IIMB-WP N0. 648/2021

WORKING PAPER NO: 648

A new proxy of financial texts' readability

Abhinav Anand
Assistant Professor
Finance & Accounting
Indian Institute of Management Bangalore
Bannerghatta Road, Bangalore – 5600 76
abhinav.anand@iimb.ac.in

Sankarshan Basu
Professor
Finance & Accounting
Indian Institute of Management Bangalore
Bannerghatta Road, Bangalore – 5600 76
sankarshan.basu@iimb.ac.in

Jalaj Pathak
Assistant Professor
Finance & Accounting
Great Lakes Institute of Management
Chennai
jalaj.p@greatlakes.edu.in

Year of Publication – November 2021

A new proxy of financial texts' readability*

Abhinav Anand †

Sankarshan Basu‡

Jalaj Pathak§

Abstract

We define and construct the 'semantic complexity index' (SCI) which captures the marginal contribution of multi-clausal phrases (e.g., 'slowdown in business activity'); and adjectives, adverbs and adversative conjunctions (e.g., 'greater', 'slightly', 'however') which alter the connotation of financial texts. More semantic complexity makes text difficult to interpret, which leads to poor readability, increased ambiguity and more investor uncertainty. We show that during 1994–2018, yearly SCIs of US firms' MD&A section display significantly positive association with firms' subsequent volatility and standardized unexpected earnings. We find that firms downplay negative information and exaggerate positive information, but markets react to such firms by associating higher subsequent volatility to them. We also show that the SEC Plain English Rule (October 1998) has reduced the semantic complexity of US firms' 10-Ks, and hence has improved readability—observations at odds with other popular readability metrics.

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

JEL Classification: G12, G14, G18, M41

*We thank seminar participants at the Indian Institutes of Management Ahmedabad and Bangalore (IIMA and IIMB) for their constructive criticism on prior versions of the paper.

†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 proxy of financial texts' readability—the 'semantic complexity index' (SCI)—which isolates and quantifies text connotation due to the additional contribution of multi-clausal phrases (e.g., 'buoyancy in animal spirits') and the impact of 'valence shifters': adjectives and adverbs which alter the meaning of sentences (e.g., 'barely', 'cannot', 'however' etc.) [Anand et al., 2021a,b]. All else equal, higher prevalence of multi-clausal phrases and valence shifters tends to make the text-interpretation process more difficult which leads to more ambiguity regarding the text's connotation; and creates more uncertainty for investors and analysts alike. In principle, such prolix, nuanced writing could be used to obfuscate, prevaricate or create ambiguity with regard to the connotation of the underlying text. Thus, we argue that (all else equal) higher semantic complexity in texts leads to poorer readability, and lower semantic complexity leads to enhanced readability.

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 interpreting 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 of semantic complexity. 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,

[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.

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 as in the whole 10-K document.

Our approach, however, is different from that of the techniques outlined above. We rely on the marginal connotation of that part of a sentence which originates from the usage of multi-clausal phrases (e.g., 'enhancement in business profitability') as well as that due to 'valence shifters': adjectives, adverbs and (adversative) conjunctions (e.g., 'slightly', 'massively', 'despite', 'but' etc.) which modify the connotation of noun-forms with which they are used. Anand et al. [2021a,b] propose a technique wherein they implement a dictionary-based 'ngram' approach (n words at a time), where the n is determined endogenously for each separate sentence; and the dictionary is the standard financial lexicon [Loughran and McDonald, 2011] augmented with the collection of valence shifters. On the other hand, the well-known study of Loughran and McDonald [2011] is based on a dictionary-based 'unigram' approach ($n = 1$, one word at a time/bag-of-words) which ignores the contribution of both valence shifters and collections of more than one words (e.g., 'enhanced profitability') in assigning connotation to financial texts. We calculate a text's semantic complexity as the absolute difference between the connotation of the financial text calculated according to Anand et al. [2021a,b] and that obtained from the usage of the LM dictionary and bag-of-words approach. In other words, semantic complexity is the absolute difference between texts' connotation with and without multi-clausal phrases and valence shifters. It is quite straightforward to notice that the difference in connotation between the two approaches is precisely the marginal contribution of multi-clausal phrases and valence shifters in ascribing connotation to the whole text. Hence, a high level of semantic complexity implies high prevalence of valence shifters and multi-clausal phrases, which, all else equal, contribute towards more difficulty in interpretation, leading to more ambi-

[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.

guity, higher uncertainty, and hence to poorer readability.

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 proxy 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 connotation assignment 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' perspective. Further, our proposed readability proxy is 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 proxy ensures that this dictum is obeyed since it is able to assign proper weights to 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 the text's semantic complexity since it explicitly quantifies the effect of adjectives, adverbs and (adversative) conjunctions 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 semantic complexity overcomes this challenge. Moreover, since SCI 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 SCI is built upon capturing the connotation 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) 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 volatility—measured as the market model's RMSE [Loughran and McDonald, 2014b]—is significantly associated with their semantic complexity, over and above the impact of other popular measures of readability and relevant controls. We find that the semantic complexity of firms' MD&A 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 SCI. Both these findings are in line with our hypothesis that higher semantic complexity—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.

We also find that firms in our sample tend to exaggerate positive information and understate negative information in their MD&A by the usage of prolix writing, consistent with recent findings in Koonce et al. [2021]. However, we show that the markets are not taken in by the semantic complexity of the MD&A section and punish them by associating with such firms, a higher level of subsequent volatility. On the other hand, for cautious, sedate firms which underplay positive information and over-emphasize negative information in their MD&A section, the markets react by rewarding them with lower subsequent volatility. In other words, we find that firms cannot fool markets by using more difficult-to-interpret language in their financial disclosures, and for firms which indulge in such behavior, the markets react by attaching higher idiosyncratic risk to them.

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-interpret 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 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 polarity (sign) of the connotation, recompute the semantic complexity, 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, SCI 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 improvement 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', Fog index, BOG index 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 evaluates if firms hide bad news, section 7 outlines trends in readability, and finally section 8 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 due to reduced activity by small investors. Lehavy et al. [2011] also use Fog Index as a 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 2018 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]

[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.)

4.2 Measures based on vocabulary and size

The LM 'vocab' measure is calculated as the number of unique words in the MD&A section of the 10-K 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].

4.3 Financial texts' 'semantic complexity index' (SCI)

We define and construct a new measure of financial texts' semantic complexity which we name 'semantic complexity index', or 'SCI'. The semantic complexity index of a financial text captures the marginal connotation of that part of a sentence which originates from the usage of multi-clausal phrases (e.g., 'enhancement in business profitability') as well as that due to 'valence shifters': adjectives, adverbs and (adversative) conjunctions (e.g., 'slightly', 'massively', 'despite', 'but' etc.) which modify the connotation of noun-forms with which they are used. All else equal, increased usage of multi-clausal phrases and valence shifters makes ascribing meaning to sentences more difficult, and therefore, makes the text harder to read. Thus, (all else equal) higher semantic complexity in texts leads to poorer readability, and lower semantic complexity leads to enhanced readability. In this way, the semantic complexity index (SCI) of a text captures its degree of readability by being inversely related to it. In this way, we operationalize SCI as a proxy of financial texts' readability.

Financial disclosure documents which display high levels of semantic complexity necessarily employ high levels of multi-clausal phrases and/or usage of adjectives, adverbs and (adversative) conjunctions, which alter the connotation of text. In principle, such complex, nuanced writing could be used to obfuscate, prevaricate or create ambiguity with regard to the connotation of the underlying text. From this perspective, financial disclosure documents with high SCI—and hence poor readability—can create ambiguity and uncertainty among investors, analysts, as well retail traders who are primary readers of such documents.

In two recent papers, Anand et al. [2021a,b] outline a methodology which captures the contribution of multi-clausal phrases and valence shifters in financial texts' connotation. From a methodological standpoint, they implement a dictionary-based 'ngram' approach (n words at a time), where n is determined endogenously for each separate sentence; and the dictionary is the standard financial lexicon [Loughran and McDonald, 2011] augmented with the collection of valence shifters.[4] On the other hand, the well-known study of Loughran and McDonald [2011] is based on a dictionary-based 'unigram' approach ($n = 1$, one word at a time/bag-of-words) which ignores the contribution of both valence shifters and collections of more than one words (e.g., 'enhanced profitability') in assigning connotation to financial texts.

We calculate the semantic complexity index as the absolute value of the difference between the connotation of the financial text calculated according to Anand et al. [2021a,b] and that obtained from the usage of the LM dictionary and bag-of-words approach. Clearly, the difference in connotation between the two approaches is precisely the marginal contribution of multi-clausal phrases and valence shifters in ascribing connotation to the whole text.

In order to explicitly show how connotation is derived and the SCI computed, we produce a collection of five hypothetical sentences and one taken from the MD&A section of the 10-K report.

[4] The valence shifters can themselves 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').

1. We expect to witness an increase in business activity.

2. We expect to witness a *slight* increase in business activity.

3. We expect to witness a *major* increase in business activity.

4. We expect to witness *not much* increase in business activity.

5. We expect to witness a *large* increase in business activity *in spite of* Covid.

Clearly, while superficially similar, all sentences enumerated above are quite different in their connotation. For all hypothetical example sentences presented above, the unigram LM dictionary methodology assigns a score of 0. This is because valence shifters are ignored, and words like 'increase' are assigned zero weight since 'profit increase' has positive connotation, while 'unemployment increase' has a negative connotation; and hence a unigram approach is incapable of assigning polarity to it. However, the modified approach outlined in Anand et al. [2021a,b] is successfully able to distinguish between the five example sentences and assigns them scores ranging from 0.02 to 0.26 as specified in table 1.[5]

For a more realistic example, we present the following sentence taken from the MD&A section of the 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 score of this sentence using the bag-of-words approach and LM dictionary is:

$$\frac{(-1)[\text{=declined}]}{14} = -0.0714$$

[5]We assign valence shifters a weight of 0.8 in line with Anand et al. [2021a,b] but note that changing the weights from 0.5–0.9 reproduces the same set of results.

Table 1: Example sentences' connotation

	LM score	MCVS score
1.	0	+0.16
2.	0	+0.02
3.	0	+0.25
4.	0	+0.02
5.	0	+0.26

Note: This table presents the tone calculated using the LM dictionary and bag of words approach and the MCVS approach. 'MCVS' denotes connotation according to the 'multi-clausal phrases with valence shifter' methodology outlined in Anand et al. [2021a,b]. 'LM' denotes the methodology taken from Loughran and McDonald [2011].

However, the sentence has one valence shifter: 'slightly' which is a de-amplifier. Thus, the altered score using valence shifters is:

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

Hence, the new readability score, proxied by the semantic complexity index for this sentence is:

$$\text{SCI} = |-0.0714 - (-0.0125)| = 0.084$$

Table 2 illustrates SCI 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. Table 3 compares the semantic complexity index with prior complexity-based readability measures for some sentences in our sample.

Insert table 2 here.

Insert table 3 here.

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 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 4 presents the sample creation process for our study. We start with all 10-K files from 1994 to 2018 (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. We use 19 different search terms to extract MD&A section from the 10-K files. [6] 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.

[6] The search terms are specified in the appendix.

Table 4 presents the details below.

Insert table 4 here.

Table 5 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 the semantic complexity index, SCI: the readability proxy introduced in this study. The values of SCI 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.

Insert table 5 here.

Table 6 presents the summary statistics for all control variables used in this study. The mean values for all variables are similar to Loughran and McDonald [2014b].

Insert table 6 here.

Table 7 presents correlations among readability measures. SCI 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 moderately low negative correlation with the 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).

Insert table 7 here.

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

$$\begin{aligned} RMSE_{tk} = a_0 + a_1 SCI_{tk} + a_2 Readability\_Measures_{tk} + \\ a_3 Controls_{tk} + u_{tk} \end{aligned} \tag{1}$$

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 semantic complexity on RMSE

Table 8 presents the results of the regression which evaluates the hypothesis that for the MD&A section of the 10-K report, the market model's residual size (RMSE) has a significant association with SCI 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 other popular readability measures: Fog Index, FK Index and SMOG Index. The first column presents the benchmark association between post-filing RMSE and control variables in the absence of any readability measure. For this benchmark column, we observe that the higher are the pre-filing performance and size, the lower is the subsequent volatility. Further, firms with low book-to-market, higher pre-filing stock return volatility and with larger absolute returns on the filing date have significantly higher RMSE. The coefficients and signs of the control variables are in line with Loughran and McDonald [2014b].

Insert table 8 here.

Column 2 presents the impact of the new readability proxy, the semantic complexity index and it is persistently positive and significant in all specifications. In the presence of controls, the SCI displays significant impact on the market residual RMSE above and beyond that attributable to prior complex-words based readability measures. Further, the SCI renders the Fog and SMOG indices insignificant in its presence. The positive sign of the SCI coefficient in all regression specifications suggests that all else equal, a rise (fall) in SCI—i.e., an increase (decrease) in the semantic complexity due to high usage of multi-clausal phrases and valence shifters—leads to a rise (fall) in subsequent volatility (RMSE).

Table 9 presents the results of the regression which evaluates the hypothesis that the market model's residual size (RMSE) has a significant association with SCI over and above that for other vocab and size-based readability scores.

Insert table 9 here.

SCI shows persistent, positive significance in all specifications. In the presence of controls, the SCI displays significant impact on the market residual RMSE above and beyond that attributable to prior vocab and size-based readability measures. The LM based 'vocab' and SCI are both significant when tested together (so is log of total words). The positive sign of the SCI coefficient in all regression specifications suggests that all else equal, a rise (fall) in SCI—i.e., an increase (decrease) in the marginal contribution of valence shifters and multi-clausal phrases to the connotation of firms' MD&A—leads to a rise (fall) in RMSE.

In table 10 we present the results of regressions in which all readability scores—from both the complex words and vocab and size-based methods—are included to test whether there is any incremental significance for the SCI measure. We test the impact of SCI in the presence of all combinations of complex-words based, and vocab and size-based readability scores and find that for each model, the SCI continues to display significance over and above other readability measures. In particular, the semantic complexity index renders the Fog and SMOG indices, as well the 'Financial Term' index insignificant in its presence. Among other readability scores, the LM-based 'vocab', the size-based log of words and FK Index retain their significance in the presence of SCI. The coefficient of SCI is positive in all specifications suggesting that all else equal, a rise in the usage of multi-clausal phrases and valence shifters lead to an increase in subsequent volatility.

Insert table 10 here.

5.3 Controlling for file size of the 10-K document

In the well-known paper of Loughran and McDonald [2014b] the authors argue that instead of parsing financial text, one could simply take the file size of the 10-K document as a proxy of its readability. To test whether the semantic complexity index retains its significant association with subsequent volatility, we rerun the regression specification outline in equation 1 in the presence of the gross file size of the 10-K document. Table 11 presents our results.

Insert table 11 here.

As in previous tables, SCI retains positive significance in all specifications, and as before, renders the Fog and SMOG indices, as well the variable 'financial term' insignificant. On the other hand, as before, the FK index, LM vocab and the log words metrics retain their significance. The log of gross file size displays positive significance in all specifications as well.

5.4 Impact of SCI on Standardized Unexpected Earnings

The central thesis of this paper is that the incremental effect of using multi-clausal phrases and valence shifters on financial texts' readability is to make the text more difficult to read, which in turn leads to greater ambiguity and more investor uncertainty. More ambiguity and uncertainty among readers of financial disclosures should also increase standardized unexpected earnings (SUE). To test this formally, we examine the impact of SCI on SUE, with the measure calculated according to Loughran and McDonald [2014b]. Further, in line with Bonsall IV et al. [2017] who show that quantity of disclosures is significantly associated with information intermediaries such as analysts, we control for log words in all specifications.

Insert table 12 here.

Table 12 displays the results of the regression in which the putative impact of SCI 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 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].

The first column shows the benchmark specification without any readability scores. In all other specifications, we introduce other readability scores one at a time, with the exception of log of words which we retain in all regression specifications in line with Bonsall IV et al. [2017]. We find that both SCI and log words display significantly positive association in all regression models, with log words showing a stronger association than SCI. Among other readability scores, the Fog, FK and SMOG indices, as well as the variable 'financial term' are rendered insignificant, while the LM vocab retains its positive significance, consistent with results compiled in previous tables.

The finding that the semantic complexity index is positively and significantly associated with SUE is consistent with the finding in Loughran and McDonald [2014b]. This implies that all else equal, an increase (decrease) in SCI—i.e., an increase (decrease) in semantic complexity of the financial disclosure—is associated with higher (lower) earnings surprises.

5.5 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 SCI using a modified dictionary; and finally, use only adversative conjunctions and negators to test the impact of SCI on RMSE.

5.5.1 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, 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]. It 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 higher firm-specific complexity.

Insert table 13 here.

SCI shows a significantly positive coefficient for all model specifications, consistent with previous results. Among other readability scores, the LM based vocab measure, log of words and FK index are positively significant.

The 'financial term' variable also displays significance but its sign is negative which is in contrast with results for all other readability scores. The new control variable: business complexity also displays positive significance in all regression specifications, consistent with analogous results obtained in Loughran and McDonald [2014b]. The persistently positive values of SCI suggest that all else equal, higher levels of SCI, i.e., lower levels of readability, lead to increased values of RMSE.

5.5.2 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.).

We add such verb-noun combinations to the list of LM polar phrases to constitute a new dictionary and rerun the regression specifications to evaluate its effects. The results for the impact of MD&A SCI when the underlying dictionary is augmented by the addition of such verb-noun combinations are included in table 14.

Insert table 14 here.

Again, similar to the benchmark results, we see that SCI displays persistence in its significantly positive association with subsequent volatility of firms over and above that due to other readability metrics. Among existing readability measures, we find that log of words, LM vocab, and the FK Index are also significant associated with the post-filing RMSE.

5.5.3 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 interpret, 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 the semantic complexity index has a significantly positive association with firms' subsequent volatility.

Insert table 15 here.

However, a natural counter to our approach is as follows. Among the class of valence shifters, not all adjectives and adverbs make text more complex to interpret. 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 semantic complexity of firms' financial disclosures.

To allay such concerns we isolate two components of valence shifters: negators and adversative conjunctions. These two special categories of valence shifters alter the polarity of the connotation during the text interpretation procedure. We use only these two special categories of valence shifters to reconstruct the semantic complexity index. The results of our analysis are included in table 15.

The modified SCI 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.[7] Among other readability scores, the LM vocab, log words and the FK index show statistical significance. Further, the uniformly positive values of SCI suggests that all else equal, as the SCI—and hence the MD&A unreadability—rises, the RMSE increases.

6 Do firms exaggerate positive information?

Do firms overemphasize positive developments in their MD&A section by using more prolix, highfalutin writing? Or do they understate negative information by prevaricating and employing more complicated writing which

[7]The coefficient magnitudes reduce in size though.

is hard to interpret? Our methodology is able to answer such questions by considering a variant of the semantic complexity index. The semantic complexity index is the absolute difference in the connotation of text with and without multi-clausal phrases and valence shifters. In this section, we will instead, use the signed values of the SCI to answer questions posed above. For example, when the sentence-based ngram analysis with valence shifters and the unigram LM dictionary based method both yield positive scores, with the LM score being smaller, it implies that by the usage of more prolix writing, the positive news is being exaggerated. On the other hand, when both scores are negative, and the LM score is smaller, it suggests that semantic complexity is being introduced to understate the impact of the negative connotation of the text.

To test whether firms in their MD&A section are in fact, systematically overstating the positive and/or understating the negative, we use the signed value of semantic complexity. For our full sample, we find that 12.22% of the firm-years have positive MD&A connotation, and 84.85% have negative connotation, according to both multi-clausal valence shifter and unigram LM methodology.[8] Table 16 compiles the results of our investigation.

Insert table 16 here.

The table shows that cases in which the unigram LM dictionary methodology score is numerically smaller than that of the ngram multi-clausal phrases with valence shifter (MCVS) technique, the signed semantic complexity index is positive, and the corresponding T test for equality of means yields a p-value indistinguishable from 0. This result holds when both techniques' scores are positive, as well as when both are negative. When both scores are positive and MCVS score exceeds the LM score, it suggests that by usage of multi-clausal phrases, as well as by employing valence shifters, the management is trying to overstate the positive connotation of the MD&A section—in effect,

[8]The residual $\sim$3% firm-years have opposite connotation signs for the two methodologies. We ignore them in our initial analysis.

amplifying the positive tone. Similarly, when both MCVS and LM scores are negative, and the MCVS score is higher in magnitude, it suggests that by employing more multi-clausal phrases, as well as adjectives, adverbs and (adversative) conjunctions—all of which, the LM technique ignores—the negative connotation of the MD&A section is being understated, to minimize its putative detrimental impact on readers' perception regarding the state of the firm. The same effect is absent when analyzing cases in which the LM score exceeds the MCVS score, as the corresponding p-values suggest. Thus we conclude that for our sample, positive developments are exaggerated, and negative developments are downplayed by making the financial disclosure more unreadable. We note that our results are consistent with the recent findings of Koonce et al. [2021] which is based on experimental evidence.

We next move to enquire if markets are able to discern this tendency in firms' to exaggerate or underplay information in the MD&A section. Presumably, if the markets can distinguish such behavior, the exaggeration of positive MD&A tone, and downplaying of negative information should not reduce subsequent volatility. We test these hypotheses and compile results in table 17 by using signed SCI and its interaction with the signed SCI dummy which takes the value 1 when it is positive and 0 otherwise.

Insert table 17 here.

As table 17 shows, in all regression specifications, the signed SCI has a negative sign, but the interaction (signed) SCI*dummy has a positive sign whose magnitude uniformly exceeds that of (signed) SCI coefficient. We interpret this in the following way. The signed SCI dummy is positive only when the usage of multi-clausal phrases and valence shifters pushes up the score of MD&A section to either overstate positive information (when both MCVS and LM scores are positive) or understate negative information (when both are negative) corresponding to the first two rows of the table 16. In this case, the dummy takes value 1, and the aggregate impact of the signed complexity index is positive. This in turn implies that when firms either overstate the positive or understate negative, the RMSE is positive—i.e., the markets react negatively to it by increasing such firms' subsequent volatility. On the other hand, when signed SCI is negative, corresponding to rows 3 and 4 of table 16, the dummy takes value 0 and the negative sign of the signed SCI coefficient implies that RMSE decreases. In other words, when firms understate the positive and overstate the negative, the markets react positively by reducing firms' subsequent volatility.

What this result implies is that financial markets are not taken in by the semantic complexity of the MD&A section. When firms try to hide, or underplay negative information, or when they try to overstate positive information, instead of reducing their subsequent firm-specific volatility, it ends up getting increased. In other words, firms cannot fool the market by using prolix writing featuring valence shifters and multi-clausal phrases in their financial disclosures, and in fact, the markets punish such behavior by assigning higher subsequent volatility to its stocks. On the other hand, when sedate, cautious firms underplay positive news, and are eager to overemphasize negative news in their MD&A section, the markets reward them by assigning lower subsequent volatility to their stocks.

7 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? This has previously been studied by Dyer et al. [2017] who show that the length of 10-K in general as well as its various sections has increased over the years. Similar trends are reported in Bonsall IV et al. [2017], where they show that log of total number of words, file size, Fog Index as well as their BOG Index has increased over the years. Loughran and McDonald [2014a] show that post-1998, firms' financial disclosures show improved stylistic conventions consistent with SEC's plain English initiative. We reinvestigate trends in firms' financial disclosures' readability based on 10-Ks and its MD&A

section on the basis of their semantic complexity and compare it to that according to other popular readability measures.

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.

7.1 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' SCI from 1994 to 2018, evaluate whether the plain English rule had any impact on SCI; and compare it to distributions and trends in other popular measures of readability.

7.1.1 SCI trend

Insert figure 1 here.

Figure 1 presents the yearly time series of the median, the 75$^{\text{th}}$ percentile, and the 25$^{\text{th}}$ percentile of SCI. From visual inspection we see that for the median firm, its SCI 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 SCI 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 SCI firm (75$^{\text{th}}$ percentile) displays the same behavior: increasing SCI 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 SCI firm than for the median firm, suggesting that the rule impacted firms with high SCIs more strongly than those with low SCI values. This hypothesis finds more evidence in its favor when we observe the yearly time series of the median low SCI firm (25^{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-interpret language prior to the imposition of the rule seem to have changed their style so as to conform to SEC's initiative. This result seems to be in line with that of Loughran and McDonald [2014a].

High levels of SCI are accompanied with high incidence of multi-clausal phrases as well as adjectives and adverbs, which indicates more semantic complexity in texts and hence low levels of readability. Negative trends and falling SCI levels for the median firm and the median high SCI 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 SCI for all firms in our sample which have 4 or more years of SCI data and present the results in table 18. 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 SCI—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 SCI. Similarly, before 1999, there are 297 firms (43% of 678) which show a negative trend in SCI. But this number increases to 1435 (58% of 2491 firms) after the SEC Plain English rule comes into effect.

Insert table 18 here.

In figure 2 we present trends in the SCI of firms based on their size (market equity). We plot the median small firm's (25^{th} percentile) and the median large firm's (75^{th} percentile) SCI over 1994–2018. Both the median small and large firms show positive trends (increasing levels) in SCI from 1994–1998, but this trend reverses after the implementation of the plain English rule. The median large firm displays the peak in its SCI 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 SCI values start falling. In particular, both firms exhibit a negative trend in their SCI after 1999, especially the median large firm which indicates that the SEC rule impacts large firms more than it does smaller firms.

Insert figure 2 here.

This is also borne out by table 18 in which we calculate linear time trends for all firms in the top quartile (≥ 0.75 quantile) and the bottom quartile (≤ 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.[9] Before 1999, there are 354 small firms, out of which, 196 (55%) show increasing SCI 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 SCI 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 SCI 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).

7.1.2 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

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

in this study and compare their evolution over time. In order to facilitate 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.[10]

Insert figure 3 here.

Figure 3 presents boxplots of SCI 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 SCI 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 SCI 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 SCI is not isolated to a few select firms but encompasses the entire collection of firms in the US. To the extent that lower levels of SCI indicate high readability and low semantic complexity, it indicates that the MD&A section has become more readable over time, especially after the imposition of the SEC rule.

Insert figure 4 here.

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.

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

The SEC rule imposition in 1999 seems to not have reduced the average words per sentence measure. In fact, the trend 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^{th} and the 25^{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 interpretation of the text more complex, and hence lead to poor readability, this suggests that over time, the MD&A readability—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.

Insert figure 5 here.

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.

Insert figure 6 here.

Figure 6 presents yearly boxplots of the distribution of the (natural) log of words for the MD&A section over the years.[11] 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.

7.1.3 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.

Insert figures 7 and 8 here.

In figure 7, we compare the yearly time series of median SCI 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 SCI 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

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

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 SCI 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 SCI 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.

Insert figure 9 here.

Finally, in figure 9 we plot together the median SCI 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 SCI 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.

Insert table 19 here.

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 19. 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 SCI, 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 SCI 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.

7.2 10-K readability over time

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

7.2.1 Distribution of 10-K readability over the years

Insert figure 10 here.

Figure 10 presents the yearly boxplots of the SCI for US firms' 10-Ks over the years 1995–2018. As visual inspection makes it clear, the median SCI 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 SCI 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 SCI distribution.

Formal statistical tests for differences in mean 10-K SCI before and after 1999 provide corroborating evidence for this phenomenon. As table 19 shows, we can summarily reject the null hypothesis of equal means in SCI pre- and post-1999 in favor of the alternative hypothesis of lower SCIs 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.

Insert figure 11 here.

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 19. As the table indicates, the null hypothesis of equal means before and after 1999 can be summarily rejected in favor of the alternative 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.

Insert figure 12 here.

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 19. 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 interpreting. 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.

Insert figure 13 here.

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.[12]

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 19. 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.

7.2.2 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 SCI, its net file size and its gross file size.

Insert figures 14 and 15 here.

Figure 14 presents trends for the median firm's 10-K readability in terms of SCI 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 makes it clear, the median firm's 10-K's SCI has steadily fallen over the duration of the sample 1995–2018. There are some upticks in median 10-K SCI in 2001 and 2008 but overall the negative trend in SCI is prominent and unmistakable. The SEC Plain English rule in 1999 accelerates the fall in median SCI and by the end of our sample period in 2018 we observe historically lowest levels of median SCI in firms' 10-K. On the other hand,

[12]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`.

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.

Insert figures 16 and 17 here.

Figure 16 presents trends for the median firm's 10-K readability in terms of the BOG index [Bonsall IV et al., 2017]. As a cursory glance suggests, there has been a positive trend in the median firm's 10-K's BOG index, whose rate of growth is dampened by the introduction of the SEC initiative in 1998. This suggests that the plain English initiative has not negatively impacted the BOG index and hence readability has not become progressively better. In the same way, figure 17 index displays trends for the FOG index which also exhibits positive trends whose rate of growth decreases somewhat after the introduction of the SEC initiative.

8 Concluding Remarks

We introduce a new proxy of financial texts' readability: the semantic complexity index—which captures the incremental impact of multi-clausal phrases, adjectives, adverbs and (adversative) conjunctions—the effect of which is to quantify the effect of hard-to-interpret text, higher values of which, lead to increased ambiguity and investor uncertainty. This manifests in significant positive associations of SCI with firms' subsequent volatility and standardized unexpected earnings. We also show that firms understate negative information in the MD&A section and overemphasize positive developments; but this leads to an increase in their subsequent volatility, which suggests that markets are not taken in by semantically complex writing in the MD&A section of the 10-K report. 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' semantic complexity 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.

Samuel B Bonsall IV, Andrew J Leone, Brian P Miller, and Kristina Rennekamp. A plain english measure of financial reporting readability. *Journal of Accounting and Economics*, 63(2-3):329–357, 2017.

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

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.

Travis Dyer, Mark Lang, and Lorien Stice-Lawrence. The evolution of 10-k textual disclosure: Evidence from latent dirichlet allocation. *Journal of Accounting and Economics*, 64(2-3):221–245, 2017.

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.

Lisa Koonce, Zheng Leitter, and Brian White. The effect of a warning on investors' reactions to disclosure readability. *Review of Accounting Studies*, 2021.

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.

9 Tables and Figures

9.1 Tables

Table 2: MD&A Valence Shifters

Company (CIK) &Date	Type	Word	Sentence	Score New	Score LM	SCI	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 connotation 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 negative connotation of the 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 connotation 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 'SCI' 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.

Table 3: Comparison with Prior Readability Measures

Sentence	AWPS	Per_CW	LM Vocab	SCI
"the company and our industry have gone through a very difficult cycle over the past five years."	17	0.17	0.06	0.146
"the gross profit margin percentage declined slightly from the prior year primarily due to start up activities at the indianapolis airframe maintenance facility."	23	0.26	0.04	0.05
"the impact of the divestiture and exits had a negative effect on net sales during 2008, but a positive impact on margins."	21	0.14	0.16	0.16
"gross profit decreased in 2000 to 22.4% compared to 23.3% in 1999, but was considerably greater than the margins of 18.1% in 1998."	17	0.05	0.05	0.35

Note: This table presents specific examples for sentences with comparison to SCI.

Table 4: Sample Creation

	Dropped	Sample Size
SEC 10-K files 1994:2018		242,181
MD&A extracted from SEC 10-K files 1994:2018	76,565	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
Reported on CRSP as ordinary common equity	3,901	53,617
Price on filling date minus one less than 3 USD	9,008	44,609

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 5 refers to the CRSP and COMPUSTAT databases.

Table 5: 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
SCI	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 6: Summary Statistics (Control Variables)

Control Measures	Mean (1994-2018)	Mean (Pre 1999)	Mean (Post 1999)
Size (in million USD)	2892.5	845.4	3168.3
BM	0.64	0.53	0.65
Filling Period Return	0.03	0.03	0.03
Pre filling alpha	0.05	0.06	0.05
Pre filling RMSE	2.86	3.35	2.79
Post filling RMSE	2.38	3.09	2.29
SUE	0.47	0.33	0.47
No. of Analyst	5.99	4.51	6.06

Note: This table presents the summary statistics for all controls used in this study. Log transformations are used for Size and Book-to-market in the subsequent regressions.

Table 7: Correlation among Readability Measures

Variable	Fog_Index	FK_Index	SMOG_Index	AWPS	per_CW	log_words	Vocab	Fin_Term	SCI
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	
SCI	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 measures of readability along with the new measure SCI.

Table 8: SCI impact on RMSE [in presence of complex-word based measures]

Readability Variables	(1)	(2)	(3)	(4)	(5)
Fog_Index			0.006		
			(0.008)		
FK_Index				0.030***	
				(0.011)	
SMOG_Index					0.015
					(0.013)
SCI		3.689**	3.457*	2.971*	3.279*
		(1.834)	(1.902)	(1.693)	(1.877)
Control Variables					
Pre-filing alpha	−0.548**	−0.548**	−0.547**	−0.543**	−0.546**
	(0.247)	(0.253)	(0.251)	(0.249)	(0.250)
Pre-filing RMSE	0.435***	0.435***	0.435***	0.433***	0.434***
	(0.051)	(0.052)	(0.054)	(0.055)	(0.055)
Filing Period Return	4.295***	4.296***	4.294***	4.293***	4.294***
	(0.779)	(0.777)	(0.778)	(0.774)	(0.779)
Size	−0.104***	−0.102***	−0.103***	−0.106***	−0.104***
	(0.024)	(0.025)	(0.029)	(0.027)	(0.031)
BM	−0.094*	−0.094*	−0.094*	−0.094*	−0.094*
	(0.048)	(0.049)	(0.048)	(0.048)	(0.048)
NASDAQ Dummy	0.215**	0.214**	0.215**	0.216**	0.215**
	(0.087)	(0.088)	(0.087)	(0.086)	(0.087)
Adjusted R^2	0.461	0.462	0.462	0.462	0.462
N	44,609	44,609	44,609	44,609	44,609

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.

Table 9: SCI impact on RMSE [in presence of vocab and size-based measures]

Readability Variables	(1)	(2)	(3)	(4)
Vocab	0.239***		0.240***	
	(0.087)		(0.087)	
Financial_Term		−78.329	−82.801	
		(60.599)	(61.702)	
Log Words				0.063**
				(0.029)
SCI	4.724**	3.658**	4.696**	5.073***
	(1.856)	(1.833)	(1.854)	(1.867)
Control Variables				
Pre-filing alpha	−0.541**	−0.549**	−0.542**	−0.541**
	(0.254)	(0.253)	(0.254)	(0.254)
Pre-filing RMSE	0.432***	0.435***	0.432***	0.433***
	(0.052)	(0.052)	(0.053)	(0.054)
Filing Period Return	4.281***	4.298***	4.284***	4.279***
	(0.775)	(0.777)	(0.775)	(0.774)
Size	−0.115***	−0.102***	−0.115***	−0.111***
	(0.025)	(0.025)	(0.025)	(0.028)
BM	−0.104**	−0.094**	−0.104**	−0.099**
	(0.049)	(0.047)	(0.048)	(0.049)
NASDAQ Dummy	0.214**	0.215**	0.215**	0.214**
	(0.087)	(0.084)	(0.084)	(0.086)
Adjusted R^2	0.461	0.461	0.461	0.461
N	44,609	44,609	44,609	44,609

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.

Table 10: Impact of SCI on RMSE [in presence of all readability measures]

Readability Variables	(1)	(2)	(3)	(4)	(5)
Vocab				0.270***	
				(0.095)	
Financial_Term					−79.812
					(61.459)
Log Words	0.063**	0.068**	0.062**	−0.012	0.063**
	(0.028)	(0.030)	(0.029)	(0.038)	(0.030)
Fog_Index	0.006				
	(0.007)				
FK_Index		0.032***			
		(0.011)			
SMOG_Index			0.014		
			(0.013)		
SCI	4.847**	4.414**	4.672**	4.597**	5.044***
	(1.922)	(1.732)	(1.916)	(1.871)	(1.864)
Control Variables					
Pre-filing alpha	−0.540**	−0.535**	−0.539**	−0.542**	−0.541**
	(0.252)	(0.251)	(0.252)	(0.253)	(0.255)
Pre-filing RMSE	0.433***	0.431***	0.432***	0.432***	0.433***
	(0.055)	(0.055)	(0.055)	(0.054)	(0.054)
Filing Period Return	4.278***	4.276***	4.278***	4.282***	4.282***
	(0.776)	(0.771)	(0.776)	(0.775)	(0.775)
Size	−0.111***	−0.115***	−0.112***	−0.115***	−0.111***
	(0.030)	(0.028)	(0.030)	(0.028)	(0.028)
BM	−0.099**	−0.100**	−0.099**	−0.104**	−0.099**
	(0.049)	(0.048)	(0.049)	(0.048)	(0.048)
NASDAQ Dummy	0.214**	0.216**	0.215**	0.214**	0.215**
	(0.086)	(0.085)	(0.086)	(0.087)	(0.083)
Adjusted R^2	0.461	0.461	0.461	0.461	0.461
N	44,609	44,609	44,609	44,609	44,609

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.

Table 11: SCI impact on RMSE controlling for file size

Readability Variables	(1)	(2)	(3)	(4)	(5)
Fog_Index	0.004				
	(0.007)				
FK_Index		0.032***			
		(0.011)			
SMOG_Index			0.012		
			(0.014)		
vocab				0.246**	
				(0.096)	
Financial_Term					−80.285
					(60.488)
Log_Words	0.053*	0.059**	0.053*	−0.014	0.053*
	(0.029)	(0.029)	(0.029)	(0.038)	(0.029)
log(GrossFileSize)	0.045**	0.045**	0.044**	0.042*	0.047**
	(0.021)	(0.023)	(0.021)	(0.022)	(0.022)
SCI	4.803**	4.332**	4.645**	4.559**	4.954***
	(1.919)	(1.739)	(1.916)	(1.871)	(1.866)
Control variables					
Pre-filing RMSE	0.432***	0.430***	0.432***	0.431***	0.432***
	(0.054)	(0.055)	(0.054)	(0.054)	(0.054)
BM	−0.104**	−0.105**	−0.104**	−0.108**	−0.105**
	(0.049)	(0.048)	(0.049)	(0.048)	(0.047)
Size	−0.118***	−0.122***	−0.119***	−0.121***	−0.118***
	(0.030)	(0.029)	(0.030)	(0.028)	(0.028)
Filing Period Return	4.275***	4.272***	4.275***	4.278***	4.279***
	(0.777)	(0.772)	(0.778)	(0.776)	(0.777)
Pre-filing alpha	−0.538**	−0.533**	−0.537**	−0.540**	−0.539**
	(0.252)	(0.252)	(0.252)	(0.253)	(0.255)
NASDAQ_Dummy	0.217**	0.219**	0.217**	0.217**	0.217***
	(0.087)	(0.086)	(0.087)	(0.088)	(0.084)
Adjusted R^2	0.461	0.462	0.461	0.461	0.461
N	44,609	44,609	44,609	44,609	44,609

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 12: SCI impact on standardized unexpected earnings (SUE)

Readability	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fog_Index			0.01				
			(-0.008)				
FK_Index				0.004			
				(-0.01)			
SMOG_Index					0.015		
					(-0.012)		
vocab						0.231^{**}	
						(-0.114)	
Financial_Term							28.803
							(61.821)
Log Words		0.084^{***}	0.079^{***}	0.084^{***}	0.079^{***}	0.007	0.083^{***}
		(-0.019)	(-0.02)	(-0.022)	(-0.02)	(-0.045)	(0.019)
SCI		2.642^{**}	2.233^{*}	2.566^{*}	2.250^{*}	1.972	2.648^{**}
		(-1.281)	(-1.276)	(-1.318)	(-1.283)	(-1.448)	(1.284)
Control Variables							
Pre-filing RMSE	0.145^{***}	0.141^{***}	0.141^{***}	0.141^{***}	0.141^{***}	0.141^{***}	0.141^{***}
	(-0.028)	(-0.03)	(-0.03)	(-0.028)	(-0.03)	(-0.03)	(0.030)
BM	0.136^{***}	0.127^{***}	0.127^{***}	0.127^{***}	0.127^{***}	0.123^{***}	0.127^{***}
	(-0.042)	(-0.041)	(-0.042)	(-0.04)	(-0.042)	(-0.04)	(0.041)
Size	-0.119^{***}	-0.128^{***}	-0.129^{***}	-0.129^{***}	-0.129^{***}	-0.131^{***}	-0.129^{***}
	(-0.022)	(-0.024)	(-0.024)	(-0.023)	(-0.025)	(-0.024)	(0.024)
Filing Period Return	0.944^{***}	0.930^{***}	0.926^{***}	0.930^{***}	0.927^{***}	0.937^{***}	0.928^{***}
	(-0.342)	(-0.335)	(-0.337)	(-0.335)	(-0.338)	(-0.336)	(0.335)
Pre-filing alpha	-0.410^{***}	-0.405^{***}	-0.404^{***}	-0.404^{***}	-0.404^{***}	-0.408^{***}	-0.405^{***}
	(-0.086)	(-0.085)	(-0.085)	(-0.084)	(-0.084)	(-0.086)	(0.086)
NASDAQ_Dummy	-0.150^{***}	-0.147^{***}	-0.147^{***}	-0.147^{***}	-0.147^{***}	-0.148^{***}	-0.147^{***}
	(-0.053)	(-0.052)	(-0.051)	(-0.053)	(-0.051)	(-0.052)	(0.052)
Number_of_Analyst	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003
	(-0.004)	(-0.004)	(-0.004)	(-0.004)	(-0.004)	(-0.004)	(0.004)
Adjusted R^2	0.124	0.126	0.126	0.126	0.126	0.126	0.126
N	31427	31427	31427	31427	31427	31427	31,427

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.

Table 13: Impact of SCI on RMSE controlling for business complexity

Readability	(1)	(2)	(3)	(4)	(5)
Vocab				0.356***	
				(0.130)	
Financial_Term					−123.086**
					(60.120)
Log Words	0.094**	0.098**	0.093**	0.004	0.097**
	(0.039)	(0.042)	(0.039)	(0.043)	(0.040)
Fog_Index	0.009				
	(0.006)				
FK_Index		0.031**			
		(0.014)			
SMOG_Index			0.015		
			(0.010)		
SCI	5.764**	5.429**	5.723**	5.559**	6.071**
	(2.667)	(2.458)	(2.645)	(2.667)	(2.674)
Control Variables					
Business Complexity	0.152***	0.151***	0.152***	0.162***	0.152***
	(0.048)	(0.049)	(0.048)	(0.049)	(0.050)
Pre-filing alpha	−0.486**	−0.482**	−0.486**	−0.489**	−0.488**
	(0.234)	(0.234)	(0.234)	(0.233)	(0.233)
Pre-filing RMSE	0.432***	0.430***	0.432***	0.432***	0.432***
	(0.060)	(0.062)	(0.061)	(0.060)	(0.060)
Filing Period Return	4.407***	4.408***	4.408***	4.405***	4.405***
	(0.905)	(0.901)	(0.905)	(0.901)	(0.905)
Size	−0.107***	−0.110***	−0.107***	−0.110***	−0.106***
	(0.031)	(0.029)	(0.031)	(0.030)	(0.030)
BM	−0.108**	−0.108**	−0.108**	−0.113**	−0.108*
	(0.054)	(0.054)	(0.054)	(0.054)	(0.055)
NASDAQ Dummy	0.237***	0.236***	0.237***	0.240***	0.238***
	(0.092)	(0.091)	(0.091)	(0.093)	(0.092)
Adjusted R^2	0.481	0.481	0.481	0.481	0.481
N	28,331	28,331	28,331	28,331	28,331

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.

Table 14: Impact of SCI on RMSE [modified dictionary]

Readability Variables	(1)	(2)	(3)	(4)	(5)
Vocab				0.232**	
				(0.095)	
Financial_Term					−77.596
					(66.113)
Log Words	0.078**	0.085**	0.077**	0.008	0.082**
	(0.036)	(0.036)	(0.038)	(0.049)	(0.036)
Fog_Index	0.009				
	(0.009)				
FK_Index		0.031***			
		(0.011)			
SMOG_Index			0.016		
			(0.015)		
SCI	5.046**	4.841***	4.927**	4.768**	5.495***
	(2.084)	(1.817)	(2.074)	(1.943)	(1.944)
Control Variables					
Pre-filing alpha	−0.508**	−0.504**	−0.508**	−0.511**	−0.510**
	(0.242)	(0.242)	(0.242)	(0.242)	(0.244)
Pre-filing RMSE	0.440***	0.438***	0.440***	0.440***	0.440***
	(0.053)	(0.053)	(0.053)	(0.053)	(0.053)
Filing Period Return	4.178***	4.175***	4.178***	4.182***	4.181***
	(0.767)	(0.763)	(0.768)	(0.767)	(0.766)
Size	−0.112***	−0.115***	−0.112***	−0.114***	−0.111***
	(0.029)	(0.027)	(0.029)	(0.027)	(0.027)
BM	−0.098*	−0.099**	−0.098*	−0.102**	−0.098**
	(0.050)	(0.050)	(0.050)	(0.050)	(0.050)
NASDAQ Dummy	0.208**	0.209**	0.208**	0.208**	0.208**
	(0.085)	(0.085)	(0.085)	(0.087)	(0.084)
Adjusted R^2	0.462	0.463	0.463	0.463	0.462
N	43,621	43,621	43,621	43,621	43,621

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.

Table 15: Impact of negators and adversative conjunctions on RMSE

Readability Variables	(1)	(2)	(3)	(4)	(5)
Vocab				0.279***	
				(0.096)	
Financial_Term					−80.840
					(61.350)
Log Words	0.056*	0.063**	0.056*	−0.021	0.057*
	(0.029)	(0.030)	(0.029)	(0.038)	(0.030)
Fog_Index	0.006				
	(0.007)				
FK_Index		0.033***			
		(0.011)			
SMOG_Index			0.015		
			(0.013)		
SCI	3.209**	2.960***	3.036**	2.941**	3.413***
	(1.259)	(1.119)	(1.247)	(1.197)	(1.217)
Control Variables					
Pre-filing alpha	−0.540**	−0.535**	−0.539**	−0.542**	−0.542**
	(0.251)	(0.249)	(0.250)	(0.252)	(0.254)
Pre-filing RMSE	0.433***	0.431***	0.433***	0.433***	0.433***
	(0.056)	(0.056)	(0.056)	(0.054)	(0.054)
Filing Period Return	4.281***	4.278***	4.281***	4.285***	4.285***
	(0.781)	(0.774)	(0.783)	(0.778)	(0.779)
Size	−0.111***	−0.115***	−0.112***	−0.115***	−0.110***
	(0.030)	(0.028)	(0.031)	(0.027)	(0.028)
BM	−0.099**	−0.100**	−0.099**	−0.104**	−0.099**
	(0.049)	(0.049)	(0.049)	(0.049)	(0.049)
NASDAQ Dummy	0.214**	0.216**	0.215**	0.214**	0.215**
	(0.086)	(0.085)	(0.086)	(0.087)	(0.084)
Adjusted R^2	0.461	0.461	0.461	0.461	0.461
N	44,609	44,609	44,609	44,609	44,609

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.

Table 16: Signed SCI and its relation to MD&A under- or overstatement

Case	MCVS	LM	% of sample	Prop	signed SCI	p-value
MCVS > LM	Positive	Positive	12.22%	61.10%	Positive	0
MCVS > LM	Negative	Negative	83.63%	57.38%	Positive	0
MCVS < LM	Positive	Positive	12.22%	33.15%	Negative	0.99
MCVS < LM	Negative	Negative	83.63%	40.90%	Negative	0.99

Note: This table presents signed SCI and its relation to whether firms overstate or undertsate information in the MD&A section. 'MCVS' denotes connotation according to the 'multi-clausal phrases and valence shifter' methodology outlined in Anand et al. [2021a,b]. 'LM' denotes the methodology taken from Loughran and McDonald [2011]. 'Prop' denotes the proportion of the '% of sample' column. The p-value is that for the T test for equality of means.

Table 17: Signed SCI impact on RMSE

Readability	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fog_Index			0.010				
			(0.009)				
FK_Index				0.025**			
				(0.010)			
SMOG_Index					0.017		
					(0.015)		
vocab						0.258**	
						(0.118)	
Financial_Term							−85.252
							(70.955)
Log Words		0.084**	0.080**	0.087**	0.079**	0.002	0.085**
		(0.036)	(0.037)	(0.037)	(0.038)	(0.053)	(0.035)
signed SCI	−11.321**	−12.500***	−12.187**	−11.699**	−12.119**	−12.017**	−12.471***
	(4.765)	(4.769)	(4.793)	(4.596)	(4.795)	(4.717)	(4.770)
signed SCI*Dummy	11.541**	15.353***	14.411***	14.156***	14.255***	13.904***	15.308***
	(4.940)	(4.979)	(5.104)	(4.819)	(5.104)	(4.876)	(4.988)
Control Variables							
Pre-filing RMSE	0.449***	0.446***	0.446***	0.444***	0.445***	0.445***	0.446***
	(0.053)	(0.055)	(0.055)	(0.055)	(0.054)	(0.054)	(0.055)
BM	−0.091*	−0.099*	−0.099*	−0.099*	−0.098*	−0.103**	−0.099*
	(0.050)	(0.051)	(0.051)	(0.051)	(0.050)	(0.051)	(0.050)
Size	−0.098***	−0.110***	−0.110***	−0.113***	−0.111***	−0.113***	−0.109***
	(0.023)	(0.026)	(0.027)	(0.026)	(0.027)	(0.025)	(0.026)
Filing Period Return	4.341***	4.322***	4.321***	4.320***	4.322***	4.327***	4.324***
	(0.801)	(0.797)	(0.799)	(0.796)	(0.799)	(0.799)	(0.797)
Pre-filing alpha	−0.549**	−0.541**	−0.539**	−0.536**	−0.538**	−0.542**	−0.541**
	(0.238)	(0.238)	(0.237)	(0.238)	(0.237)	(0.237)	(0.239)
NASDAQ_Dummy	0.202**	0.203**	0.204**	0.205**	0.204**	0.203**	0.204**
	(0.089)	(0.088)	(0.086)	(0.086)	(0.086)	(0.089)	(0.085)
Adjusted R^2	0.468	0.468	0.468	0.469	0.468	0.468	0.468
N	44,609	44,609	44,609	44,609	44,609	44,609	44,609

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 root mean squared error (RMSE). 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 18: SCI time trends

	Pre 1999			Post 1999		
SCI	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 SCI 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 pre and post 1999.

Table 19: Impact of the SEC plain English rule on mean readability

Readability measure	Name of the test	Alt:$\mathbb{H}_1$	p-value
SCI (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
SCI (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'.

9.2 Figures

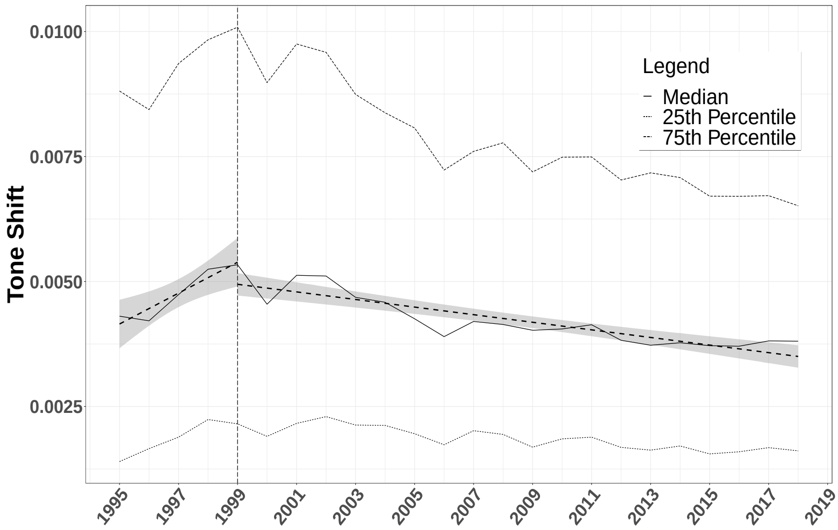

Figure 1: The plot presents the movement of SCI in 1994–2018 for the median firm (quantile 0.50), the median high SCI firm (quantile 0.75); and the median low SCI firm (quantile 0.25). The dashed vertical line in 1999 denotes the implementation of the SEC Plain English rule (October 1998). The grey band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

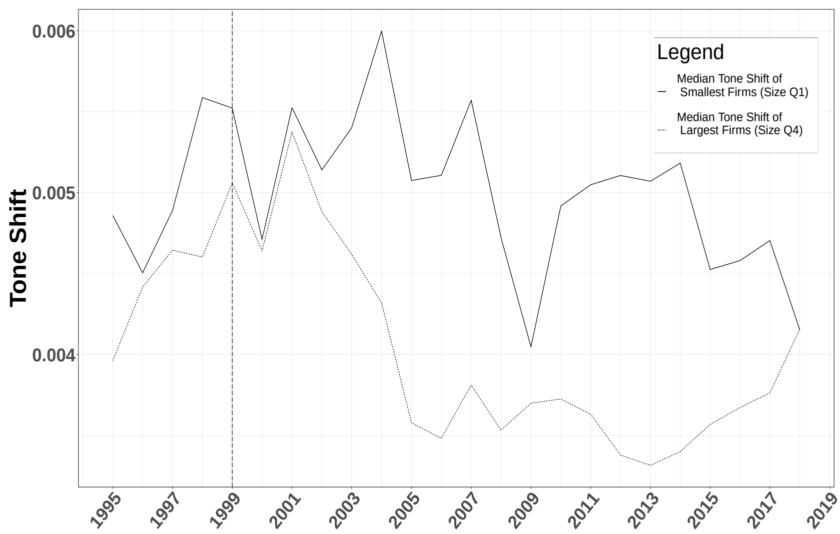

Figure 2: The plot presents the movement of SCI in 1994–2018 for the median large firm (size quantile 0.75); and that for the median small firm (size quantile 0.25). The dashed vertical line in 1999 denotes the implementation of the SEC Plain English rule (October 1998).

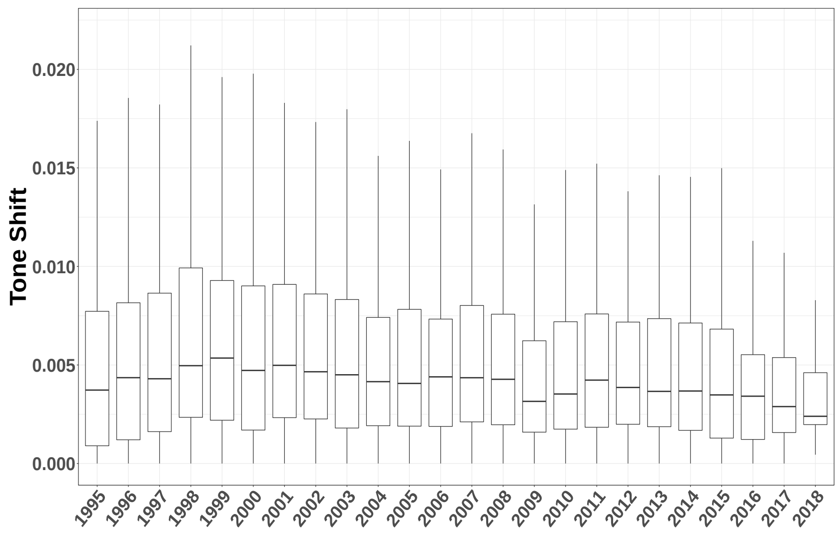

Figure 3: The boxplots of yearly MD&A SCI distribution.

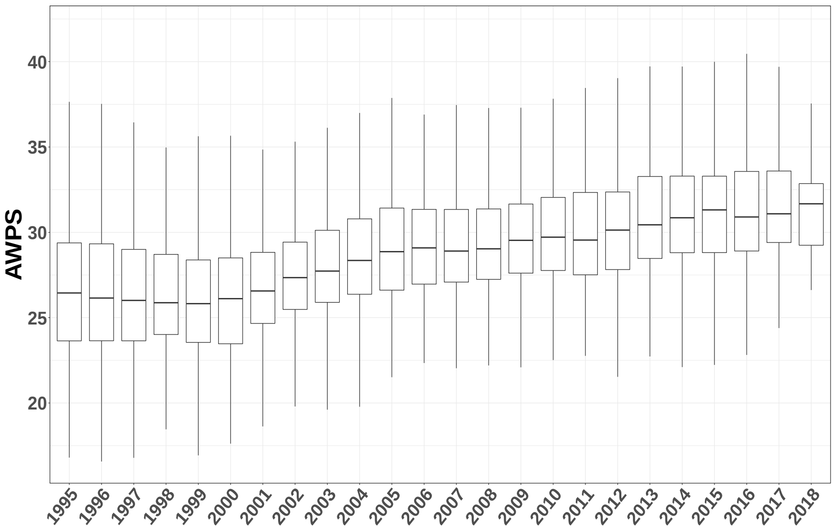

Figure 4: The boxplots of yearly MD&A ‘average words per sentence’ distribution.

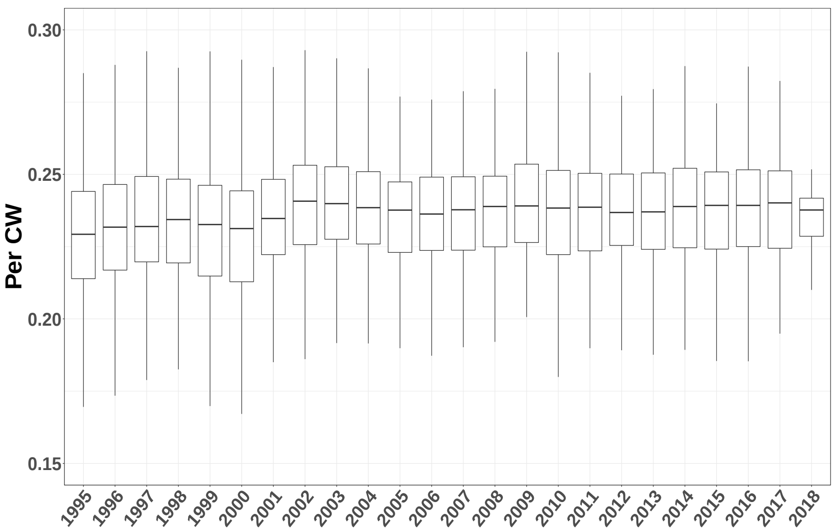

Figure 5: The boxplots of yearly MD&A ‘percentage of complex words’ distribution.

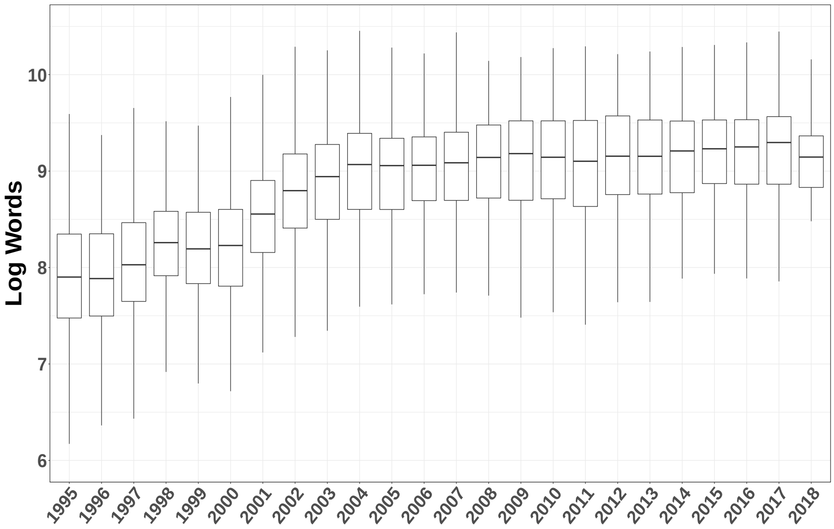

Figure 6: The boxplots of yearly MD&A log of words distribution.

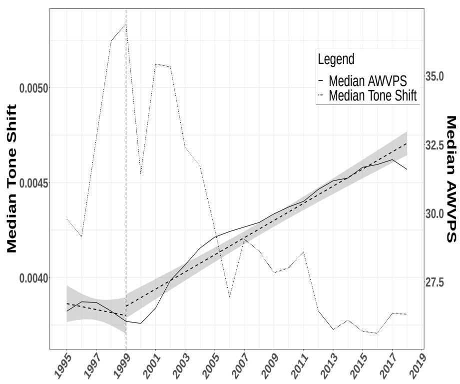

Figure 7: The plot presents the movement of SCI in 1994–2018 for the median firm. The dashed vertical line in 1999 denotes the implementation of the SEC Plain English rule (October 1998). The grey band around the trend denotes the 95% confidence interval. The solid line denotes the median 'average words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

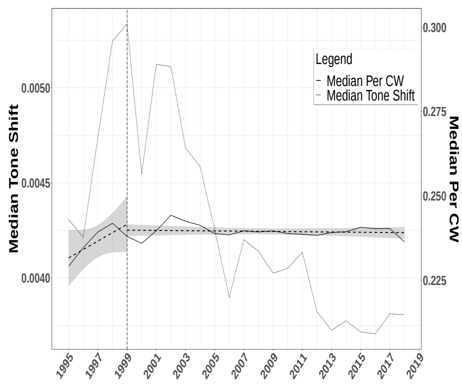

Figure 8: The plot presents the movement of SCI in 1994–2018 for the median firm. The dashed vertical line in 1999 denotes the implementation of the SEC Plain English rule (October 1998). The grey band around the trend denotes the 95% confidence interval. The solid line denotes the median '% of complex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

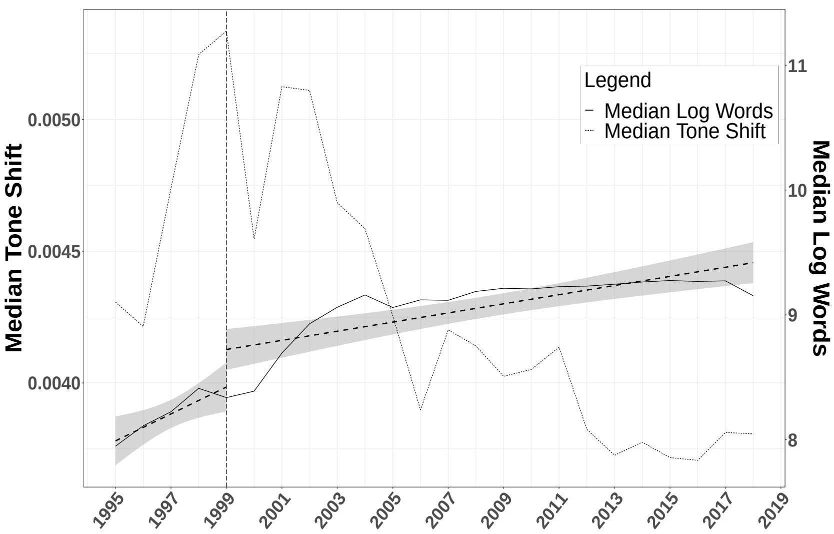

Figure 9: The plot presents the movement of SCI in 1994–2018 for the median firm. The dashed vertical line in 1999 denotes the implementation of the SEC Plain English rule (October 1998). The grey band around the trend denotes the 95% confidence interval. The solid line denotes the median of log of words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

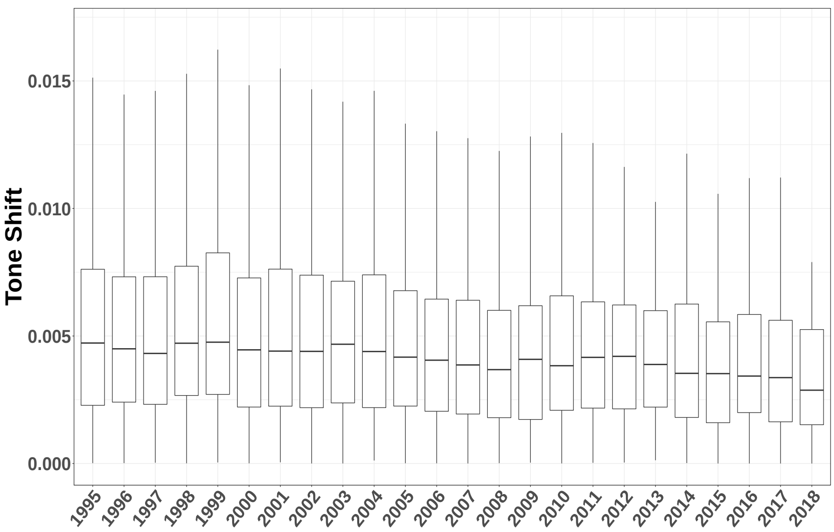

Figure 10: The boxplots of yearly 10-K SCI distribution.

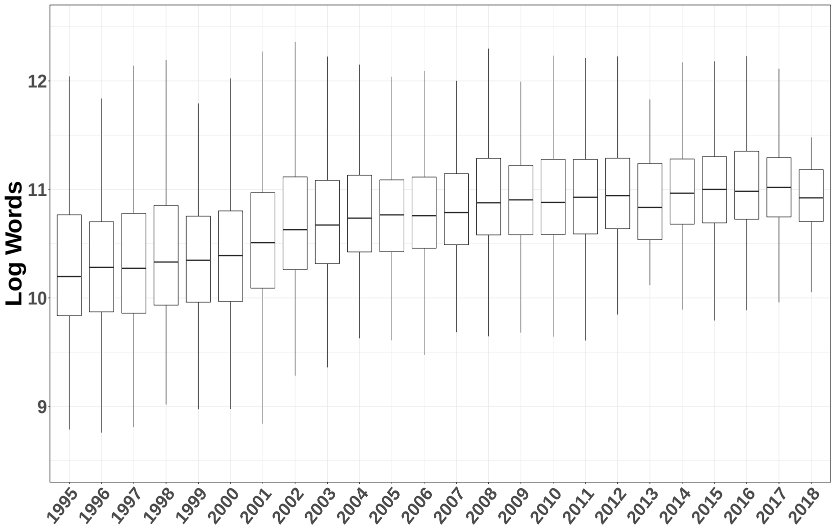

Figure 11: The boxplots of yearly 10-K log of words distribution.

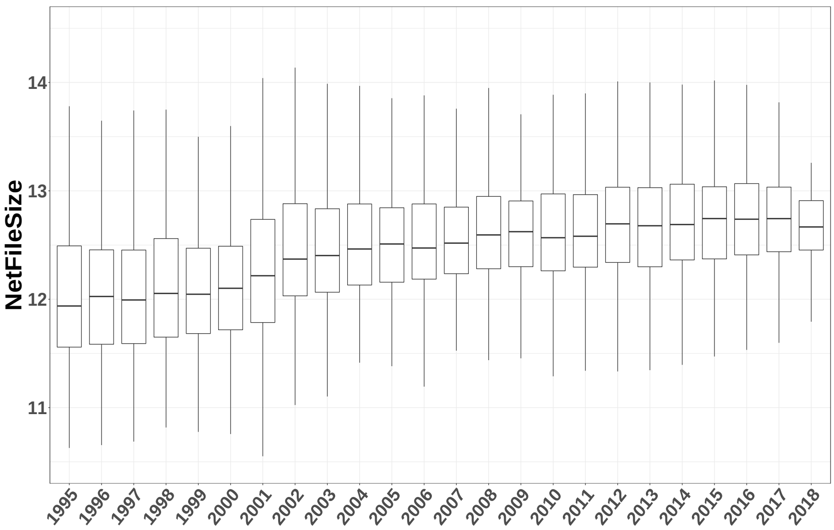

Figure 12: The boxplots of yearly 10-K net file size distribution, obtained after removing the graphics, XBRL and HTML elements from the size of the original 10-K documents.

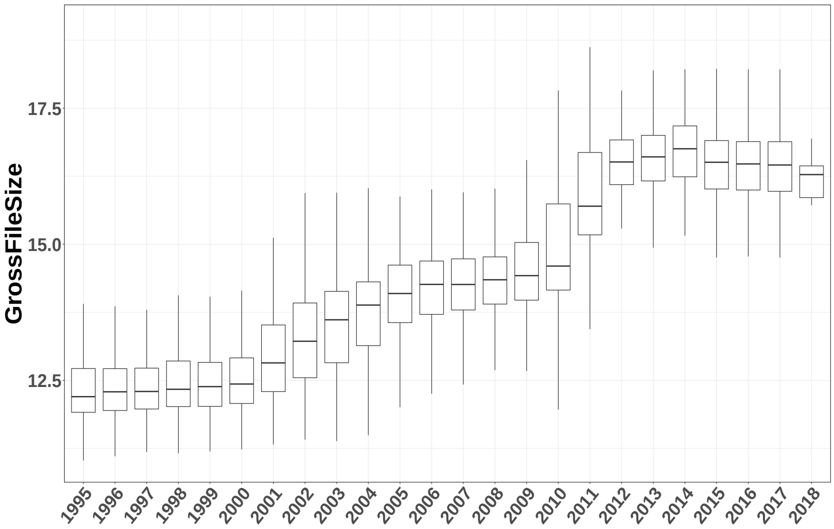

Figure 13: The boxplots of yearly 10-K gross file size distribution.

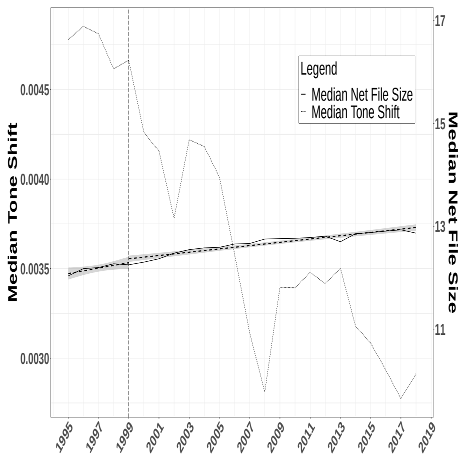

Figure 14: The plot presents the movement of SCI in 1995–2018 for the median firm. The dashed vertical line in 1999 denotes the implementation of the SEC Plain English rule (October 1998). The solid line denotes the median of net file size, obtained after removing the graphics, XBRL and HTML elements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

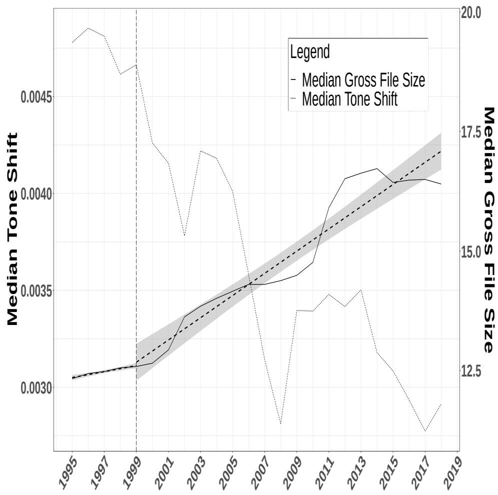

Figure 15: The plot presents the movement of SCI in 1995–2018 for the median firm. The dashed vertical line in 1999 denotes the implementation of the SEC Plain English rule (October 1998). The solid line denotes the median of gross file size in the 10-K section. The grey band around the trend denotes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

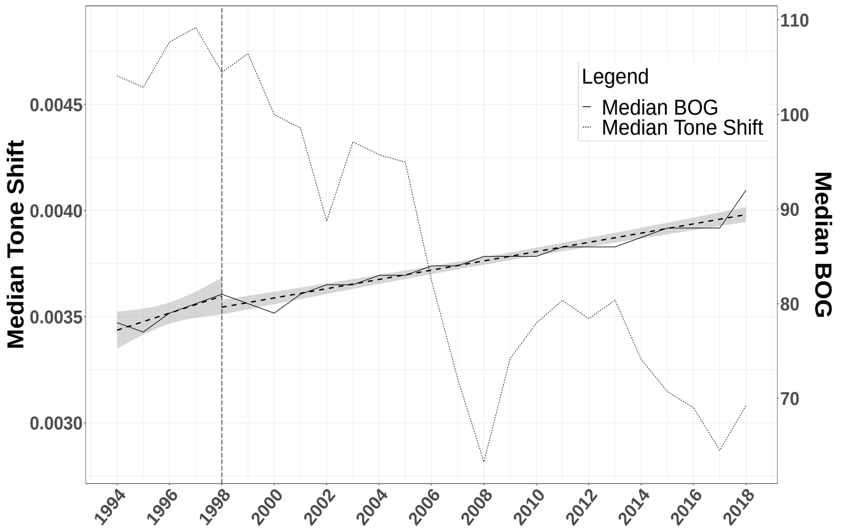

Figure 16: The timeseries movement of yearly 10-K BOG Index and SCI.

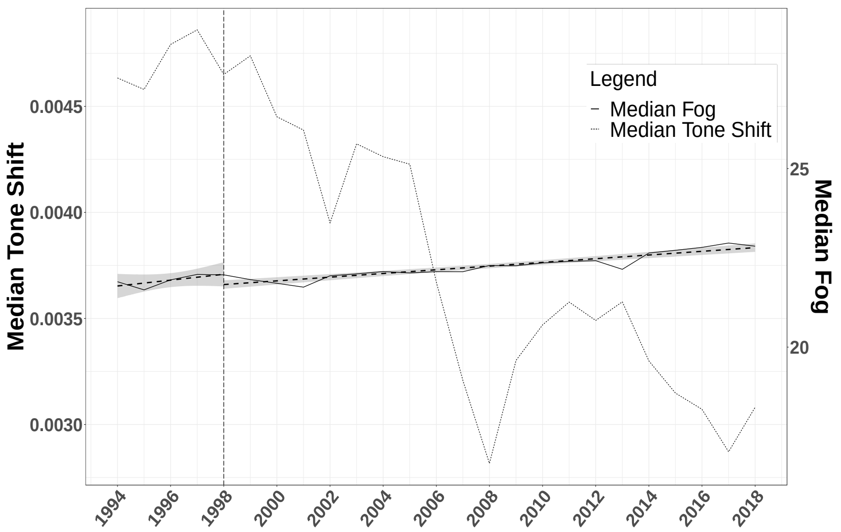

Figure 17: The timeseries movement of yearly 10-K Fog Index and SCI.

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.