TEXTUAL DISCLOSURE IN SEC FILINGS AND LITIGATION RISK

by

Arup Ganguly

Submitted to the Graduate Faculty of

the Joseph M. Katz Graduate School of Business

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH

JOSEPH M. KATZ GRADUATE SCHOOL OF BUSINESS

This dissertation was presented

by

Arup Ganguly

It was defended on

August 10th, 2018

and approved by

Dr. Kenneth M. Lehn
Samuel A. McCullough Professor of Finance
Katz Graduate School of Business

Dr. David J. Denis
Roger S. Ahlbrandt, Sr. Chair and Professor of Finance
Katz Graduate School of Business

Dr. Sara B. Moeller
Associate Professor of Finance
Katz Graduate School of Business

Dr. Chad J. Zutter
Associate Professor of Finance
Katz Graduate School of Business

Dr. Douglas M. Branson
W. Edward Sell Professor of Business Law
University of Pittsburgh

Dissertation Director: Dr. Kenneth M. Lehn, Katz Graduate School of Business

ABSTRACT

Textual Disclosure in SEC Filings and Litigation Risk

Arup Ganguly

Katz Graduate School of Business

University of Pittsburgh

Prior studies are quite ambivalent on the relation between disclosure and litigation risk since greater disclosure can be perceived as either *ex-ante deterrent* or *ex-post misleading*. I hypothesize that more information is disclosed in the non-numerical narratives in SEC filings than that has been analyzed in the extant literature. Using comprehensive hand-collected data on federal securities class action lawsuits spanning nearly two decades, matched peers, and widely used measures in natural language processing (NLP) that capture *degree*, *readability,* and *sentiments* in textual disclosures, I find results consistent with the theoretical view that argues that more and difficult to comprehend disclosure is often perceived as *ex-post misleading*, hence, increasing the odds of litigations. After controlling for other explanatory numerical variables, these results are robust to various empirical specifications using difference-in-differences (DiD), principal component analyses (PCA), and market response, across different types of shareholder class action litigations. Finally, using the Ninth Circuit Court of Appeals ruling, *Re: Silicon Graphics Inc.,* that led to an unexpected and sudden reduction in the threat of litigation for firms headquartered in the Ninth Circuit, I find that firms that are headquartered in the Ninth Circuit tend to use more uncertainty words in their filings post-shock, which is consistent with my main results. Such findings indicate that there is a need to distinguish between more versus better disclosures.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

I would like to thank my dissertation committee members, Professor Kenneth Lehn (Chair), Professor Douglas Branson, Professor David Denis, Professor Sara Moeller and Professor Chad Zutter, without whose guidance and helpful comments this thesis would not have been possible. I am also very grateful to the intellectual influence of my other professors, both at the University of Pittsburgh and at the Carnegie Mellon University who have taught me different Ph.D. courses at various points during my doctoral studies.

I thank Ms. Carrie Woods and Professor Dennis Galletta for their encouragement and support throughout the process. I also thank my colleague, Thuy Bui for her helpful conversations about the topic.

Finally, I would like to thank my wife, my parents and my younger brother for being my biggest supporters.

1. INTRODUCTION

The extant corporate finance and accounting literature, despite the extensive research conducted on the relation between information disclosure and securities litigation risk, is still starkly divided on the nature of this association. The *ex-ante deterrence* school of thought predicts a negative association between disclosure and litigation risk, arguing that securities law and regulations create deterring incentives that encourage greater disclosure, increasing transparency and value-relevant information and hence reducing the likelihood of lawsuits. Another school of thought deduces a positive association, contending that greater disclosure can often be perceived as overly optimistic and *ex-post misleading*, hence triggering lawsuits. While these opposing viewpoints provide researchers and policymakers with useful theoretical frameworks for understanding the role of disclosures in precipitating or abating shareholder litigations, the critical question of how disclosures affect litigation risk is ultimately an empirical one.

However, the empirical literature addressing this question is also split similarly in terms of their findings. While, empirical studies such as Skinner (1994) and Field, Lowry and Shu (2005) document that disclosure lessens the probability of a lawsuit, researchers such as Francis, Philbrick and Schipper (1994) and Rogers, Van Buskirk and Zechman (2011) indicate that greater disclosure results in more lawsuits. Lowry (2009) sums up this tension in the literature when she notes:

"... we are still left with the same question: what is the nature of the relation between disclosure and litigation risk?" (p. 159).

Empirical researchers in this field encounter two key challenges: First, prior empirical work is mainly focused on disclosures that are in numerical forms, such as earnings forecasts or announcements, various accounting variables, large stock-price drops, sales and earnings growth

and others. However, there is certainly more information and disclosure in the form of textual data (non-numerical narrative) that researchers have not yet fully explored in the context of litigation risk.[1] Second, any attempt to empirically analyze the relation between disclosure and litigation risk and to claim a directional causality is prone to issues associated with endogeneity and identification. This paper attempts to address both of these challenges. First, it focusses on textual disclosure, while controlling for all the non-textual explanatory factors that have been identified in the prior literature. And, second, although this paper does not claim causality, it addresses the endogeneity issues to some extent in several ways indicating that there is perhaps an underlying predictive relation between textual disclosure in SEC filings and the likelihood of securities class action litigations.

More precisely, I investigate the nature of the relation between information disclosure and securities class action lawsuits by first extracting textual information disclosed by public corporations that have been sued and their matched peers, in their main SEC filings, i.e., the 10-Ks and 10-Qs, and then employing panel data methods using various fixed effects, difference-in-differences (DiD) methodology, principal component analyses (PCA), event studies and a quasi-natural experiment to address endogeneity issues to a certain extent.

I pose three central questions: First, does more disclosure through text in SEC filings (10-Ks and 10-Qs) deter or trigger the incidence of securities class action lawsuits? Second, is readability, that is the ease with which a typical reader can comprehend the intended disclosed

---

[1] Text analytics experts have long claimed that 80-85% of business-relevant information is in textual form. See
https://breakthroughanalysis.com/2008/08/01/unstructured-data-and-the-80-percent-rule/
https://www.ibm.com/blogs/watson/2016/05/biggest-data-challenges-might-not-even-know/

message, associated with the probability of class action litigations? And third, are various disclosure tones or sentiments portrayed in the choice of words used in the narrative in SEC filings associated with the likelihood of shareholders' class action litigations?

I find a significant positive association between the degree of textual disclosure in both 10-Ks and 10-Qs and the risk of securities class action litigations. More specifically, I document that increasing the number of words by 2.718 times (approximate value of e = 2.718) in a 10-K (10-Q) filing on average results in 59% (120%) increase in odds of being litigated. In other words, if the odds that a particular firm would be litigated happens to be 1 to 1, then, all else equal, nearly tripling the number of words in a 10-K (10-Q) would increase the odds to more than 1.59 (2.20) to 1. Such results are significant at the 1% level and are robust to seven different proxies for textual measures of the degree of disclosure.

I further find, it is not just the degree of textual disclosure but also the difficulty level in comprehending the text used, i.e., its readability, that has significant predictive power in explaining the incidence of class action litigations. Using seven different readability measures as the main explanatory variable, I document a strong positive relation between the difficulty level of comprehending or understanding the text used and the probability of litigations. These results are also statistically significant at the 1% level and are robust to various controls used in the prior literature. Finally, I find that different sentiments induced with the choice of words in SEC filings can also be associated with the risk of being litigated in a class action. For instance, nearly tripling the number of uncertain words in a 10-K (10-Q) filing results in more than 79% (89%) increase in

odds of being litigated, significant at the 1% level. Such results are robust to using proportions instead of raw word counts and to different control variables.

These results are also consistent throughout, both in the case of 10-Ks and 10-Qs, with stronger results in the case of 10-Qs, indicating that textual disclosures in 10-Qs have a greater predictive power in explaining the incidence of litigations, possibly because they are more frequent and contain more up-to-date information at the time of their release. I also find that the use of words that proxy for *ex-ante* uncertainty in the 10-Ks and 10-Qs are significantly and positively associated with longer case periods, and more negative cumulative abnormal returns around the case filing dates. Finally, I use the Ninth Circuit Court of Appeals ruling, *Re: Silicon Graphics Inc.,* that led to an unexpected reduction in the threat of litigation for firms headquartered in the Ninth Circuit, and find that firms that are headquartered in the Ninth Circuit tend to use more uncertainty words in their filings post-shock, which is in line with my main results. Overall, such results support the theoretical view that argues more and difficult to comprehend disclosure is often perceived as *ex-post misleading*, hence, precipitating litigations.

This paper contributes to at least four different strands of literature. First, it contributes to the literature on the relation between voluntary information disclosure and litigation risk by incorporating non-numerical, textual form of disclosure that has largely been ignored in the earlier literature (Core, 2001). Managers use not only financial, and accounting numbers for disclosure but also use natural language and narrative to communicate information to their shareholders. To the best of my knowledge, this is a first such comprehensive study, spanning nearly two decades of data, on the relation between narratives in SEC disclosures and securities class action litigations,

post Private Securities Litigation Reform Act ("PSLRA") of 1995, that arguably made filing frivolous lawsuits difficult.

Second, this paper is closely related to a growing body of literature in finance and accounting that uses textual analysis to answer questions in corporate finance and accounting research.[2] Tetlock (2007) is one of the earliest studies in finance to have applied textual content analysis to a popular daily Wall Street Journal column to measure investor sentiments. Tetlock (2007) uses a widely-used Harvard's General Inquirer word list to measure sentiments. I employ the same idea to compute various sentiment measures based on the texts used in 10-Ks and 10-Qs, and I have also used Harvard's General Inquirer master lexicon. Loughran and McDonald (2011) recognize that word lists such as the Harvard's General Inquirer word list are inadequate and potentially misleading when used in the context of corporate filings as they note that almost three-fourths of the negative words in the Harvard's General Inquirer word list do not have a pessimistic connotation in the context of SEC filings. Therefore, they created six different word lists that are arguably more suited for textual analysis of financial documents and are freely available at Professor McDonald's website.[3] I use both Loughran and McDonald's (2011) and the Harvard's General Inquirer word lists for the word-content analyses in this paper.

Third, this paper provides some direct tests for the behavioral finance theories that apply psychology to finance and predict that manipulation of disclosures by firm's management can provoke different reactions from investors resulting in over- or under-valuation (Hirshleifer and

---

[2] Loughran and McDonald (2016), Das (2014) and Kearney and Liu (2014) provide excellent surveys on the use of textual analysis in finance.

[3] http://www.nd.edu/~mcdonald/Word_Lists.html

Teoh, 2003). More recently, Hirshleifer (2015) notes that *"Verbal communication, such as misleading disclosures......., can also be used to incite misvaluation." (p. 149).* This paper measures the various sentiments used by firms' management through their textual disclosures in filings and its influence on the likelihood of being litigated.

Finally, this paper is also related to the heated academic and policy debates on the need and optimum level of regulation for financial disclosure as it indicates that it is not just the amount but also the form, comprehensibility, and quality of disclosure that matters.

The paper has the following organization. The next section discusses the related literature. Following it, section 3 develops the hypotheses and section 4 describes the data and presents the summary statistics. The main results are presented in section 5. Section 6 uses a quasi-natural experiment to test how firms respond to a sudden reduction in litigation risk with respect to their textual disclosure in SEC filings. I conduct several robustness tests and address some potential endogeneity issues in section 7. Finally, I conclude in section 8.

2. RELATED LITERATURE

Securities class actions are typically triggered by stock price drops and filed when a publicly listed firm or its managers make an (alleged) untrue statement of material fact or (supposedly) omit a critical piece of information in their disclosures. Such false statements or intentional omissions can adversely impact firm valuation. However, evidence on the relation between information disclosure and shareholder litigation is mixed in the extant literature. While on the one hand it has been argued that forward-looking voluntary disclosure can prove to be costly if perceived as overly optimistic and sometimes *ex-post misleading*, on the other hand, greater disclosure could also reduce the probability of shareholder litigations *ex-ante* by reducing the chances of omission of a material fact or negative news. The *ex-ante deterrence* theory is also prescribed by regulators, who often work under this premise that more information is better than less, especially after the corporate accounting scandals like Enron and WorldCom, which resulted in a knee-jerk reaction of Sarbanes-Oxley-Act of 2002.

The voluntary nature of disclosure through texts further complicates this relation between disclosure and litigation, as firms can strategically and selectively choose to reveal information. Also, a significant portion of information revealed in textual narratives in filings often suffers from the non-verifiability problem and can be akin to "cheap talk" (Crawford and Sobel, 1982), which could be useless to the court. At the same time, the verifiable section of information revealed in narratives can be used as "signaling" (Spence, 1973) by high-type firms to differentiate themselves from the crowd.

7

Disclosure of positive versus negative news can also have distinctive impacts on the incidence of litigation. Skinner (1994) investigates earnings-related disclosures of a random sample of 93 NASDAQ firms during the period 1981-1990 and finds that firms take precautionary measures of voluntarily disclosing negative news to reduce the likelihood of shareholder litigation. Using data on 45 firms that were litigated during the period of January 1988 – September 1992, Francis, Philbrick and Schipper (1994) document an opposite result when they find that in their sample of litigated firms, early earnings warnings seemed to have precipitated shareholder litigations. Following these two influential papers, several academics have found evidence on both sides of the argument as Healy and Palepu (2001) point out in their survey paper, *"The empirical evidence on the litigation hypothesis is mixed." (p. 423).*

More recently, Field, Lowry and Shu (2005) recognize that the endogenous relation between information disclosure and shareholder litigation could be the potential cause of opposite results documented in the extant literature. They use a sample of 78 securities litigations that were filed between 1996 and 2000 and document a negative association between disclosure and litigation. Rogers and Van Buskirk (2009) take a time-series approach and investigate the change in disclosure behavior of firms after they have been litigated. Using a sample of 827 class action securities litigation cases filed during the period between 1996 and 2005, the authors report a significant decrease in the magnitude and precision of disclosures post-litigation and conclude that fear of litigation abets firms to reduce disclosure. However, the potential concern here is the generalizability of the results as their results are based on a sample of sued firms.

To the best of my knowledge, the only two papers that have analyzed textual content in relation to litigation risks are Rogers, Van Buskirk and Zechman (2011) and Hanley and Hoberg (2012). While Rogers, Van Buskirk and Zechman (2011) investigated the disclosure tone of a random sample of 20 firms that were litigated and conclude that the use of positive language in disclosures accentuates litigation risk, Hanley and Hoberg (2012) focussed their attention on IPO related litigations and utilizing word content analyses of IPO prospectuses document that greater disclosure is a substitute for underpricing and is efficacious in reducing the likelihood of all types of IPO related lawsuits. Interestingly, both these studies using textual analysis documented diametrically opposite results. In contrast, in this paper, I analyze a comprehensive sample of federal securities class action lawsuits during the time-period 1996-2014, and also examine the readability, besides degree and sentiments, in SEC filings by these firms as Hwang and Kim (2017) have recently documented that low readability of disclosure documents can cause investors to doubt and discount a firm's value.

Even after fifteen years, since the publication of the survey paper by Healy and Palepu (2001), in another comprehensive and more recent survey paper on the economics of disclosure, Leuz and Wysocki (2016) observed that, *"…the evidence regarding the effects of litigation on disclosure is mixed and also quite subtle or nuanced…" (p.552).*

Overall, researchers are still divided on the nature of the association between the degree of information disclosure and the risk of securities litigations. Therefore, this paper systematically examines all the federal cases of securities class action lawsuits filed after the Private Securities Litigation Reform Act of 1995, between January 1, 1996 and December 31, 2014, and tracked by

the Securities Class Action Clearinghouse (SCAC) database and textually analyzes 10Ks and 10Qs

of 2,137 litigated firms and 2,137 matched peer sample based on industry and firm characteristics.[4]

---

[4] I start with 3,899 securities class action litigations filed in the Federal Court during the 1996-2014 period and after matching with stock price data from CRSP and accounting data from Compustat, my final sample comprises of 2,335 cases of securities class action lawsuits.

3. HYPOTHESES DEVELOPMENT

The empirical predictions from theory on the association between disclosure and litigation risk are not always clear-cut. Similar to the economics of any other law enforcement, the *deterrence theory* on securities class action litigations hypothesizes that managers and executives of publicly listed firms should (*ex-ante*) respond proactively by enhancing voluntary disclosure to the deterring incentives created by securities law and regulations. Such line of thinking predicts a negative association between disclosure and incidence of litigations since with greater disclosure and transparency there is less likelihood of omission of value-relevant information and consequently lower litigation risk. More disclosure also makes it increasingly difficult for plaintiffs to establish "loss causation."[5] However, another theoretical perspective takes an *ex-post* view on this issue, arguing that greater disclosure can often be perceived as overly optimistic or overconfident and *ex-post* misleading that could potentially precipitate securities class action litigations. Banerjee, Humphery-Jenner, Nanda and Tham (2018) also found empirical evidence that the presence of an overconfident CEO or a senior executive in the firm increases its likelihood of being litigated in a securities class action. This line of literature predicts a positive association between disclosure and incidence of litigations. Therefore, it is ultimately an empirical question, and because of this ambiguity in the relation between disclosure and risk of litigations, I do not provide a directional hypothesis and ask:

---

[5] "Loss causation" is a legal requirement for plaintiffs in securities class action lawsuits to show that the damage to shareholders was inflicted by information omission.

Does more disclosure through text in SEC filings deter or trigger the incidence of

securities class action lawsuits?

Another essential aspect of word content analysis is readability, which is often defined as the ease with which a typical reader can understand the intended message. It can be argued that it is not just the quantity of textual disclosure that can impact the incidence of class action lawsuits but also the ease of its interpretation that can influence the likelihood of lawsuits. Complex and difficult to comprehend language can potentially cause divergence in opinions on the same text, increasing the likelihood of litigations. Greater disclosure in abstruse language can often be perceived as confusing noise and can lessen transparency. Shareholders are constricted by bounded rationality (Simon, 1955) and limited cognitive ability to process information. Even experts and institutional investors can be prone to information overload (Hirshleifer and Teoh, 2003; Biggs, Bedard, Gaber and Linsmeier, 1985). Note that unstructured textual data can only be viewed as useful information, once it has been analyzed and interpreted, which requires time and effort. Hence, borrowing from the extant literature in natural language processing (NLP), computational linguistics and stylometry, I compute the commonly used readability indices for the SEC filings (10-Ks and 10-Qs) and hypothesize:

Other things equal, there will be a significant positive association between the

difficulty level in readability and the incidence of class action lawsuits.

Besides the degree of disclosure and disclosure readability, disclosure sentiments or tone could potentially shape the relation between disclosure and litigations. Textual sentiments or tone analysis has been widely used in finance research, in which certain word lists have been created

from dictionaries with financial text in mind that convey sentiments such as positive words, negative words, uncertain words, litigious words, etc. Kearney and Liu (2014) have summarized the different techniques used in textual sentiment in the finance literature. One can also think of sentiments as a common cognitive error. The basic idea here is that both in the world of rational (Angeletos and La'O, 2013; Benhabib, Wang and Wen, 2015) and behavioral finance (Akerlof and Shiller, 2010; Shiller, 2015), the interpretation of textual sentiments can have a profound impact on shareholders' behavior. Hence, I conjecture that:

Ceteris paribus, various disclosure tones or sentiment measures will have differential effects on the likelihood of shareholders' class action litigations.

The main thrust of these hypotheses is to explore the direction and magnitude of the association between the three different dimensions (i.e., *degree*, *readability,* and *sentiments*) of textual disclosures in SEC filings and the risk of securities class action lawsuits.

4. DATA AND SUMMARY STATISTICS

4.1 Litigation Data

I manually collect data on all securities class actions litigations filed in the Federal Court for the years 1996 to 2014. My primary source of litigation data is the Securities Class Action Clearinghouse (SCAC), a free, online database hosted by Stanford Law School in collaboration with Cornerstone Research.[6] SCAC is one of the most widely used and prominent databases (Karpoff, Koester, Lee, and Martin, 2014) on securities class action lawsuits and encompasses information on federal civil securities class action lawsuits starting from 1996. I use SCAC database to collect data on indicator variable for the securities class action and other case details such as the case filing date, case status, case end date, case docket number, beginning of the class period and end of the class period. I also manually collect data on the settlement amount of the class actions if available from case summaries, 10-Ks, 10-Qs or 8-Ks, and media articles.

Although I begin with 3,899 cases of federal civil securities class action litigations filed during the period 1996-2014, post-matching with stock price data from CRSP and accounting data from Compustat, my final sample constitutes of 2,335 cases of which 1,285 cases have been settled (including the ones adjudicated at trial), 917 cases were dismissed, and 133 cases were still active at the stage of data collection. Table 1 provides summary statistics on litigation data of my sample. Note that during the sample period under study, the number of securities litigations peaked in 2001,

---

[6] http://securities.stanford.edu/

which was partly due to the dot-com bubble crash of 2000 and it declined over time post-Sarbanes-Oxley-Act of 2002. The mean case period, which is the time between the case filing date and case end date is 1,157 days and the mean class period, which is the time between the class start date and class end date is 506 days. The mean settlement amount is $28.2 million with a maximum of $3.2 billion and a minimum of $37,500. Appendix A provides the distribution of the litigations based on two-digit SIC industry code. The top three most frequently sued industries in my sample are business services, chemicals and allied services and electronic and other electrical equipment & components.

[Insert Table 1 here]

4.2 Matched Sample

Matching on observable characteristics is one of the popular statistical techniques used to address certain endogeneity issues in empirical corporate finance research (Roberts and Whited, 2013). I start with 2,335 litigated firms and search the remaining population of firms not litigated that best match my "treated" firms in the following dimensions: 2-digit SIC code (perfectly matched), year (perfectly matched), size (measured by market value), return on assets (ROA), loss indicator, earnings growth and sales growth. For the set of analyses including IBES variables, I also match control sample with treated sample in terms of negative earnings surprise and number of analysts following. I select the nearest matched control firm for each of my "treated" firms, and the matching process is done without replacement to ensure independence among control firms

and to avoid multiple appearances by control firms.[7] Treated firms are dropped if no matched

control firms are found. Finally, I have 2,137 litigated firms as "treated" and 2,137 non-litigated

firms as a matched peer sample. Table 2 checks the covariate balance to ensure that the "treated"

and the "control" firms are similar in terms of the explanatory characteristics that influence

disclosure.

[Insert Table 2 here]

Not surprisingly, given my matching criteria, the observable characteristics of treated

(firms that are litigated) and control (firms that are not litigated) sample are not significantly

different as shown in Table 2. The differences both in the means (p-values reported in the last

column of Table 2) and the medians (untabulated) are not significant.

4.3 Textual Analysis Data

I use a web crawler to download the 10Ks and 10Qs from SEC's EDGAR (Electronic Data

Gathering, Analysis, and Retrieval) system.[8] To clean the filings before creating various textual

measures of disclosure, i.e., degree, readability and sentiments, I closely follow the methodologies

of Li (2008), Miller (2010), Loughran and McDonald (2011) and Hwang and Kim (2017) with

minor differences.

First, I convert the pdfs into ASCII format. Then, I remove the graphics, XBRL and the

unwanted markup tags (XML). I also remove the content between <SEC-HEADER> and the

---

[7] Results are also robust to matching with replacement.
[8] https://www.sec.gov/edgar.shtml

</SECHEADER> tags as it simply contains firm's information such as name, address, year, etc. Since tables used in filings may or may not contain text, I only include the tables that have more than 65% alphabetical characters. I re-encode special characters like "&AMP" and remove the obvious proper nouns.

Finally, I extract and parse texts from all the 10-Ks (and its variant:10KSB) and 10-Qs (and its variant:10QSB) SEC filings of my sample of litigated firms and the propensity-score matched sample from 1994 to 2014 using the programming language Python to create various variables used in degree of disclosure, readability and sentiments.[9] I focus on two main SEC filings, i.e., 10-Ks and 10-Qs as they are the two most frequently cited SEC filings in securities class action litigation complaints (Rogers, Van Buskirk and Zechman, 2011). I analyze all the 10-Ks and 10-Qs of my sample firms and their matched sample, not only 365 days pre- and post- the case filing date, but also, a year before and after the year of the case filing date, as a robustness test. I also use two word lists or lexicons, namely, the Harvard's General Inquirer word list and Loughran and McDonald (L-M) textual sentiments word list that are freely available and have been extensively used in the extant accounting and finance literature (Das, 2014; Loughran and McDonald, 2016) to construct textual sentiment variables.[10,11] While the Harvard's General Inquirer word list has been used widely for language analytics in many different fields, L-M word list is more suited to finance research, especially for textual analyses of SEC filings. The reasoning here is that certain

---

[9] 10KSB and 10QSB used to be filed by smaller companies. SEC removed such option of 10QSB on October 31st, 2008 and the option of 10KSB on March 15th, 2009. Results are qualitatively similar without the inclusion of 10KSB and 10QSB filings.

[10] http://www.wjh.harvard.edu/~inquirer/homecat.htm

[11] http://www3.nd.edu/~mcdonald/

words that may have negative connotations in other areas may not have the same meaning in financial documents as Loughran and McDonald (2011) argue, *"Words such as tax, cost, capital, board, liability, foreign, and vice are on the Harvard list. These words also appear with great frequency in the vast majority of 10-Ks, yet often do no more than name a board of directors or a company's vice-presidents. Other words on the Harvard list, such as mine, cancer, crude (oil), tire, or capital, are more likely to identify a specific industry segment than reveal a negative financial event." (p.36).*

After matching with Compustat, CRSP and litigation database and pre-screening the matched sample based on several observable characteristics, I textually analyze the 10-Ks and 10-Qs of 2,137 litigated firms and 2,137 matched peer firms. Since companies file multiple 10-Qs and in some cases multiple 10-Ks, I take the average of textual variables in a particular firm-year-filing type.

4.3.1 Disclosure Variables

Disclosures in SEC filings are not just quantitative but are also text-based or narrative in nature. I construct textual measures of the degree of disclosure by using seven different proxies such as the file size, word count, complex word count, sentence count, average words per sentence count, paragraph count and average words per paragraph. Appendix B provides the variable definitions for each of these disclosure variables. It can be argued that the bigger the file size or, the higher the various word counts, the greater is the degree of disclosure. Table 3.1 Panel A provides summary statistics of the textual disclosure variables for the 10-K filings from 1994 to

18

2014 by litigated firms versus matched non-litigated firms. The results in Panel A show that the degree of textual disclosure is significantly higher for firms that are litigated as compared to the matched firms that are not litigated in most of the measures (except when measured by "file size" and "Average No. of Words per Paragraph", where the differences are not significant with p-values of 0.4323 and 0.5128 respectively). For instance, while the mean word count of 10-K filings for firms that were litigated is approximately 43,391, the mean word count of firms that were not litigated is around 38,563. The p-values for the differences in means are reported in the last column. The p-values (untabulated) for the differences in medians are qualitatively similar. I document similar results for the 10-Q filings in Table 3.2 Panel A. Such results indicate that *ex-post* there seems to be a positive association between the degree of disclosure and incidence of litigations.

<div align="center">

[Insert Table 3.1 here]

[Insert Table 3.2 here]

</div>

4.3.2 Readability Variables

Next, I construct seven different readability variables, namely Flesch Reading Ease Index, Flesch-Kincaid Readability Index, RIX Readability Index, Gunning Fog Readability Index, Automated Readability Index, Smog Readability Index and Lasbarhets Readability Index for the 10-K and 10-Q filings following extensive literature in computational linguistics. The details on the construction of these seven variables have been provided in the Appendix B. The main goal of all these readability measures is to come up with a scale, often using a linear combination of

sentence and (or) words characteristics that would indicate the degree of difficulty in comprehending a textual document. Except for the Flesch Reading Ease Index, the higher is the value of the readability variable, the greater is the degree of difficulty in understanding the intended message of the text. Panels B of Tables 3.1 and 3.2 present the summary statistics of the readability variables for 10-Ks and 10-Qs respectively. As hypothesized, for both 10-Ks and 10-Qs, these readability measures indicate that readability is significantly more difficult for firms that have been litigated versus the matched peer firms that were not litigated indicated by the p-values of their differences in means and medians (untabulated).

4.3.3 Sentiment or Tone Variables

Finally, following the previous accounting and finance literature (Das, 2014; Loughran and McDonald, 2016) in textual analysis, I construct fourteen different sentiment measures. The definitions of all these measures have been provided in the Appendix B. Two of these measures, the Harvard Negative Word Count and the Harvard Negative Word Proportion have been created using Harvard's General Inquirer word list, while twelve L-M sentiment variables use Loughran and McDonald's (L-M) textual sentiments word lists. Panel C (sentiment variables are measured in proportion) and Panel D (sentiment variables are measured by count) of Tables 3.1 and 3.2 present the summary statistics of the sentiment variables for both 10-Ks and 10-Qs respectively.

Some key points to note from the summary statistics. First, note that the negative word count and also the negative word percentage (proportion) used in 10-Ks and 10-Qs are significantly higher for the firms that have been litigated. This is true for both L-M Negative Word (Count and

Percentage) and Harvard Negative Word (Count and Percentage). These results suggest a positive association between the negative tone set in the filings and the incidence of litigation.

Second, the positive word count and word percentage (proportion) used in 10-Ks and 10-Qs are also significantly higher for the litigated firms in comparison to the non-litigated firms that indicates that positive tone or sentiments are perhaps construed as overly optimistic or *ex-post* misleading.

Third, there seems to be a significantly higher use of L-M weak modal words (e.g., may, might, could, etc.) in 10-Ks and 10-Qs of the litigated firms. It is plausible that weak modal words signal trouble or wrong-doing in a firm.

Fourth, the number and percentage of the L-M uncertainty words (e.g., depend, uncertain, indefinite, etc.) used in 10-K and 10-Q filings of litigated firms are also significantly higher than the non-litigated firms suggesting that uncertainty or ambiguity in tone could increase the likelihood of being litigated in a securities class action.

Last but not the least, I use L-M's litigious word list and find that firms that have been litigated have used significantly higher number and proportion of litigious words (e.g., claimant, testimony, tort, etc.) which could be possibly signaling a more litigious environment and hence increasing the probability of litigations. Again, the p-values (untabulated) for the differences in medians are qualitatively similar. Overall, such univariate tests indicate a positive association between the degree or difficulty level in comprehension of textual disclosure and incidence of securities class action lawsuits. Moreover, these univariate results also suggest a positive

association between the use of uncertain or ambiguous words and the likelihood of securities class action lawsuits.

4.4 Other Independent Variables

The other independent or control variables used have been selected based on the extant literature studying the relation between disclosure and litigation risk. Daily stock price data used to compute volatility comes from CRSP. Accounting data such as the firm size, market-to-book ratio, return on assets, earnings growth, sales growth and auditor quality have been taken from Compustat. Data on analyst following and negative earnings surprise is gathered from the IBES database. Appendix B describes these control variables. The next section discusses the main results in multivariate settings.

5. MAIN RESULTS

5.1 Litigation Risk

5.1.1 Degree of Disclosure

So far univariate tests on the degree of textual disclosure have revealed a key finding that the degree of textual disclosure is significantly higher for the litigated firms as compared to the matched sample of non-litigated firms. I further test these results in a multivariate setting controlling for different firm and performance characteristics that have been found to be correlated with the incidence of litigation in the extant literature. I use a logit model where the regressand is a dichotomous variable indicating the incidence of securities class action lawsuit (1 for litigated and 0 for not litigated) and seven different proxies of textual measures of degree of disclosure as the main explanatory variable. I use the following empirical specification:

$$\textit{Litigation Dummy }_{i,t} = \beta_0 + \delta * Disclosure_{i,t-1} + \beta_1 * X_{i,t} + \beta_2 * X_{i,t-1} + \varepsilon_{i,t} \qquad (1)$$

The seven different proxies that have been used to measure the degree of textual disclosure are Ln (File Size), Ln (Word Count), Ln (Complex Word Count), Ln (Sentence Count), Ln (Average Words per Sentence), Ln (Paragraph Count) and Ln (Average Words per Paragraph). δ captures the effect of the degree of textual disclosure on the probability of the firm being litigated. $X_{i,t}$ are the firm-level control variables, some of which are lagged, as suggested in the extant literature. The results for 10K and 10Q filings are shown in Table 4.

[Insert Table 4 here]

23

As shown in the table, I run separate regressions (as suggested by Loughran and McDonald, 2013) for each of these measures of textual disclosure given the high correlations (as shown in Appendix D) between the different measures. All the regressions have year and industry fixed effects to control for unobserved heterogeneity across time and industries, and the standard errors have been clustered at the firm level. I also include as controls, variables that have been found significantly related to the likelihood of getting litigated in the extant literature, such as the firm size, return on assets (ROA), loss indicator (a dummy variable that equals to 1 if net income for the year is negative, and 0 otherwise), earnings and sales growth, market-to-book, big-8 auditor dummy, in addition to the lagged values of firm size, ROA, Tobin's Q, loss indicator, stock volatility (measured as the standard deviation of daily stock returns, measured over a 365-day period), stock return and institutional ownership.[12] In additional tests (Appendix C), I also include analyst following and negative earnings surprise, and the results stay the same qualitatively. Since the inclusion of IBES variables considerably reduces the sample size, I do not include them in the main results. While firm size and Tobin's Q have been used as control variables in several textual analysis research (Tetlock, Saar-Tsechansky, and Macskassy, 2008; Loughran and McDonald, 2011), earnings and sales growth, return on assets, analyst following, negative earnings surprise, volatility, and loss indicator have been found to be correlated with both tone and litigation risk (Rogers, Van Buskirk and Zechman, 2011). The extant literature has shown that auditors can also influence the quality and content of disclosures, which can impact the likelihood of being litigated. DeAngelo (1981) argues that auditor quality is associated with auditor size as bigger auditors with

---

[12] For brevity, the coefficients on the vector of firm-level controls have been reported in Appendix E.

numerous clients are less dependent on "client specific quasi-rents" and hence provide better audit quality. Therefore, I control for audit quality using the Big 8.[13] More recently, Bird and Karolyi (2016) show that institutional ownership can also impact firm disclosure and, hence, I control for it in all the specifications.

The results show that there is a significant positive association between the degree of textual disclosure in both 10-Ks and 10-Qs and the incidence of securities class action litigations. Six out of the seven models depict this positive association in the case of 10-Ks, and all seven models illustrate this relation for 10-Qs, all significant at the 1% level. For instance, the coefficient on the ln (word count) is 0.463, which is significant at the 1% level. Therefore, the odds of being litigated (or the odds ratio) is $e^{0.463} = 1.59$, which shows that there is a 59-percentage change in odds of being litigated. In other words, increasing the number of words by 2.718 times (approximate value of e = 2.718) would result in 59 percentage change in odds of being litigated. That is, simply tripling the number of words used in a 10-K would result in more than 59 percentage change in odds of being litigated which is both statistically and economically significant.

Similarly, by tripling the number of complex words used in the 10-K filing, we would expect to see more than 65 percentage increase in the odds of being litigated.[14] I also find a significantly positive association between litigation and file size, sentence count, average words per sentence count and paragraph count. The results for 10-Q filings are even stronger as tripling

---

[13] In untabulated results, I also tried Big 6, Big 5 and Big 4 auditors and the results are robust.
[14] Words containing three or more syllables.

the number of words would result in more than 120 percentage change in odds of being litigated. In the case of 10-Qs, the ln (Average Words per Paragraph) is also positively associated with the risk of litigation, significant at the 1% level. Such stronger results make sense as it can be argued that the disclosure released in 10-Qs are more timely, proximate, frequent, and have been more recently updated. Such results are also robust to the inclusion of IBES control variables as presented in Appendix C.1. Overall, the multivariate results show a strong positive association between the degree of textual disclosure and the incidence of securities class action lawsuits, robust to different proxies of disclosure and various controls that have been used in the literature.

5.1.2 Readability

My next set of results answer the question whether there is an association between the difficulty level in readability, as measured by various readability indices, and the incidence of class action lawsuits. Note that readability is a different feature of textual analysis and is distinct from the degree of disclosure. Greater textual disclosure may not necessarily mean better readability. However, it can be argued that more readable 10-Ks and 10-Qs should be more informative to investors (Loughran and MacDonald, 2014; Hwang and Kim, 2017). As explained in the hypotheses development section, I conjecture a positive association between the incidence of litigation and readability difficulty. My empirical specification is as follows:

$$\textit{Litigation Dummy }_{i,t} = \beta_0 + \delta * \textit{Readability}_{i,t-1} + \beta_1 * X_{i,t} + \beta_2 * X_{i,t-1} + \varepsilon_{i,t} \qquad (2)$$

The dependent variable, that is the litigation dummy, takes the value of 1 for the incidence of litigation, and 0 otherwise. The main independent variable in this specification is the readability

measure, which measures the level of difficulty in comprehending the intended message of the text. I construct and use seven different readability measures borrowed from the extant literature.[15] In this model, δ captures the effect of readability of the text used in 10-K and 10-Q filings on the probability of the firm being litigated. The results are shown in Table 5. There is a strong positive and significant association between the degree of difficulty measured by various readability indices and the probability of litigation. For example, in the second regression in Table 5 Panel A, where the main predictor variable is Flesch-Kincaid Readability Index, the coefficient is positive and statistically significant (t-statistic of 3.68). The negative sign of the coefficient of Flesch Reading Ease Index, which is opposite to the sign on coefficients of all the other readability indices, is because of the way it is measured. In case of Flesch Reading Ease Index, the higher the score, the easier the text is to read. For all the other indices, the higher the score, the more difficult the text is to read. Details on how these seven different indices have been created are provided in the Appendix B. All models include year and industry (2-digit SIC code) fixed effects, with standard errors clustered at the firm level and the same controls as used in the prior literature. These results, albeit a bit weaker in terms of significance levels in the case of 10-Ks, are also robust to the inclusion of analyst following and negative earnings surprises as shown in the Appendix C.2.

[Insert Table 5 here]

---

[15] Flesch Reading Ease Index, Flesch-Kincaid Readability Index, RIX Readability Index, Gunning Fog Readability Index, Automated Readability Index, Smog Readability Index, and Lasbarhets Readability Index.

Such results indicate that readability of 10-K and 10-Q filings have significant predictive power in explaining the incidence of class action litigations.

5.1.3 Sentiments

My subsequent set of main results focuses on the influence of common sentiments generated by choice of words in the texts. I rely on prior literature to measure tone or sentiments of 10-Ks and 10-Qs. Using fourteen different commonly used sentiment measures as explained in variable definitions in Appendix B, I conduct the following test:

$$\textit{Litigation Dummy }_{i,t} = \beta_0 + \delta * Sentiments_{i,t-1} + \beta_1 * X_{i,t} + \beta_2 * X_{i,t-1} + \varepsilon_{i,t} \qquad (3)$$

In this empirical specification, my main covariate of interest is the measure of sentiments. The response variable is again a dummy variable indicating the incidence of litigation. The control variables are the same as used in the previous specifications. The results are shown in Table 6.1 and Table 6.2. All models include year and industry (2-digit SIC code) fixed effects, with clustering done at the firm level. Standard errors have been reported in parentheses. In order to construct these sentiments variables, I have relied on two sources of word lists, namely, Loughran and McDonald's (L-M) textual sentiments word lists and Harvard's General Inquirer word list. Due to the high correlations (Appendix D) and word overlap in these lists, I run the above specification separately for each list.[16] While Table 6.1 examines the link between the probability of being litigated and the various sentiment measures as a raw word count used in 10-Ks and 10-

---

[16] Loughran and McDonald (2013) also recommend running the regressions separately for each word list.

Qs respectively, Table 6.2 repeats the same regressions using word proportion or percentage as the main independent variables.

[Insert Table 6.1 here]

[Insert Table 6.2 here]

Table 6.1 presents the relation between the sentiments generated, measured as a raw count and the probability of a firm being litigated. The first model shows a strong positive association between the use of uncertain words and the likelihood of being litigated. Loughran and McDonald's uncertain word list contains words such as "ambiguity," "anomalous," "confusing," "contingent," etc., that signal indecision or lack of conviction. I find that the coefficient on the ln (uncertainty word count) is 0.580, which is significant at the 1% level. To put it differently, tripling the number of uncertainty words in 10-K filing results in more than 79% percentage increase in odds of being litigated, significant at the 1% level. Model 2, in Table 6.1 analyzes the influence of the use of weak modal words on the chances of being a target of litigators. Weak modal words are words such as "depending," "possibly," "sometimes," "maybe," etc., that are associated with management's inability to have a clear vision, also significantly increase the probability of being litigated. The coefficient of interest on the ln (modal weak word count) is 0.527, significant at the 1% level, which means that tripling the number of weak modal words used in 10-K filings would result in more than 69 percentage increase in odds of being litigated. Model 3 shows the results of using negative words, that is, words such as "fails," "flaw," "exaggeration," "loss," etc., that have a negative connotation also have a similar positive association with the incidence of securities class action lawsuits. The coefficient is 0.484, significant at the 1% level. It is possible that the negative

sentiments generated in the minds of shareholders with the use of negative words in 10-K filings are increasing the chances of being litigated. Such results are robust to using alternative word lists such as the Harvard negative word count, which also gives similar significant results as shown in model 7. In model 5, as expected, the use of litigious words, also significantly increases the chances of being litigated. Litigious words refer to words like "settlement," "contracts," "acquit," "indemnify," etc., that have a legal connotation. Model 5 documents that the coefficient on the ln (litigious word count) is 0.323, which is significant at the 1% level. In other words, simply tripling the number of litigious words in 10-K filing would result in more than 38 percentage change in odds of being litigated. However, surprisingly, the use of positive words, that arguably generate positive sentiments, also has a positive association with the likelihood of litigation as shown in model 4. Repeating the tests in the context of 10-Q filings in Table 6.1 (Panel B) yields similar significant results with larger effects. This is because 10-Qs are more frequently updated and contain more up-to-date information.

Loughran and McDonald (2016) argue, *"In most instances we do not want to use the raw count, since this is obviously strongly tied to document length." (p.26).* Therefore, in the next set of tests in Table 6.2, I use word proportions as suggested by Loughran and McDonald (2016) as a solution to this issue as the main covariate. The results are still significant with both 10-K and 10-Q filings except for positive words where the sign flips and is only significant in the case of 10-Qs. Such a result makes more sense as it can be argued that the greater the percentage of positive words used in the filings (that generate positive sentiments), the lesser are the chances of being litigated. More specifically, I find that in case of 10-Q filings, for each one percentage point

increase in the proportion of positive words, the log odds of being litigated (versus not-litigated) decreases by (39.918/100) % or 0.3992%, significant at the 5% level. Note that besides the opposite sign, both significance and magnitude of the coefficients for positive word proportion is less than the coefficients for negative word proportion which supports the view of asymmetric effects of positive and negative news as predicted by theoretical models such as Veronesi (1999) and Epstein and Schneider (2008) and empirically documented by both Tetlock (2007) and Loughran and McDonald (2011). People tend to be affected asymmetrically more by the negative news as compared to positive news.

Finally, I test the robustness of these findings by including analyst following and negative earnings surprises in my regressions, and the results stay the same qualitatively as shown in Appendix C.3 and C.4. Overall, such statistically significant and economically meaningful results indicate that sentiment measures created by textual analytics have significant predictive power in explaining the incidence of shareholder class action litigations.

5.1.4 Principal Component Analyses (PCA)

Are the proxies used for measuring degree, readability, and sentiments accurately gauging the targeted characteristics of textual disclosure? To address this question, I employ principal component analysis (PCA), a statistical procedure, in which the idea is to extract the principal components from these proxies by reducing its dimensionality but retaining most of the variation in the original factors. It is akin to creating an index of the different proxies that are highly correlated, as is the case here (Appendix D), by retaining their uncorrelated and normalized components, using vector space transformation.

I extract the principal components from the seven proxies for the degree of textual disclosure and the seven proxies for readability of textual disclosure. For sentiments, I follow Loughran and McDonald (2013), and group uncertain, weak modal, and negative word proportion and count, as an *ex-ante* measure of uncertainty. The results are reported in Table 7.

[Insert Table 7 here]

Note that such principal components not only enable me to capture the common essence of different textual variables, but it also improves the exposition of the results. The results are similar to the main results reported in the previous section. Table 7 shows that both in the case of 10-Ks (Panel A) and 10-Qs (Panel B) filings, the first component is highly significant in all the specifications (i.e., for degree of disclosure, readability, sentiments (count) and sentiments %) with higher correlations for the 10-Q filings, which is also consistent with the earlier results. The next set of tests explore the plausible impacts of textual disclosure in SEC filings on other aspects of shareholder class action litigations.

5.2 Case Period

The time between the case filing date and the settlement/dismissal date is defined as the "case period" or simply the time to the verdict. The "case period" is a function of several factors such as the severity and complexity of the alleged fraud, the quantity, and quality of the evidence, litigiousness of the industry and others. In this section, I ask whether textual disclosures in the filings have any association with the length of the "case period" after controlling for other factors. The idea here is that the text that is difficult to comprehend or is ambiguous will be likely to increase the "case period" and hence the costs to both the litigants and the defendants.

To formally test it for both 10-Ks and 10-Qs, I use OLS models with the dependent variable as the length of the "case period" in days and the principal components of either degree, readability, or sentiments of a firm's textual disclosure as the main independent variable of interest. The results are presented in Table 8.

[Insert Table 8 here]

Panels A and B of Table 8 present the results for 10-K and 10-Q filings, respectively. All the eight models in Table 8 also control for size (natural logarithm of market value), return on assets (ROA), loss indicator (negative net income), earnings growth, sales growth, market-to-book, big-8 auditor dummy, lag size, lag ROA, lag loss indicator, lag stock volatility, lag stock return, and lag institutional ownership, besides year and industry (2-digit SIC code) fixed effects. The results show that there is a significant positive relation (significant at 1% level) between the use of words that proxy for ex-ante uncertainty (i.e., uncertain, weak modal, and negative word proportion and count) and the length of the "case period" for both 10-Ks and 10-Qs, as anticipated. Although there is no significant relation between the degree of disclosure and the length of the "case period," readability is negatively associated with the length of the "case period," (significant at the 10% level for 10-Qs), indicating that the use of text that is difficult to comprehend likely increases the length of the "case period."

5.3 Market Response

In this section, I test the association between the degree, readability, and sentiments of a firm's textual disclosure in 10-Ks and 10-Qs and the cumulative abnormal returns (CARs) around

the case filing date. The extant literature on litigation in corporate finance has used event study methodologies to determine the stock market reaction at the time of lawsuit filings. Several studies over the years such as Karpoff and Lott (1993), Bizjak and Coles (1995), Bhagat, Bizjak and Coles (1998), Gande and Lewis (2009) and most recently Klock (2015) have documented a significantly negative stock price reaction at the lawsuit filing date. Studies have attributed such stock price declines to reputational costs, increased probability of financial distress costs and higher chances of subsequent lawsuits. I confirm such negative cumulative abnormal returns (CARs) around the case filing date in my sample of securities class action lawsuits, and the results are presented in Table 9.

[Insert Table 9 here]

The CARs have been computed as the difference between the actual stock price return and the expected market model (CRSP Value Weighted Index) return over the windows listed. Market model has been estimated using 255 days of daily returns ending 46 days before the case filing date. Note that in all the six different windows (days), the CARs are significantly negative. Furthermore, I ask if the association between textual disclosure in SEC filings and the likelihood of shareholder class action lawsuits is true, does the market perceive it as such? To formally answer this question, I run OLS regressions with cumulative abnormal returns (CARs) around the case filing date as a dependent variable, and the proxies of textual disclosure created a year prior to the case filing date as the main independent variable of interest. The regressions also control for size (natural logarithm of market value), return on assets (ROA), loss indicator (negative net income), earnings growth, sales growth, market-to-book, big-8 auditor dummy, lag size, lag ROA, lag loss

indicator, lag stock volatility, lag stock return, lag institutional ownership, and dummy indicating whether the litigation is either merger-related or accounting related. All models include year and industry (2-digit SIC code) fixed effects, with clustering at the firm level. Table 10 presents the results of such tests.

[Insert Table 10 here]

For brevity, I only present the results showing the association between the principal components of degree, readability, and sentiments of a firm's textual disclosure and the cumulative abnormal returns (CARs) around the case filing date for CAR windows [-1, +1] and [-2, +2], using the market model. Panel A presents the results from analyses of firms' 10-K filings, while Panel B provides the results from analyses of firms' 10-Q filings. The results show that there is a significant negative relation between the use of words that proxy for *ex-ante* uncertainty (i.e., uncertain, weak modal, and negative word proportion and count) and market reaction on the lawsuit filing, indicating that the market also perceives the association between textual disclosure using words that proxy for *ex-ante* uncertainty in SEC filings and the likelihood of shareholder class action lawsuits. However, I do not find any significant results for either degree of disclosure or readability. A plausible reason for such results could be that often in shareholder litigations, the market becomes aware of the alleged fraud in stages rather than on one single day like the filing date or the class end date. So, the stock price reaction at the lawsuit filing date is muted to some extent.

5.4 Settlement Amount

Approximately half of the cases in my sample of shareholder class action litigations are settled. Although there are many indirect costs of litigation such as reputational costs, loss of customers and suppliers, loss of management's time, etc., that are hard to measure, one direct and measurable cost of litigation is the settlement disbursement.

In this section, I explore the relation between textual disclosure in the texts of 10-Ks and 10-Qs and the settlement amounts. I hand-collect the settlement amounts from various sources such as court documents, firm's disclosures, and media articles. I further divide the settlements as a cash-only portion of the settlement and total settlement, which also includes the non-cash (securities) portion of the settlements. Then I scale the settlement amounts by the Disclosure Dollar Loss (DDL), which is the dollar value change in the litigated firm's market capitalization between the trading day immediately prior to the class end date and the trading day immediately following the class end date.

Following it, I repeat the main tests using the principal components of degree, readability, and sentiments of a firm's textual disclosure as the main independent variables and the litigation settlement amount (cash amount and total amount) scaled by Disclosure Dollar Loss (DDL) as the dependent variables. So, the dependent variable can be thought of as a rough proxy for the recovery rate of the potential damages. Table 11 presents the results of such specification.

[Insert Table 11]

Panel A of Table 11 presents the results from the OLS models of 10-K filings, and Panel B repeats the same regressions for 10-Q filings. Note that only the coefficients on the Uncertainty

% are significantly negative. In other words, the use of higher proportion of words that proxy for uncertainty leads to lower recovery rates in settled cases. One plausible reason for such finding can be that, ex-post, the use of uncertain words in filings can also be used as an effective tool for defense from the defendant firm's side. One common defense argument from a legal perspective that is often made in such cases is that uncertain words were used not to hide any material information but rather to make the investors aware of the potential risks.

6. A QUASI-NATURAL EXPERIMENT

Crane and Koch (2018) provided evidence that in 1999, the Ninth Circuit Court of Appeals ruling, *Re: Silicon Graphics Inc.,* led to an unexpected and sudden reduction in the threat of litigation for firms headquartered in the Ninth Circuit that resulted in the number of shareholder class action lawsuits in the Ninth Circuit plunging 43% as compared to a 14% rise in other circuits. This ruling arguably made it harder for the shareholders of the firms located in the Ninth Circuit (i.e., the firms headquartered in the US states: Alaska, Washington, Oregon, Idaho, Montana, California, Nevada, Arizona, and Hawaii) to litigate firms in a shareholder class action lawsuit as it required the litigants to establish that the defendants acted with "deliberate recklessness." Note that "deliberate recklessness" requires the plaintiffs to prove the "intent" of the defendant to cause harm to the shareholders' wealth and is a stricter requirement than simply "recklessness."

Similar to Crane and Koch (2018), in this section I use the Ninth Circuit Court of Appeals ruling, *Re: Silicon Graphics Inc.,* as a quasi-natural experiment and an exogenous shock to the threat of litigation for firms headquartered in the Ninth Circuit, as it is unlikely that the firms chose their headquarters in anticipation of this ruling. I conjecture that if the relation between textual disclosure and shareholder class action litigation documented in the earlier sections have a causal link, then one should expect that after the Ninth Circuit Court of Appeals ruling, *Re: Silicon Graphics Inc.,* firms headquartered in the Ninth Circuit would disclose more in their 10-Ks and 10-Qs, use a greater proportion of words that proxy for *ex-ante* measures of uncertainty, and their disclosures will be less readable relative to disclosures of firms across other circuits. This is

because the sudden reduction in the threat of litigation for the firms headquartered in the Ninth Circuit would allow them more flexibility and room to adjust their disclosure in response to such reduced shareholder litigation risk.

To formally test such a hypothesis, I use the following difference-in-differences specification with year and firm fixed effects:

$$\textit{Textual Disclosure }_{i,k,t} = \beta_0 + \delta * \textit{Treatment }_{k,t} + \beta_1 * X_{i,t-1} + \alpha_i + \alpha_t + \varepsilon_{i,k,t}$$

$$\text{where, } \textit{Treatment = Post 1999 Dummy * Ninth Circuit Dummy}$$

The dependent variable in the specification above is the textual disclosure measure of either degree of disclosure, sentiments or readability. The subscripts i, k, and t indicate the firm i, location of the firm headquarters in state k, and time t, respectively. The main independent variable of interest is *Treatment,* which is an interaction of two dummies, i.e., whether it is pre- vs. post- 1999 and whether the firm is headquartered in the Ninth Circuit (i.e., headquartered in the US states: Alaska, Washington, Oregon, Idaho, Montana, California, Nevada, Arizona, and Hawaii). Since Compustat only reports current headquarters location, I programmatically extract the information on historical headquarters of firms in 1998 from the header sections of the 10-Ks and 10-Qs. The specification also controls for other firm-level observable characteristics, $X_{i,t-1}$ and year (α_t) and firm (α_i) fixed effects to control for time-variant and time-invariant unobservable factors, respectively. The other observable firm-level characteristics include size (natural logarithm of market value), return on assets (ROA), earnings growth, sales growth, market-to-book, big-8

auditor dummy, stock volatility, and institutional ownership.[17] Here the treated firms are the Ninth Circuit firms and the control firms are the non-Ninth Circuit firms exactly matched on years[18] and industry (2-digit SIC code) and also matched on the closest estimated propensity score based on size (measured by market value) and market-to-book (M/B) ratio. The results are reported in Table 12.

[Insert Table 12 here]

For brevity, I only present the results with Principal Component Analyses (PCA) for both 10-K (Panel A) and 10-Q (Panel B) filings. As it can be seen from the coefficients on the interaction of Post 1999 Dummy x Ninth Circuit Dummy, firms in the Ninth Circuit significantly increase their use of words that proxy for *ex-ante* uncertainty (proxied with uncertain, weak modal, and negative word proportion and count) and their degree of textual disclosure (only for 10-Qs) after the Ninth Circuit Court of Appeals ruling as compared to matched non-Ninth Circuit firms. The results for readability (untabulated) are not significantly different. One probable reason for this could be that institutional owners can systematically choose to own firms where it is more difficult for small shareholders to litigate them (Crane and Koch, 2018) and institutional owners demand more transparent disclosures from their investee firms as recently shown by Boone and White (2015) and Bird and Karolyi (2016).

The next section conducts several robustness tests for the main results and addresses some of the concerns of endogeneity.

---

[17] The results (untabulated) are also robust to the use of state-level macro variables such as state GDP, population, net tax receipts, and unemployment.

[18] I omit the year 1999 as the ruling occurred in the middle of 1999, and analyze three pre-years (1996, 1997, 1998) and three post-years (2000, 2001, 2002).

7. ENDOGENEITY CONCERNS AND ROBUSTNESS TESTS

In this section, I discuss the possible endogeneity concerns and perform a series of additional robustness tests to address some of them.

It is likely that disclosure and the risk of litigation are endogenously determined. According to Roberts and Whited (2013), there are three specific sources of endogeneity: Simultaneity, measurement error, and omitted variables. This paper addresses all these three sources of endogeneity to some extent. The first source of endogeneity in this set-up is simultaneity or reverse causality. Here the main concern is that disclosure may not be causing class action litigations but instead, certain types of firms that are more likely to face litigations, disclose in a certain manner. I address this concern by creating textual disclosure variables from 10-Ks and 10-Qs, which were filed prior to the litigation filing date in all the specifications. Figure 1 presents the typical timeline of securities class action lawsuits and demonstrates the two empirical strategies used throughout the paper. Although I have used *empirical strategy 1* for the results shown in the paper, in unreported analysis, I also replicate the main results by omitting the year of case filing and only analyzing the textual content of the filings a year before and after the case filing year to avoid the noise created in filings due to litigations as shown in *empirical strategy 2*. Also, the use of lagged covariates in all my empirical specifications and the use of the matched sample should mitigate such a concern to a certain extent.

[Insert Figure 1 here]

41

The second source of endogeneity is measurement error which arises from discrepancies between the proxy used and the true value of the explanatory variable. In other words, the concern here is that we do not measure the true values of degree of textual disclosure, textual readability, and textual sentiments. The paper addresses this concern to a great degree by using seven different textual disclosure variables, seven different textual readability variables, fourteen different textual sentiments measures and two different and widely used sources of word lists in textual analysis from the extant literature. Moreover, I also use principal component analysis (PCA) to address this concern further as shown in the prior sections.

The third source of endogeneity is omitted variable bias, where the argument is that there might be something unobservable which has not been included in the vector of covariates but might be driving both incidences of litigations and disclosure. Although I have tried to address this concern by including control variables used in prior literature and fixed effects, it is still possible that the results suffer from omitted variable bias. It is plausible that there is a third variable that affects both the textual content in SEC filings and litigation risk. One might also argue that complex or more litigious situations necessitate a certain type of disclosure. The following sub-sections attempt to address some of these concerns.

7.1 The Class Period (Alleged Damage Period)

In this section, I hypothesize that if there is a causal connection between the nature of disclosure and litigation risk, then it should also be indicated in the disclosure behavior, once the firms enter the "class period" or the alleged damage period. Hence, I ask whether something

changed in the nature of textual disclosure, once these firms enter the class period, after controlling

for other numerical explanatory variables. More specifically, I ask, is that change different from

the very similar firms that did not get litigated?

To answer this question, I separate the sample into dismissed and settled cases and

investigate the textual content of the filings for each sub-sample pre- vs. post-class start date using

a standard difference-in-differences (DiD) framework and a matched sample. The empirical

strategy is depicted in Figure 2.

<center>*[Insert Figure 2 here]*</center>

And, the DiD specification used is shown below:

$$\textit{Textual Variable }_{i,t} = \beta_0 + \beta_1 * D(\textit{Litigated Firm}_i) * D(\textit{Post Class Start Date}_{i,t}) + \beta_2 * D(\textit{Post Class}$$

$$\textit{Start Date}_{i,t}) + \beta_3 * D(\textit{Litigated Firm}_i) + \beta * X_{i,t} + \varepsilon_{i,t} \qquad (4)$$

The predicted variable in the above DiD specification is one of the twenty-eight textual

variables described before and the main coefficient of interest is β_1, i.e., the coefficient on the

interaction term. The results are reported in Tables 13.1 (degree of disclosure), 13.2 (readability),

13.3 (sentiments) and 13.4 (sentiments%).

<center>*[Insert Table 13.1 here]*</center>

<center>*[Insert Table 13.2 here]*</center>

<center>*[Insert Table 13.3 here]*</center>

<center>*[Insert Table 13.4 here]*</center>

While panels A and C of these tables provide the results for 10-K filings, panels B and D show the results for 10-Q filings. As depicted in Table 13.1, the main coefficient of interest, β_1 is negative and significant in most of the specifications for both 10-Ks and 10-Qs, indicating that the firms that have been litigated (both settled and dismissed) significantly reduce their degree of textual disclosure during the class period as compared to the matched sample. One likely interpretation of such results is that on average, rational managers of sued firms understand the causal relation between the degree of disclosure and the risk of litigation and try to intentionally reduce disclosure especially during the period when the (alleged) financial misconduct is being committed. However, as seen in the main results, such differential reduction in textual disclosure is not enough to prevent litigations.

Following it, I test whether managers of sued firms improve the readability of their disclosures during the class period. The idea here is that if there is a causal connection between readability and litigation risk, rational managers will attempt to improve the readability of their disclosures, especially during the class period to deter litigation. The results are reported in Table 13.2, where I find that there is a differential reduction in readability scores (i.e., improvement in readability), especially for the settled cases and 10-Qs. Note that I do not find significant results for the dismissed cases, though the signs of the coefficients are similar, probably because many of the dismissed cases are frivolous, to begin with.

Finally, in Tables 13.3 and 13.4, I test whether managers of the litigated firms change the textual sentiments delineated in filings during the class period. As reported in table 8.3, I find that managers of the sued firms significantly reduce their use of uncertainty, weak modal, negative

44

(both Harvard and L&M), litigious words in both 10-K and 10-Q filings as compared to the matched firms post class start date. Such results are robust to the use of different controls and fixed effects and are similar for both settled and dismissed cases. Table 8.4 further confirms the results in terms of word proportions. Overall, the results in this section show that sued firms *ex-ante* change the nature of their textual disclosure during the class period in an attempt to deter litigations, indicating that there might be a causal connection between the nature of textual disclosure and litigation risk.

7.2 Pre- vs. Post- Litigation Textual Analysis

Following the empirical analyses above, I examine whether and how firms that are sued in securities class action lawsuits change their behavior regarding textual disclosure post-litigation. It is important to answer this question as it indicates how managers react in terms of textual disclosure after their experience of being litigated. It can be argued that if managers are rational and they perceive that greater textual disclosure or certain types of disclosure are triggering class action litigations, they would take steps to alter such disclosures post-litigation. It can also be argued that such possible changes in disclosure behavior post-litigation would differ for settled versus dismissed cases as managers of firms with settled class action lawsuits would perceive the cost-benefit analysis of disclosure differently from the managers of firms with dismissed cases, as the costs of settled cases are significantly higher than those of dismissed or frivolous cases.

Therefore, I again divide the sample of sued firms into dismissed and settled cases and examine pre- vs. post-litigation textual content of each sub-sample using a standard difference-in-

differences (DiD) framework. Matched control firms are selected as described earlier in the paper and the year of litigation is considered as the pseudo-event year for the matched firm in the DiD specification, as shown below:

$$\textit{Textual Variable}_{i,t} = \beta_0 + \beta_1 * D(\textit{Litigated Firm}_i) * D(\textit{Post Litigation}_{i,t}) + \beta_2 * D(\textit{Post Litigation}_{i,t})$$

$$+ \beta_3 * D(\textit{Litigated Firm}_i) + \beta * X_{i,t} + \varepsilon_{i,t} \qquad (5)$$

The dependent variable in the above DiD model is one of the twenty-eight textual variables that proxy for either degree of textual disclosure, readability or sentiments as described earlier in the paper. The main coefficient of interest on the right-hand side is β_1, i.e., the coefficient on the interaction term where, D (Post Litigation$_{i,t}$) is the dummy variable which is equal to one if the year is two-years post-litigation year and D (Litigated Firm$_i$) is the dummy variable which is equal to one if the firm has been litigated. β_1 denotes the differential change in textual variables for litigated firms pre- (i.e., two-years prior to the litigation year) and post- (i.e., two-years post-litigation year) litigation, compared to the matched sample. Focusing on pre- and post- years enables this test to measure the impact of litigation on textual disclosure behavior without getting entangled in confounding events. $X_{i,t}$ is a vector of control variables as described in earlier tests. I also include year and industry fixed effects, with clustering at the firm level. The results are reported in Tables 14.1, 14.2, 14.3 and 14.4.

[Insert Table 14.1 here]

[Insert Table 14.2 here]

[Insert Table 14.3 here]

[Insert Table 14.4 here]

46

While panels A and C provide the results for 10-Ks (for both settled and dismissed cases), panels B and D show the results for 10-Qs (for both settled and dismissed cases). If it is true that greater textual disclosure precipitates securities class action litigations, then one can expect that managers of the litigated firms would likely reduce disclosure post-litigation. We can see from Table 14.1 that the coefficient on the interaction term shows that there is a significant decrease in the degree of textual disclosure post-litigation. Also, such differential reduction in the textual disclosure variables is a lot higher in magnitude and significance for firms that were litigated, and there was a settlement. One interpretation of such a finding could be that since firms that settle bear significantly greater costs of litigation as compared to firms in which the cases were dismissed, the managers react more vigorously.

In the case of readability variables (Table 14.2), the differential change is a lot weaker as most specifications indicate that there is no change pre- and post- litigation for both litigated/settled and litigated/dismissed cases suggesting that managers do not make readability significantly harder post-litigation as they probably learn on being sued, that there is a positive association between difficulty level in readability and incidence of litigation. Note that, one of the proxies of readability, i.e., Automated Readability Index (ARI) shows an increase in the difficulty level of readability. Such a result is surprising at first; however, sometimes, it could merely be a result of legally binding settlement clause to improve disclosure and explain a firm's litigious situation, that might lead to higher readability scores.

Finally, in the case of sentiment variables (Tables 14.3 and 14.4), I document that managers of the sued firms significantly reduce their use of uncertainty, weak modal, negative words (both

LM and Harvard) pre- and post- litigation for both the settled and dismissed cases. However, note that in the case of 10-Qs, that are more frequent and up-to-date, the proportion the use of negative and litigious words increases post litigation. This is puzzling because if the use of negative and litigious words increases the likelihood of being litigated as shown in prior tests, why do managers tend to increase the use of proportion of negative and litigious words post-litigation in their 10-Qs? One probable explanation could be that managers tend to disclose bad news post-litigation to avoid litigation since they treat disclosure of bad news differently from the disclosure of good news. Such results are corroborated by Graham, Harvey and Rajgopal (2005), who surveyed more than 400 executives and found that 76.8% of executives agree or strongly agree that disclosing bad news faster not only enhances their reputation for transparency but also reduces the risk of potential lawsuits. Moreover, a recent working paper, Billings, Cedergren and Dube (2016), also documented similar results when the authors found, *"…our evidence indicates that the nature of disclosure matters: while managers reduce and delay their forecasts of positive news, bad news warnings actually increase and become more timely following litigation." (p.31).*

Overall, the results presented in this section show that managers of the sued firms change the disclosure behavior in filings, post-litigation as compared to a matched sample, suggesting a causal connection between the nature of textual disclosure and litigation risk.

7.3 Largest Daily Stock Price Drop

Anecdotal evidence suggests that shareholder litigations are often a result of a sharp stock price drop that grabs investor attention. Therefore, in the following empirical tests, I control for the largest daily stock price drop during the alleged damage period.

Table 15 presents the relation between the principal components of degree, readability, and sentiments of a firm's textual disclosure and the probability of the firm being litigated. While Panel A shows the results from analyses of firms' 10K filings, Panel B presents the results from analyses of firms' 10Q filings. All models control for size (natural logarithm of market value), return on assets (ROA), loss indicator (negative net income), earnings growth, sales growth, market-to-book, big-8 auditor dummy, lag size, lag ROA, lag loss indicator, lag stock volatility, lag stock return, lag institutional ownership, and the largest daily stock price drop during the alleged damage period. It is worth noting that although the largest daily stock return during the alleged damage period has an expected negative relation which is significant at the 1% level, the coefficients on the first components of disclosure, readability, and uncertainty (both count and %) stay qualitatively similar even after controlling for largest daily stock price drop.

[Insert Table 15 here]

7.4 Shareholder Class Action Litigation Type

In this robustness sub-section, I explore if the results vary with the type of shareholder class action litigations. To test it formally, I manually read the First Identified Complaint (FIC) filings, other case documents, press releases and identify whether the litigations are: (1) Mergers

and Acquisitions (M&A) Related; or, (2) Disclosure Related. While M&A related litigations are filed when the litigant claims that there was a breach of fiduciary duty in relation to an M&A deal or tender offer, disclosure related litigations include, but are not limited to, alleged GAAP violations, misstatement of financial reports and other alleged material misrepresentations or omissions.

[Insert Table 16 here]

Table 16 shows that the main results stay qualitatively similar and are robust at 1% significance level, controlling for the type of litigation, both in the case of 10-K and 10-Q filings. Moreover, the coefficients on these litigation-type dummies are not significantly different from zero, indicating that the results are not driven by one specific type of litigation.

7.5 Plain English Initiative

The SEC published a guide titled, "A Plain English Handbook: How to Create Clear SEC Disclosure Documents," in 1998 and implemented the Plain English Initiative in October 1998.[19] The idea behind this initiative of SEC was to make the disclosure documents such as 10-Ks and 10-Qs more readable. Therefore, it can be argued that a part of the results on readability of 10-Ks and 10-Qs could be because of such a regulatory shock.

To address such concerns, I split the sample pre- and post- Plain English Initiative in October 1998 and test the following specification on readability:

---

[19] https://www.sec.gov/pdf/handbook.pdf

$$Litigation\ Dummy_{i,t} = \beta_0 + \delta*Readability_{i,t-1}* Post\ PEI + \beta_1* Readability_{i,t-1}+ \beta_2* Post\ PEI +$$

$$\beta*X_{i,t} + \varepsilon_{i,t} \qquad (6)$$

The results are reported in Table 17.

[Insert Table 17 here]

Two points can be noted here. First, the relation between readability and litigation risk documented in the main results holds both pre- and post- Plain English Initiative. Second, the coefficient of the interaction term, i.e., δ shows that the relation weakens post- Plain English Initiative as this regulatory shock is supposed to standardize the readability of disclosure documents.

7.6 Sarbanes-Oxley Act ("SOX")

The 2002 adoption of the Sarbanes-Oxley Act ("SOX") coerced publicly listed firms to have greater transparency (Cohen, Dey and Lys, 2005), enhanced disclosure (Beneish, Billings and Hodder, 2008) and discouraged risk-taking (Bargeron, Lehn and Zutter, 2010) by increasing the likelihood and severity of punishment on fraudulent and misleading disclosures. Hence, one potential concern is that the main results could be driven by the adoption of SOX on July 30, 2002.

To address such a concern, I split the sample pre- and post- SOX and test the following specification on the degree of textual disclosure:

$$Litigation\ Dummy_{i,t} = \beta_0 + \delta*Disclosure_{i,t-1}* Post\ SOX + \beta_1* Disclosure_{i,t-1}+ \beta_2* Post\ SOX +$$

$$\beta*X_{i,t} + \varepsilon_{i,t} \qquad (7)$$

The results are reported in Table 18.

[Insert Table 18 here]

The results presented in Table 18 not only indicate that the relation between the degree of disclosure and litigation risk documented in the main results holds both pre- and post- SOX but also shows that such relation weakens post-SOX (especially for 10-Qs) as SOX standardized the degree of disclosure in filings to a great extent.

7.7 Other Tests

I conduct several other robustness tests (untabulated) that further corroborate my main results. A disproportionate amount of securities litigations occurs in the business services (particularly, technology firms) and chemicals and allied products (notably, pharmaceutical companies). I have tried to address this using industry fixed effects in all my specifications (Gormley and Matsa, 2014) and I also find that the (untabulated) results stay qualitatively consistent when I exclude industries with a high number of litigations.[20]

Finally, I conduct falsification tests (unreported) by repeating difference-in-differences (DiD) analyses two years prior to the class start date and two years before the litigation filing date by falsely assuming these dates two years before their actual dates, and do not find any significant results, which reinforces the difference-in-differences (DiD) results.

---

[20] The results are robust to the exclusion of high litigation risk industries, both at two SIC digit code level (73,28,36,35 & 38) and more granular four SIC digit code level (Computers, 3570-3577 and 7370-7374; Electronics, 3600-3674; 2833-2836 etc.).

8. CONCLUSION

Employing well-established big data text analytics techniques, I study the relation between non-numeric textual disclosure in the narratives of SEC filings, and litigation risk after controlling for explanatory numerical variables. Using degree of textual disclosure, readability of disclosures and sentiments generated through the choice of words used in 10-Ks and 10-Qs, and matched sample, I find that greater textual disclosure, readability or comprehensibility of texts used and sentiments produced through the choice of words have a significant predictive power in explaining the likelihood of being sued by shareholders in class actions. While I find that more disclosure through texts in filings and the difficulty level of comprehending text used in filings increases the risk of litigation, this paper also shows that it is not just the degree of textual disclosure but also the sentiments portrayed in disclosures that are associated with the incidence of shareholder class action litigations. For instance, the use of words that generate negative sentiments such as uncertainty words, weak modal words, litigious words and words having negative connotations in 10-K and 10-Q filings increase the likelihood of being litigated.

Finally, I show how managers alter their behavior with respect to textual disclosures pre- and post-class start date and pre- versus post-litigation using a standard difference-in-differences (DiD) framework. These results are robust to the use of different controls, matched sample and empirical specifications, including principal component analysis (PCA) and a quasi-natural experiment that address several concerns of endogeneity. Overall, the results presented in the paper

demonstrate that there is a need to recognize the difference between merely more versus better disclosure.

REFERENCES

Akerlof, G.A. and Shiller, R.J., 2010. *Animal spirits: How human psychology drives the economy, and why it matters for global capitalism*. Princeton University Press.

Anderson, J., 1983. LIX and RIX: Variations on a little-known readability index. *Journal of Reading*, *26*(6), pp.490-496.

Angeletos, G.M. and La'O, J., 2013. Sentiments. *Econometrica*, *81*(2), pp.739-779.

Banerjee, S., Humphery-Jenner, M., Nanda, V.K. and Tham, T.M., 2018. Executive overconfidence and securities class actions. *Journal of Financial and Quantitative Analysis,* (forthcoming).

Bargeron, L.L., Lehn, K.M. and Zutter, C.J., 2010. Sarbanes-Oxley and corporate risk-taking. *Journal of Accounting and Economics*, *49*(1), pp.34-52.

Beneish, M.D., Billings, M.B. and Hodder, L.D., 2008. Internal control weaknesses and information uncertainty. *The Accounting Review*, *83*(3), pp.665-703.

Benhabib, J., Wang, P. and Wen, Y., 2015. Sentiments and aggregate demand fluctuations. *Econometrica*, *83*(2), pp.549-585.

Bhagat, S., Bizjak, J. and Coles, J.L., 1998. The shareholder wealth implications of corporate lawsuits. *Financial Management*, pp.5-27.

Biggs, S.F., Bedard, J.C., Gaber, B.G. and Linsmeier, T.J., 1985. The effects of task size and similarity on the decision behavior of bank loan officers. *Management Science, 31(8),* pp.970-987.

Billings, M.B., Cedergren, M.C. and Dube, S., 2016. Do Managers Respond to Litigation with Silence? Working Paper., Leonard N. Stern School of Business, New York University.

Bird, A. and Karolyi, S.A., 2016. Do institutional investors demand public disclosure? *The Review of Financial Studies*, *29*(12), pp.3245-3277.

Bizjak, J.M. and Coles, J.L., 1995. The effect of private antitrust litigation on the stock-market valuation of the firm. *The American Economic Review*, pp.436-461.

Björnsson, C.H., 1968. *Läsbarhet*. Liber.

Bodnaruk, A., Loughran, T. and McDonald, B., 2015. Using 10-k text to gauge financial constraints. *Journal of Financial and Quantitative Analysis*, *50*(04), pp.623-646.

Boone, A.L. and White, J.T., 2015. The effect of institutional ownership on firm transparency and information production. *Journal of Financial Economics*, *117*(3), pp.508-533.

Cohen, D.A., Dey, A. and Lys, T.Z., 2005. Trends in earnings management and informativeness of earnings announcements in the pre-and post-Sarbanes Oxley periods. *Available at SSRN 658782*.

Core, J.E., 2001. A review of the empirical disclosure literature: discussion. *Journal of Accounting and Economics*, *31*(1), pp.441-456.

Crane, A.D. and Koch, A., 2018. Shareholder litigation and ownership structure: Evidence from a natural experiment. *Management Science*, *64*(1), pp.5-23.

Crawford, V.P. and Sobel, J., 1982. Strategic information transmission. *Econometrica: Journal of the Econometric Society*, pp.1431-1451.

Das, S.R., 2014. Text and Context: Language Analytics in Finance. *Foundations and Trends (R) in Finance*, *8*(3), pp.145-261.

DeAngelo, L.E., 1981. Auditor size and audit quality. *Journal of Accounting and Economics*, *3*(3), pp.183-199.

Epstein, L.G. and Schneider, M., 2008. Ambiguity, information quality, and asset pricing. *The Journal of Finance*, *63*(1), pp.197-228.

Field, L., Lowry, M. and Shu, S., 2005. Does disclosure deter or trigger litigation? *Journal of Accounting and Economics*, *39*(3), pp.487-507.

Flesch, R., 1948. A new readability yardstick. *Journal of Applied Psychology*, *32*(3), p.221.

Francis, J., Philbrick, D. and Schipper, K., 1994. Shareholder litigation and corporate disclosures. *Journal of Accounting Research*, pp.137-164.

Gande, A. and Lewis, C.M., 2009. Shareholder-initiated class action lawsuits: Shareholder wealth effects and industry spillovers. *Journal of Financial and Quantitative Analysis*, *44*(4), pp.823-850.

Gillick, D., 2009, May. Sentence boundary detection and the problem with the US. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers* (pp. 241-244). Association for Computational Linguistics.

Gormley, T.A. and Matsa, D.A., 2014. Common errors: How to (and not to) control for unobserved heterogeneity. *Review of Financial Studies*, *27*(2), pp.617-661.

Graham, J.R., Harvey, C.R. and Rajgopal, S., 2005. The economic implications of corporate financial reporting. *Journal of Accounting and Economics*, *40*(1), pp.3-73.

Gunning, R., 1952. The technique of clear writing. *McGraw-Hill*, 1952

Hanley, K.W. and Hoberg, G., 2012. Litigation risk, strategic disclosure and the underpricing of initial public offerings. *Journal of Financial Economics*, *103*(2), pp.235-254.

Healy, P.M. and Palepu, K.G., 2001. Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics*, *31*(1), pp.405-440.

Hirshleifer, D., 2015. Behavioral finance. *Annual Review of Financial Economics*, *7*, pp.133-159.

Hirshleifer, D. and Teoh, S.H., 2003. Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics*, *36*(1), pp.337-386.

Hwang, B.H. and Kim, H.H., 2017. It pays to write well. *Journal of Financial Economics*, *124*(2), pp.373-394.

Karpoff, J.M., Koester, A., Lee, D.S. and Martin, G.S., 2014. Database challenges in financial misconduct research. *Georgetown McDonough School of Business Research Paper*, (2012-15).

Karpoff, J.M. and Lott Jr, J.R., 1993. The reputational penalty firms bear from committing criminal fraud. *The Journal of Law and Economics*, *36*(2), pp.757-802.

Kearney, C. and Liu, S., 2014. Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, *33*, pp.171-185.

Kincaid, J.P., Fishburne Jr, R.P., Rogers, R.L. and Chissom, B.S., 1975. *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel* (No. RBR-8-75). Naval Technical Training Command Millington TN Research Branch.

Klock, M., 2015. Do class action filings affect stock prices? The stock market reaction to securities class actions post PSLRA. *Journal of Business & Securities Law*, *15*(2), p.109.

Leuz, C. and Wysocki, P.D., 2016. The economics of disclosure and financial reporting regulation: Evidence and suggestions for future research. *Journal of Accounting Research*, *54*(2), pp.525-622.

Li, F., 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, *45*(2), pp.221-247.

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

Loughran, T. and McDonald, B., 2013. IPO first-day returns, offer price revisions, volatility, and form S-1 language. *Journal of Financial Economics*, *109*(2), pp.307-326.

Loughran, T. and McDonald, B., 2014. Measuring readability in financial disclosures. *The Journal of Finance*, *69*(4), pp.1643-1671.

Loughran, T. and McDonald, B., 2016. Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*.

Lowry, M., 2009. Discussion of 'Shareholder litigation and changes in disclosure behavior.' *Journal of Accounting and Economics*, *47*(1), pp.157-159.

Mc Laughlin, G.H., 1969. SMOG grading-a new readability formula. *Journal of Reading*, *12*(8), pp.639-646.

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

Roberts, M.R. and Whited, T.M., 2013. Endogeneity in Empirical Corporate Finance 1. *Handbook of the Economics of Finance*, *2*, pp.493-572.

Rogers, J.L. and Van Buskirk, A., 2009. Shareholder litigation and changes in disclosure behavior. *Journal of Accounting and Economics*, *47*(1), pp.136-156.

Rogers, J.L., Van Buskirk, A. and Zechman, S.L., 2011. Disclosure tone and shareholder litigation. *The Accounting Review*, *86*(6), pp.2155-2183.

Senter, R.J. and Smith, E.A., 1967. Automated readability index. *Cincinnati University, OH*.

Shiller, R.J., 2015. Irrational exuberance. *Princeton University Press.*

Simon, H.A., 1955. A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1), pp.99-118.

Skinner, D.J., 1994. Why firms voluntarily disclose bad news. *Journal of Accounting Research*, *32*(1), pp.38-60.

Spence, M., 1973. Job market signaling. *The Quarterly Journal of Economics*, *87*(3), pp.355-374.

Tetlock, P.C., 2007. Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, *62*(3), pp.1139-1168.

Tetlock, P.C., Saar-Tsechansky, M. and Macskassy, S., 2008. More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, *63*(3), pp.1437-1467.

Veronesi, P., 1999. Stock market overreactions to bad news in good times: a rational expectations equilibrium model. *Review of Financial Studies*, *12*(5), pp.975-1007.

Figure 1. Timeline of Shareholder Class Action Lawsuits and the Empirical Strategies 1 and 2
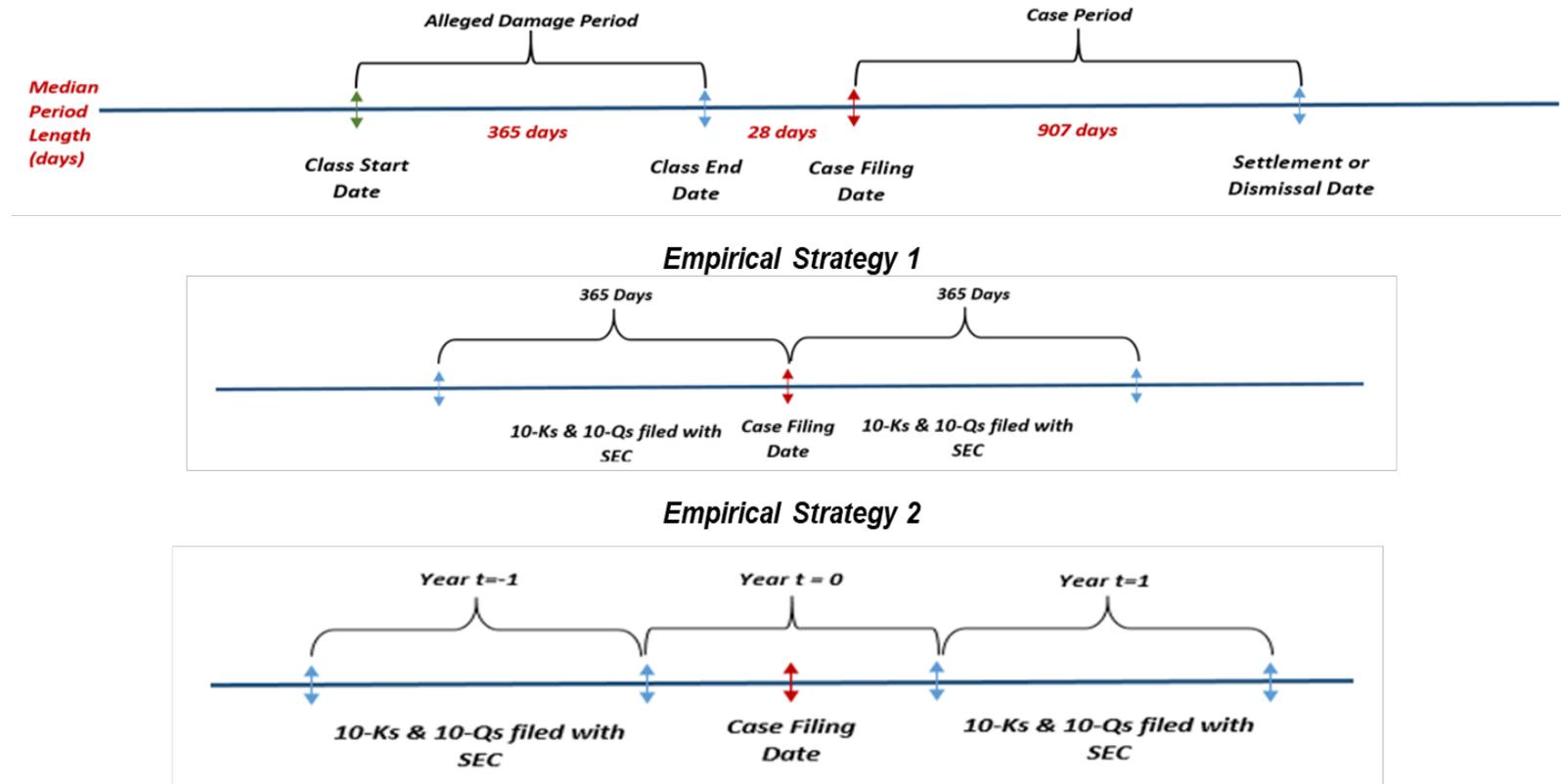
Figure 2. Timeline of Shareholder Class Action Lawsuits and the Empirical Strategy 3
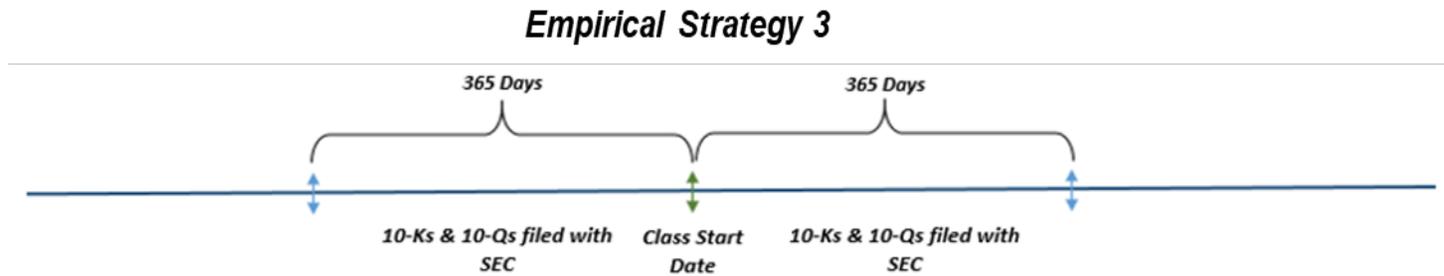
Table 1 Summary Statistics on Litigation

This table summarizes the number of litigations filed each year from 1996-2014 in my sample (Panel A) and the status of those litigation cases (Panel B). Panel C reports the summary statistics on the case period and class period for the cases, and total settlement amount for those cases that are settled. Total Settlement Amount includes both the cash amount and noncash amount.

<table>
<tr><th colspan="2">Panel A:
No of litigations from 1996-2014</th><th colspan="2">Panel B:
No of litigations by Status</th></tr>
<tr><td>Year</td><td>No. of Litigations</td><td>Case Status</td><td>No. of Litigations</td></tr>
<tr><td>1996</td><td>73</td><td>Active</td><td>133</td></tr>
<tr><td>1997</td><td>121</td><td>Dismissed</td><td>917</td></tr>
<tr><td>1998</td><td>160</td><td>Settled (including Adjudicated at Trial)</td><td>1285</td></tr>
<tr><td>1999</td><td>158</td><td>**Total**</td><td>**2,335**</td></tr>
<tr><td>2000</td><td>136</td><td></td><td></td></tr>
<tr><td>2001</td><td>331</td><td></td><td></td></tr>
<tr><td>2002</td><td>150</td><td></td><td></td></tr>
<tr><td>2003</td><td>140</td><td></td><td></td></tr>
<tr><td>2004</td><td>160</td><td></td><td></td></tr>
<tr><td>2005</td><td>127</td><td></td><td></td></tr>
<tr><td>2006</td><td>79</td><td></td><td></td></tr>
<tr><td>2007</td><td>106</td><td></td><td></td></tr>
<tr><td>2008</td><td>110</td><td></td><td></td></tr>
<tr><td>2009</td><td>82</td><td></td><td></td></tr>
<tr><td>2010</td><td>95</td><td></td><td></td></tr>
<tr><td>2011</td><td>86</td><td></td><td></td></tr>
<tr><td>2012</td><td>83</td><td></td><td></td></tr>
<tr><td>2013</td><td>107</td><td></td><td></td></tr>
<tr><td>2014</td><td>31</td><td></td><td></td></tr>
<tr><td>**Total**</td><td>**2,335**</td><td></td><td></td></tr>
</table>

Panel C: Case Period, Class Period and Settlement Amount of Litigations

	Mean	Median	Std. Dev.	Min.	Max.
Case Period (days)	1,157	907	856	1	5,036
Class Period (days)	506	365	493	1	5,513
Settlement Amount	$28,238,076	$5,000,000	$149,955,942	$37,500	$3,200,000,000

64

Table 2 Matched Sample – Covariate Balance

The table reports the summary statistics for the independent variables used in my analysis for both "treated" (litigated firms) and "control" (non-litigated firms) sample. The last column reports the p-value of the mean differences. Variable definitions are included in Appendix B.

Variables	Firms litigated				Firms not litigated				Differences
	N	Mean	Median	Std. Dev.	N	Mean	Median	Std. Dev.	p-value
Size (Market Cap)	2,137	6,897.93	425.46	24,432.93	2,137	5,963.89	437.95	24,673.48	0.2138
Return on Assets (ROA)	2,137	-32.87%	-1.75%	236.11%	2,137	-83.18%	-0.96%	3247.86%	0.4752
Earnings Growth	2,137	23.31%	-1.49%	1365.89%	2,137	45.81%	-0.21%	2231.36%	0.6910
Sales Growth	2,137	128.03%	105.03%	267.81%	2,137	122.93%	106.21%	140.16%	0.4356
Loss Indicator	2,137	0.56	1.00	0.50	2,137	0.57	1.00	0.49	0.3085
Market-to-Book	2,069	2.46	1.60	8.26	2,072	3.96	1.66	51.72	0.1921
Stock Return	2,070	1.12	0.93	1.22	2,072	1.13	1.73	0.89	0.7953
Institutional Ownership	2,070	39.59%	36.62%	36.47%	2,072	37.84%	34.24%	35.62%	0.1193
Negative Earnings Surprise	1,092	0.66	1.00	0.47	1,092	0.66	1.00	0.47	0.8213
No. of Analyst Following	1,092	16.20	14.00	10.33	1,092	15.64	13.00	10.73	0.2129

Table 3.1 Summary Statistics on Textual Analysis variables for Litigated vs. Non-litigated Firms (10K)

This table reports the univariate comparisons in disclosure, readability, and sentiment in 10K (and its variants) filings between litigated firms and their matched non-litigated firms. The last column reports the p-value of the differences. Variable definitions are included in Appendix B.

Panel A: Disclosure Variables

Variables	Firms litigated				Firms not litigated				Differences
	N	Mean	Median	Std. Dev.	N	Mean	Median	Std. Dev.	p-value
File Size	1,929	4,144,937.00	1,178,344.00	9,456,263.00	1,929	4,378,273.00	1,218,502.00	8,991,052.00	0.4323
Word Count	1,928	43,390.69	35,643.00	35,271.02	1,929	38,562.74	33,351.00	27,688.19	< 0.0000
Complex word count	1,928	11,026.50	9,089.00	8,468.44	1,929	9,803.02	8,485.00	6,861.20	< 0.0000
Sentence count	1,928	1,700.51	1,473.50	1,103.03	1,929	1,534.13	1,378.00	964.69	< 0.0000
Paragraph Count	1,928	571.82	486.00	413.44	1,929	514.18	455.00	308.52	< 0.0000
Average No. of words per sentence	1,928	24.74	24.33	2.79	1,929	24.53	24.22	2.76	0.0191
Average No. of words per paragraph	1,928	229.52	71.02	4,090.52	1,929	162.16	70.12	1,920.71	0.5128

Panel B: Readability Variables

Variables	Firms litigated				Firms not litigated				Differences
	N	Mean	Median	Std. Dev.	N	Mean	Median	Std. Dev.	p-value
Flesch Reading Ease Index *	1,928	24.84	24.82	3.61	1,929	25.66	25.46	3.96	< 0.0000
Flesch-Kincaid Readability Index	1,928	15.94	15.83	1.05	1,929	15.77	15.73	1.08	< 0.0000
RIX Readability Index	1,928	8.92	8.83	0.88	1,929	8.77	8.73	0.88	< 0.0000
Gunning Fog Readability Index	1,928	20.14	20.03	1.11	1,929	20.03	19.98	1.11	0.0015
Automated Readability Index (ARI)	1,928	22.53	22.33	1.33	1,929	22.40	22.27	1.28	0.0019
Smog Readability Index	1,928	17.48	17.42	0.76	1,929	17.40	17.38	0.77	0.0008
LIX Readability Index	1,928	60.86	60.68	2.65	1,929	60.37	60.35	2.71	< 0.0000

Panel C: Sentiment Variables (%)

Variables	Firms litigated				Firms not litigated				Differences
	N	Mean	Median	Std. Dev.	N	Mean	Median	Std. Dev.	p-value
LM Negative Word (%)	1,928	1.74%	1.72%	0.52%	1,929	1.61%	1.59%	0.53%	< 0.0000
LM Positive Word (%)	1,928	0.82%	0.80%	0.19%	1,929	0.81%	0.79%	0.20%	0.0370
LM Modal Weak Word (%)	1,928	0.60%	0.58%	0.24%	1,929	0.56%	0.54%	0.24%	< 0.0000
LM Litigious Word (%)	1,928	1.13%	0.97%	0.57%	1,929	1.09%	0.90%	0.62%	0.0279
LM Uncertainty Word (%)	1,928	1.46%	1.47%	0.34%	1,929	1.40%	1.42%	0.35%	< 0.0000
LM Modal Strong Word (%)	1,928	0.32%	0.30%	0.11%	1,929	0.30%	0.28%	0.11%	< 0.0000
Harvard Negative Word (%)	1,928	3.99%	3.99%	0.65%	1,929	3.96%	3.99%	0.68%	0.1541

Panel D: Sentiment Variables (Count)

Variables	Firms litigated				Firms not litigated				Differences
	N	Mean	Median	Std. Dev.	N	Mean	Median	Std. Dev.	p-value
LM Negative Word (Count)	1,928	723.74	572.50	662.02	1,929	605.44	491.00	530.66	< 0.0000
LM Positive Word (Count)	1,928	315.23	267.00	229.80	1,929	280.39	251.00	194.20	< 0.0000
LM Modal Weak Word (Count)	1,928	237.20	206.50	180.30	1,929	205.19	175.00	156.46	< 0.0000
LM Litigious Word (Count)	1,928	505.73	312.00	831.93	1,929	418.60	270.00	603.30	0.0002
LM Uncertainty Word (Count)	1,928	570.51	497.00	412.27	1,929	500.90	442.00	352.84	< 0.0000
LM Modal Strong Word (Count)	1,928	122.77	103.00	111.11	1,929	104.66	89.00	80.14	< 0.0000
Harvard Negative Word (Count)	1,928	1,599.43	1,312.00	1,265.92	1,929	1,423.41	1,233.00	1,077.84	< 0.0000

* For Flesch Reading Ease Index, the higher is the value, the lower is the degree of difficulty in understanding the intended message of the text.

Table 3.2 Summary Statistics on Textual Analysis variables for Litigated vs. Non-litigated Firms (10Q)

This table reports the univariate comparisons in disclosure, readability, and sentiment in 10Q (and its variants) filings between litigated firms and their matched non-litigated firms. The last column reports the p-value of the differences. Variable definitions are included in Appendix B.

Panel A: Disclosure Variables

Variables	Firms litigated				Firms not litigated				Differences
	N	Mean	Median	Std. Dev.	N	Mean	Median	Std. Dev.	p-value
File Size	2,137	1,963,734.00	421,818.70	5,090,269.00	2,137	1,960,586.00	389,034.30	4,838,285.00	0.9835
Word Count	2,137	15,674.84	12,571.33	14,129.99	2,136	12,625.94	10,329.33	9,593.17	< 0.0000
Complex word count	2,137	3,830.94	3,081.67	3,382.03	2,136	3,071.43	2,467.17	2,386.53	< 0.0000
Sentence count	2,137	597.84	502.33	455.97	2,136	490.76	420.00	341.17	< 0.0000
Paragraph Count	2,137	200.33	164.00	169.29	2,136	165.93	137.67	114.30	< 0.0000
Average No. of words per sentence	2,137	25.37	25.15	2.49	2,136	24.91	24.86	2.76	< 0.0000
Average No. of words per paragraph	2,137	129.15	73.46	1,089.78	2,136	133.58	70.41	1,213.55	0.9001

Panel B: Readability Variables

Variables	Firms litigated				Firms not litigated				Differences
	N	Mean	Median	Std. Dev.	N	Mean	Median	Std. Dev.	p-value
Flesch Reading Ease Index *	2,137	26.90	26.55	4.31	2,136	28.33	28.15	5.02	< 0.0000
Flesch-Kincaid Readability Index	2,137	15.81	15.79	1.08	2,136	15.50	15.52	1.25	< 0.0000
RIX Readability Index	2,137	8.96	8.90	0.88	2,136	8.72	8.70	1.00	< 0.0000
Gunning Fog Readability Index	2,137	19.91	19.87	1.17	2,136	19.62	19.64	1.32	< 0.0000
Automated Readability Index (ARI)	2,137	22.67	22.54	1.31	2,136	22.48	22.39	1.42	< 0.0000
Smog Readability Index	2,137	17.31	17.30	0.81	2,136	17.10	17.13	0.92	< 0.0000
LIX Readability Index	2,137	60.73	60.61	2.73	2,136	59.95	59.85	3.08	< 0.0000

Panel C: Sentiment Variables (%)

Variables	Firms litigated				Firms not litigated				Differences
	N	Mean	Median	Std. Dev.	N	Mean	Median	Std. Dev.	p-value
LM Negative Word (%)	2,137	2.04%	1.95%	0.80%	2,136	1.63%	1.50%	0.72%	< 0.0000
LM Positive Word (%)	2,137	0.73%	0.71%	0.22%	2,136	0.73%	0.70%	0.24%	0.5640
LM Modal Weak Word (%)	2,137	0.60%	0.45%	0.40%	2,136	0.50%	0.39%	0.34%	< 0.0000
LM Litigious Word (%)	2,137	1.17%	1.01%	0.62%	2,136	0.97%	0.81%	0.60%	< 0.0000
LM Uncertainty Word (%)	2,137	1.56%	1.48%	0.50%	2,136	1.40%	1.33%	0.48%	< 0.0000
LM Modal Strong Word (%)	2,137	0.32%	0.30%	0.14%	2,136	0.29%	0.27%	0.14%	< 0.0000
Harvard Negative Word (%)	2,137	4.31%	4.27%	0.77%	2,136	4.04%	4.02%	0.81%	< 0.0000

Panel D: Sentiment Variables (Count)

Variables	Firms litigated				Firms not litigated				Differences
	N	Mean	Median	Std. Dev.	N	Mean	Median	Std. Dev.	p-value
LM Negative Word (Count)	2,137	311.17	230.33	308.78	2,136	214.07	140.50	226.29	< 0.0000
LM Positive Word (Count)	2,137	105.77	85.00	94.00	2,136	86.85	66.33	77.50	< 0.0000
LM Modal Weak Word (Count)	2,137	91.14	54.67	94.46	2,136	66.26	36.00	82.63	< 0.0000
LM Litigious Word (Count)	2,137	188.19	118.00	302.16	2,136	129.58	78.00	185.01	< 0.0000
LM Uncertainty Word (Count)	2,137	225.50	174.33	190.41	2,136	171.83	125.50	156.86	< 0.0000
LM Modal Strong Word (Count)	2,137	45.29	35.00	41.09	2,136	34.47	26.00	31.86	< 0.0000
Harvard Negative Word (Count)	2,137	625.14	501.33	549.21	2,136	483.08	373.33	407.44	< 0.0000

For Flesch Reading Ease Index, the higher is the value, the lower is the degree of difficulty in understanding the intended message of the text.

Table 4 Disclosure and Litigation – 10K and 10Q Filings

The table shows the association between disclosure by a firm (using seven different measures) and the probability of the firm being litigated. Panel A has the results from analysis of firms' 10K filings, while Panel B has the results from analysis of firms' 10Q filings. All models control for size (natural logarithm of market value), return on assets (ROA), loss indicator (negative net income), earnings growth, sales growth, market-to-book, big-8 auditor dummy, lag size, lag ROA, lag loss indicator, lag stock volatility, lag stock return, and lag institutional ownership. All models include year and industry (2-digit SIC code) fixed effects, with clustering at the firm level. Standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels. Variable definitions are explained in Appendix B.

Variables	Panel A: 10K Filings (Litigation Logit Model)							Panel B: 10Q Filings (Litigation Logit Model)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	D_Litigation	D_Litigation	D_Litigation	D_Litigation	D_Litigation	D_Litigation	D_Litigation	D_Litigation	D_Litigation	D_Litigation	D_Litigation	D_Litigation	D_Litigation	D_Litigation
LN (File Size)	0.152***							0.310***						
	(0.058)							(0.050)						
LN (Word Count)		0.463***							0.791***					
		(0.090)							(0.078)					
LN (Complex Word Count)			0.501***							0.792***				
			(0.094)							(0.076)				
LN (Sentence Count)				0.510***							0.827***			
				(0.101)							(0.084)			
LN (Average Words per Sentence)					0.995***							2.236***		
					(0.357)							(0.362)		
LN (Paragraph Count)						0.311***							0.580***	
						(0.095)							(0.097)	
LN (Average Words per Paragraph)							0.125							0.312***
							(0.101)							(0.119)
Constant	-4.241***	-6.656***	-6.283***	-5.508***	-5.425***	-3.960***	-2.834***	-5.187***	-7.865***	-6.685***	-5.551***	-9.062***	-4.023***	-3.356***
	(1.187)	(1.307)	(1.245)	(1.177)	(1.511)	(1.122)	(1.085)	(0.706)	(0.747)	(0.653)	(0.595)	(1.224)	(0.586)	(0.662)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	3732	3732	3732	3732	3732	3732	3732	4193	4193	4193	4193	4193	4193	4193
Pseudo R-squared	0.140	0.145	0.145	0.144	0.140	0.142	0.139	0.141	0.155	0.156	0.153	0.141	0.145	0.137

Table 5 Readability and Litigation – 10K and 10Q Filings

The table shows the association between readability of a firm's filings (using seven different measures) and the probability of the firm being litigated. Panel A has the results from analysis of firms' 10K filings, while Panel B has the results from analysis of firms' 10Q filings. All models control for size (natural logarithm of market value), return on assets (ROA), loss indicator (negative net income), earnings growth, sales growth, market-to-book, big-8 auditor dummy, lag size, lag ROA, lag loss indicator, lag stock volatility, lag stock return, and lag institutional ownership. All models include year and industry (2-digit SIC code) fixed effects, with clustering at the firm level. Standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels. Variable definitions are explained in Appendix B.

Variables	Panel A: 10K Filings (Litigation Logit Model)							Panel B: 10Q Filings (Litigation Logit Model)						
	(1) D_Litigation	(2) D_Litigation	(3) D_Litigation	(4) D_Litigation	(5) D_Litigation	(6) D_Litigation	(7) D_Litigation	(1) D_Litigation	(2) D_Litigation	(3) D_Litigation	(4) D_Litigation	(5) D_Litigation	(6) D_Litigation	(7) D_Litigation
Automated Readability Index (ARI)	0.047* (0.027)							0.091*** (0.025)						
Flesch-Kincaid Readability Index		0.136*** (0.037)							0.260*** (0.035)					
Gunning Fog Readability			0.102*** (0.034)							0.216*** (0.031)				
Smog Readability Index				0.158*** (0.050)							0.323*** (0.044)			
Flesch Reading Ease Index					-0.050*** (0.011)							-0.071*** (0.009)		
LIX Readability Index						0.054*** (0.014)							0.083*** (0.013)	
RIX Readability Index							0.159*** (0.043)							0.262*** (0.040)
Constant	-3.313*** (1.159)	-4.231*** (1.148)	-4.198*** (1.190)	-4.878*** (1.296)	-0.584 (1.101)	-5.369*** (1.301)	-3.512*** (1.076)	-4.112*** (0.748)	-5.631*** (0.676)	-5.987*** (0.740)	-7.238*** (0.858)	0.605 (0.579)	-6.760*** (0.879)	-4.084*** (0.570)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	3732	3732	3732	3732	3732	3732	3732	4193	4193	4193	4193	4193	4193	4193
Pseudo R-squared	0.139	0.141	0.140	0.140	0.142	0.141	0.141	0.137	0.145	0.143	0.144	0.146	0.142	0.142

For Flesch Reading Ease Index, the higher is the value, the lower is the degree of difficulty in understanding the intended message of the text.

Table 6.1 Sentiment (Count) and Litigation – 10K and 10Q Filings

The table shows the association between sentiment of a firm's filings (measured by count) and the probability of the firm being litigated. Panel A has the results from analysis of firms' 10K filings, while Panel B has the results from analysis of firms' 10Q filings. All models control for size (natural logarithm of market value), return on assets (ROA), loss indicator (negative net income), earnings growth, sales growth, market-to-book, big-8 auditor dummy, lag size, lag ROA, lag loss indicator, lag stock volatility, lag stock return, and lag institutional ownership. All models include year and industry (2-digit SIC code) fixed effects, with clustering at the firm level. Standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels. Variable definitions are explained in Appendix B.

Variables	Panel A: 10K Filings (Litigation Logit Model)							Panel B: 10Q Filings (Litigation Logit Model)						
	(1) D_Litigation	(2) D_Litigation	(3) D_Litigation	(4) D_Litigation	(5) D_Litigation	(6) D_Litigation	(7) D_Litigation	(1) D_Litigation	(2) D_Litigation	(3) D_Litigation	(4) D_Litigation	(5) D_Litigation	(6) D_Litigation	(7) D_Litigation
LN (Uncertainty Word Count)	0.580*** (0.088)							0.634*** (0.057)						
LN (Modal Weak Word Count)		0.527*** (0.067)							0.387*** (0.041)					
LN (Negative Word Count)			0.484*** (0.069)							0.658*** (0.050)				
LN (Positive Word Count)				0.366*** (0.077)							0.464*** (0.059)			
LN (Litigious Word Count)					0.323*** (0.059)							0.581*** (0.047)		
LN (Modal Strong Word Count)						0.431*** (0.065)							0.484*** (0.049)	
LN (Harvard Negative Word Count)							0.449*** (0.084)							0.751*** (0.068)
Constant	-4.883*** (1.082)	-3.971*** (1.019)	-4.516*** (1.099)	-3.894*** (1.062)	-3.739*** (1.043)	-3.865*** (1.060)	-4.875*** (1.125)	-3.420*** (0.500)	-2.344*** (0.491)	-3.661*** (0.504)	-2.960*** (0.501)	-3.480*** (0.502)	-2.363*** (0.489)	-4.926*** (0.552)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	3732	3730	3732	3732	3732	3731	3732	4192	4191	4193	4193	4193	4187	4193
Pseudo R-squared	0.150	0.152	0.150	0.143	0.145	0.147	0.145	0.157	0.151	0.168	0.146	0.166	0.153	0.158

Table 6.2 Sentiment (%) and Litigation – 10K and 10Q Filings

The table shows the association between sentiment of a firm's filings (measured by percentage) and the probability of the firm being litigated. Panel A has the results from analysis of firms' 10K filings, while Panel B has the results from analysis of firms' 10Q filings. All models control for size (natural logarithm of market value), return on assets (ROA), loss indicator (negative net income), earnings growth, sales growth, market-to-book, big-8 auditor dummy, lag size, lag ROA, lag loss indicator, lag stock volatility, lag stock return, and lag institutional ownership. All models include year and industry (2-digit SIC code) fixed effects, with clustering at the firm level. Standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels. Variable definitions are explained in Appendix B.

Variables	Panel A: 10K Filings (Litigation Logit Model)							Panel B: 10Q Filings (Litigation Logit Model)						
	(1) D_Litigation	(2) D_Litigation	(3) D_Litigation	(4) D_Litigation	(5) D_Litigation	(6) D_Litigation	(7) D_Litigation	(1) D_Litigation	(2) D_Litigation	(3) D_Litigation	(4) D_Litigation	(5) D_Litigation	(6) D_Litigation	(7) D_Litigation
Uncertainty Word Proportion	65.656*** (14.387)							41.290*** (8.190)						
Modal Weak Word Proportion		116.855*** (22.684)							50.508*** (11.232)					
Negative Word Proportion			57.946*** (8.786)							56.498*** (5.337)				
Positive Word Proportion				-14.859 (19.331)							-39.918** (17.056)			
Litigious Word Proportion					13.668* (6.977)							51.871*** (6.573)		
Modal Strong Word Proportion						132.997*** (38.484)							136.683*** (29.820)	
Harvard Negative Word Proportion							13.708** (6.543)							25.689*** (4.993)
Constant	-2.837*** (1.002)	-2.545** (1.000)	-2.828*** (1.069)	-2.218** (1.012)	-2.436** (1.005)	-2.805*** (1.032)	-2.716*** (1.031)	-2.336*** (0.482)	-2.121*** (0.485)	-2.591*** (0.489)	-1.903*** (0.489)	-2.635*** (0.486)	-2.348*** (0.488)	-3.016*** (0.527)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	3732	3732	3732	3732	3732	3732	3732	4193	4193	4193	4193	4193	4193	4193
Pseudo R-squared	0.143	0.145	0.147	0.138	0.139	0.141	0.139	0.139	0.139	0.157	0.136	0.147	0.139	0.139

Table 7: Principal Component Analysis – Litigation Risk

The table shows the association between the principal components of degree, readability, and sentiments of a firm's textual disclosure and the probability of the firm being litigated. Panel A has the results from analysis of firms' 10K filings, while Panel B has the results from analysis of firms' 10Q filings. All models control for size (natural logarithm of market value), return on assets (ROA), loss indicator (negative net income), earnings growth, sales growth, market-to-book, big-8 auditor dummy, lag size, lag ROA, lag loss indicator, lag stock volatility, lag stock return, and lag institutional ownership. All models include year and industry (2-digit SIC code) fixed effects, with clustering at the firm level. Standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels. Variable definitions are explained in Appendix B.

	Panel A: 10K Filings (Logit Model)				Panel B: 10Q Filings (Logit Model)			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Variables	D_litigation	D_litigation	D_litigation	D_litigation	D_litigation	D_litigation	D_litigation	D_litigation
Disclosure (PCA)	0.140***				0.280***			
	(0.027)				(0.028)			
Readability (PCA)		0.053***				0.103***		
		(0.016)				(0.016)		
Uncertainty % (PCA)			0.204***				0.214***	
			(0.033)				(0.026)	
Uncertainty Count (PCA)				0.163***				0.188***
				(0.036)				(0.028)
Constant	-1.803*	-2.126**	-1.656	-1.692	-0.413	-1.747***	-1.496***	-1.283**
	(1.019)	(1.025)	(1.030)	(1.034)	(0.517)	(0.489)	(0.494)	(0.500)
Controls	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y	Y
Observations	3732	3732	3732	3732	4095	4095	4095	4095
Pseudo R-squared	0.145	0.140	0.147	0.145	0.156	0.143	0.148	0.146

Table 8: Principal Component Analysis – Case Period

The table shows the association between the principal components of degree, readability, and sentiments of a firm's textual disclosure and the length of case period in days (the difference between the case end date and the case filing date). Panel A has the results from analysis of firms' 10K filings, while Panel B has the results from analysis of firms' 10Q filings. All models control for size (natural logarithm of market value), return on assets (ROA), loss indicator (negative net income), earnings growth, sales growth, market-to-book, big-8 auditor dummy, lag size, lag ROA, lag loss indicator, lag stock volatility, lag stock return, and lag institutional ownership. All models include year and industry (2-digit SIC code) fixed effects, with clustering at the firm level. Standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels. Variable definitions are explained in Appendix B.

Variables	Panel A: 10K Filings (OLS Model)				Panel B: 10Q Filings (OLS Model)			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Case Period (days)				Case Period (days)			
Disclosure (PCA)	8.387				15.885			
	(11.655)				(12.520)			
Readability (PCA)		-10.599				-11.813*		
		(7.457)				(7.108)		
Uncertainty % (PCA)			39.673***				35.622***	
			(14.758)				(12.485)	
Uncertainty Count (PCA)				10.485				8.705
				(9.582)				(10.210)
Size (ln_MV)	-17.873	-18.259	-17.271	-17.849	11.511	10.568	12.984	11.401
	(18.812)	(18.830)	(18.697)	(18.789)	(18.880)	(18.912)	(18.691)	(18.889)
ROA	-3.866	-3.688	-5.507	-4.059	0.031	0.215	-0.959	-0.125
	(13.255)	(13.276)	(12.301)	(13.198)	(9.092)	(9.226)	(9.396)	(9.239)
Earnings Growth	22.990	23.941	20.471	22.756	1.599***	1.697***	1.608***	1.620***
	(25.268)	(25.127)	(24.910)	(25.243)	(0.314)	(0.317)	(0.314)	(0.316)
Sales Growth	-2.271	-2.474	-1.918	-2.228	-1.594	-1.890	-0.818	-1.623
	(2.507)	(2.621)	(2.301)	(2.489)	(2.615)	(2.799)	(2.668)	(2.680)
Loss Indicator	112.330***	114.920***	102.315**	111.956***	128.286***	136.702***	124.504***	130.876***
	(43.376)	(43.013)	(43.253)	(43.247)	(42.021)	(41.421)	(41.888)	(41.978)
Lag Size (ln_MV)	80.259***	83.893***	81.038***	80.058***	40.916*	48.618**	41.458*	43.575**
	(22.852)	(22.630)	(22.469)	(22.882)	(22.306)	(21.913)	(21.746)	(22.149)
Lag_Market-to-Book	-0.429	-0.435	-0.196	-0.422	-1.791	-1.879	-1.934	-1.872
	(1.118)	(1.113)	(1.115)	(1.115)	(1.617)	(1.728)	(1.616)	(1.663)
Lag_Volatility	41.241	33.725	24.487	39.045	44.661	42.020	32.817	43.594
	(52.701)	(52.846)	(51.557)	(52.589)	(46.044)	(47.270)	(45.923)	(46.584)
Lag_Return on Assets (ROA)	-1.956	-2.561	-1.875	-1.947	-4.747	-5.738	-5.001	-5.039
	(9.949)	(9.960)	(9.227)	(9.905)	(6.928)	(7.008)	(7.133)	(7.026)
Lag_Loss Indicator	73.964	81.855*	66.293	73.017	31.738	45.780	24.543	34.719
	(45.216)	(45.393)	(45.712)	(45.548)	(43.658)	(43.195)	(43.353)	(43.842)
Stock Return	-61.969***	-63.240***	-62.636***	-62.229***	-71.356***	-73.309***	-70.919***	-72.184***
	(16.944)	(17.115)	(16.687)	(16.917)	(15.319)	(15.559)	(14.741)	(15.401)
Big-8 Auditor	-83.084	-79.306	-89.143*	-83.935	-79.184	-71.183	-87.100*	-77.530
	(52.028)	(52.013)	(51.453)	(52.092)	(48.708)	(48.553)	(48.690)	(48.967)
Institutional Ownership	-41.599	-45.698	-52.481	-42.067	-77.180	-82.499	-89.207*	-79.666
	(55.690)	(55.788)	(56.115)	(55.668)	(53.332)	(53.222)	(53.657)	(53.185)
Constant	366.714	319.850	490.396*	383.675	593.560**	452.750**	631.469***	552.334**
	(283.927)	(287.001)	(290.593)	(285.779)	(231.643)	(224.062)	(228.040)	(230.005)
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y	Y
Observations	1625	1625	1625	1625	1833	1833	1833	1833
Pseudo R-squared	0.324	0.324	0.327	0.324	0.355	0.355	0.358	0.354

Table 9: Event Study on Case Filing Date

Table presents the mean cumulative abnormal returns (CARs) around the case filing date for six different event windows (days): (-30, -2), (-1, +1), -2, +2), (0, 0), (0, +1), and (0, +2). All CARs are computed as the difference between the actual stock price return and the expected market model (CRSP Value Weighted Index) return over the windows listed. Market model has been estimated using 255 days of daily returns ending 46 days prior to the case filing date. The table also reports both the Patell Z-statistic and the corresponding p-value.

Windows (days)	N	Mean CAR	Patell Z	p-value
(-30, -2)	1888	-13.06%	-35.995	<.0001
(-1, +1)	1871	-4.14%	-30.866	<.0001
(-2, +2)	1874	-5.92%	-35.131	<.0001
(0, 0)	1867	-1.42%	-18.164	<.0001
(0, +1)	1867	-2.27%	-18.321	<.0001
(0, +2)	1867	-2.85%	-18.851	<.0001

Table 10: Principal Component Analysis – Market Response

The table shows the association between the principal components of degree, readability, and sentiments of a firm's textual disclosure and the cumulative abnormal return (CAR) around the case filing date. Panel A has the results from analysis of firms' 10K filings, while Panel B has the results from analysis of firms' 10Q filings. All models control for size (natural logarithm of market value), return on assets (ROA), loss indicator (negative net income), earnings growth, sales growth, market-to-book, big-8 auditor dummy, lag size, lag ROA, lag loss indicator, lag stock volatility, lag stock return, lag institutional ownership, and dummy indicating whether the litigation is either merger-related or accounting related. All models include year and industry (2-digit SIC code) fixed effects, with clustering at the firm level. Standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels. Variable definitions are explained in Appendix B.

Variables	Panel A: 10K Filings (OLS Model)								Panel B: 10Q Filings (OLS Model)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	CAR Window (-1,+1)				CAR Window (-2,+2)				CAR Window (-1,+1)				CAR Window (-2,+2)			
Disclosure (PCA)	-0.006				-0.005				-0.003				-0.004			
	(0.010)				(0.013)				(0.009)				(0.012)			
Readability (PCA)		0.012				0.015				-0.009				-0.013		
		(0.009)				(0.011)				(0.009)				(0.011)		
Uncertainty % (PCA)			-0.019**				-0.025**				-0.021**				-0.021**	
			(0.009)				(0.011)				(0.008)				(0.011)	
Uncertainty Count (PCA)				-0.021**				-0.031**				-0.004				-0.011
				(0.009)				(0.012)				(0.009)				(0.011)
Constant	0.099**	0.083*	0.116***	0.110**	0.045	0.025	0.069	0.062	0.089**	0.097***	0.109***	0.090***	0.070	0.081*	0.088**	0.075*
	(0.043)	(0.042)	(0.044)	(0.044)	(0.044)	(0.043)	(0.044)	(0.043)	(0.035)	(0.035)	(0.035)	(0.035)	(0.045)	(0.045)	(0.045)	(0.044)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	1251	1251	1251	1251	1253	1253	1253	1253	1436	1436	1436	1436	1438	1438	1438	1438
Pseudo R-squared	0.090	0.091	0.093	0.094	0.087	0.088	0.091	0.092	0.070	0.071	0.074	0.070	0.088	0.089	0.091	0.089

Table 11: Principal Component Analysis – Settlement Amount

The table shows the association between the principal components of degree, readability, and sentiments of a firm's textual disclosure and the litigation settlement amount (cash amount and total amount) scaled by Disclosure Dollar Loss (DDL). Panel A has the results from analysis of firms' 10K filings, while Panel B has the results from analysis of firms' 10Q filings. All models control for size (natural logarithm of market value), return on assets (ROA), loss indicator (negative net income), earnings growth, sales growth, market-to-book, big-8 auditor dummy, lag size, lag ROA, lag loss indicator, lag stock volatility, lag stock return, and lag institutional ownership. All models include year and industry (2-digit SIC code) fixed effects, with clustering at the firm level. Standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels. Variable definitions are explained in Appendix B.

Variables	Panel A: 10K Filings (OLS Model)								Panel B: 10Q Filings (OLS Model)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Cash Settlement Amount				Total Settlement Amount				Cash Settlement Amount				Total Settlement Amount			
Disclosure (PCA)	-0.035				-0.034				-0.003				-0.002			
	(0.039)				(0.039)				(0.034)				(0.034)			
Readability (PCA)		0.003				0.002				-0.034				-0.034		
		(0.026)				(0.026)				(0.022)				(0.022)		
Uncertainty % (PCA)			-0.071**				-0.072**				-0.053*				-0.055*	
			(0.029)				(0.029)				(0.028)				(0.028)	
Uncertainty Count (PCA)				-0.017				-0.017				-0.014				-0.013
				(0.040)				(0.039)				(0.031)				(0.031)
Constant	-1.831***	-2.087***	-2.288***	-1.986***	-1.774***	-2.012***	-2.221***	-1.918***	-2.058***	-2.175***	-2.162***	-2.106***	-1.965***	-2.090***	-2.080***	-2.019***
	(0.477)	(0.450)	(0.390)	(0.428)	(0.479)	(0.455)	(0.398)	(0.432)	(0.388)	(0.396)	(0.353)	(0.345)	(0.392)	(0.403)	(0.360)	(0.352)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	568	568	568	568	569	569	569	569	564	564	564	564	565	565	565	565
Pseudo R-squared	0.062	0.059	0.067	0.060	0.060	0.058	0.066	0.059	0.055	0.060	0.061	0.055	0.053	0.058	0.059	0.053

Table 12: Litigation Risk and Disclosure (1999 Ninth Circuit Decision)

The table shows the association between litigation risk and disclosure by a firm. Panel A has the results from analysis of the 10K filings, while Panel B has the results from analysis of the 10Q filings. Using the 1999 Ninth Circuit Decision as a shock to a firm's litigation risk, all models are conducted in a difference-in-difference setup, controlling for size (natural logarithm of market value), return on assets (ROA), earnings growth, sales growth, market-to-book, big-8 auditor dummy, stock volatility, and institutional ownership. All models include firm and year fixed effects, with clustering at the firm level. Standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels. Variable definitions are explained in Appendix B.

Variables	Panel A: 10K Filings (OLS Model)			Panel B: 10Q Filings (OLS Model)		
	(1) Disclosure (PCA)	(2) Uncertainty % (PCA)	(3) Uncertainty Count (PCA)	(1) Disclosure (PCA)	(2) Uncertainty % (PCA)	(3) Uncertainty Count (PCA)
9thCircuit x Post1999	-0.009	0.206**	0.147	0.240*	0.507***	0.570***
	(0.109)	(0.095)	(0.109)	(0.124)	(0.110)	(0.121)
Size (ln_MV)	0.093***	-0.042*	0.052*	0.080***	-0.017	0.035
	(0.026)	(0.023)	(0.029)	(0.027)	(0.025)	(0.033)
ROA	0.005	-0.004	0.008	0.002	0.003	0.010
	(0.005)	(0.007)	(0.009)	(0.006)	(0.002)	(0.012)
Earnings Growth	-0.011*	0.001	-0.006*	-0.011**	-0.008	-0.010**
	(0.006)	(0.004)	(0.004)	(0.005)	(0.006)	(0.005)
Sales Growth	-0.012***	-0.009**	-0.012***	0.017***	0.007	0.005
	(0.003)	(0.004)	(0.005)	(0.003)	(0.004)	(0.006)
Market-to-Book	-0.015	-0.000	-0.016	0.001	0.002	-0.001
	(0.023)	(0.016)	(0.018)	(0.004)	(0.002)	(0.003)
Volatility	0.032	0.072**	0.060	0.027	0.035	0.068*
	(0.040)	(0.033)	(0.044)	(0.037)	(0.042)	(0.041)
Stock Return	-0.044***	0.006	-0.040***	-0.038***	-0.021	-0.059***
	(0.013)	(0.012)	(0.014)	(0.013)	(0.017)	(0.019)
Big-8 Auditor	0.068	0.212**	0.096	0.228**	0.065	0.157*
	(0.125)	(0.099)	(0.109)	(0.107)	(0.102)	(0.092)
Institutional Ownership	-0.206	0.192	0.229	-0.051	0.187	0.168
	(0.220)	(0.184)	(0.336)	(0.250)	(0.227)	(0.289)
Constant	-0.607***	-0.223*	-0.473***	-1.285***	-0.162	-0.560***
	(0.171)	(0.131)	(0.172)	(0.155)	(0.143)	(0.166)
Firm Fixed Effect	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y
Clusters at Firm Level	Y	Y	Y	Y	Y	Y
Observations	10,694	10,696	10,696	11,598	11,600	11,600
Adjusted R-squared	0.695	0.784	0.673	0.760	0.756	0.752

Table 13.1: Disclosure Analysis – Pre- vs. Post-Class Start Date (10K and 10Q Filings)

The table shows the change in the textual disclosure (degree) of a firm's filings pre- vs. post-class start date. Panel A and C have the results from analysis of firms' 10K filings for settled and dismissed cases respectively. Panel B and D have the results from analysis of firms' 10Q filings for settled and dismissed cases respectively. All models control for size (natural logarithm of market value), return on assets (ROA), loss indicator (negative net income), earnings growth, sales growth, market-to-book, big-8 auditor dummy, lag size, lag ROA, lag loss indicator, lag stock volatility, lag stock return, and lag institutional ownership. All models include year and industry (2-digit SIC code) fixed effects, with clustering at the firm level. Standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels. Variable definitions are explained in Appendix B.

	Panel A: 10K Filings (Settled Cases)							Panel B: 10Q Filings (Settled Cases)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variables	LN (File Size)	LN (Word Count)	LN (Complex Word Count)	LN (Sentence Count)	LN (Average Words per Sentence)	LN (Paragraph Count)	LN (Average Words per Paragraph)	LN (File Size)	LN (Word Count)	LN (Complex Word Count)	LN (Sentence Count)	LN (Average Words per Sentence)	LN (Paragraph Count)	LN (Average Words per Paragraph)
Post-ClassStartDate x Litigated	-0.258***	-0.077**	-0.073**	-0.073**	-0.003	-0.068*	-0.008	-0.348***	-0.143***	-0.146***	-0.124***	-0.018***	-0.066**	-0.087***
	(0.040)	(0.033)	(0.032)	(0.029)	(0.008)	(0.037)	(0.026)	(0.043)	(0.032)	(0.032)	(0.028)	(0.007)	(0.032)	(0.030)
Litigated	0.031	0.019	0.021	0.020	-0.001	-0.020	0.040	0.071*	0.047	0.051	0.057**	-0.007	0.068**	-0.020
	(0.039)	(0.031)	(0.030)	(0.028)	(0.007)	(0.039)	(0.031)	(0.043)	(0.031)	(0.031)	(0.028)	(0.007)	(0.032)	(0.029)
Post-ClassStartDate	-0.316***	-0.114***	-0.114***	-0.108***	-0.006*	-0.067***	-0.047***	-0.432***	-0.217***	-0.224***	-0.196***	-0.021***	-0.140***	-0.090***
	(0.019)	(0.013)	(0.013)	(0.012)	(0.003)	(0.013)	(0.008)	(0.020)	(0.015)	(0.016)	(0.013)	(0.014)	(0.014)	(0.012)
Constant	11.038***	8.261***	6.837***	5.186***	3.075***	4.609***	3.601***	10.326***	8.026***	6.623***	4.863***	3.167***	3.913***	4.107***
	(0.176)	(0.152)	(0.146)	(0.125)	(0.044)	(0.150)	(0.097)	(0.153)	(0.140)	(0.137)	(0.119)	(0.031)	(0.126)	(0.073)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	2581	2581	2581	2581	2581	2581	2581	2707	2707	2707	2707	2707	2707	2707
Pseudo R-squared	0.650	0.432	0.458	0.461	0.073	0.293	0.101	0.713	0.567	0.577	0.597	0.111	0.460	0.103

	Panel C: 10K Filings (Dismissed Cases)							Panel D: 10Q Filings (Dismissed Cases)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variables	LN (File Size)	LN (Word Count)	LN (Complex Word Count)	LN (Sentence Count)	LN (Average Words per Sentence)	LN (Paragraph Count)	LN (Average Words per Paragraph)	LN (File Size)	LN (Word Count)	LN (Complex Word Count)	LN (Sentence Count)	LN (Average Words per Sentence)	LN (Paragraph Count)	LN (Average Words per Paragraph)
Post-ClassStartDate x Litigated	-0.385***	-0.074**	-0.072**	-0.069**	-0.005	-0.070**	-0.003	-0.485***	-0.137***	-0.140***	-0.133***	-0.001	-0.122***	-0.014
	(0.041)	(0.030)	(0.029)	(0.027)	(0.006)	(0.028)	(0.013)	(0.045)	(0.035)	(0.035)	(0.032)	(0.007)	(0.031)	(0.013)
Litigated	0.038	0.031	0.032	0.034	-0.003	0.013	0.018	0.047	0.036	0.039	0.033	0.006	0.019	0.019
	(0.040)	(0.029)	(0.028)	(0.026)	(0.006)	(0.027)	(0.013)	(0.044)	(0.034)	(0.034)	(0.031)	(0.006)	(0.030)	(0.013)
Post-ClassStartDate	-0.410***	-0.106***	-0.105***	-0.104***	-0.002	-0.084***	-0.022***	-0.524***	-0.177***	-0.183***	-0.167***	-0.010***	-0.144***	-0.034***
	(0.019)	(0.012)	(0.012)	(0.011)	(0.002)	(0.011)	(0.005)	(0.020)	(0.013)	(0.013)	(0.012)	(0.003)	(0.011)	(0.005)
Constant	11.562***	9.102***	7.705***	6.025***	3.077***	5.129***	3.944***	9.418***	6.922***	5.427***	3.895***	3.026***	3.030***	3.837***
	(0.530)	(0.247)	(0.232)	(0.216)	(0.043)	(0.200)	(0.114)	(0.462)	(0.163)	(0.174)	(0.160)	(0.040)	(0.138)	(0.079)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	2540	2540	2540	2540	2540	2540	2540	2598	2598	2598	2598	2598	2598	2598
Pseudo R-squared	0.745	0.440	0.458	0.460	0.055	0.363	0.093	0.789	0.539	0.544	0.561	0.115	0.510	0.208

Table 13.2: Readability Analysis – Pre- vs. Post-Class Start Date (10K and 10Q Filings)

The table shows the change in the textual disclosure (readability) of a firm's filings pre- vs. post-class start date. Panel A and C have the results from analysis of firms' 10K filings for settled and dismissed cases respectively. Panel B and D have the results from analysis of firms' 10Q filings for settled and dismissed cases respectively. All models control for size (natural logarithm of market value), return on assets (ROA), loss indicator (negative net income), earnings growth, sales growth, market-to-book, big-8 auditor dummy, lag size, lag ROA, lag loss indicator, lag stock volatility, lag stock return, and lag institutional ownership. All models include year and industry (2-digit SIC code) fixed effects, with clustering at the firm level. Standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels. Variable definitions are explained in Appendix B.

	Panel A: 10K Filings (Settled Cases)							Panel B: 10Q Filings (Settled Cases)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variables	Automated Readability Index (ARI)	Flesch-Kincaid Readability Index	Gunning Fog Readability	Smog Readability Index	Flesch Reading Ease Index	LIX Readability Index	RIX Readability Index	Automated Readability Index (ARI)	Flesch-Kincaid Readability Index	Gunning Fog Readability	Smog Readability Index	Flesch Reading Ease Index	LIX Readability Index	RIX Readability Index
Post-ClassStartDate x Litigated	-0.090	-0.039	0.005	0.001	0.129	0.003	-0.014	-0.241***	-0.249***	-0.211***	-0.146***	0.970***	-0.508***	-0.175***
	(0.093)	(0.071)	(0.075)	(0.051)	(0.221)	(0.176)	(0.059)	(0.093)	(0.071)	(0.078)	(0.053)	(0.253)	(0.185)	(0.061)
Litigated	-0.135	0.014	-0.002	0.003	-0.184	0.027	0.001	-0.214**	-0.056	-0.058	-0.029	0.031	-0.202	-0.070
	(0.090)	(0.067)	(0.071)	(0.048)	(0.211)	(0.168)	(0.057)	(0.090)	(0.069)	(0.073)	(0.051)	(0.246)	(0.175)	(0.058)
Post-ClassStartDate	-0.013	-0.089***	-0.043	-0.040*	0.429***	-0.137*	-0.053**	-0.145***	-0.271***	-0.243***	-0.180***	1.124***	-0.502***	-0.175***
	(0.041)	(0.032)	(0.034)	(0.023)	(0.096)	(0.081)	(0.027)	(0.048)	(0.039)	(0.044)	(0.031)	(0.127)	(0.099)	(0.032)
Constant	21.500***	13.461***	18.411***	16.196***	37.691***	54.359***	7.041***	21.572***	14.406***	19.402***	16.936***	34.372***	56.986***	7.877***
	(0.520)	(0.426)	(0.433)	(0.281)	(1.202)	(0.999)	(0.325)	(0.407)	(0.314)	(0.341)	(0.222)	(1.030)	(0.758)	(0.240)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	2581	2581	2581	2581	2581	2581	2581	2707	2707	2707	2707	2707	2707	2707
Pseudo R-squared	0.106	0.142	0.119	0.127	0.286	0.154	0.120	0.060	0.222	0.171	0.188	0.347	0.153	0.136

	Panel C: 10K Filings (Dismissed Cases)							Panel D: 10Q Filings (Dismissed Cases)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variables	Automated Readability Index (ARI)	Flesch-Kincaid Readability Index	Gunning Fog Readability	Smog Readability Index	Flesch Reading Ease Index	LIX Readability Index	RIX Readability Index	Automated Readability Index (ARI)	Flesch-Kincaid Readability Index	Gunning Fog Readability	Smog Readability Index	Flesch Reading Ease Index	LIX Readability Index	RIX Readability Index
Post-ClassStartDate x Litigated	0.019	-0.030	-0.018	-0.018	0.052	-0.004	-0.017	0.091	-0.064	-0.052	-0.033	0.364	-0.011	-0.007
	(0.082)	(0.064)	(0.068)	(0.046)	(0.207)	(0.162)	(0.053)	(0.084)	(0.068)	(0.076)	(0.053)	(0.258)	(0.174)	(0.057)
Litigated	-0.039	0.018	-0.020	-0.015	-0.264	0.066	0.006	0.055	0.079	0.058	0.048	-0.370	0.177	0.055
	(0.080)	(0.062)	(0.067)	(0.045)	(0.205)	(0.158)	(0.052)	(0.083)	(0.066)	(0.074)	(0.052)	(0.249)	(0.171)	(0.056)
Post-ClassStartDate	0.067**	-0.053**	-0.008	-0.009	0.299***	-0.078	-0.027	0.020	-0.168***	-0.151***	-0.109***	0.794***	-0.246***	-0.082***
	(0.027)	(0.022)	(0.024)	(0.017)	(0.081)	(0.057)	(0.018)	(0.034)	(0.029)	(0.033)	(0.023)	(0.111)	(0.075)	(0.024)
Constant	20.826***	14.061***	18.665***	16.374***	33.187***	55.837***	7.330***	20.838***	12.746***	17.206***	15.413***	40.152***	53.851***	6.804***
	(0.496)	(0.471)	(0.482)	(0.329)	(2.384)	(1.455)	(0.395)	(0.577)	(0.495)	(0.542)	(0.380)	(2.088)	(1.397)	(0.411)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	2540	2540	2540	2540	2540	2540	2540	2598	2598	2598	2598	2598	2598	2598
Pseudo R-squared	0.060	0.127	0.097	0.105	0.264	0.139	0.103	0.046	0.212	0.154	0.165	0.300	0.138	0.126

Table 13.3: Sentiment Analysis (Count) – Pre- vs. Post-Class Start Date (10K and 10Q Filings)

The table shows the change in the textual disclosure (sentiment count) of a firm's filings pre- vs. post-class start date. Panel A and C have the results from analysis of firms' 10K filings for settled and dismissed cases respectively. Panel B and D have the results from analysis of firms' 10Q filings for settled and dismissed cases respectively. All models control for size (natural logarithm of market value), return on assets (ROA), loss indicator (negative net income), earnings growth, sales growth, market-to-book, big-8 auditor dummy, lag size, lag ROA, lag loss indicator, lag stock volatility, lag stock return, and lag institutional ownership. All models include year and industry (2-digit SIC code) fixed effects, with clustering at the firm level. Standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels. Variable definitions are explained in Appendix B.

Panel A: 10K Filings (Settled Cases)

Variables	(1) LN (Uncertainty Word Count)	(2) LN (Modal Weak Word Count)	(3) LN (Negative Word Count)	(4) LN (Positive Word Count)	(5) LN (Litigious Word Count)	(6) LN (Modal Strong Word Count)	(7) LN (Harvard Negative Word Count)
Post-ClassStartDate x Litigated	-0.124***	-0.102**	-0.186***	-0.073*	-0.093*	-0.020	-0.139***
	(0.035)	(0.044)	(0.045)	(0.039)	(0.049)	(0.044)	(0.035)
Litigated	0.055	0.095**	0.000	0.048	-0.020	0.081*	0.003
	(0.034)	(0.042)	(0.042)	(0.037)	(0.046)	(0.042)	(0.033)
Post-ClassStartDate	-0.199***	-0.221***	-0.214***	-0.127***	-0.109***	-0.134***	-0.157***
	(0.015)	(0.019)	(0.019)	(0.016)	(0.021)	(0.019)	(0.015)
Constant	3.051***	1.365***	2.293***	2.848***	3.475***	1.460***	4.456***
	(0.174)	(0.242)	(0.212)	(0.193)	(0.247)	(0.215)	(0.161)
Controls	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y
Observations	2581	2581	2581	2581	2581	2581	2581
Pseudo R-squared	0.530	0.470	0.479	0.396	0.378	0.376	0.497

Panel B: 10Q Filings (Settled Cases)

Variables	(1) LN (Uncertainty Word Count)	(2) LN (Modal Weak Word Count)	(3) LN (Negative Word Count)	(4) LN (Positive Word Count)	(5) LN (Litigious Word Count)	(6) LN (Modal Strong Word Count)	(7) LN (Harvard Negative Word Count)
Post-ClassStartDate x Litigated	-0.229***	-0.250***	-0.262***	-0.175***	-0.186***	-0.118**	-0.217***
	(0.042)	(0.055)	(0.051)	(0.039)	(0.054)	(0.050)	(0.036)
Litigated	0.064	0.026	0.024	0.051	0.014	0.108**	0.015
	(0.041)	(0.055)	(0.050)	(0.038)	(0.054)	(0.049)	(0.036)
Post-ClassStartDate	-0.331***	-0.327***	-0.334***	-0.233***	-0.235***	-0.257***	-0.264***
	(0.022)	(0.026)	(0.024)	(0.017)	(0.026)	(0.023)	(0.017)
Constant	3.309***	2.173***	3.388***	2.878***	2.508***	2.589***	4.275***
	(0.183)	(0.240)	(0.203)	(0.157)	(0.229)	(0.219)	(0.148)
Controls	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y
Observations	2707	2703	2707	2707	2707	2699	2707
Pseudo R-squared	0.563	0.433	0.527	0.529	0.401	0.425	0.585

Panel C: 10K Filings (Dismissed Cases)

Variables	(1) LN (Uncertainty Word Count)	(2) LN (Modal Weak Word Count)	(3) LN (Negative Word Count)	(4) LN (Positive Word Count)	(5) LN (Litigious Word Count)	(6) LN (Modal Strong Word Count)	(7) LN (Harvard Negative Word Count)
Post-ClassStartDate x Litigated	-0.108***	-0.114***	-0.165***	-0.078**	-0.138***	-0.050	-0.128***
	(0.032)	(0.039)	(0.040)	(0.034)	(0.046)	(0.039)	(0.032)
Litigated	0.067**	0.071*	0.024	0.042	-0.033	0.065*	0.023
	(0.031)	(0.037)	(0.039)	(0.033)	(0.045)	(0.038)	(0.032)
Post-ClassStartDate	-0.180***	-0.200***	-0.200***	-0.127***	-0.104***	-0.123***	-0.156***
	(0.014)	(0.016)	(0.018)	(0.014)	(0.016)	(0.016)	(0.013)
Constant	3.917***	3.036***	4.110***	4.052***	4.303***	3.103***	5.278***
	(0.315)	(0.372)	(0.356)	(0.292)	(0.341)	(0.321)	(0.278)
Controls	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y
Observations	2540	2540	2540	2540	2540	2540	2540
Pseudo R-squared	0.551	0.529	0.502	0.416	0.361	0.412	0.506

Panel D: 10Q Filings (Dismissed Cases)

Variables	(1) LN (Uncertainty Word Count)	(2) LN (Modal Weak Word Count)	(3) LN (Negative Word Count)	(4) LN (Positive Word Count)	(5) LN (Litigious Word Count)	(6) LN (Modal Strong Word Count)	(7) LN (Harvard Negative Word Count)
Post-ClassStartDate x Litigated	-0.180***	-0.170***	-0.296***	-0.170***	-0.267***	-0.120**	-0.209***
	(0.040)	(0.054)	(0.051)	(0.038)	(0.054)	(0.049)	(0.037)
Litigated	0.068*	0.081	-0.021	0.018	-0.038	0.079	0.006
	(0.039)	(0.053)	(0.049)	(0.037)	(0.053)	(0.049)	(0.036)
Post-ClassStartDate	-0.261***	-0.264***	-0.273***	-0.191***	-0.242***	-0.207***	-0.216***
	(0.019)	(0.023)	(0.023)	(0.016)	(0.021)	(0.022)	(0.015)
Constant	1.472***	0.288	1.607***	1.627***	1.631***	0.294	3.147***
	(0.270)	(0.405)	(0.287)	(0.230)	(0.255)	(0.344)	(0.175)
Controls	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y
Observations	2596	2596	2596	2596	2596	2586	2596
Pseudo R-squared	0.564	0.451	0.496	0.525	0.433	0.400	0.575

Table 13.4: Sentiment Analysis (%) – Pre- vs. Post-Class Start Date (10K and 10Q Filings)

The table shows the change in the textual disclosure (sentiment %) of a firm's filings pre- vs. post-class start date. Panel A and C have the results from analysis of firms' 10K filings for settled and dismissed cases respectively. Panel B and D have the results from analysis of firms' 10Q filings for settled and dismissed cases respectively. All models control for size (natural logarithm of market value), return on assets (ROA), loss indicator (negative net income), earnings growth, sales growth, market-to-book, big-8 auditor dummy, lag size, lag ROA, lag loss indicator, lag stock volatility, lag stock return, and lag institutional ownership. All models include year and industry (2-digit SIC code) fixed effects, with clustering at the firm level. Standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels. Variable definitions are explained in Appendix B.

Panel A: 10K Filings (Settled Cases)

Variables	(1) Uncertainty Word Proportion	(2) Modal Weak Word Proportion	(3) Negative Word Proportion	(4) Positive Word Proportion	(5) Litigious Word Proportion	(6) Modal Strong Word Proportion	(7) Harvard Negative Word Proportion
Post-ClassStartDate x Litigated	-0.059***	-0.016	-0.158***	-0.002	-0.022	0.019***	-0.240***
	(0.018)	(0.012)	(0.029)	(0.012)	(0.042)	(0.007)	(0.035)
Litigated	0.035*	0.026**	-0.051*	0.009	-0.061	0.015**	-0.096***
	(0.018)	(0.012)	(0.029)	(0.012)	(0.041)	(0.007)	(0.035)
Post-ClassStartDate	-0.095***	-0.042***	-0.121***	-0.002	0.024	0.001	-0.135***
	(0.007)	(0.004)	(0.013)	(0.005)	(0.017)	(0.003)	(0.017)
Constant	0.511***	-0.078	-0.149	0.480***	1.016***	0.067**	2.498***
	(0.081)	(0.057)	(0.133)	(0.055)	(0.215)	(0.033)	(0.172)
Controls	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y
Observations	2581	2581	2581	2581	2581	2581	2581
Pseudo R-squared	0.369	0.332	0.356	0.161	0.190	0.257	0.345

Panel B: 10Q Filings (Settled Cases)

Variables	(1) Uncertainty Word Proportion	(2) Modal Weak Word Proportion	(3) Negative Word Proportion	(4) Positive Word Proportion	(5) Litigious Word Proportion	(6) Modal Strong Word Proportion	(7) Harvard Negative Word Proportion
Post-ClassStartDate x Litigated	-0.076***	-0.021	-0.150***	-0.015	-0.050	0.016*	-0.267***
	(0.025)	(0.017)	(0.041)	(0.013)	(0.036)	(0.008)	(0.045)
Litigated	0.022	0.004	-0.054	-0.002	-0.057	0.016**	-0.142***
	(0.025)	(0.017)	(0.041)	(0.013)	(0.036)	(0.008)	(0.044)
Post-ClassStartDate	-0.106***	-0.031***	-0.121***	-0.001	-0.005	-0.002	-0.142***
	(0.011)	(0.006)	(0.017)	(0.006)	(0.016)	(0.004)	(0.019)
Constant	1.065***	0.356***	0.983***	0.668***	0.449***	0.638***	2.539***
	(0.086)	(0.053)	(0.132)	(0.054)	(0.148)	(0.029)	(0.184)
Controls	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y
Observations	2707	2707	2707	2707	2707	2707	2707
Pseudo R-squared	0.303	0.244	0.304	0.141	0.163	0.213	0.280

Panel C: 10K Filings (Dismissed Cases)

Variables	(1) Uncertainty Word Proportion	(2) Modal Weak Word Proportion	(3) Negative Word Proportion	(4) Positive Word Proportion	(5) Litigious Word Proportion	(6) Modal Strong Word Proportion	(7) Harvard Negative Word Proportion
Post-ClassStartDate x Litigated	-0.037**	-0.013	-0.116***	0.004	-0.055	0.010*	-0.187***
	(0.017)	(0.011)	(0.026)	(0.011)	(0.038)	(0.006)	(0.036)
Litigated	0.038**	0.020**	-0.015	0.008	-0.080**	0.008	-0.044
	(0.017)	(0.011)	(0.026)	(0.011)	(0.036)	(0.006)	(0.035)
Post-ClassStartDate	-0.082***	-0.042***	-0.111***	-0.008**	0.027**	-0.001	-0.153***
	(0.006)	(0.004)	(0.010)	(0.004)	(0.013)	(0.003)	(0.015)
Constant	0.576***	0.257***	0.710***	0.741***	1.044***	0.290***	2.401***
	(0.147)	(0.066)	(0.156)	(0.063)	(0.236)	(0.051)	(0.313)
Controls	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y
Observations	2540	2540	2540	2540	2540	2540	2540
Pseudo R-squared	0.382	0.372	0.396	0.199	0.198	0.285	0.346

Panel D: 10Q Filings (Dismissed Cases)

Variables	(1) Uncertainty Word Proportion	(2) Modal Weak Word Proportion	(3) Negative Word Proportion	(4) Positive Word Proportion	(5) Litigious Word Proportion	(6) Modal Strong Word Proportion	(7) Harvard Negative Word Proportion
Post-ClassStartDate x Litigated	-0.024	-0.002	-0.163***	-0.007	-0.083**	0.009	-0.192***
	(0.025)	(0.018)	(0.041)	(0.012)	(0.040)	(0.008)	(0.046)
Litigated	0.043*	0.021	-0.049	-0.007	-0.043	0.009	-0.070
	(0.024)	(0.018)	(0.041)	(0.012)	(0.039)	(0.008)	(0.045)
Post-ClassStartDate	-0.077***	-0.029***	-0.104***	-0.001	-0.047***	-0.000	-0.111***
	(0.011)	(0.007)	(0.016)	(0.005)	(0.015)	(0.003)	(0.018)
Constant	0.338***	-0.004	0.141	0.601***	0.469***	0.204***	2.413***
	(0.120)	(0.121)	(0.189)	(0.133)	(0.140)	(0.070)	(0.285)
Controls	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y
Observations	2598	2598	2598	2598	2598	2598	2598
Pseudo R-squared	0.284	0.247	0.257	0.165	0.138	0.243	0.251

Table 14.1: Disclosure Analysis – Pre- vs. Post-Litigation (10K and 10Q Filings)

The table shows the change in the textual disclosure (degree) of a firm's filings pre- vs. post-litigation. Panel A and C have the results from analysis of firms' 10K filings for settled and dismissed cases respectively. Panel B and D have the results from analysis of firms' 10Q filings for settled and dismissed cases respectively. All models control for size (natural logarithm of market value), return on assets (ROA), loss indicator (negative net income), earnings growth, sales growth, market-to-book, big-8 auditor dummy, lag size, lag ROA, lag loss indicator, lag stock volatility, lag stock return, and lag institutional ownership. All models include year and industry (2-digit SIC code) fixed effects, with clustering at the firm level. Standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels. Variable definitions are explained in Appendix B.

Panel A: 10K Filings (Settled Cases)

Variables	(1) LN (File Size)	(2) LN (Word Count)	(3) LN (Complex Word Count)	(4) LN (Sentence Count)	(5) LN (Average Words per Sentence)	(6) LN (Paragraph Count)	(7) LN (Average Words per Paragraph)
Post-CaseFilingDate x Litigated	-0.752***	-0.161***	-0.162***	-0.166***	0.005	-0.117***	-0.042**
	(0.042)	(0.030)	(0.029)	(0.027)	(0.006)	(0.029)	(0.017)
Litigated	0.123***	0.090***	0.088***	0.072***	0.018***	0.072**	0.018
	(0.039)	(0.030)	(0.029)	(0.027)	(0.006)	(0.031)	(0.020)
Post-CaseFilingDate	-0.815***	-0.236***	-0.238***	-0.231***	-0.005	-0.177***	-0.057***
	(0.024)	(0.017)	(0.017)	(0.016)	(0.004)	(0.016)	(0.007)
Constant	11.596***	9.237***	7.742***	6.112***	3.125***	5.149***	4.069***
	(0.176)	(0.221)	(0.205)	(0.176)	(0.058)	(0.204)	(0.078)
Controls	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y
Observations	2956	2956	2956	2956	2956	2956	2956
Pseudo R-squared	0.689	0.382	0.407	0.402	0.085	0.293	0.044

Panel B: 10Q Filings (Settled Cases)

Variables	(1) LN (File Size)	(2) LN (Word Count)	(3) LN (Complex Word Count)	(4) LN (Sentence Count)	(5) LN (Average Words per Sentence)	(6) LN (Paragraph Count)	(7) LN (Average Words per Paragraph)
Post-CaseFilingDate x Litigated	-0.892***	-0.202***	-0.198***	-0.204***	-0.001	-0.130***	-0.079***
	(0.043)	(0.029)	(0.030)	(0.027)	(0.006)	(0.030)	(0.024)
Litigated	0.173***	0.161***	0.172***	0.132***	0.027***	0.127***	0.036
	(0.039)	(0.028)	(0.028)	(0.025)	(0.005)	(0.029)	(0.024)
Post-CaseFilingDate	-1.020***	-0.381***	-0.391***	-0.368***	-0.016***	-0.300***	-0.086***
	(0.024)	(0.016)	(0.017)	(0.015)	(0.003)	(0.015)	(0.011)
Constant	9.856***	7.590***	6.147***	4.449***	3.161***	3.383***	4.142***
	(0.208)	(0.224)	(0.224)	(0.190)	(0.050)	(0.201)	(0.094)
Controls	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y
Observations	3380	3380	3380	3380	3380	3380	3380
Pseudo R-squared	0.726	0.513	0.521	0.539	0.101	0.415	0.059

Panel C: 10K Filings (Dismissed Cases)

Variables	(1) LN (File Size)	(2) LN (Word Count)	(3) LN (Complex Word Count)	(4) LN (Sentence Count)	(5) LN (Average Words per Sentence)	(6) LN (Paragraph Count)	(7) LN (Average Words per Paragraph)
Post-CaseFilingDate x Litigated	-0.888***	-0.124***	-0.120***	-0.119***	-0.005	-0.096***	-0.028
	(0.042)	(0.030)	(0.029)	(0.027)	(0.007)	(0.037)	(0.029)
Litigated	0.065*	0.115***	0.117***	0.107***	0.008	0.078**	0.037
	(0.038)	(0.027)	(0.026)	(0.025)	(0.006)	(0.031)	(0.031)
Post-CaseFilingDate	-0.928***	-0.236***	-0.234***	-0.223***	-0.013***	-0.208***	-0.028**
	(0.023)	(0.016)	(0.015)	(0.015)	(0.003)	(0.017)	(0.013)
Constant	12.508***	9.084***	7.643***	6.034***	3.050***	5.113***	3.951***
	(0.502)	(0.188)	(0.173)	(0.152)	(0.059)	(0.178)	(0.077)
Controls	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y
Observations	2602	2602	2602	2602	2602	2602	2602
Pseudo R-squared	0.722	0.443	0.465	0.461	0.104	0.260	0.012

Panel D: 10Q Filings (Dismissed Cases)

Variables	(1) LN (File Size)	(2) LN (Word Count)	(3) LN (Complex Word Count)	(4) LN (Sentence Count)	(5) LN (Average Words per Sentence)	(6) LN (Paragraph Count)	(7) LN (Average Words per Paragraph)
Post-CaseFilingDate x Litigated	-0.875***	-0.113***	-0.113***	-0.120***	0.005	-0.118***	0.005
	(0.044)	(0.030)	(0.031)	(0.028)	(0.006)	(0.031)	(0.020)
Litigated	0.255***	0.215***	0.219***	0.187***	0.027***	0.141***	0.074***
	(0.040)	(0.027)	(0.028)	(0.025)	(0.005)	(0.031)	(0.023)
Post-CaseFilingDate	-1.115***	-0.322***	-0.329***	-0.296***	-0.026***	-0.260***	-0.062***
	(0.025)	(0.017)	(0.017)	(0.015)	(0.003)	(0.015)	(0.008)
Constant	9.782***	7.044***	5.525***	4.041***	2.996***	3.103***	3.928***
	(0.269)	(0.225)	(0.230)	(0.202)	(0.035)	(0.200)	(0.103)
Controls	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y
Observations	2717	2717	2717	2717	2717	2717	2717
Pseudo R-squared	0.769	0.572	0.572	0.578	0.161	0.465	0.050

Table 14.2: Readability Analysis - Pre vs. Post-Litigation (10K and 10Q Filings)

The table shows the change in the textual disclosure (readability) of a firm's filings pre- vs. post-litigation. Panel A and C have the results from analysis of firms' 10K filings for settled and dismissed cases respectively. Panel B and D have the results from analysis of firms' 10Q filings for settled and dismissed cases respectively. All models control for size (natural logarithm of market value), return on assets (ROA), loss indicator (negative net income), earnings growth, sales growth, market-to-book, big-8 auditor dummy, lag size, lag ROA, lag loss indicator, lag stock volatility, lag stock return, and lag institutional ownership. All models include year and industry (2-digit SIC code) fixed effects, with clustering at the firm level. Standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels. Variable definitions are explained in Appendix B.

| | Panel A: 10K Filings (Settled Cases) | | | | | | | Panel B: 10Q Filings (Settled Cases) | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
Variables	Automated Readability Index (ARI)	Flesch-Kincaid Readability Index	Gunning Fog Readability	Smog Readability Index	Flesch Reading Ease Index	LIX Readability Index	RIX Readability Index	Automated Readability Index (ARI)	Flesch-Kincaid Readability Index	Gunning Fog Readability	Smog Readability Index	Flesch Reading Ease Index	LIX Readability Index	RIX Readability Index
Post-CaseFilingDate x Litigated	0.185**	-0.023	0.048	0.025	0.434**	0.121	0.037	0.152**	-0.058	0.026	0.019	0.340	0.253	0.054
	(0.080)	(0.061)	(0.066)	(0.046)	(0.208)	(0.153)	(0.052)	(0.077)	(0.061)	(0.068)	(0.047)	(0.235)	(0.161)	(0.052)
Litigated	0.107	0.153**	0.147**	0.110**	-0.339*	0.368**	0.140***	0.261***	0.330***	0.372***	0.265***	-1.198***	0.824***	0.270***
	(0.078)	(0.060)	(0.063)	(0.043)	(0.190)	(0.149)	(0.050)	(0.074)	(0.058)	(0.063)	(0.043)	(0.211)	(0.150)	(0.049)
Post-CaseFilingDate	0.133***	-0.135***	-0.040	-0.046*	0.886***	-0.219**	-0.072**	0.097**	-0.311***	-0.224***	-0.167***	1.565***	-0.259***	-0.109***
	(0.047)	(0.039)	(0.040)	(0.028)	(0.126)	(0.093)	(0.031)	(0.042)	(0.034)	(0.038)	(0.027)	(0.133)	(0.088)	(0.028)
Constant	20.936***	13.899***	18.207***	16.046***	36.629***	55.536***	7.416***	23.252***	14.134***	19.094***	16.687***	36.312***	57.652***	8.097***
	(0.729)	(0.594)	(0.585)	(0.367)	(1.496)	(1.363)	(0.463)	(0.600)	(0.496)	(0.501)	(0.335)	(1.662)	(1.262)	(0.383)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	2956	2956	2956	2956	2956	2956	2956	3380	3380	3380	3380	3380	3380	3380
Pseudo R-squared	0.093	0.142	0.133	0.140	0.288	0.150	0.121	0.074	0.203	0.155	0.170	0.319	0.145	0.129

| | Panel C: 10K Filings (Dismissed Cases) | | | | | | | Panel D: 10Q Filings (Dismissed Cases) | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
Variables	Automated Readability Index (ARI)	Flesch-Kincaid Readability Index	Gunning Fog Readability	Smog Readability Index	Flesch Reading Ease Index	LIX Readability Index	RIX Readability Index	Automated Readability Index (ARI)	Flesch-Kincaid Readability Index	Gunning Fog Readability	Smog Readability Index	Flesch Reading Ease Index	LIX Readability Index	RIX Readability Index
Post-CaseFilingDate x Litigated	0.129	-0.033	0.011	-0.006	0.122	0.086	0.001	0.180**	0.012	0.059	0.036	0.147	0.367**	0.101*
	(0.089)	(0.066)	(0.071)	(0.048)	(0.209)	(0.170)	(0.056)	(0.079)	(0.064)	(0.070)	(0.049)	(0.249)	(0.163)	(0.053)
Litigated	0.096	0.136**	0.096	0.067	-0.625***	0.365**	0.108**	0.233***	0.320***	0.301***	0.211***	-1.109***	0.732***	0.248***
	(0.081)	(0.059)	(0.063)	(0.043)	(0.186)	(0.151)	(0.050)	(0.074)	(0.057)	(0.062)	(0.043)	(0.220)	(0.147)	(0.048)
Post-CaseFilingDate	0.031	-0.181***	-0.090***	-0.074***	0.822***	-0.285***	-0.107***	-0.091**	-0.368***	-0.317***	-0.228***	1.492***	-0.504***	-0.192***
	(0.042)	(0.031)	(0.033)	(0.023)	(0.110)	(0.079)	(0.026)	(0.038)	(0.033)	(0.036)	(0.025)	(0.132)	(0.084)	(0.027)
Constant	20.367***	13.494***	17.807***	15.820***	35.453***	54.546***	6.942***	20.003***	12.063***	16.512***	14.942***	43.204***	51.772***	6.174***
	(0.579)	(0.495)	(0.448)	(0.299)	(1.211)	(1.119)	(0.403)	(0.410)	(0.385)	(0.412)	(0.276)	(1.538)	(0.906)	(0.277)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	2602	2602	2602	2602	2602	2602	2602	2717	2717	2717	2717	2717	2717	2717
Pseudo R-squared	0.093	0.178	0.148	0.155	0.307	0.167	0.144	0.068	0.258	0.195	0.204	0.315	0.176	0.170

Table 14.3: Sentiment (count) Analysis - Pre vs. Post-Litigation (10K and 10Q Filings)

The table shows the change in the textual disclosure (sentiment count) of a firm's filings pre- vs. post-litigation. Panel A and C have the results from analysis of firms' 10K filings for settled and dismissed cases respectively. Panel B and D have the results from analysis of firms' 10Q filings for settled and dismissed cases respectively. All models control for size (natural logarithm of market value), return on assets (ROA), loss indicator (negative net income), earnings growth, sales growth, market-to-book, big-8 auditor dummy, lag size, lag ROA, lag loss indicator, lag stock volatility, lag stock return, and lag institutional ownership. All models include year and industry (2-digit SIC code) fixed effects, with clustering at the firm level. Standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels. Variable definitions are explained in Appendix B.

	Panel A: 10K Filings (Settled Cases)							Panel B: 10Q Filings (Settled Cases)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variables	LN (Uncertainty Word Count)	LN (Modal Weak Word Count)	LN (Negative Word Count)	LN (Positive Word Count)	LN (Litigious Word Count)	LN (Modal Strong Word Count)	LN (Harvard Negative Word Count)	LN (Uncertainty Word Count)	LN (Modal Weak Word Count)	LN (Negative Word Count)	LN (Positive Word Count)	LN (Litigious Word Count)	LN (Modal Strong Word Count)	LN (Harvard Negative Word Count)
Post-CaseFilingDate x Litigated	-0.264***	-0.261***	-0.240***	-0.205***	-0.055	-0.137***	-0.256***	-0.256***	-0.219***	-0.124***	-0.249***	0.016	-0.084*	-0.224***
	(0.033)	(0.040)	(0.040)	(0.033)	(0.044)	(0.038)	(0.033)	(0.039)	(0.052)	(0.045)	(0.037)	(0.049)	(0.046)	(0.034)
Litigated	0.109***	0.151***	0.164***	0.059*	0.172***	0.119***	0.085**	0.175***	0.180***	0.325***	0.138***	0.402***	0.236***	0.179***
	(0.033)	(0.037)	(0.039)	(0.033)	(0.042)	(0.036)	(0.033)	(0.036)	(0.047)	(0.043)	(0.035)	(0.047)	(0.043)	(0.032)
Post-CaseFilingDate	-0.393***	-0.460***	-0.398***	-0.279***	-0.177***	-0.264***	-0.338***	-0.536***	-0.564***	-0.522***	-0.432***	-0.388***	-0.404***	-0.460***
	(0.020)	(0.024)	(0.024)	(0.020)	(0.026)	(0.022)	(0.020)	(0.022)	(0.029)	(0.025)	(0.020)	(0.027)	(0.026)	(0.018)
Constant	4.255***	2.986***	4.175***	4.348***	3.945***	3.512***	5.607***	2.978***	1.934***	2.634***	2.511***	2.401***	1.504***	3.812***
	(0.252)	(0.262)	(0.325)	(0.240)	(0.377)	(0.253)	(0.229)	(0.246)	(0.304)	(0.413)	(0.220)	(0.388)	(0.266)	(0.260)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	2956	2952	2956	2956	2956	2954	2956	3378	3376	3380	3380	3380	3374	3380
Pseudo R-squared	0.464	0.481	0.432	0.362	0.336	0.363	0.436	0.518	0.404	0.474	0.466	0.356	0.372	0.527

	Panel C: 10K Filings (Dismissed Cases)							Panel D: 10Q Filings (Dismissed Cases)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variables	LN (Uncertainty Word Count)	LN (Modal Weak Word Count)	LN (Negative Word Count)	LN (Positive Word Count)	LN (Litigious Word Count)	LN (Modal Strong Word Count)	LN (Harvard Negative Word Count)	LN (Uncertainty Word Count)	LN (Modal Weak Word Count)	LN (Negative Word Count)	LN (Positive Word Count)	LN (Litigious Word Count)	LN (Modal Strong Word Count)	LN (Harvard Negative Word Count)
Post-CaseFilingDate x Litigated	-0.214***	-0.241***	-0.175***	-0.148***	-0.044	-0.082**	-0.182***	-0.125***	-0.081	-0.038	-0.154***	0.052	-0.037	-0.125***
	(0.033)	(0.041)	(0.039)	(0.035)	(0.044)	(0.040)	(0.033)	(0.039)	(0.055)	(0.046)	(0.038)	(0.054)	(0.048)	(0.035)
Litigated	0.144***	0.173***	0.200***	0.121***	0.189***	0.176***	0.143***	0.255***	0.324***	0.368***	0.203***	0.424***	0.264***	0.250***
	(0.030)	(0.036)	(0.036)	(0.032)	(0.041)	(0.037)	(0.030)	(0.036)	(0.049)	(0.041)	(0.033)	(0.047)	(0.042)	(0.031)
Post-CaseFilingDate	-0.363***	-0.422***	-0.378***	-0.273***	-0.219***	-0.265***	-0.324***	-0.400***	-0.424***	-0.404***	-0.352***	-0.363***	-0.310***	-0.373***
	(0.019)	(0.021)	(0.021)	(0.018)	(0.021)	(0.022)	(0.019)	(0.023)	(0.031)	(0.027)	(0.020)	(0.027)	(0.027)	(0.019)
Constant	4.239***	3.027***	4.219***	3.933***	4.082***	3.255***	5.369***	1.895***	0.659	1.864***	1.989***	1.877***	1.031**	3.335***
	(0.199)	(0.237)	(0.313)	(0.214)	(0.368)	(0.278)	(0.216)	(0.322)	(0.412)	(0.340)	(0.256)	(0.380)	(0.426)	(0.229)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	2602	2602	2602	2602	2602	2602	2602	2717	2717	2717	2717	2717	2712	2717
Pseudo R-squared	0.532	0.498	0.480	0.422	0.391	0.356	0.499	0.536	0.428	0.504	0.520	0.444	0.406	0.569

Table 14.4: Sentiment (%) Analysis - Pre vs. Post-Litigation (10K and 10Q Filings)

The table shows the change in the textual disclosure (sentiment %) of a firm's filings pre- vs. post-litigation. Panel A and C have the results from analysis of firms' 10K filings for settled and dismissed cases respectively. Panel B and D have the results from analysis of firms' 10Q filings for settled and dismissed cases respectively. All models control for size (natural logarithm of market value), return on assets (ROA), loss indicator (negative net income), earnings growth, sales growth, market-to-book, big-8 auditor dummy, lag size, lag ROA, lag loss indicator, lag stock volatility, lag stock return, and lag institutional ownership. All models include year and industry (2-digit SIC code) fixed effects, with clustering at the firm level. Standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels. Variable definitions are explained in Appendix B.

| | Panel A: 10K Filings (Settled Cases) | | | | | | | Panel B: 10Q Filings (Settled Cases) | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
Variables	Uncertainty Word Proportion	Modal Weak Word Proportion	Negative Word Proportion	Positive Word Proportion	Litigious Word Proportion	Modal Strong Word Proportion	Harvard Negative Word Proportion	Uncertainty Word Proportion	Modal Weak Word Proportion	Negative Word Proportion	Positive Word Proportion	Litigious Word Proportion	Modal Strong Word Proportion	Harvard Negative Word Proportion
Post-CaseFilingDate x Litigated	-0.117***	-0.047***	-0.099***	-0.024**	0.154***	0.014**	-0.318***	-0.043*	0.009	0.138***	-0.030**	0.200***	0.037***	-0.071
	(0.019)	(0.013)	(0.028)	(0.011)	(0.033)	(0.006)	(0.036)	(0.025)	(0.019)	(0.045)	(0.012)	(0.036)	(0.007)	(0.044)
Litigated	0.020	0.023*	0.113***	-0.033***	0.078**	0.007	-0.031	0.008	0.001	0.236***	-0.030***	0.212***	0.014**	0.030
	(0.018)	(0.012)	(0.026)	(0.011)	(0.032)	(0.006)	(0.033)	(0.023)	(0.016)	(0.041)	(0.011)	(0.034)	(0.007)	(0.041)
Post-CaseFilingDate	-0.177***	-0.098***	-0.209***	-0.015**	0.120***	0.005	-0.316***	-0.171***	-0.085***	-0.189***	-0.016**	0.018	0.005	-0.237***
	(0.010)	(0.006)	(0.015)	(0.007)	(0.021)	(0.004)	(0.019)	(0.014)	(0.009)	(0.022)	(0.007)	(0.019)	(0.004)	(0.024)
Constant	0.697***	0.136***	0.483***	0.879***	0.626**	0.390***	2.875***	1.025***	0.339***	0.710***	0.739***	0.703***	0.278***	2.562***
	(0.095)	(0.046)	(0.143)	(0.079)	(0.311)	(0.040)	(0.160)	(0.117)	(0.065)	(0.200)	(0.072)	(0.228)	(0.033)	(0.279)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	2956	2956	2956	2956	2956	2956	2956	3380	3380	3380	3380	3380	3380	3380
Pseudo R-squared	0.333	0.360	0.351	0.172	0.226	0.246	0.308	0.277	0.278	0.248	0.177	0.128	0.236	0.209

| | Panel C: 10K Filings (Dismissed Cases) | | | | | | | Panel D: 10Q Filings (Dismissed Cases) | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
Variables	Uncertainty Word Proportion	Modal Weak Word Proportion	Negative Word Proportion	Positive Word Proportion	Litigious Word Proportion	Modal Strong Word Proportion	Harvard Negative Word Proportion	Uncertainty Word Proportion	Modal Weak Word Proportion	Negative Word Proportion	Positive Word Proportion	Litigious Word Proportion	Modal Strong Word Proportion	Harvard Negative Word Proportion
Post-CaseFilingDate x Litigated	-0.102***	-0.057***	-0.058**	-0.010	0.123***	0.018***	-0.172***	0.002	0.015	0.156***	-0.020	0.162***	0.030***	-0.022
	(0.018)	(0.013)	(0.027)	(0.012)	(0.036)	(0.006)	(0.035)	(0.027)	(0.019)	(0.042)	(0.013)	(0.040)	(0.008)	(0.043)
Litigated	0.032*	0.025**	0.119***	-0.004	0.068**	0.016***	0.099***	0.039	0.042**	0.257***	-0.014	0.181***	0.014**	0.126***
	(0.016)	(0.011)	(0.027)	(0.011)	(0.032)	(0.006)	(0.032)	(0.025)	(0.018)	(0.039)	(0.011)	(0.035)	(0.007)	(0.041)
Post-CaseFilingDate	-0.144***	-0.090***	-0.180***	-0.015**	0.069***	-0.001	-0.271***	-0.072***	-0.038***	-0.091***	-0.006	-0.007	0.014***	-0.147***
	(0.009)	(0.006)	(0.014)	(0.006)	(0.018)	(0.004)	(0.019)	(0.015)	(0.011)	(0.022)	(0.007)	(0.018)	(0.004)	(0.024)
Constant	0.902***	0.267***	0.804***	0.689***	0.744**	0.363***	2.821***	0.580***	0.079	0.357	0.814***	0.752***	0.324***	2.712***
	(0.089)	(0.061)	(0.204)	(0.078)	(0.341)	(0.045)	(0.281)	(0.140)	(0.075)	(0.309)	(0.080)	(0.266)	(0.047)	(0.344)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	2602	2602	2602	2602	2602	2602	2602	2717	2717	2717	2717	2717	2717	2717
Pseudo R-squared	0.337	0.375	0.340	0.222	0.252	0.225	0.330	0.181	0.199	0.243	0.185	0.176	0.229	0.233

Table 15: Principal Component Analysis – Controlling for Largest Stock Price Drop

The table shows the association between the principal components of degree, readability, and sentiments of a firm's textual disclosure and the probability of the firm being litigated. Panel A has the results from analysis of firms' 10K filings, while Panel B has the results from analysis of firms' 10Q filings. All models control for size (natural logarithm of market value), return on assets (ROA), loss indicator (negative net income), earnings growth, sales growth, market-to-book, big-8 auditor dummy, lag size, lag ROA, lag loss indicator, lag stock volatility, lag stock return, lag institutional ownership, and the largest daily stock price drop during the alleged damage period. All models include year and industry (2-digit SIC code) fixed effects, with clustering at the firm level. Standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels. Variable definitions are explained in Appendix B.

	Panel A: 10K Filings (Logit Model)				Panel B: 10Q Filings (Logit Model)			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Variables	D_litigation	D_litigation	D_litigation	D_litigation	D_litigation	D_litigation	D_litigation	D_litigation
Disclosure (PCA)	0.123***				0.267***			
	(0.028)				(0.028)			
Readability (PCA)		0.044***				0.100***		
		(0.017)				(0.016)		
Uncertainty % (PCA)			0.183***				0.199***	
			(0.034)				(0.027)	
Uncertainty Count (PCA)				0.136***				0.168***
				(0.035)				(0.029)
Size (ln_MV)	-1.017***	-1.011***	-1.010***	-1.012***	-0.879***	-0.870***	-0.865***	-0.872***
	(0.067)	(0.067)	(0.067)	(0.066)	(0.059)	(0.058)	(0.059)	(0.058)
ROA	0.024	0.023	0.018	0.022	0.029	0.027	0.025	0.025
	(0.030)	(0.030)	(0.031)	(0.030)	(0.038)	(0.034)	(0.036)	(0.035)
Earnings Growth	-0.014	-0.012	-0.023	-0.016	0.009	0.027	0.007	0.008
	(0.064)	(0.065)	(0.065)	(0.065)	(0.065)	(0.079)	(0.031)	(0.050)
Sales Growth	0.036	0.035	0.035	0.036	0.016	0.013	0.018	0.016
	(0.031)	(0.032)	(0.030)	(0.031)	(0.014)	(0.012)	(0.015)	(0.015)
Loss Indicator	-0.509***	-0.494***	-0.528***	-0.511***	-0.608***	-0.505***	-0.550***	-0.556***
	(0.098)	(0.098)	(0.099)	(0.098)	(0.098)	(0.096)	(0.097)	(0.097)
Lag Size (ln_MV)	1.189***	1.210***	1.212***	1.187***	1.000***	1.064***	1.056***	1.038***
	(0.072)	(0.072)	(0.073)	(0.072)	(0.063)	(0.063)	(0.063)	(0.063)
Lag_Market-to-Book	0.000	-0.000	0.001	0.000	-0.001	-0.001	-0.001	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)
Lag_Volatility	0.307*	0.331*	0.255	0.281	0.422***	0.514***	0.414***	0.444***
	(0.182)	(0.188)	(0.181)	(0.184)	(0.135)	(0.136)	(0.136)	(0.136)
Lag_Return on Assets (ROA)	-0.020	-0.027	-0.026	-0.023	-0.004	-0.012	-0.016	-0.010
	(0.038)	(0.053)	(0.044)	(0.043)	(0.031)	(0.030)	(0.031)	(0.030)
Lag_Loss Indicator	-0.035	0.016	-0.035	-0.036	-0.037	0.022	-0.014	-0.007
	(0.101)	(0.100)	(0.100)	(0.102)	(0.097)	(0.097)	(0.097)	(0.097)
Stock Return	0.006	0.001	-0.006	0.003	-0.030	-0.045	-0.036	-0.037
	(0.038)	(0.037)	(0.037)	(0.037)	(0.042)	(0.043)	(0.041)	(0.041)
Big-8 Auditor	0.040	0.040	0.023	0.034	-0.075	0.000	-0.027	-0.030
	(0.122)	(0.123)	(0.123)	(0.122)	(0.119)	(0.117)	(0.118)	(0.119)
Institutional Ownership	-0.058	-0.089	-0.122	-0.062	0.154	0.095	0.076	0.127
	(0.117)	(0.116)	(0.118)	(0.117)	(0.127)	(0.125)	(0.123)	(0.124)
Largest Daily Stock Price Drop	-5.355***	-5.467***	-5.278***	-5.321***	-4.189***	-4.416***	-4.339***	-4.287***
	(0.601)	(0.608)	(0.603)	(0.601)	(0.537)	(0.536)	(0.531)	(0.531)
Constant	-2.253**	-2.540**	-2.099*	-2.172**	-1.113**	-2.396***	-2.160***	-1.994***
	(1.092)	(1.098)	(1.105)	(1.106)	(0.540)	(0.511)	(0.513)	(0.520)
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y	Y
Observations	3624	3624	3624	3624	3987	3987	3987	3987
Pseudo R-squared	0.166	0.163	0.168	0.166	0.173	0.161	0.165	0.163

Table 16: Principal Component Analysis – Controlling for Type of Litigations

The table shows the association between the principal components of degree, readability, and sentiments of a firm's textual disclosure and the probability of the firm being litigated. Panel A has the results from analysis of firms' 10K filings, while Panel B has the results from analysis of firms' 10Q filings. All models control for size (natural logarithm of market value), return on assets (ROA), loss indicator (negative net income), earnings growth, sales growth, market-to-book, big-8 auditor dummy, lag size, lag ROA, lag loss indicator, lag stock volatility, lag stock return, lag institutional ownership, and dummy indicating whether the litigation is either merger-related or accounting related. All models include year and industry (2-digit SIC code) fixed effects, with clustering at firm level. Standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels. Variable definitions are explained in Appendix B.

Variables	Panel A: 10K Filings (Logit Model)				Panel B: 10Q Filings (Logit Model)			
	(1) D_litigation	(2) D_litigation	(3) D_litigation	(4) D_litigation	(1) D_litigation	(2) D_litigation	(3) D_litigation	(4) D_litigation
Disclosure (PCA)	0.137***				0.285***			
	(0.028)				(0.028)			
Readability (PCA)		0.052***				0.108***		
		(0.017)				(0.016)		
Uncertainty % (PCA)			0.207***				0.215***	
			(0.034)				(0.027)	
Uncertainty Count (PCA)				0.157***				0.186***
				(0.036)				(0.029)
Size (ln_MV)	-0.982***	-0.975***	-0.975***	-0.978***	-0.878***	-0.868***	-0.860***	-0.869***
	(0.065)	(0.065)	(0.066)	(0.065)	(0.058)	(0.057)	(0.057)	(0.057)
ROA	0.018	0.016	0.013	0.016	0.026	0.021	0.021	0.021
	(0.030)	(0.030)	(0.029)	(0.029)	(0.035)	(0.034)	(0.034)	(0.034)
Earnings Growth	-0.009	-0.008	-0.018	-0.012	0.015	0.033	0.012	0.013
	(0.057)	(0.058)	(0.058)	(0.057)	(0.073)	(0.072)	(0.049)	(0.060)
Sales Growth	0.029	0.027	0.028	0.029	0.013	0.010	0.015	0.013
	(0.024)	(0.025)	(0.023)	(0.024)	(0.012)	(0.010)	(0.013)	(0.012)
Loss Indicator	-0.482***	-0.466***	-0.504***	-0.487***	-0.591***	-0.477***	-0.527***	-0.538***
	(0.098)	(0.097)	(0.099)	(0.098)	(0.098)	(0.095)	(0.096)	(0.096)
Lag Size (ln_MV)	1.109***	1.131***	1.135***	1.107***	0.954***	1.021***	1.010***	0.991***
	(0.070)	(0.070)	(0.071)	(0.070)	(0.062)	(0.061)	(0.062)	(0.062)
Lag_Market-to-Book	-0.000	-0.000	0.000	-0.000	-0.001	-0.001	-0.001	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)
Lag_Volatility	0.988***	1.042***	0.917***	0.954***	0.915***	1.050***	0.941***	0.954***
	(0.194)	(0.197)	(0.192)	(0.195)	(0.135)	(0.137)	(0.136)	(0.136)
Lag_Return on Assets (ROA)	-0.006	-0.015	-0.014	-0.010	0.008	0.002	-0.004	0.003
	(0.042)	(0.066)	(0.048)	(0.049)	(0.029)	(0.029)	(0.029)	(0.028)
Lag_Loss Indicator	0.027	0.080	0.024	0.023	0.006	0.071	0.033	0.037
	(0.101)	(0.100)	(0.100)	(0.101)	(0.097)	(0.096)	(0.096)	(0.096)
Stock Return	-0.085**	-0.093**	-0.096**	-0.087**	-0.106**	-0.129***	-0.117***	-0.117**
	(0.039)	(0.040)	(0.039)	(0.039)	(0.046)	(0.048)	(0.044)	(0.046)
Big-8 Auditor	0.094	0.095	0.072	0.085	-0.034	0.049	0.020	0.009
	(0.121)	(0.122)	(0.123)	(0.121)	(0.117)	(0.115)	(0.116)	(0.117)
Institutional Ownership	-0.046	-0.081	-0.118	-0.051	0.172	0.110	0.091	0.147
	(0.116)	(0.115)	(0.118)	(0.116)	(0.114)	(0.113)	(0.112)	(0.113)
Merger Related	0.079	0.128	0.131	0.098	0.084	0.088	0.150	0.121
	(0.191)	(0.190)	(0.189)	(0.189)	(0.191)	(0.187)	(0.187)	(0.188)
Disclosure Related	-0.002	-0.007	-0.002	0.005	0.013	0.010	0.047	0.030
	(0.082)	(0.082)	(0.083)	(0.082)	(0.077)	(0.076)	(0.076)	(0.076)
Constant	-1.664	-1.957*	-1.489	-1.558	-0.451	-1.799***	-1.571***	-1.359***
	(1.080)	(1.086)	(1.094)	(1.097)	(0.535)	(0.508)	(0.510)	(0.516)
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y	Y
Observations	3,624	3,624	3,624	3,624	3,987	3,987	3,987	3,987
Pseudo R-squared	0.145	0.141	0.148	0.145	0.158	0.144	0.149	0.147

Table 17: Readability Analysis - Pre vs. Post-Plain English Initiative (10K and 10Q Filings)

The table compares the relationship between textual disclosure (readability) and the probability of a firm being litigated pre- vs. post-Plain English Initiative passed in 1998. Panel A has the results from analysis of firms' 10K filings. Panel B has the results from analysis of firms' 10Q filings. All models control for size (natural logarithm of market value), return on assets (ROA), loss indicator (negative net income), earnings growth, sales growth, market-to-book, big-8 auditor dummy, lag size, lag ROA, lag loss indicator, lag stock volatility, lag stock return, and lag institutional ownership. All models include year and industry (2-digit SIC code) fixed effects, with clustering at the firm level. Standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels. Variable definitions are explained in Appendix B.

Variables	Panel A: 10K Filings (Litigation Logit Model)							Panel B: 10Q Filings (Litigation Logit Model)						
	(1) D_Litigation	(2) D_Litigation	(3) D_Litigation	(4) D_Litigation	(5) D_Litigation	(6) D_Litigation	(7) D_Litigation	(1) D_Litigation	(2) D_Litigation	(3) D_Litigation	(4) D_Litigation	(5) D_Litigation	(6) D_Litigation	(7) D_Litigation
Automated Readability Index (ARI)	0.128*							0.178***						
	(0.071)							(0.065)						
Automated Readability Index (ARI) x Post-PEI	-0.096							-0.108						
	(0.077)							(0.070)						
Flesch-Kincaid Readability Index		0.183**							0.407***					
		(0.088)							(0.075)					
Flesch-Kincaid Readability Index x Post-PEI		-0.060							-0.185**					
		(0.096)							(0.083)					
Gunning Fog Readability			0.173**							0.363***				
			(0.084)							(0.066)				
Gunning Fog Readability x Post-PEI			-0.091							-0.189**				
			(0.091)							(0.074)				
Smog Readability Index				0.282**							0.536***			
				(0.123)							(0.092)			
Smog Readability Index x Post-PEI				-0.157							-0.276***			
				(0.133)							(0.104)			
Flesch Reading Ease Index					-0.073***							-0.113***		
					(0.025)							(0.018)		
Flesch Reading Ease Index x Post-PEI					0.030							0.052**		
					(0.027)							(0.020)		
LIX Readability Index						0.113***							0.136***	
						(0.037)							(0.029)	
LIX Readability Index x Post-PEI						-0.073*							-0.068**	
						(0.039)							(0.031)	
RIX Readability Index							0.313***							0.454***
							(0.112)							(0.090)
RIX Readability Index x Post-PEI							-0.190							-0.241**
							(0.121)							(0.099)
Post-PEI	2.852	1.575	2.445	3.352	-0.173	5.036**	2.298**	2.350	2.665**	3.570**	4.580**	-1.748**	3.985**	2.002**
	(1.760)	(1.517)	(1.831)	(2.318)	(0.787)	(2.382)	(1.084)	(1.617)	(1.316)	(1.487)	(1.811)	(0.708)	(1.937)	(0.944)
Constant	-5.123***	-4.949***	-5.616***	-7.024***	0.030	-8.900***	-4.829***	-6.036***	-7.714***	-8.673***	-10.680***	2.032**	-9.800***	-5.592***
	(1.850)	(1.659)	(1.900)	(2.295)	(1.279)	(2.356)	(1.361)	(1.516)	(1.161)	(1.286)	(1.555)	(0.808)	(1.711)	(0.858)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	3732	3732	3732	3732	3732	3732	3732	4095	4095	4095	4095	4095	4095	4095
Pseudo R-squared	0.140	0.142	0.141	0.141	0.143	0.143	0.142	0.138	0.147	0.145	0.146	0.148	0.143	0.144

Table 18: Disclosure Analysis - Pre vs. Post-Sarbanes-Oxley Act (10K and 10Q Filings)

The table compares the relationship between textual disclosure (degree) and the probability of a firm being litigated pre- vs. post-Sarbanes-Oxley Act passed in 2002. Panel A has the results from analysis of firms' 10K filings. Panel B has the results from analysis of firms' 10Q filings. All models control for size (natural logarithm of market value), return on assets (ROA), loss indicator (negative net income), earnings growth, sales growth, market-to-book, big-8 auditor dummy, lag size, lag ROA, lag loss indicator, lag stock volatility, lag stock return, and lag institutional ownership. All models include year and industry (2-digit SIC code) fixed effects, with clustering at the firm level. Standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels. Variable definitions are explained in Appendix B.

	Panel A: 10K Filings (Litigation Logit Model)							Panel B: 10Q Filings (Litigation Logit Model)						
Variables	(1) D_Litigation	(2) D_Litigation	(3) D_Litigation	(4) D_Litigation	(5) D_Litigation	(6) D_Litigation	(7) D_Litigation	(1) D_Litigation	(2) D_Litigation	(3) D_Litigation	(4) D_Litigation	(5) D_Litigation	(6) D_Litigation	(7) D_Litigation
LN (File Size)	0.164*							0.384***						
	(0.092)							(0.076)						
LN (File Size) x Post-SOX	-0.014							-0.120						
	(0.109)							(0.093)						
LN (Word Count)		0.414***							0.920***					
		(0.109)							(0.111)					
LN (Word Count) x Post-SOX		0.105							-0.243*					
		(0.162)							(0.134)					
LN (Complex Word Count)			0.478***							0.916***				
			(0.113)							(0.106)				
LN (Complex Word Count) x Post-SOX			0.048							-0.239*				
			(0.165)							(0.129)				
LN (Sentence Count)				0.485***							1.019***			
				(0.121)							(0.121)			
LN (Sentence Count) x Post-SOX				0.052							-0.346**			
				(0.178)							(0.143)			
LN (Average Words per Sentence)					0.493							2.127***		
					(0.459)							(0.476)		
LN (Average Words per Sentence) x Post-SOX					1.246*							0.265		
					(0.713)							(0.724)		
LN (Paragraph Count)						0.328***							0.768***	
						(0.114)							(0.137)	
LN (Paragraph Count) x Post-SOX						-0.029							-0.337**	
						(0.171)							(0.162)	
LN (Average Words per Paragraph)							0.019							0.287
							(0.119)							(0.196)
LN (Average Words per Paragraph) x Post-SOX							0.220							0.055
							(0.225)							(0.250)
Post-SOX	-0.325	-1.603	-0.938	-0.867	-4.510*	-0.365	-1.456	1.306	2.152*	1.782*	2.014**	-1.078	1.519*	-0.407
	(1.533)	(1.719)	(1.526)	(1.327)	(2.308)	(1.099)	(1.035)	(1.226)	(1.291)	(1.071)	(0.932)	(2.355)	(0.881)	(1.120)
Constant	-4.388***	-6.155***	-6.082***	-5.334***	-3.783***	-4.066***	-2.407**	-5.988***	-8.960***	-7.550***	-6.565***	-8.697***	-4.858***	-3.229***
	(1.490)	(1.456)	(1.369)	(1.272)	(1.787)	(1.194)	(1.114)	(0.948)	(0.999)	(0.833)	(0.749)	(1.562)	(0.729)	(0.902)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	3732	3732	3732	3732	3732	3732	3732	4095	4095	4095	4095	4095	4095	4095
Pseudo R-squared	0.140	0.145	0.146	0.145	0.141	0.142	0.139	0.143	0.157	0.157	0.155	0.142	0.147	0.138

Appendix A: Industry Distribution of Litigations

The table below reports the industry distribution of 2,335 litigations from 1996-2014 in my sample. Industry code is reported as 2-digit SIC codes and their corresponding names. The list is organized in descending order in terms of the number of litigations filed during the sample period. For example, business services is listed at the top with 465 litigations from 1996-2014.

Industry Code	Industry Name	No. of Litigations	Industry Code	Industry Name	No. of Litigations
73	Business Services	465	72	Personal Services	10
28	Chemicals and Allied Products	225	15	Construction - General Contractors & Operative Builders	9
36	Electronic & Other Electrical Equipment & Components	209	57	Home Furniture, Furnishings and Equipment Stores	9
35	Industrial and Commercial Machinery and Computer Equipment	144	78	Motion Pictures	9
38	Measuring, Photographic, Medical, & Optical Goods, & Clocks	127	12	Coal Mining	8
48	Communications	100	26	Paper and Allied Products	8
60	Depository Institutions	81	27	Printing, Publishing and Allied Industries	8
63	Insurance Carriers	77	42	Motor Freight Transportation	8
80	Health Services	66	32	Stone, Clay, Glass, and Concrete Products	7
62	Security & Commodity Brokers, Dealers, Exchanges & Services	65	65	Real Estate	7
49	Electric, Gas and Sanitary Services	58	31	Leather and Leather Products	6
59	Miscellaneous Retail	50	44	Water Transportation	6
61	Nondepository Credit Institutions	50	55	Automotive Dealers and Gasoline Service Stations	6
67	Holding and Other Investment Offices	45	10	Metal Mining	5
13	Oil and Gas Extraction	41	16	Heamy Construction, Except Building Construction, Contractor	5
37	Transportation Equipment	33	24	Lumber and Wood Products, Except Furniture	5
87	Engineering, Accounting, Research, and Management Services	33	47	Transportation Services	5
20	Food and Kindred Products	31	75	Automotive Repair, Services and Parking	5
99	Services, Not Elsewhere Classified	29	17	Construction - Special Trade Contractors	4
51	Wholesale Trade - Nondurable Goods	26	22	Textile Mill Products	4
82	Educational Services	26	45	Transportation by Air	4
50	Wholesale Trade - Durable Goods	25	70	Hotels, Rooming Houses, Camps, and Other Lodging Places	4
56	Apparel and Accessory Stores	23	25	Furniture and Fixtures	3
58	Eating and Drinking Places	22	29	Petroleum Refining and Related Industries	3
23	Apparel, Finished Products from Fabrics & Similar Materials	21	83	Social Services	3
39	Miscellaneous Manufacturing Industries	17	07	Agricultural Services	2
53	General Merchandise Stores	14	14	Mining and Quarrying of Nonmetallic Minerals, Except Fuels	2
33	Primary Metal Industries	13	21	Tobacco Products	2
34	Fabricated Metal Products	13	52	Building Materials, Hardware, Garden Supplies & Mobile Homes	2
79	Amusement and Recreation Services	13	01	Agricultural Production - Crops	1
64	Insurance Agents, Brokers and Service	11	40	Railroad Transportation	1
30	Rubber and Miscellaneous Plastic Products	10	41	Local & Suburban Transit & Interurban Highway Transportation	1
54	Food Stores	10	**Total**		**2,335**

Appendix B: Variable Definitions

TEXTUAL DISCLOSURE VARIABLES

Name	Definition	Source
File Size	Size of the filing (bytes)	SEC Filings
Word Count	Number of words in the filing	
Complex Word Count	Number of words containing three or more syllables in the filing	SEC Filings
Sentence Count*	Number of sentences in the filing	SEC Filings
Avg. Words Per Sentence	Average number of words per sentence in the filing	SEC Filings
Paragraph Count**	Number of paragraphs in the filing	SEC Filings
Avg. Words Per Paragraph	Average number of words per paragraph in the filing	SEC Filings

* Minimum number of words needed to be considered a sentence is five. I follow the methodology provided by Gillick (2009)[21] to identify sentence boundaries.

**Minimum number of words needed to be considered a paragraph is ten.

---

[21] Gillick, D., 2009, May. Sentence boundary detection and the problem with the US. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume* (pp. 241-244). Association for Computational Linguistics.

Variable Definitions (Contd.)

TEXTUAL READABILITY VARIABLES

Name	Definition	Source
Flesch Reading Ease Index[22]	Originally developed by Rudolph Flesch in 1948, the Flesch Reading Ease Index has been computed using the formula: ***206.835 – 1.015 (the number of words divided by the number of sentences) – 84.6 (the number of syllables divided by the number of words***) The Flesch Reading scores vary from 0 and 100. The higher the score, the easier the text is to read. For instance, while scores between 90 and 100 are considered comprehensible by an average 5th grader, scores between 0 and 30 are considered understandable by an average college graduate.	SEC Filings
Flesch-Kincaid Readability Index[23]	Flesch-Kincaid Readability Index modifies the original Flesch Reading Ease Index and has been computed using the following formula: ***0.39 (the number of words divided by the number of sentences) + 11.8(the number of syllables divided by the number of words) – 15.59*** The higher the Flesch-Kincaid score, the more difficult the text is to read. For example, a score of 12 is interpreted as a text that a 12th grader would be able to understand.	SEC Filings

---

[22] Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, *32*(3), 221.

[23] Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel* (No. RBR-8-75). Naval Technical Training Command Millington TN Research Branch.

TEXTUAL READABILITY VARIABLES

Name	Definition	Source
RIX Readability Index[24]	RIX Readability Index, which is another widely used readability measure has been computed using the formula: ***Number of words of length 7 characters or more divided by the number of sentences.*** The higher the RIX Readability score, the more difficult the text is to read.	SEC Filings
Gunning Fog Readability Index[25]	Gunning Fog Readability Index was developed by Robert Gunning in 1952 and uses the following formula: ***0.4 (the number of words divided by the number of sentences) + 100 (the number of complex words divided by the number of words)*** The higher the Gunning Fog score, the more difficult the text is to read.	SEC Filings
Automated Readability Index[26]	Automated Readability Index computes the grade-level readability and has been calculated using the formula: ***4.71 (the number of characters divided by the number of words) + 0.5 (the number of words divided by the number of sentences) – 21.43*** The higher the Automated Readability score, the more difficult the text is to read.	SEC Filings

---

[24] Anderson, J. (1983). LIX and RIX: Variations on a little-known readability index. *Journal of Reading*, *26*(6), 490-496.

[25] Gunning, R. (1952). The Technique of Clear Writing. *McGraw-Hill. pp. 36–37.*

[26] Senter, R. J., & Smith, E. A. (1967). Automated readability index. *Cincinnati University, OH.*

TEXTUAL READABILITY VARIABLES

Name	Definition	Source
Smog Readability Index[27]	Smog Readability Index was created by G. Harry McLaughlin in 1969 and uses the following formula: *1.043 x Sqrt (number of complex words x 30/number of sentences) + 3.1291* The higher the Smog Readability score, the more difficult is the text to comprehend.	SEC Filings
Lasbarhets Readability Index[28]	Also known as LIX Readability Index and has been widely used to estimate readability of western European languages including English. Lasbarhets Readability Index has been calculated using the formula: *(the number of words divided by the number of sentences) + (the number of words over 6 letters multiplied by 100 and then divided by the number of words)* The higher the Lasbarhets Readability score, the more difficult the text is to read.	SEC Filings

---

[27] Mc Laughlin, G. H. (1969). SMOG grading-a new readability formula. *Journal of Reading*, *12*(8), 639-646.
[28] Björnsson, C.H. (1968) Lasbarhet. *Stockholm, Sweden: Bokförlaget Liber.*

Variable Definitions (Contd.)

TEXTUAL SENTIMENT VARIABLES

Name	Definition	Source
Harvard Negative Word Count[29]	Number of Harvard General Inquirer Negative words in the filing	SEC Filings
Harvard Negative Word Proportion	Proportion of Harvard General Inquirer Negative words in the filing	SEC Filings
L-M Negative Word Count[30]	Number of L-M Negative words in the filing	SEC Filings
L-M Negative Word Proportion	Proportion of L-M Negative words in the filing	SEC Filings
L-M Positive Word Count	Number of L-M Positive words in the filing	SEC Filings
L-M Positive Word Proportion	Proportion of L-M Positive words in the filing	SEC Filings
L-M Financial Weak Modal Word Count	Number of L-M Financial Weak Modal words in the filing	SEC Filings
L-M Financial Weak Modal Word Proportion	Proportion of L-M Weak Modal words in the filing	SEC Filings
L-M Financial Strong Modal Word Count	Number of L-M Financial Strong Modal words in the filing	SEC Filings
L-M Financial Strong Modal Word Proportion	Proportion of L-M Financial Strong Modal words in the filing	SEC Filings
L-M Litigious Word Count	Number of L-M Litigious words in the filing	SEC Filings
L-M Litigious Word Proportion	Proportion of L-M Litigious words in the filing	SEC Filings
L-M Uncertainty Word Count	Number of L-M Uncertainty words in the filing	SEC Filings
L-M Uncertainty Word Proportion	Proportion of L-M Uncertainty words in the filing	SEC Filings

[29] I use Harvard's General Inquirer word list to compute Harvard Negative Word Count and Harvard Negative Word Proportion. This word list has been widely used in the extant accounting and finance literature and can be found at http://www.wjh.harvard.edu/~inquirer/homecat.htm

[30] Loughran and McDonald (L-M) textual sentiments word lists is freely available at Bill McDonald's website at http://www3.nd.edu/~mcdonald/. The details of these lists can be found in Loughran and McDonald (2011) and Bodnaruk, Loughran and McDonald (2015).

Variable Definitions (Contd.)

LITIGATION VARIABLES

Name	Definition	Source
Securities Class Action	An indicator variable for the incidence of securities class action litigation	SCAC[31]
Case Filing Date	Date the class action was filed in Federal court	SCAC
Case End Date	Date the class action ended	SCAC
Case Status	Settled, Dismissed or Ongoing	SCAC
Class Start Date	Beginning of the class period	SCAC
Class End Date	End of the class period	SCAC
Settlement Amount	Actual amount in dollars paid by the defendant firm	Case Summaries, 10Ks, 10Qs or 8Ks

---

[31] Securities Class Action Clearinghouse (SCAC), http://securities.stanford.edu/

Variable Definitions (Contd.)

CONTROL VARIABLES

Name	Definition	Source
Firm Size	Natural logarithm of equity market value	Compustat
Market-to-Book Ratio	(Total Debt + Market Value of Equity) / (Total Debt + Book Value of Equity)	Compustat
Analyst Following	Natural logarithm of 1 plus the number of analyst following the firm	IBES
Negative Earnings Surprise	Dummy variable equals to one if SUE Score is negative, and zero otherwise[32]	IBES
Volatility	Standard deviation of daily stock returns, measured over a 365-day period	CRSP
Return on Assets	Net income scaled by total assets	Compustat
Earnings Growth	Change in net income relative to the previous year, scaled by total assets	Compustat
Sales Growth	Percentage growth in sales relative to the previous year	Compustat
Loss Indicator	Dummy variable equals to one if net income for the year is negative, and zero otherwise	Compustat
Auditor Quality	Dummy variable equals to one if the auditor codes are between 1 and 8, and zero otherwise	Compustat
Institutional Ownership	Total institutional ownership as a percentage of shares outstanding	Institutional (13f) Holdings
Stock Return	Natural log of annualized stock return adjusted by inflation	CRSP

Note: Regressions also include lag values (by one year) of size, ROA, Loss Indicator, Stock Volatility, Stock Return, and Institutional Ownership.

---

[32] SUE (Standardized Unanticipated Earnings) Score = (Actual EPS – Surprise Mean) / Standard Deviation. Actual EPS is the actual reported earnings. Surprise Mean is the arithmetic average of all estimates on earnings in IBES for a given period when a company announces its earnings. Standard deviation measures the dispersion of those estimates for a given period.

Appendix C.1 Disclosure and Litigation (IBES Controls) – 10K and 10Q Filings

The table shows the association between disclosure by a firm and the probability of the firm being litigated. Panel A has the results from analysis of firms' 10K filings, while Panel B has the results from analysis of firms' 10Q filings. All models control for size (natural logarithm of market value), return on assets (ROA), loss indicator (negative net income), earnings growth, sales growth, market-to-book, big-8 auditor dummy, number of analyst following, negative earnings surprise, lag size, lag ROA, lag loss indicator, lag stock volatility, lag stock return, and lag institutional ownership. All models include year and industry (2-digit SIC code) fixed effects, with clustering at the firm level. Standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels. Variable definitions are explained in Appendix B.

Variables	Panel A: 10K Filings (Litigation Logit Model)							Panel B: 10Q Filings (Litigation Logit Model)						
	(1) D_Litigation	(2) D_Litigation	(3) D_Litigation	(4) D_Litigation	(5) D_Litigation	(6) D_Litigation	(7) D_Litigation	(1) D_Litigation	(2) D_Litigation	(3) D_Litigation	(4) D_Litigation	(5) D_Litigation	(6) D_Litigation	(7) D_Litigation
LN (File Size)	0.105 (0.079)							0.255*** (0.070)						
LN (Word Count)		0.479*** (0.121)							0.694*** (0.109)					
LN (Complex Word Count)			0.499*** (0.127)							0.691*** (0.106)				
LN (Sentence Count)				0.532*** (0.137)							0.735*** (0.116)			
LN (Average Words per Sentence)					0.938** (0.468)							1.903*** (0.524)		
LN (Paragraph Count)						0.417*** (0.128)							0.455*** (0.129)	
LN (Average Words per Paragraph)							0.009 (0.103)							0.306** (0.149)
Constant	-2.024 (1.532)	-5.115*** (1.634)	-4.597*** (1.556)	-3.982*** (1.468)	-3.528* (1.852)	-2.807** (1.371)	-0.717 (1.289)	-1.664 (1.124)	-4.188*** (1.147)	-3.120*** (1.026)	-2.148** (0.956)	-5.194*** (1.868)	-0.587 (0.946)	-0.428 (1.038)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	2153	2153	2153	2153	2153	2153	2153	2224	2224	2224	2224	2224	2224	2224
Pseudo R-squared	0.132	0.139	0.139	0.139	0.133	0.137	0.131	0.123	0.134	0.134	0.133	0.123	0.126	0.121

Appendix C.2 Readability and Litigation (IBES Controls) – 10K and 10Q Filings

The table shows the association between readability of a firm's filings and the probability of the firm being litigated. Panel A has the results from analysis of firms' 10K filings, while Panel B has the results from analysis of firms' 10Q filings. All models control for size (natural logarithm of market value), return on assets (ROA), loss indicator (negative net income), earnings growth, sales growth, market-to-book, big-8 auditor dummy, number of analyst following, negative earnings surprise, lag size, lag ROA, lag loss indicator, lag stock volatility, lag stock return, and lag institutional ownership. All models include year and industry (2-digit SIC code) fixed effects, with clustering at the firm level. Standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels. Variable definitions are explained in Appendix B.

Variables	Panel A: 10K Filings (Litigation Logit Model)							Panel B: 10Q Filings (Litigation Logit Model)						
	(1) D_Litigation	(2) D_Litigation	(3) D_Litigation	(4) D_Litigation	(5) D_Litigation	(6) D_Litigation	(7) D_Litigation	(1) D_Litigation	(2) D_Litigation	(3) D_Litigation	(4) D_Litigation	(5) D_Litigation	(6) D_Litigation	(7) D_Litigation
Automated Readability Index (ARI)	0.049 (0.034)							0.069* (0.036)						
Flesch-Kincaid Readability Index		0.101** (0.048)							0.209*** (0.051)					
Gunning Fog Readability			0.073 (0.045)							0.177*** (0.045)				
Smog Readability Index				0.100 (0.066)							0.266*** (0.065)			
Flesch Reading Ease Index					-0.021 (0.016)							-0.055*** (0.013)		
LIX Readability Index						0.041** (0.019)							0.067*** (0.019)	
RIX Readability Index							0.127** (0.056)							0.212*** (0.058)
Constant	-1.682 (1.386)	-2.073 (1.372)	-1.997 (1.450)	-2.282 (1.595)	0.020 (1.332)	-2.982* (1.610)	-1.611 (1.275)	-0.705 (1.167)	-2.112* (1.109)	-2.423** (1.191)	-3.471*** (1.347)	2.844*** (0.980)	-2.967** (1.355)	-0.828 (0.960)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	2153	2153	2153	2153	2153	2153	2153	2224	2224	2224	2224	2224	2224	2224
Pseudo R-squared	0.132	0.133	0.132	0.132	0.132	0.133	0.133	0.120	0.125	0.124	0.124	0.124	0.123	0.123

Appendix C.3 Sentiment (Count) and Litigation (IBES Controls) – 10K and 10Q Filings

The table shows the association between sentiment (count) of a firm's filings and the probability of the firm being litigated. Panel A has the results from analysis of firms' 10K filings, while Panel B has the results from analysis of firms' 10Q filings. All models control for size (natural logarithm of market value), return on assets (ROA), loss indicator (negative net income), earnings growth, sales growth, market-to-book, big-8 auditor dummy, number of analyst following, negative earnings surprise, lag size, lag ROA, lag loss indicator, lag stock volatility, lag stock return, and lag institutional ownership. All models include year and industry (2-digit SIC code) fixed effects, with clustering at the firm level. Standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels. Variable definitions are explained in Appendix B.

Variables	Panel A: 10K Filings (Litigation Logit Model)							Panel B: 10Q Filings (Litigation Logit Model)						
	(1) D_Litigation	(2) D_Litigation	(3) D_Litigation	(4) D_Litigation	(5) D_Litigation	(6) D_Litigation	(7) D_Litigation	(1) D_Litigation	(2) D_Litigation	(3) D_Litigation	(4) D_Litigation	(5) D_Litigation	(6) D_Litigation	(7) D_Litigation
LN (Uncertainty Word Count)	0.521*** (0.122)							0.539*** (0.081)						
LN (Modal Weak Word Count)		0.488*** (0.089)							0.312*** (0.058)					
LN (Negative Word Count)			0.448*** (0.094)							0.569*** (0.070)				
LN (Positive Word Count)				0.378*** (0.106)							0.429*** (0.084)			
LN (Litigious Word Count)					0.378*** (0.078)							0.515*** (0.066)		
LN (Modal Strong Word Count)						0.447*** (0.085)							0.466*** (0.070)	
LN (Harvard Negative Word Count)							0.444*** (0.116)							0.644*** (0.094)
Constant	-2.980** (1.297)	-2.431** (1.238)	-2.673** (1.269)	-2.322* (1.294)	-2.197* (1.215)	-2.310* (1.248)	-3.176** (1.370)	-0.179 (0.849)	0.609 (0.866)	-0.368 (0.854)	0.136 (0.846)	-0.252 (0.849)	0.699 (0.886)	-1.490 (0.909)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	2153	2151	2153	2153	2153	2152	2153	2223	2222	2224	2224	2224	2222	2224
Pseudo R-squared	0.141	0.143	0.142	0.137	0.141	0.141	0.139	0.134	0.130	0.143	0.128	0.142	0.135	0.136

Appendix C.4 Sentiment (%) and Litigation (IBES Controls) – 10K and 10Q Filings

The table shows the association between sentiment (%) of a firm's filings and the probability of the firm being litigated. Panel A has the results from analysis of firms' 10K filings, while Panel B has the results from analysis of firms' 10Q filings. All models control for size (natural logarithm of market value), return on assets (ROA), loss indicator (negative net income), earnings growth, sales growth, market-to-book, big-8 auditor dummy, number of analyst following, negative earnings surprise, lag size, lag ROA, lag loss indicator, lag stock volatility, lag stock return, and lag institutional ownership. All models include year and industry (2-digit SIC code) fixed effects, with clustering at the firm level. Standard errors are reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels. Variable definitions are explained in Appendix B.

Variables	Panel A: 10K Filings (Litigation Logit Model)							Panel B: 10Q Filings (Litigation Logit Model)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	D_Litigation	D_Litigation	D_Litigation	D_Litigation	D_Litigation	D_Litigation	D_Litigation	D_Litigation	D_Litigation	D_Litigation	D_Litigation	D_Litigation	D_Litigation	D_Litigation
Uncertainty Word Proportion	34.305*							24.075**						
	(19.620)							(12.214)						
Modal Weak Word Proportion		65.293**							26.097*					
		(30.703)							(15.477)					
Negative Word Proportion			41.965***							44.393***				
			(11.651)							(7.234)				
Positive Word Proportion				-14.756							-30.676			
				(25.803)							(23.626)			
Litigious Word Proportion					27.488***							48.494***		
					(9.616)							(8.824)		
Modal Strong Word Proportion						146.120***							158.614***	
						(52.572)							(45.457)	
Harvard Negative Word Proportion							8.119							18.183***
							(8.596)							(6.895)
Constant	-1.012	-0.980	-1.040	-0.554	-0.785	-1.265	-0.927	0.696	0.774	0.612	0.951	0.418	0.523	0.291
	(1.225)	(1.225)	(1.216)	(1.228)	(1.190)	(1.240)	(1.251)	(0.849)	(0.863)	(0.848)	(0.869)	(0.857)	(0.874)	(0.889)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	2153	2153	2153	2153	2153	2153	2153	2224	2224	2224	2224	2224	2224	2224
Pseudo R-squared	0.133	0.133	0.136	0.132	0.134	0.134	0.132	0.120	0.120	0.132	0.119	0.129	0.123	0.121

Appendix D Correlations between Textual Variables

The table shows the correlations between the textual variables for the sample. ***, **, and * indicate significance at 1%, 5%, and 10% levels. Variable definitions are explained in Appendix B.

Disclosure	File Size	Word Count	Complex word count	Sentence count	Paragraph Count	Avg. No. of words per sentence	Avg. No. of words per paragraph
File Size	1.0000						
Word Count	0.4213***	1.0000					
Complex word count	0.4317***	0.9947***	1.0000				
Sentence count	0.451***	0.972***	0.9788***	1.0000			
Paragraph Count	0.4565***	0.9385***	0.9349***	0.9302***	1.0000		
Avg. No. of words per sentence	0.1392***	0.4400***	0.4291***	0.3313***	0.3155***	1.0000	
Avg. No. of words per paragraph	-0.0134	0.0471***	0.0458***	0.0408***	-0.0563***	0.1190***	1.0000

Readability	Flesch Reading Ease Index	Flesch-Kincaid Readability Index	RIX Readability Index	Gunning Fog Readability Index	Automated Readability Index (ARI)	Smog Readability Index	LIX Readability Index
Flesch Reading Ease Index	1.0000						
Flesch-Kincaid Readability Index	-0.8982***	1.0000***					
RIX Readability Index	-0.808***	0.9467***	1.0000				
Gunning Fog Readability Index	-0.8652***	0.9647***	0.9415***	1.0000			
Automated Readability Index (ARI)	-0.6238***	0.8205***	0.8882***	0.8242***	1.0000		
Smog Readability Index	-0.8763***	0.9592***	0.9357***	0.9966***	0.8055***	1.0000	
LIX Readability Index	-0.8349***	0.9215***	0.9873***	0.9248***	0.8658***	0.9202***	1.0000

Continued

Sentiment (count)	LM Negative Word (Count)	LM Positive Word (Count)	LM Modal Weak Word (Count)	LM Litigious Word (Count)	LM Uncertainty Word (Count)	LM Modal Strong Word (Count)	Harvard Negative Word (Count)
LM Negative Word (Count)	1.0000						
LM Positive Word (Count)	0.8716***	1.0000					
LM Modal Weak Word (Count)	0.7983***	0.7992***	1.0000				
LM Litigious Word (Count)	0.7806***	0.6934***	0.5435***	1.0000			
LM Uncertainty Word (Count)	0.8986***	0.8871***	0.8965***	0.6032***	1.0000		
LM Modal Strong Word (Count)	0.8145***	0.8366***	0.8042***	0.6603***	0.8311***	1.0000	
Harvard Negative Word (Count)	0.9564***	0.9063***	0.7638***	0.7589***	0.9204***	0.8123***	1.0000

Sentiment (%)	LM Negative Word (%)	LM Positive Word (%)	LM Modal Weak Word (%)	LM Litigious Word (%)	LM Uncertainty Word (%)	LM Modal Strong Word (%)	Harvard Negative Word (%)
LM Negative Word (%)	1.0000						
LM Positive Word (%)	0.1952***	1.0000					
LM Modal Weak Word (%)	0.5161***	0.4431***	1.0000				
LM Litigious Word (%)	0.5601***	-0.0813***	0.0398***	1.0000			
LM Uncertainty Word (%)	0.4648***	0.3385***	0.832***	-0.0152	1.0000		
LM Modal Strong Word (%)	0.2847***	0.3341***	0.4878***	0.0303**	0.4181***	1.0000	
Harvard Negative Word (%)	0.7347***	0.1617***	0.3848***	0.3057***	0.4028***	0.1351***	1.0000

Appendix E Disclosure and Litigation – 10K (With Coefficients on Controls Reported)

Variables	Panel A: 10K Filings (Litigation Logit Model)						
	(1) D_Litigation	(2) D_Litigation	(3) D_Litigation	(4) D_Litigation	(5) D_Litigation	(6) D_Litigation	(7) D_Litigation
LN (File Size)	0.152***						
	(0.058)						
LN (Word Count)		0.463***					
		(0.090)					
LN (Complex Word Count)			0.501***				
			(0.094)				
LN (Sentence Count)				0.510***			
				(0.101)			
LN (Average Words per Sentence)					0.995***		
					(0.357)		
LN (Paragraph Count)						0.311***	
						(0.095)	
LN (Average Words per Paragraph)							0.125
							(0.101)
LN (Market Value)	-0.973***	-0.975***	-0.976***	-0.976***	-0.967***	-0.969***	-0.968***
	(0.065)	(0.065)	(0.065)	(0.065)	(0.065)	(0.065)	(0.066)
ROA	0.002	0.004	0.004	0.005	0.002	0.004	0.002
	(0.032)	(0.031)	(0.031)	(0.030)	(0.033)	(0.032)	(0.033)
Loss Indicator	-0.475***	-0.495***	-0.499***	-0.493***	-0.474***	-0.473***	-0.468***
	(0.096)	(0.096)	(0.096)	(0.096)	(0.096)	(0.095)	(0.096)
Earnings Growth	-0.021	-0.027	-0.026	-0.026	-0.026	-0.027	-0.024
	(0.054)	(0.051)	(0.051)	(0.051)	(0.055)	(0.052)	(0.055)
SalesGrowth	0.010	0.011	0.011	0.012	0.009	0.010	0.009
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Market-to-Book	0.001	0.001	0.001	0.001	0.000	0.001	0.000
	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)
Big-8 Auditor	0.074	0.068	0.060	0.073	0.078	0.073	0.092
	(0.121)	(0.120)	(0.120)	(0.120)	(0.120)	(0.121)	(0.121)
Lag LN (Market Value)	1.120***	1.106***	1.103***	1.105***	1.131***	1.115***	1.135***
	(0.071)	(0.070)	(0.070)	(0.070)	(0.070)	(0.070)	(0.070)
Lag ROA	-0.008	0.000	0.002	0.000	-0.013	-0.004	-0.017
	(0.079)	(0.051)	(0.047)	(0.050)	(0.095)	(0.064)	(0.103)
Lag Loss Indicator	0.099	0.051	0.044	0.049	0.106	0.074	0.115
	(0.098)	(0.099)	(0.099)	(0.099)	(0.099)	(0.099)	(0.099)
Lag Stock Volatility	1.022***	0.983***	0.982***	0.976***	1.036***	1.000***	1.031***
	(0.193)	(0.189)	(0.189)	(0.190)	(0.192)	(0.191)	(0.193)
Lag Stock Return	-0.093**	-0.087**	-0.086**	-0.087**	-0.095**	-0.092**	-0.097**
	(0.040)	(0.039)	(0.039)	(0.039)	(0.040)	(0.040)	(0.041)
Lag Institutional Ownership	-0.065	-0.035	-0.034	-0.036	-0.065	-0.053	-0.072
	(0.114)	(0.114)	(0.114)	(0.115)	(0.114)	(0.115)	(0.114)
Constant	-4.241***	-6.656***	-6.283***	-5.508***	-5.425***	-3.960***	-2.834***
	(1.187)	(1.307)	(1.245)	(1.177)	(1.511)	(1.122)	(1.085)
Year Fixed Effect	Y	Y	Y	Y	Y	Y	Y
Industry Fixed Effect	Y	Y	Y	Y	Y	Y	Y
Clusters at firm level	Y	Y	Y	Y	Y	Y	Y
Observations	3732	3732	3732	3732	3732	3732	3732
Pseudo R-squared	0.140	0.145	0.145	0.144	0.140	0.142	0.139