
| per_CW | 0.010 | | | | 0.009 | 0.008 |
| | (0.009) | | | | (0.009) | (0.009) |
| Tone Shift | $0.058^{***}$ | $0.060^{***}$ | $0.059^{***}$ | $0.060^{***}$ | $0.052^{**}$ | $0.052^{**}$ |
| | (0.020) | (0.021) | (0.021) | (0.021) | (0.020) | (0.020) |
| Control Variables | | | | | | |
| Business Complexity | $-0.0002$ | $-0.0002$ | $-0.0002$ | $-0.0002$ | $-0.0000$ | $-0.0000$ |
| | (0.0005) | (0.0004) | (0.0005) | (0.0004) | (0.0005) | (0.0005) |
| Pre-filing alpha | $-1.531^{***}$ | $-1.532^{***}$ | $-1.531^{***}$ | $-1.532^{***}$ | $-1.532^{***}$ | $-1.532^{***}$ |
| | (0.322) | (0.323) | (0.322) | (0.323) | (0.322) | (0.322) |
| Pre-filing RMSE | $0.549^{***}$ | $0.549^{***}$ | $0.549^{***}$ | $0.549^{***}$ | $0.548^{***}$ | $0.548^{***}$ |
| | (0.054) | (0.054) | (0.054) | (0.054) | (0.054) | (0.054) |
| Filing Period Return | $-0.020^{***}$ | $-0.020^{***}$ | $-0.020^{***}$ | $-0.020^{***}$ | $-0.020^{**}$ | $-0.020^{**}$ |
| | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) |
| Size | $-0.001^{***}$ | $-0.0016^{***}$ | $-0.0017^{***}$ | $-0.0016^{***}$ | $-0.0017^{***}$ | $-0.0017^{***}$ |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| BM | $-0.001^{***}$ | $-0.0016^{***}$ | $-0.0016^{***}$ | $-0.0016^{***}$ | $-0.0017^{***}$ | $-0.0017^{***}$ |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| NASDAQ Dummy | $0.002^{***}$ | $0.002^{***}$ | $0.0020^{***}$ | $0.002^{***}$ | $0.002^{***}$ | $0.002^{***}$ |
| | (0.0007) | (0.0007) | (0.0007) | (0.0007) | (0.0007) | (0.0007) |
| Adjusted $R^2$ | 43.95% | 43.95% | 43.95% | 43.95% | 43.98% | 43.98% |
| $N$ | 42,356 | 42,357 | 42,357 | 42,357 | 42,355 | 42,354 |

Note: This table presents the results from the regression of RMSE on complex-words based as well as vocab and size based readability measures along with business complexity as an additional control. The dependent variable is the RMSE for trading days [6,28] (post-filing date market model root mean square error). The regression includes an intercept, calender year dummies, and Fama and French 48-Industry dummies. The results are reported in line with equation 1. The standard errors (reported in parentheses) are clustered by industry and year. ***, ** and * indicate that the coefficient estimate are significantly different from zero at the 1 percent, 5 percent and 10 percent levels respectively.

of tone shift. The persistently positive values of tone shift suggest that all else equal, higher levels of tone shift, i.e., lower levels of readability lead to increased values of RMSE. Business complexity is almost 0 in all regression specifications and does not exhibit any significance in our results.

### 5.6.2 Limited Valence Shifters

Valence shifters play a major role in tone shift calculations. In this section, we focus on only the least frequently used valence shifters to allay concerns that over-representation of a few valences shifters could skew our results. To rule this out, we select only those valence shifters which are individually

used fewer than 2% of the time in the 10-K reports. There are a total of 78 different valence shifters used in the MD&A section for our sample of US firms.[5] However, out of these 78 words, 8 are used with very high frequency ("certain", "significant", "more", "but", "however", "most", "significantly" and "only"). Taken together, these 8 words constitute 36% of the total usage of the valence shifters in our study. We remove the high-frequency valence shifters, recalculate the tone shift and test its impact on RMSE in table 13.

As with other benchmark tables, tone shift displays uniform, persistence significance across all models and readability controls. Except for the LM based variable 'vocab', all other readability measures—whether from 'complex words based', FOG/FK/SMOG indices, or vocab based ('Financial_Term')—are rendered insignificant in the presence of tone shift. All this suggests that even with the least frequently used valence shifters, tone shift is large enough to positively impact firms' idiosyncratic volatility. Further, the uniformly positive sign of tone shift indicates that increased unreadability in terms of tone shift leads to higher RMSE for firms.

### 5.6.3 Modified dictionaries

Certain verb-noun combinations such as 'increased profits' or 'decreased stability' etc. cannot be granted weights according to the LM dictionary based bag-of-words approach since the verb (e.g., 'decreased') can have either a positive or a negative weight based on the noun following it (e.g., negative for 'decreased confidence' but positive for 'decreased losses' etc.).

To incorporate the connotation of such words in our dictionary, we add these verb-noun combinations to the list of LM polar phrases to constitute a new dictionary. The results for the impact of MD&A tone shift when the underlying dictionary is augmented by the addition of such verb-noun combinations are included in table 14.

Again, similar to the benchmark results, we see that tone shift displays

[5]Tables A.1, and A.2 contain the full list of valence shifters encountered in the MD&A section of the 10-K reports for firms in our sample.

Table 13: Tone shift impact with only the least used valence shifters

| Readability Variables | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| awps | | 0.00001 | | | | | |
| | | (0.00003) | | | | | |
| per_CW | | 0.011 | | | | | |
| | | (0.007) | | | | | |
| Fog_Index | | | 0.0001 | | | | |
| | | | (0.0001) | | | | |
| FK_Index | | | | 0.0001 | | | |
| | | | | (0.0001) | | | |
| SMOG_Index | | | | | 0.0001 | | |
| | | | | | (0.0001) | | |
| vocab | | | | | | $0.002^{**}$ | |
| | | | | | | (0.001) | |
| Financial_Term | | | | | | | $-0.434$ |
| | | | | | | | (0.591) |
| Tone Shift | $0.077^{**}$ | $0.068^{*}$ | $0.074^{**}$ | $0.070^{**}$ | $0.073^{*}$ | $0.083^{**}$ | $0.076^{**}$ |
| | (0.038) | (0.035) | (0.038) | (0.035) | (0.038) | (0.037) | (0.038) |
| Control Variables | | | | | | | |
| Pre-filing RMSE | $0.581^{***}$ | $0.580^{***}$ | $0.580^{***}$ | $0.580^{***}$ | $0.580^{***}$ | $0.578^{***}$ | $0.581^{***}$ |
| | (0.048) | (0.049) | (0.049) | (0.048) | (0.049) | (0.048) | (0.048) |
| BM | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| Size | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ | $-0.001^{***}$ |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| Filing Period Return | $-0.016^{***}$ | $-0.016^{***}$ | $-0.016^{***}$ | $-0.016^{***}$ | $-0.016^{***}$ | $-0.016^{***}$ | $-0.016^{***}$ |
| | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) |
| Pre-filing alpha | $-1.565^{***}$ | $-1.564^{***}$ | $-1.565^{***}$ | $-1.564^{***}$ | $-1.564^{***}$ | $-1.559^{***}$ | $-1.565^{***}$ |
| | (0.310) | (0.310) | (0.310) | (0.310) | (0.310) | (0.310) | (0.310) |
| NASDAQ_Dummy | $0.002^{**}$ | $0.002^{**}$ | $0.002^{**}$ | $0.002^{**}$ | $0.002^{**}$ | $0.002^{**}$ | $0.002^{**}$ |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Adjusted $R^2$ | 0.439 | 0.440 | 0.439 | 0.440 | 0.440 | 0.440 | 0.439 |
| *N* | 57,514 | 57,514 | 57,514 | 57,514 | 57,514 | 57,514 | 57,514 |

Note: This table presents the results from the regression of RMSE on complex-words based as well as vocab and size based readability measures. The dependent variable is the RMSE for trading days [6,28] (post-filing date market model root mean square error). The tone shift is computed using only the least frequently used valence shifters. The regression includes an intercept, calender year dummies, and Fama and French 48-Industry dummies. The results are reported in line with equation 1. The standard errors (reported in parentheses) are clustered by industry and year. ***, ** and * indicate that the coefficient estimate are significantly different from zero at the 1 percent, 5 percent and 10 percent levels respectively.

persistence in its significantly positive association with idiosyncratic volatility of firms over and above those due to current readability measures. Except for log of words and the LM-based 'vocab' measure, all other readability measures such as FOG/FK/SMOG indices, 'average words per sentence' or '% of complex words' are rendered insignificant in the presence of absolute tone shift.

Table 14: Impact of tone shift on RMSE: Modified dictionary [all readability controls]

| Readability Variables | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Vocab | | | | | 0.004*** | 0.004*** |
| | | | | | (0.001) | (0.001) |
| Financial_Term | | | | | | −0.474 |
| | | | | | | (0.584) |
| Log Words | 0.0004* | 0.0004 | 0.0004 | 0.0004 | −0.0008* | −0.0008* |
| | (0.0002) | (0.0002) | (0.0002) | (0.0002) | (0.0004) | (0.0004) |
| Fog_Index | | 0.00004 | | | | |
| | | (0.00006) | | | | |
| FK_Index | | | 0.0001 | | | |
| | | | (0.0001) | | | |
| SMOG_Index | | | | 0.0001 | | |
| | | | | (0.0001) | | |
| awps | 0.000003 | | | | −0.000005 | −0.000005 |
| | (0.00003) | | | | (0.00002) | (0.00002) |
| per_CW | 0.012 | | | | 0.010 | 0.010 |
| | (0.008) | | | | (0.008) | (0.008) |
| Tone Shift | 0.044*** | 0.047*** | 0.045*** | 0.046*** | 0.037*** | 0.037*** |
| | (0.014) | (0.016) | (0.014) | (0.015) | (0.014) | (0.014) |
| Control Variables | | | | | | |
| Pre-filing alpha | −1.560*** | −1.561*** | −1.560*** | −1.561*** | −1.559*** | −1.559*** |
| | (0.310) | (0.310) | (0.310) | (0.310) | (0.309) | (0.309) |
| Pre-filing RMSE | 0.579*** | 0.579*** | 0.579*** | 0.579*** | 0.577*** | 0.577*** |
| | (0.048) | (0.048) | (0.048) | (0.048) | (0.048) | (0.048) |
| Filing Period Return | −0.015*** | −0.015*** | −0.015*** | −0.015*** | −0.015** | −0.015** |
| | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) |
| Size | −0.0014*** | −0.0014*** | −0.0014*** | −0.0014*** | −0.0015*** | −0.0015*** |
| | (0.0003) | (0.0002) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| BM | −0.0013*** | −0.0013*** | −0.0013*** | −0.0013*** | −0.0014*** | −0.0014*** |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| NASDAQ Dummy | 0.0017** | 0.0017*** | 0.0017** | 0.0017*** | 0.0017*** | 0.001*** |
| | (0.0006) | (0.0006) | (0.0006) | (0.0006) | (0.0006) | (0.0006) |
| Adjusted $R^2$ | 43.96% | 43.96% | 43.96% | 43.96% | 43.99% | 43.99% |
| $N$ | 57,431 | 57,432 | 57,432 | 57,432 | 57,430 | 57,429 |

Note: This table presents the results from the regression of RMSE on complex-words based as well as vocab and size based readability measures. The dependent variable is the RMSE for trading days [6,28] (post-filing date market model root mean square error). Tone shift is computed using a modified dictionary which imparts weights to verb-noun combinations. The regression includes an intercept, calender year dummies, and Fama and French 48-Industry dummies. The results are reported in line with equation 1. The standard errors (reported in parentheses) are clustered by industry and year. ***, ** and * indicate that the coefficient estimate are significantly different from zero at the 1 percent, 5 percent and 10 percent levels respectively.

### 5.6.4 Negators and adversative conjunctions

We repeat the central idea of our paper: increased usage of more complex and nuanced language—in terms of multi-clausal phrases and/or usage of adverbs and adjectives—can make text harder to parse, leading to increased ambiguity and more investor uncertainty. Our method calculates the incremental contribution of such components to the readability of MD&As of

10-Ks and shows that tone shift has a significantly positive association with firms' idiosyncratic volatility.

Table 15: Impact of negators & adversative conjunction on RMSE [all readability controls]

| Readability Variables | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Vocab | | | | | $0.005^{***}$ | $0.005^{***}$ |
| | | | | | (0.001) | (0.001) |
| Financial_Term | | | | | | $-0.488$ |
| | | | | | | (0.564) |
| Log Words | $0.0004^{*}$ | 0.0004 | 0.0004 | 0.0004 | $-0.0007^{*}$ | $-0.0009^{*}$ |
| | (0.0002) | (0.0002) | (0.0003) | (0.0002) | (0.0004) | (0.0004) |
| Fog_Index | | 0.00003 | | | | |
| | | (0.00006) | | | | |
| FK_Index | | | 0.0001 | | | |
| | | | (0.0001) | | | |
| SMOG_Index | | | | 0.00007 | | |
| | | | | (0.0001) | | |
| awps | $-0.000005$ | | | | $-0.00001$ | $-0.00001$ |
| | (0.00003) | | | | (0.00003) | (0.00002) |
| per_CW | 0.013 | | | | 0.011 | 0.011 |
| | (0.008) | | | | (0.008) | (0.008) |
| Tone Shift | $0.063^{***}$ | $0.065^{***}$ | $0.064^{***}$ | $0.064^{***}$ | $0.056^{**}$ | $0.056^{***}$ |
| | (0.024) | (0.024) | (0.024) | (0.024) | (0.022) | (0.022) |
| Control Variables | | | | | | |
| Pre-filing alpha | $-1.587^{***}$ | $-1.588^{***}$ | $-1.587^{***}$ | $-1.588^{***}$ | $-1.585^{***}$ | $-1.585^{***}$ |
| | (0.309) | (0.309) | (0.309) | (0.309) | (0.308) | (0.308) |
| Pre-filing RMSE | $0.573^{***}$ | $0.573^{***}$ | $0.573^{***}$ | $0.573^{***}$ | $0.572^{***}$ | $0.571^{***}$ |
| | (0.051) | (0.051) | (0.051) | (0.051) | (0.051) | (0.051) |
| Filing Period Return | $-0.015^{***}$ | $-0.015^{***}$ | $-0.015^{***}$ | $-0.015^{***}$ | $-0.015^{**}$ | $-0.015^{**}$ |
| | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) |
| Size | $-0.0014^{***}$ | $-0.0014^{***}$ | $-0.0014^{***}$ | $-0.0014^{***}$ | $-0.0015^{***}$ | $-0.0015^{***}$ |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| BM | $-0.0013^{***}$ | $-0.0013^{***}$ | $-0.0013^{***}$ | $-0.0013^{***}$ | $-0.0014^{***}$ | $-0.0014^{***}$ |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| NASDAQ Dummy | $0.0017^{**}$ | $0.0017^{***}$ | $0.0017^{**}$ | $0.0017^{***}$ | $0.0017^{***}$ | $0.001^{**}$ |
| | (0.0007) | (0.0006) | (0.0007) | (0.0006) | (0.0007) | (0.0007) |
| Adjusted $R^2$ | 43.60% | 43.59% | 43.60% | 43.59% | 43.63% | 43.63% |
| $N$ | 47,985 | 47,986 | 47,986 | 47,986 | 47,984 | 47,983 |

Note: This table presents the results from the regression of RMSE on complex-words based as well as vocab and size based readability measures. The dependent variable is the RMSE for trading days [6,28] (post-filing date market model root mean square error). Tone shift is calculated using only two special subclasses: adversative conjunctions, and negators. The regression includes an intercept, calender year dummies, and Fama and French 48-Industry dummies. The results are reported in line with equation 1. The standard errors (reported in parentheses) are clustered by industry and year. ***, ** and * indicate that the coefficient estimate are significantly different from zero at the 1 percent, 5 percent and 10 percent levels respectively.

However, a natural counter to our approach is as follows. Not all adjectives and adverbs make text more complex to read. Why should all categories of valence shifters—negators ('not'), adversative conjunctions ('despite'), amplifiers ('intensely') and de-amplifiers ('faintly')—be weighed the same? Perhaps only a subset of such valence shifters contributes to the readability (or lack thereof) of firms' financial disclosures.

To allay such concerns we isolate two components of valence shifters: negators and adversative conjunctions. These two special categories of va-

lence shifters alter the sign of the tone during the text analysis/parsing procedure. We use only these two special categories of valence shifters to reconstruct the tone shift. The results of our analysis are included in table 15.

The modified tone shift displays the same relation to RMSE as all our prior benchmark results and displays significantly positive associations over and above those of other readability measures. Except for log of words and the LM-based 'vocab' measure, all other readability measures such as FOG/FK/SMOG indices, 'average words per sentence' or '% of complex words' are rendered insignificant in the presence of absolute tone shift. Further, the uniformly positive values of tone shift suggests that all else equal, as the tone shift—and hence the unreadability—rises, the RMSE increases.

# 6 Readability over the years

Has readability of financial text, in particular, the 10-K reports and its MD&A section, become worse over time? Or has it improved? We attempt an answer to this question by analyzing trends in the readability of firms' 10-K based on tone shift and compare it to that calculated according to other popular measures. Before discussing the results, however, we briefly provide a background for an important policy decision taken by the SEC in October 1998 regarding financial texts' readability.

## 6.1 SEC Plain English Rule (October 1998)

In October 1998, the SEC implemented a rule which stipulated that firms should use 'plain English' in all communications with its shareholders. The SEC classified components of plain English in the following six categories: 'average sentence length', 'average word length', 'passive voice', 'legalese', 'personal pronouns', and 'negative/superfluous phrases'. While the rule officially applied only to prospectus filings, the SEC stated its clear preference for usage of plain English in all communication with shareholders [Loughran

and McDonald, 2014a].

In the discussion that follows, we evaluate the impact, if any, of the Plain English Rule.

## 6.2 MD&A readability over time

In this section, based on the management discussion and analysis section of 10-K reports, we compute the median trends, as well as the entire yearly distribution of firms' tone shift from 1994 to 2018, evaluate whether the plain English rule had any impact on tone shift; and compare it to distributions and trends in other popular measures of readability.

### 6.2.1 Tone shift trend

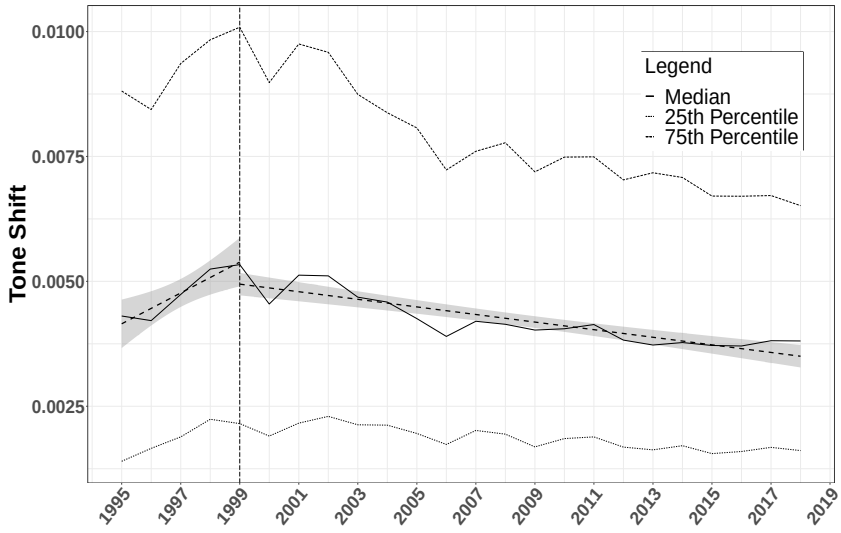


Figure 1: The plot presents the movement of tone shift in 1994–2018 for the median firm (quantile 0.50), the median high tone shift firm (quantile 0.75); and the median low tone shift firm (quantile 0.25). The dashed vertical line in 1999 denotes the implementation of the SEC Plain English rule (October 1998). The grey band around the trend denotes the 95% confidence interval.

Figure 1 presents the yearly time series of the median, the $75^{\text{th}}$ percentile, and the $25^{\text{th}}$ percentile of tone shift. From visual inspection we see that for

the median firm, its tone shift has fallen over the years and the time series exhibits a negative trend, especially post-1999, after the implementation of the SEC Plain English rule. Prior to the introduction of the rule, the tone shift for the median firm shows an increase, i.e., a positive trend in 1994–1999 which is immediately arrested by the imposition of the SEC rule. Further, the median high tone shift firm ($75^{\text{th}}$ percentile) displays the same behavior: increasing tone shift prior to the plain English rule, and falling levels with a negative trend after 1999. In particular, the trend is steeper for the median high tone shift firm than for the median firm, suggesting that the rule impacted firms with high tone shifts more strongly than those with low tone shift values. This hypothesis finds more evidence in its favor when we observe the yearly time series of the median low tone shift firm ($25^{\text{th}}$ percentile) for which there is no major change over the years. In other words, firms which did not use overly complex language in its MD&A section are not impacted by the SEC Plain English Rule but firms which featured nuanced, hard-to-parse language prior to the imposition of the rule seem to have changed their style so as to conform to SEC's initiative.

High levels of tone shift are accompanied with high incidence of multi-clausal phrases as well as adjectives and adverbs, which indicates more parsing complexity in texts and hence low levels of readability. Negative trends and falling tone shift levels for the median firm and the median high tone shift firm are indicators of improvement in the MD&A readability after the SEC mandated move to plain English.

To confirm this behavior more formally, we resort to calculating linear time trends in tone shift for all firms in our sample which have 4 or more years of tone shift data and present the results in table 16. The table shows that before 1999 i.e., up to December 1998, out of a sample of 678 firms, 381 (56%) displayed a positive trend—increasing tone shift—at the significance level of 10%. However, after the implementation of the SEC rule post-1999, out of the total sample of 2491 firms, only 1056 (42%) display increasing levels of tone shift. Similarly, before 1999, there are 297 firms (43% of 678)

which show a negative trend in tone shift. But this number increases to 1435 (58% of 2491 firms) after the SEC Plain English rule comes into effect.

Table 16: Tone shift time trends

| | **Pre 1999** | | | **Post 1999** | | |
|---|---|---|---|---|---|---|
| Tone Shift | Number | % | Total | Number | % | Total |
| Firms with pos trend | 381 | 56.19% | 678 | 1056 | 42.39% | 2491 |
| Firms with neg trend | 297 | 43.80% | 678 | 1435 | 57.56% | 2491 |
| Smallest firms (bottom quartile) | | | | | | |
| Firms with pos trend | 196 | 55.36% | 354 | 270 | 39.88% | 677 |
| Firms with neg trend | 158 | 44.63% | 354 | 407 | 60.11% | 677 |
| Largest firms (top quartile) | | | | | | |
| Firms with pos trend | 32 | 48.48% | 66 | 225 | 45.09% | 499 |
| Firms with neg trend | 34 | 51.51% | 66 | 274 | 54.90% | 499 |

Note: This table presents the trends in tone shift significant at the 10% level pre- and post-1999. To be included in this sample a firm must have at least 4 years of data.

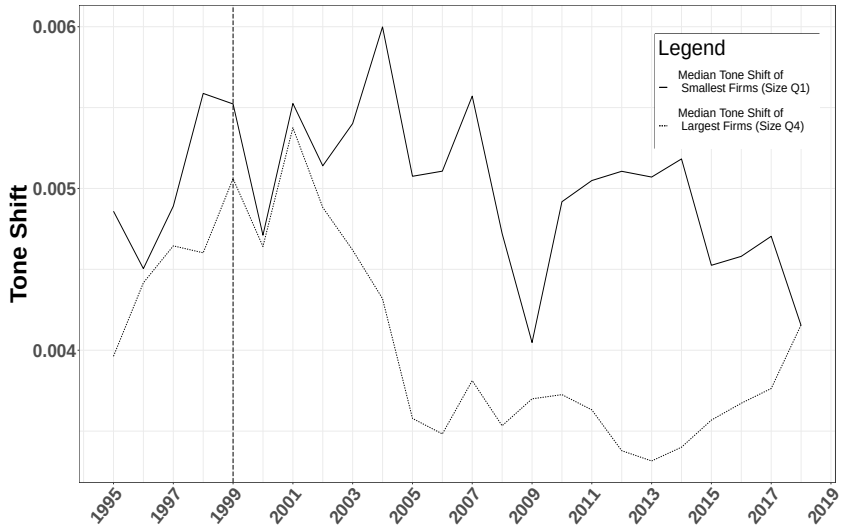


Figure 2: The plot presents the movement of tone shift in 1994–2018 for the median large firm (size quantile 0.75); and that for the median small firm (size quantile 0.25). The dashed vertical line in 1999 denotes the implementation of the SEC Plain English rule (October 1998).

In figure 2 we present trends in the tone shift of firms based on their size (market equity). We plot the median small firm's ($25^{\text{th}}$ percentile) and the

median large firm's ($75^{\text{th}}$ percentile) tone shift over 1994–2018. Both the median small and large firms show positive trends (increasing levels) in tone shift from 1994–1998, but this trend reverses after the implementation of the plain English rule. The median large firm displays the peak in its tone shift in 2001 after which it exhibits a steep negative trend, while the median small firm displays its peak in the year 2004 after which its tone shift values start falling. In particular, both firms exhibit a negative trend in their tone shift after 1999, especially the median large firm which indicates that the SEC rule impacts large firms more than it does smaller firms.

This is also borne out by table 16 in which we calculate linear time trends for all firms in the top quartile ($\geq 0.75$ quantile) and the bottom quartile ($\leq 0.25$ quantile) of size—both before and after 1999—and compare the number of firms with significant positive and/or negative trends at the 90% confidence level.[6] Before 1999, there are 354 small firms, out of which, 196 (55%) show increasing tone shift levels; but this number becomes 270, out of a total of 677 firms (40%) post-1999. Similarly, 158 small firms exhibit a negative trend pre-1999 (45% of 354) but after the SEC plain English rule, the number of small firms with falling tone shift values becomes 407 (60% of 677). The same behavior can be observed for large firms (top quartile by market equity). Pre-1999, there are 66 large firms, out of which 32 (48%) show significant positive trends, but this reverses post-1999 when out of 499 large firms, 225 (45%) show increasing tone shift levels. Similarly, pre-1999 34 large firms show a significantly negative trend (52% of 66) and this rises post-1999 to 274 firms (55% of 499).

### 6.2.2 Distribution of MD&A readability over time

In this section, we discuss the behavior of the full distribution of readability measures over the years, with a special emphasis on the impact of the SEC Plain English Rule. We compute yearly boxplots of all readability measures in this study and compare their evolution over time. In order to facilitate

[6] Or equivalently, at the 10% significance level.

such a comparison and to evaluate the effect of the SEC rule, we stipulate that only those firms be included which have available observations five years prior to, and five years after the imposition of the SEC rule in 1999.[7]

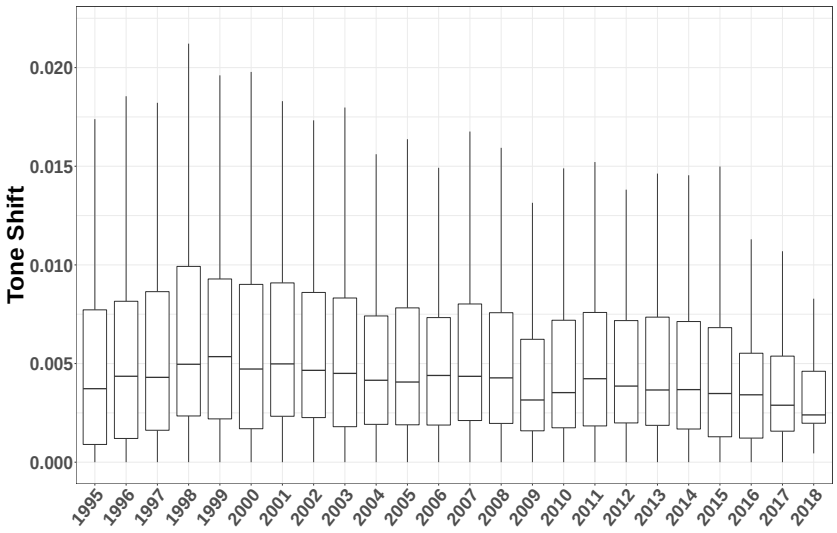


Figure 3: The boxplots of yearly MD&A tone shift distribution.

Figure 3 presents boxplots of tone shift each year from 1995 to 2018. It confirms the main finding of figure 1: i) medians rise prior to 1999 and then fall thereafter owing to the SEC initiative, and ii) the 75$^{\text{th}}$ percentiles rise during 1995–1998 then fall rapidly after 1999, even more so than the corresponding median levels.

However, the full distribution of yearly tone shift is even more informative. Not only do the median and the 75$^{\text{th}}$ percentile decrease over time, but so do the maximum values. Further, we observe that the body of the tone shift distribution has progressively shrunk over time; and the range has become more compressed as well. The shrinkage of the range (max - min) and the body (75$^{\text{th}}$-25$^{\text{th}}$ percentile) are positive signals which indicate that the progressive decrease in the tone shift is not isolated to a few select firms but encompasses the entire collection of firms in the US. To the extent that lower

[7]To focus on the median behavior and to preserve visual comparability we ignore outliers in plotting the full distribution.

levels of tone shift indicate high readability and low text-parsing complexity, it indicates that the MD&A section has become more readable over time, especially after the imposition of the SEC rule. We note that our results are in agreement with those reported in Loughran and McDonald [2014a].

Figure 4 presents boxplots of the 'average words per sentence' readability measure each year from 1995 to 2018. From visual inspection, we observe that over time, the median average words per sentence in the MD&A section has increased from around 25 words per sentence, to about 30.

The SEC rule imposition in 1999 seems to not have reduced the average words per sentence measure. In fact, the effect appears to have had been the opposite: the medians tend to rise after 1999 and continue their upward trajectory over the years till the end of the sample in 2018. The exact same behavior is manifested for the $75^{\text{th}}$ and the $25^{\text{th}}$ percentile—their levels decrease during 1995–1999 but tend to increase thereafter till 2018. Further, while it seems that the range has decreased somewhat after 1999, the body of the distribution shows no major change from year to year. Insofar as more words per sentence make parsing of the text more complex, and hence lead to poor readability, this suggests that over time, the MD&A readability—

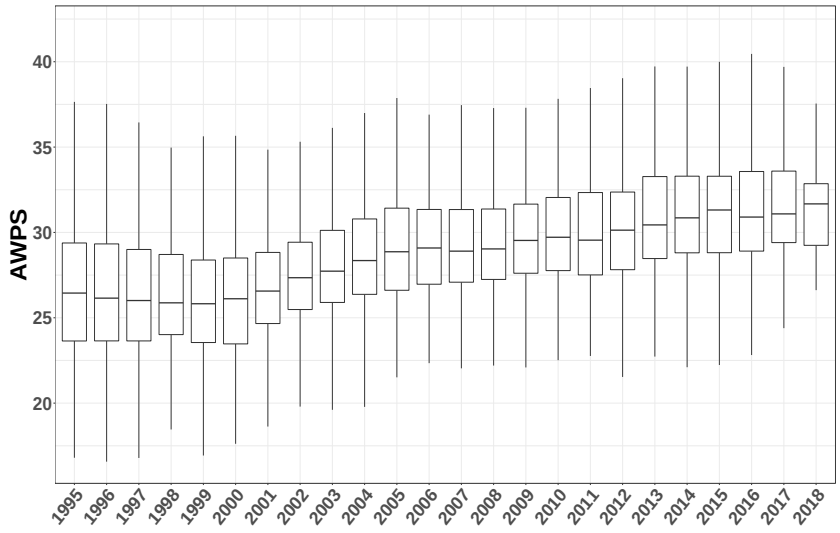


Figure 4: The boxplots of yearly MD&A 'average words per sentence' distribution.

in terms of average words per sentence—has become lower over time. In particular, we conclude that the SEC Plain English Rule has not led to a decrease in the average words per sentence which in fact, has increased over the years.

Figure 5 presents boxplots of the readability measure 'Percentage of complex words' each year from 1995 to 2018 for the MD&A section of firms' 10-K reports. From a cursory glance we can observe that this measure has not moved much over time from its initial levels at the beginning of our sample in 1995.

The median levels show a small increase over time, while there is no significant change in either the full range of the distribution or its body, except for the year 2018 where both the range and the body show significant compression. To the extent that a higher percentage of complex words signifies poor readability, we are led to believe that the SEC rule has not had an appreciable impact on this metric over the years.

Figure 6 presents yearly boxplots of the distribution of the (natural) log of words for the MD&A section over the years.[8] Visual inspection leads us to conclude that the log of words—and hence the total number of words—of the MD&A section has sharply increased over time. This is true, in particular, for the medians which show strong growth in 2001–2004 after which the rise becomes smaller.

Higher levels of this readability metric imply lengthier MD&A sections for the readers. All else equal, a shorter MD&A section is more readable and hence from this perspective the increasing length of the MD&A section over the years should denote lower readability. The figure makes it clear that the SEC rule imposition in 1999 has not made the MD&A section smaller and in fact this section has tended to become much more verbose over time—from around $e^8 \approx 3000$ words in 1995 to about $e^9 \approx 9000$ words in 2018.

[8]For example, if the median log of words equals 8 it implies that the median MD&A section has $e^8 \approx 3000$ words.

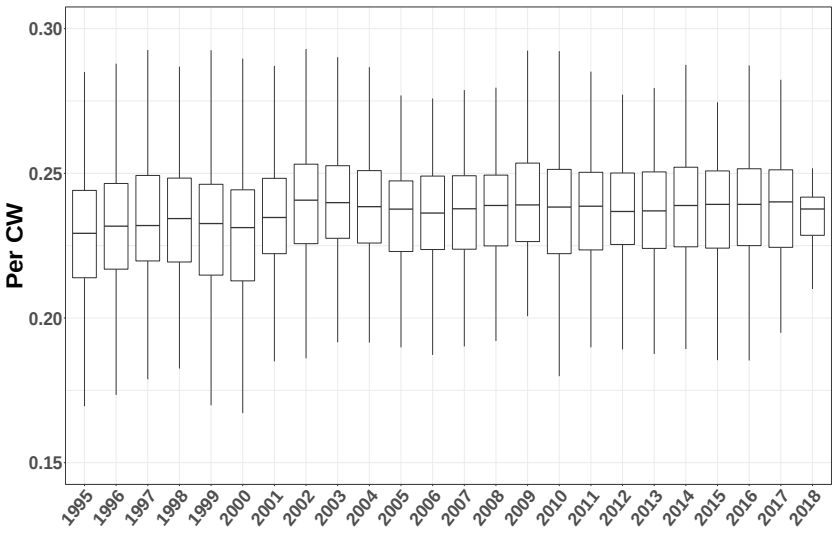


Figure 5: The boxplots of yearly MD&A 'percentage of complex words' distribution.

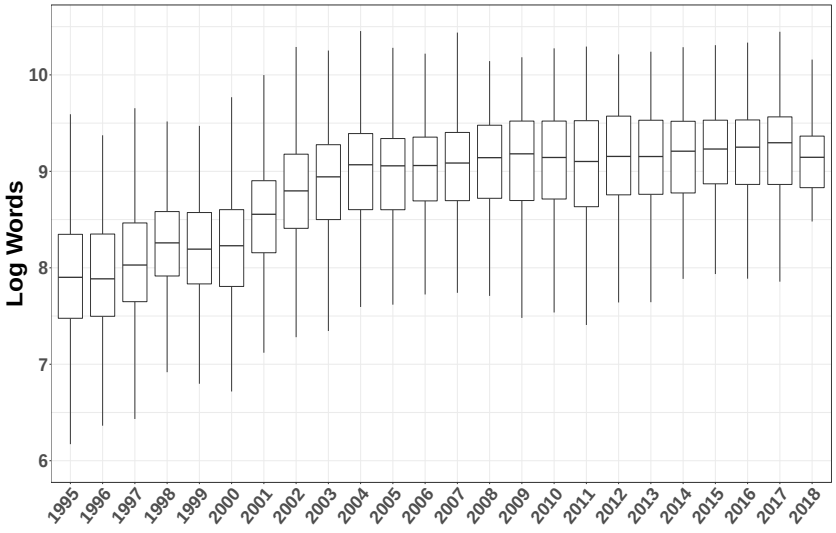


Figure 6: The boxplots of yearly MD&A log of words distribution.

### 6.2.3 Trends in other readability measures

In this section, we compare linear time trends in MD&A readability over the years. Of special interest is the effect, if any, of the SEC plain English rule.

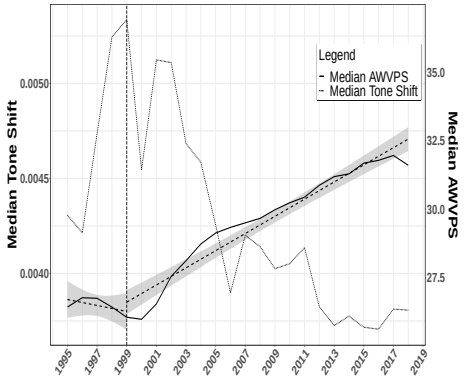


Figure 7: The plot presents the movement of tone shift in 1994–2018 for the median firm. The dashed vertical line in 1999 denotes the implementation of the SEC Plain English rule (October 1998). The grey band around the trend denotes the 95% confidence interval. The solid line denotes the median 'average words per sentence' (AWPS).

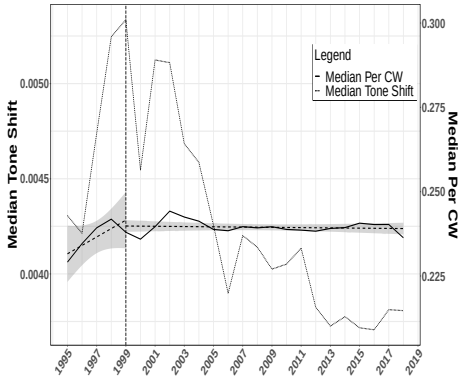


Figure 8: The plot presents the movement of tone shift in 1994–2018 for the median firm. The dashed vertical line in 1999 denotes the implementation of the SEC Plain English rule (October 1998). The grey band around the trend denotes the 95% confidence interval. The solid line denotes the median '% of complex words' (Per_CW).

In figure 7, we compare the yearly time series of median tone shift to yearly values of the median 'average words per sentence' readability measure. One important contrast between the behavior of the two time series, especially after taking into account the implementation of the SEC Plain English rule, is that while the median tone shift has continued to fall, the median level of the 'average words per sentence' exhibits the opposite behavior—negative trends prior to 1999 and positive trends after 1999. In particular, this implies that the median firm's average words per sentence in its MD&A section showed falling levels before the plain English rule but started displaying increasing levels (significantly positive trend) after its implementation. Insofar as high values of 'average words per sentence' show more complexity

and low readability, the plain English rule seems to have had the effect of increasing unreadability of the MD&A section. This behavior is opposite to that for tone shift which has shown negative trends after the SEC rule implementation which suggests improving MD&A readability after 1999.

Similarly, in figure 8, we compare the time series of yearly median tone shift to yearly median 'percentage of complex words' in the MD&A section. The median time series of percentage of complex words shows an increasing trend from 1995–1999 which is arrested by the SEC Plain English rule in 1999 after which the percentage of complex words in the MD&A section becomes almost flat. To the extent that more complex words make the text unreadable, the SEC rule appears to have had a restraining effect and has helped maintain readability levels to what they were in 1999.

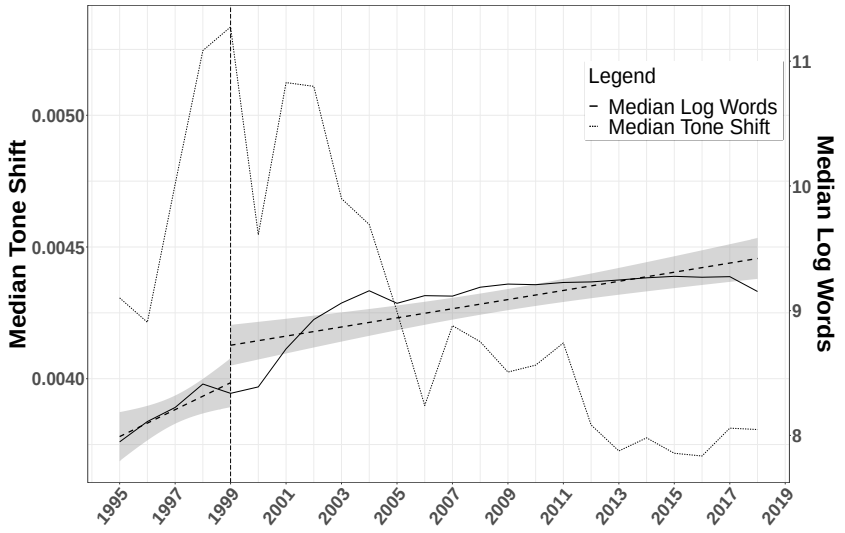


Figure 9: The plot presents the movement of tone shift in 1994–2018 for the median firm. The dashed vertical line in 1999 denotes the implementation of the SEC Plain English rule (October 1998). The grey band around the trend denotes the 95% confidence interval. The solid line denotes the median of log of words in the MD&A section.

Finally, in figure 9 we plot together the median tone shift and the median (natural) log of words for the MD&A section. The median log words shows a positive trend during 1995–1999, as well as during 1999–2018. The imposition of the SEC rule, however, appears to have damped the rate of increase of

the MD&A section's verbosity since the steepness of the positive trend falls after 1999.

The figures presented in this section indicate that the median firm's tone shift has responded to the SEC plain English rule in the way it was supposed to—by making the MD&A section more readable—as signified by its falling levels after 1999. However, it has had no effect on other measures of MD&A readability—average words per sentence, percentage of complex words and log of words—since their levels continue to increase over time. For the metric 'percentage of complex words', the SEC rule seems to have arrested the trend of increasing unreadability; and for the metric 'log of words' it seems to have reduced the high rate of unreadability after its imposition in 1999. However, for the metric 'average words per sentence', the imposition of the rule appears to have had the opposite effect: high levels of readability prior to 1999 begin to give way to a steep rise in unreadability, especially after the year 2000.

Table 17: Impact of the SEC plain English rule on mean readability

| Readability measure | Name of the test | Alt:$\mathbb{H}_1$ | $p$-value |
|---|---|---|---|
| Tone Shift (MD&A) | Welch test | Smaller post-99 | 0 |
| | Wilcoxon test | Smaller post-99 | 0.02 |
| AWPS (MD&A) | Welch test | Greater post-99 | 0 |
| | Wilcoxon test | Greater post-99 | 0 |
| Per_CW (MD&A) | Welch test | Greater post-99 | 0 |
| | Wilcoxon test | Greater post-99 | 0 |
| Log Words (MD&A) | Welch test | Greater post-99 | 0 |
| | Wilcoxon test | Greater post-99 | 0 |
| Tone Shift (10-K) | Welch test | Smaller post-99 | 0 |
| | Wilcoxon test | Smaller post-99 | 0 |
| Log Words (10-K) | Welch test | Greater post-99 | 0 |
| | Wilcoxon test | Greater post-99 | 0 |
| Netfilesize (10-K) | Welch test | Greater post-99 | 0 |
| | Wilcoxon test | Greater post-99 | 0 |
| Grossfilesize (10-K) | Welch test | Greater post-99 | 0 |
| | Wilcoxon test | Greater post-99 | 0 |

Note: This table tests for differences in mean readability by means of the Welch $T$ test and the Wilcoxon test pre- and post-1999. The null hypothesis is that the mean readability has not changed due to the introduction of the SEC plain English rule in 1999. 'AWPS' denotes 'average words per sentence', 'Per_CW' denotes 'percent of complex words'.

To verify this behavior more formally, we conduct statistical tests for the equality of means before and after the imposition of the SEC plain English rule, the results for which are presented in table 17. The null hypothesis is that the mean readability of the MD&A section has not changed due to the introduction of the SEC plain English rule in 1999.

For tone shift, the alternative hypothesis is that its mean is lower post-1999. For the Welch $T$-test, the $p$-value is 0, while that for the nonparametric Wilcoxon rank-sum test is 0.02, leading us to reject the null hypothesis of equality of means pre- and post-1999 in favor of the alternative hypothesis of lower mean tone shift after the SEC plain English rule introduction. Similarly, for the readability measures 'average words per sentence', 'percentage of complex words' and 'log of words', since the $p$-values—for both the Welch and the Wilcoxon test—are 0, we can confidently reject the null hypothesis of equality of means, in favor of the alternative hypothesis which states that the means are higher post-1999.

## 6.3 10-K readability over time

In this section we examine the readability of US firms' 10-K reports over the years according to tone shift and that computed using other popular measures of readability.

### 6.3.1 Distribution of 10-K readability over the years

Figure 10 presents the yearly boxplots of the tone shift for US firms' 10-Ks over the years 1995–2018. As visual inspection makes it clear, the median tone shift for the 10-K as a whole appears to have decreased steadily, especially after the introduction of the SEC plain English rule in 1999. Further, the range and the body of the yearly tone shift distribution also seem to have become more compact over the years—especially for the year 2018. All this is consistent with our previous finding of a reduction in the median, as well as range and body of the MD&A tone shift distribution.

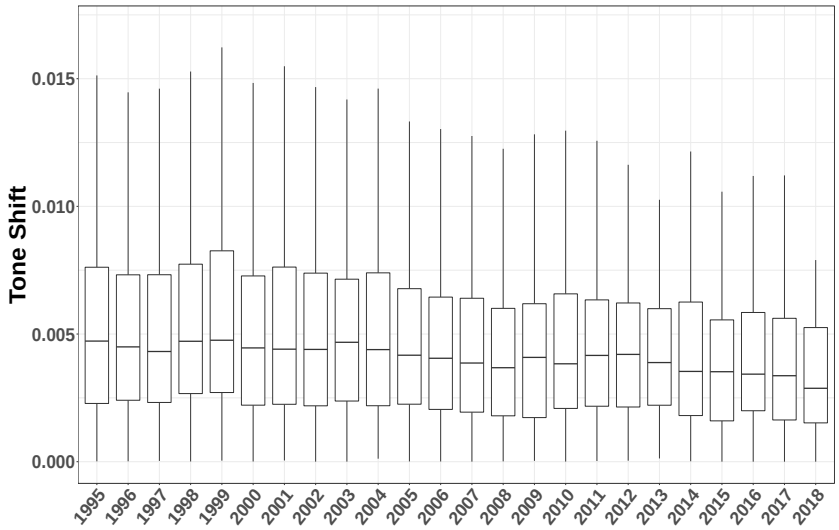


Figure 10: The boxplots of yearly 10-K tone shift distribution.

Formal statistical tests for differences in mean 10-K tone shift before and after 1999 provide corroborating evidence for this phenomenon. As table 17 shows, we can summarily reject the null hypothesis of equal means in tone shift pre- and post-1999 in favor of the alternative hypothesis of lower tone shifts post-1999 since the $p$-values for both the Welch and Wilcoxon tests are indistinguishable from 0.

All this suggests that 10-K readability for US firms has become progressively higher over time, especially after the imposition of the SEC plain English rule.

Figure 11 presents the yearly boxplots of the (natural) log of words for the 10-K document. The median log of words—and hence the median total number of words—in the 10-K document shows a steady rise over the years. The range and the body of the yearly distribution show a small decrease over time, especially for the year 2018 for which both show significant compression.

This is further corroborated by statistical tests for the differences in means—before and after the introduction of the SEC plain English rule—compiled in table 17. As the table indicates, the null hypothesis of equal means before and after 1999 can be summarily rejected in favor of the alter-

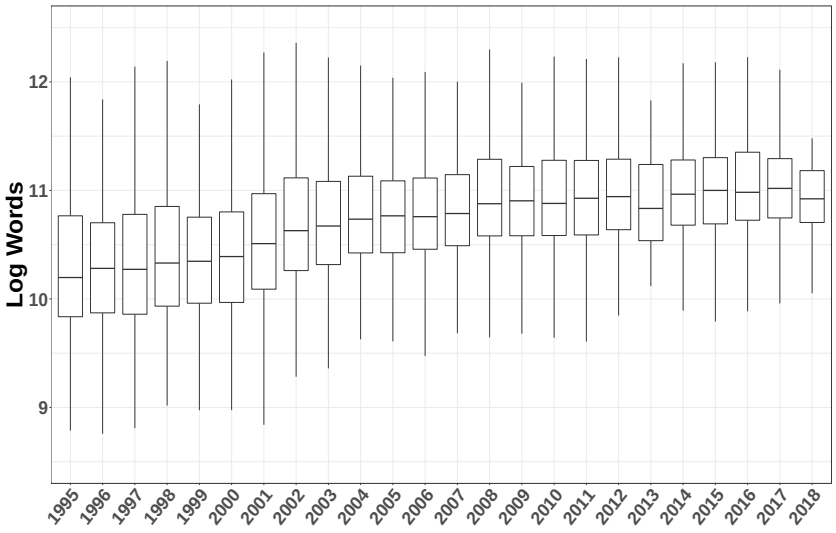


Figure 11: The boxplots of yearly 10-K log of words distribution.

native hypothesis of higher mean 10-K log of words post-1999 according to both the Welch and the Wilcoxon tests since the $p$-values are indistinguishable from 0.

To the extent that more verbose 10-K reports indicate poor readability, the plot suggests that over time, readability of 10-K documents has suffered. From this perspective, the SEC Plain English rule in 1998 seems to have not improved the readability—in terms of the 10-K's wordiness—over the past 23 years. This phenomenon of increasing verbosity in the 10-K documents mirrors our earlier discussion of the rising length of the MD&A section over time.

Figure 12 presents the yearly boxplots of US firms' 10-Ks' net file size over the years 1995–2018. The net file size of the 10-K is obtained after removing the graphics, XBRL and HTML elements from the size of the original 10-K documents. As a cursory glance at the plot suggests, the median net file size of 10-K documents has increased steadily over time and there has been a moderate reduction in the range and the body of the net file size distribution.

The increase in the median net file size is corroborated by means of standard statistical tests of equality of means pre- and post-1999 in table 17. As

the table shows, we can confidently reject the null hypothesis of equality of means in favor of the alternative hypothesis of higher levels of mean net file size post-1999 since the $p$-values for both the Welch and the Wilcoxon tests are 0.

Net file size measures the length of 10-K documents which can be used as a proxy for the amount of text content of the 10-K which needs parsing. Progressively higher levels of the net file size over the years indicate that readers have to wade through more and more amount of text to assess firms' performance and the SEC plain English initiative seems not have arrested this trend.

Figure 13 presents the yearly boxplots of US firms' 10-Ks' gross file size over the years 1995–2018. A visual inspection of the plot suggests that the median gross file size of 10-K documents has increased massively over time, in particular, during 2001–2005 but even more so after 2010 when there is a very large jump in the median 10-K file size. The sudden, substantial rise in gross file size of the 10-K reports in 2011 continues in 2012 after which it appears to stabilize somewhat. The reason for this spurt in 10-K file size is related to SEC's changes in disclosure requirements as they apply to climate change matters which were instituted on February 8, 2021.[9]

The increase in the median gross file size is corroborated by means of statistical tests of equality of means pre- and post-1999 in table 17. As the table shows, we can summarily reject the null hypothesis of equality of means in favor of the alternative hypothesis of higher levels of mean gross file size post-1999 since the $p$-values for both the Welch and the Wilcoxon tests are 0.

Since gross file size is a proxy for readability in Loughran and McDonald [2014b], progressively higher levels over the years indicate poorer readability in time which the SEC plain English initiative seems to not have had influenced in the desirable direction.

[9]The original SEC communication regarding changes in climate related disclosures can be accessed at this link: https://www.sec.gov/rules/interp/2010/33-9106.pdf.

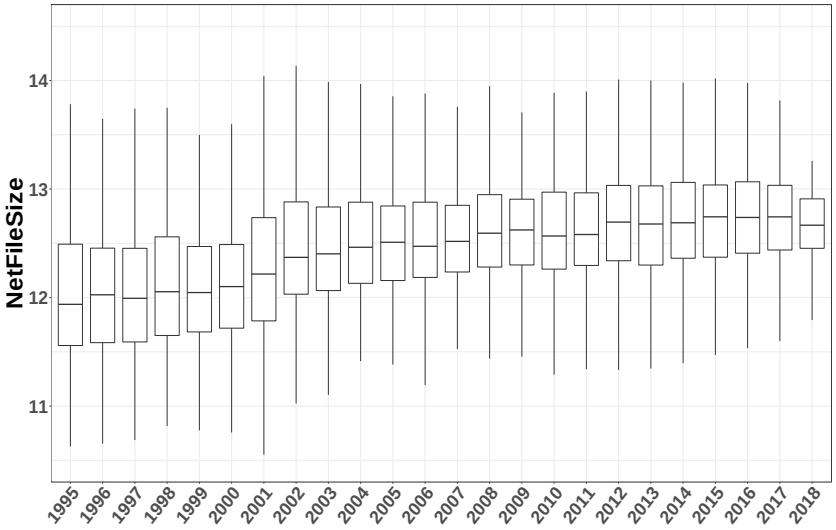


Figure 12: The boxplots of yearly 10-K net file size distribution, obtained after removing the graphics, XBRL and HTML elements from the size of the original 10-K documents.

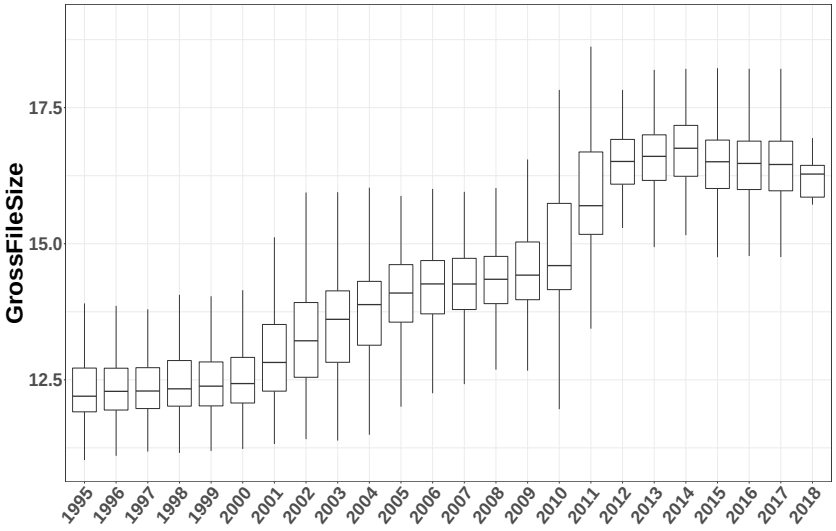


Figure 13: The boxplots of yearly 10-K gross file size distribution.

### 6.3.2 Trends in 10-K readability

In this section we compare the trends in 10-K readability over the years on the basis of the median firm's tone shift, its net file size and its gross file size.

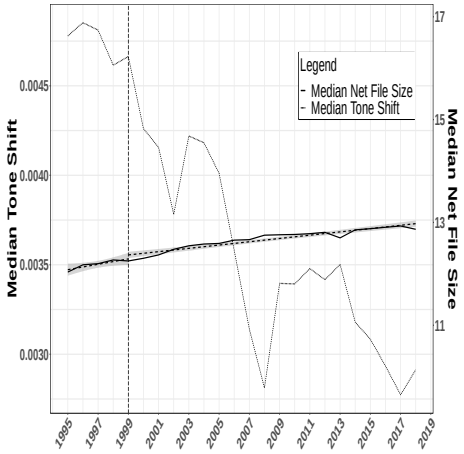


Figure 14: The plot presents the movement of tone shift in 1995–2018 for the median firm. The dashed vertical line in 1999 denotes the implementation of the SEC Plain English rule (October 1998). The solid line denotes the median of net file size, obtained after removing the graphics, XBRL and HTML elements from the size of the original 10-K documents. The grey band around the trend denotes the 95% confidence interval.

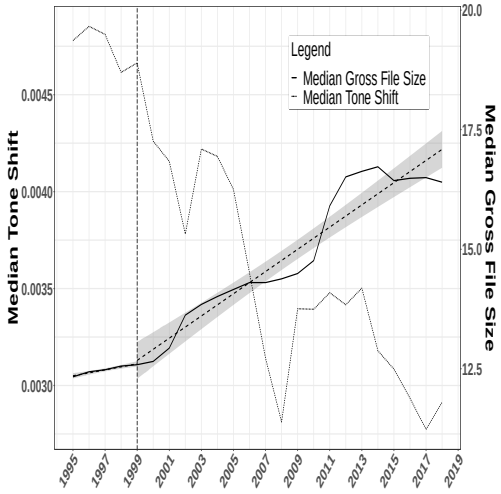


Figure 15: The plot presents the movement of tone shift in 1995–2018 for the median firm. The dashed vertical line in 1999 denotes the implementation of the SEC Plain English rule (October 1998). The solid line denotes the median of gross file size in the 10-K section. The grey band around the trend denotes the 95% confidence interval.

Figure 14 presents trends for the median firm's 10-K readability in terms of tone shift and its net file size. The net file size is obtained after removing the graphics, XBRL and HTML elements from the size of the original 10-K documents and can be considered to be a measure of the amount of text that a typical 10-K document contains.

As the plot make it clear, the median firm's 10-K's tone shift has steadily fallen over the duration of the sample 1995–2018. There are some upticks in median 10-K tone shift in 2001 and 2008 but overall the negative trend in tone shift is prominent and unmistakable. The SEC Plain English rule in 1999 accelerates the fall in median tone shift and by the end of our sample

period in 2018 we observe historically lowest levels of median tone shift in firms' 10-K. On the other hand, the median net file size of the firm shows a gradual rise over the duration of our sample and the SEC plain English rule in 1999 seems to have had no effect on curtailing its positive trend.

Figure 15 presents trends for the median firm's 10-K readability in terms of its gross file size. As visual inspection of the plot makes it apparent, there has been a steep positive trend in the median firm's 10-K gross file size owing to increased requirements for financial disclosure over the years, especially after 2010. The imposition of the SEC plain English rule has not arrested the sharp rise in firms' 10-K gross file size and in fact, the trend after 1999 seems to be far more steep than that before 1999. To the extent that larger 10-K file sizes proxy for more unreadability [Loughran and McDonald, 2014b] it suggests that for investors parsing relevant information from US firms' 10-K statements has become progressively harder over the years and the SEC rule has not improved matters in this regard.

# 7 Concluding Remarks

We introduce a new measure of financial texts' readability: 'tone shift'—which captures the incremental impact of complex multi-clausal phrases, and adjectives and adverbs—the effect of which, is to quantify the effect of hard-to-parse, complex text, higher values of which lead to increased ambiguity and investor uncertainty. This manifests in significant positive associations of 10-K tone shift with firms' subsequent idiosyncratic volatility and standardized unexpected earnings; and significant negative associations of 10-K tone shift with firms' future earnings up to two years in advance. We also show that readability of the MD&A section in particular, and the 10-K in general has been improving over time, which is reflected in the negative trend of firms' tone shift over the years, especially post-1999 after the imposition of the SEC Plain English Rule.

# References


Abhinav Anand, Sankarshan Basu, Jalaj Pathak, and Ashok Thampy. Who moved the market? Analyzing the role of Federal Reserve speeches, 2021a. IIM Bangalore Research Paper No. 622.

Abhinav Anand, Sankarshan Basu, Jalaj Pathak, and Ashok Thampy. Whose speeches impact European markets: ECBs' or the national central banks'? *European Financial Management*, 2021b. doi: https://onlinelibrary.wiley.com/doi/10.1111/eufm.12334.

Gary C Biddle, Gilles Hilary, and Rodrigo S Verdi. How does financial reporting quality relate to investment efficiency? *Journal of Accounting and Economics*, 48(2-3):112–131, 2009.

B Blanco and S Dhole. Financial statement comparability, readability and accounting fraud, 2017. AFAANZ Conference, Working Paper.

Robert Bloomfield. Discussion of "annual report readability, current earnings, and earnings persistence". *Journal of Accounting and Economics*, 45 (2-3):248–252, 2008.

Bertram Bruce et al. Why readability formulas fail. Reading education report no. 28. 1981.

Dennis Y Chung and Karel Hrazdil. Market efficiency and the post-earnings announcement drift. *Contemporary Accounting Research*, 28(3):926–956, 2011.

Bradford R Connatser. Last rites for readability formulas in technical communication. *Journal of Technical Writing and Communication*, 29(3):271–287, 1999.

C Cox. Closing remarks to the second annual corporate governance summit, 2007. Delivered at the USC Marshall School of Business, Los Angeles, CA, March.

Edward W Dolch. Fact burden and reading difficulty. *The Elementary English Review*, 16(4):135–138, 1939.

William DuBay. *Unlocking Language: The classic studies in readability*. BookSurge Publishing, 2007.

Thomas M Duffy. Readability formulas: What's the use? In Thomas M. Duffy and Robert Waller, editors, *Designing Usable Texts*, pages 113–143. Academic Press, 1985.

Eileen B Entin. Relationships of measures of interest, prior knowledge, and readability to comprehension of expository passages. 1981.

Mine Ertugrul, Jin Lei, Jiaping Qiu, and Chi Wan. Annual report readability, tone ambiguity, and the cost of borrowing. *Journal of Financial and Quantitative Analysis*, 52(2):811–836, 2017.

Eugene F Fama and Kenneth R French. Disappearing dividends: changing firm characteristics or lower propensity to pay? *Journal of Financial Economics*, 60(1):3–43, 2001.

Ronen Feldman, Suresh Govindaraj, Joshua Livnat, and Benjamin Segal. Management's tone change, post earnings announcement drift and accruals. *Review of Accounting Studies*, 15(4):915–953, 2010.

Robert Gunning. *Technique of clear writing*. McGraw-Hill, 1952.

Roderick P Hart. Redeveloping diction: Theoretical considerations. *Progress in Communication Sciences*, pages 43–60, 2001.

Leon C Hull. Measuring the readability of technical writing, 1979. 26th International Technical Communications Conference.

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

Jared N. Jennings, Hojun Seo, and Lloyd Tanlu. The effect of organizational complexity on earnings forecasting behavior. *Capital Markets: Market Efficiency eJournal*, 2014.

Chansog Kim, Ke Wang, and Liandong Zhang. Readability of 10-K reports and stock price crash risk. *Contemporary Accounting Research*, 36(2): 1184–1216, 2019.

Walter Kintsch and Douglas Vipond. Reading comprehension and readability in educational practice and psychological theory. *Perspectives on learning and memory*, pages 329–365, 2014.

David F Larcker and Anastasia A Zakolyukina. Detecting deceptive discussions in conference calls. *Journal of Accounting Research*, 50(2):495–540, 2012.

Alastair Lawrence. Individual investors and financial disclosure. *Journal of Accounting and Economics*, 56(1):130–147, 2013.

Reuven Lehavy, Feng Li, and Kenneth Merkley. The effect of annual report readability on analyst following and the properties of their earnings forecasts. *The Accounting Review*, 86(3):1087–1115, 2011.

Feng Li. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, 45(2-3):221–247, 2008.

Feng Li. The information content of forward-looking statements in corporate filings—a naïve Bayesian machine learning approach. *Journal of Accounting Research*, 48(5):1049–1102, 2010.

Kin Lo, Felipe Ramos, and Rafael Rogo. Earnings management and annual report readability. *Journal of Accounting and Economics*, 63(1):1–25, 2017.

Tim Loughran and Bill McDonald. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1):35–65, 2011.

Tim Loughran and Bill McDonald. Regulation and financial disclosure: The impact of plain English. *Journal of Regulatory Economics*, 45:94–113, 02 2014a.

Tim Loughran and Bill McDonald. Measuring readability in financial disclosures. *The Journal of Finance*, 69(4):1643–1671, 2014b.

Anthony V Manzo. Readability: A postscript. *Elementary English*, 47(7): 962–965, 1970.

Martha Maxwell. Readability: Have we gone too far? *Journal of Reading*, 21(6):525–530, 1978.

Brian P Miller. The effects of reporting complexity on small and large investor trading. *The Accounting Review*, 85(6):2107–2143, 2010.

Ralph Ojemann. The reading ability of parents and factors associated with the reading difficulty of parent education materials. *University of Iowa Studies in Child Welfare*, 8:11–32, 1934.

James W Pennebaker, Matthias R Mehl, and Kate G Niederhoffer. Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, 54(1):547–577, 2003.

Janice C Redish and Jack Selzer. The place of readability formulas in technical communication. *Technical Communication*, pages 46–52, 1985.

Qiao Xu, Guy D Fernando, and Kinsun Tam. Executive age and the readability of financial reports. *Advances in Accounting*, 43:70–81, 2018.

# A Appendix

Table A.1: List of Valence Shifters

| Word | Classification | Weight | Word | Classification | Weight |
|---|---|---|---|---|---|
| absolutely | amplifier | 0.8 | massively | amplifier | 0.8 |
| acute | amplifier | 0.8 | more | amplifier | 0.8 |
| acutely | amplifier | 0.8 | most | amplifier | 0.8 |
| almost | de-amplifier | 0.8 | much | amplifier | 0.8 |
| although | adversative-conjuction | 0.8 | neither | negator | 0.8 |
| but | adversative-conjuction | 0.8 | never | negator | 0.8 |
| cannot | negator | 0.8 | no | negator | 0.8 |
| cant | negator | 0.8 | nobody | negator | 0.8 |
| certain | amplifier | 0.8 | none | negator | 0.8 |
| certainly | amplifier | 0.8 | nor | negator | 0.8 |
| considerably | amplifier | 0.8 | not | negator | 0.8 |
| decidedly | amplifier | 0.8 | only | de-amplifier | 0.8 |
| deep | amplifier | 0.8 | particular | amplifier | 0.8 |
| deeply | amplifier | 0.8 | particularly | amplifier | 0.8 |
| definite | amplifier | 0.8 | partly | de-amplifier | 0.8 |
| definitely | amplifier | 0.8 | purpose | amplifier | 0.8 |
| doesnt | negator | 0.8 | purposely | amplifier | 0.8 |
| dont | negator | 0.8 | quite | amplifier | 0.8 |

Note: This table presents the list of valence shifters along with their classification and weight.

Table A.2: List of Valence Shifters

| Word | Classification | Weight | Word | Classification | Weight |
|---|---|---|---|---|---|
| enormous | amplifier | 0.8 | rarely | de-amplifier | 0.8 |
| especially | amplifier | 0.8 | real | amplifier | 0.8 |
| extreme | amplifier | 0.8 | really | amplifier | 0.8 |
| extremely | amplifier | 0.8 | seldom | de-amplifier | 0.8 |
| few | de-amplifier | 0.8 | serious | amplifier | 0.8 |
| greatly | amplifier | 0.8 | seriously | amplifier | 0.8 |
| havent | negator | 0.8 | severe | amplifier | 0.8 |
| heavily | amplifier | 0.8 | severely | amplifier | 0.8 |
| heavy | amplifier | 0.8 | significant | amplifier | 0.8 |
| high | amplifier | 0.8 | significantly | amplifier | 0.8 |
| highly | amplifier | 0.8 | slightly | de-amplifier | 0.8 |
| however | adversative-conjuction | 0.8 | somewhat | de-amplifier | 0.8 |
| huge | amplifier | 0.8 | sporadically | de-amplifier | 0.8 |
| hugely | amplifier | 0.8 | sure | amplifier | 0.8 |
| incredibly | de-amplifier | 0.8 | totally | amplifier | 0.8 |
| least | de-amplifier | 0.8 | true | amplifier | 0.8 |
| little | de-amplifier | 0.8 | truly | amplifier | 0.8 |
| massive | amplifier | 0.8 | uber | amplifier | 0.8 |
| vast | amplifier | 0.8 | werent | negator | 0.8 |
| vastly | amplifier | 0.8 | whereas | adversative-conjuction | 0.8 |
| very | amplifier | 0.8 | wont | negator | 0.8 |

Note: This table presents the list of valence shifters along with their classification and weight.