

Are Lengthy and Boilerplate Risk Factor Disclosures Inadequate? An Examination of Judicial and Regulatory Assessments of Risk Factor Language

August 2018

ABSTRACT

Prior research finds that lengthy and boilerplate risk factor disclosures are associated with negative capital market consequences. Yet regulators and users of financial statements continue to criticize corporate risk factor disclosures as excessively long and boilerplate. We investigate two potential sources of firms' incentives to issue lengthy, boilerplate risk factor disclosures by examining how judicial and regulatory assessments of firms' risk factor disclosures correlate with measures of disclosure length and disclosure boilerplate. Our results suggest that lengthy and more boilerplate risk factor disclosures are less likely to be considered inadequate under judicial and regulatory review. Specifically, risk factor disclosures that are lengthier and less specific are less likely to be flagged as inadequate for safe harbor purposes under the Private Securities Litigation Reform Act. In addition, more standardized risk factor disclosures are less likely to be targeted by an SEC comment letter during the SEC's filing review process.

JEL classification: D8; G38; M4

Keywords: risk disclosure; boilerplate; litigation risk; securities lawsuits; SEC comment letters; disclosure regulation

1. Introduction

The SEC mandates that firms disclose the most significant factors that make their stock speculative or risky in Item 1A of their periodic SEC filings (SEC 2005). These risk factor disclosures also play an important role under current securities laws by providing safe harbor protection for firms' forward-looking disclosures from shareholder lawsuits. Regulatory and judicial guidance explicitly warns that firms should avoid lengthy and boilerplate risk disclosure in favor of disclosure that is concise and specific to the firm (SEC 2005, SEC 2016, Cong. Rec. 28 Nov. 1995 13703). Recent research suggests that firms that deviate from this guidance by providing lengthy or boilerplate disclosure experience negative capital market consequences, such as higher cost of capital, greater stock price volatility, weaker market responses, and declines in analysts' ability to assess fundamental risk (e.g., Kravet and Muslu 2013; Campbell et al. 2014; Hope, Hu, and Lu 2016). However, practitioners and regulators lament that firms' risk factor disclosures continue to be generic and excessively long (e.g., Johnson 2010; IIRC 2016; SEC 2016; Berkman 2018). The fact that managers continue to produce lengthy and boilerplate disclosures that may expose them to negative capital market consequences suggests firms derive some offsetting benefit from this disclosure practice.

Risk factor disclosures are subject to regulatory assessment under the SEC's filing review process and judicial assessment when federal judges consider whether a sued firm's forward-looking statements should be granted safe harbor protections.¹ Firms drafting Item 1A risk factor disclosures have incentives to minimize their total costs of disclosure, which include not only capital market consequences, but also costs arising from regulatory and judicial assessments of

¹ Specifically, the Private Securities Litigation Reform Act (PSLRA) of 1995 provides a statutory safe harbor for firms' forward-looking statements accompanied by cautionary language describing the risks that could cause actual results to vary from projections.

disclosure inadequacy. Thus, even if more concise and firm-specific risk factors provide greater clarity to investors, firms may believe lengthy and boilerplate risk factor disclosures help reduce their total costs of disclosure by decreasing the likelihood of unfavorable regulatory and judicial assessments. In this study, we propose and test the notion that lengthier and more boilerplate risk factor disclosures are less likely to be flagged as inadequate under judicial and regulatory review.

Our investigation of how risk factor disclosure boilerplate and length correlate with the likelihood those disclosures will be assessed as inadequate under regulatory or judicial review builds on measures of disclosure length and boilerplate introduced by prior literature. Our primary measure of risk factor disclosure length is based on a word count of the firm's Item 1A risk factor disclosures, consistent with Campbell et al. (2014). We use two alternative approaches developed by prior research to measure risk factor disclosure boilerplate. We argue that each measure actually captures a distinct attribute of disclosure that may correlate differently with judicial and regulatory assessments of disclosure inadequacy. The first is based on the Stanford Named Entity Recognition (NER) tool, which measures the extent to which language is inherently specific versus non-specific by counting the number of specific entities named in the disclosure (Hope et al. 2016). The second is based on the extent to which language is standardized, or used so pervasively among industry peers that it is unlikely to be informative (Lang and Stice-Lawrence 2015). Although prior literature has considered both measures to be proxies for disclosure "boilerplate," we focus on the two attributes separately (hereafter *non-specific* disclosure and *standardized* disclosure) to avoid the conflation of these two distinct disclosure practices.

Official regulatory and judicial guidance explicitly indicates risk factor disclosures should be concise and tailored to the specific risks faced by the firm, and that "boilerplate" risk

disclosures are not sufficient to receive safe harbor protection (SEC 2005; SEC 2016; Cong. Rec. 28 Nov. 1995 13703).² If adjudicators adhere to this formal guidance in assessing disclosure adequacy, then lengthier, less specific, and more standardized risk disclosure language should be more likely to be flagged as inadequate under regulatory and judicial review.

In spite of this formal guidance, there are a number of reasons that lengthy, non-specific, and standardized risk factor disclosures may actually be associated with more favorable judicial and regulatory assessments. First, although the PSLRA requires risk factor disclosures to convey “substantive” firm-specific information about risks, the PSLRA does not outline how “substantive” risk disclosures are identified. Published judicial opinions suggest that disclosure length is sometimes relied on as a heuristic for determining substantiveness.³ Second, lengthy and non-specific risk factor language may provide greater legal protection if projected results fail to materialize, as risk factor disclosures that are more specific and concise inherently cover a narrower subset of adverse potential outcomes.

Third, standardized disclosure may be less likely to be flagged as inadequate because regulators and judges use peer firm disclosures to create their own benchmark of what constitutes “adequate” or “normal” disclosure. The SEC’s filing review process is conducted by staff with specialized industry expertise who review filings for a multitude of peers within the same industry.⁴ Therefore, risk disclosures that are more similar to those of industry peers may be less likely to attract regulatory scrutiny. In addition, the legal doctrine of *stare decisis* (reliance on

² The Congressional Record is the official daily record of the proceedings and debates occurring in the United States Congress (See <https://www.congress.gov/congressional-record/>). While the congressional record does not carry the weight of law, the record provides insight into the purpose for the enactment of laws and courts frequently rely upon the congressional record to interpret the law.

³ For instance, in dismissing a suit against Motorola’s allegedly false and misleading forward-looking statements, one judge reasoned that “the risk factors found in Motorola’s SEC 10-K filings are extensive and specific, extending beyond eight pages.”

⁴ See Division of Corporation Finance: Filing Review Process (SEC), available at <https://www.sec.gov/divisions/corpfin/cffilingreview.htm> (last updated June 27, 2018).

precedent) can lead judges to consider how closely a given set of risk factor disclosures for a sued firm resemble those already assessed as adequate in similar lawsuits.

Finally, a negative link between standardized language and the likelihood risk disclosures are flagged as inadequate may arise if firms adopt standardized peer firm disclosure language that has successfully passed judicial or regulatory review to increase the likelihood their own disclosures will be considered acceptable (Kahan and Klausner 1997). McMullin (2016) suggests that the copying of peer firm disclosure is rampant and that the process of disclosure standardization across industry peers is driven by firms sharing a common audit firm.⁵

We construct two separate samples to test how risk factor disclosure length, non-specificity, and standardization correlate with the likelihood risk disclosures are assessed as inadequate under judicial or regulatory review. The first consists of judicial decisions ruling on firms' motion to dismiss securities lawsuits that allege false or misleading forward-looking statements. We focus on judges' explicit evaluation of firms' risk factor language in these judicial decisions because these evaluations provide the most direct indication of how risk factor attributes relate to safe harbor protections. We read each published judicial decision to determine the judge's assessment of the defendant's risk factor language as adequate or not for purposes of the safe harbor. To construct our second sample, we use Audit Analytics' SEC Comment Letter database to identify firms that received an SEC comment letter indicating a deficiency in the firm's risk factor disclosures. To better understand the judicial assessment and regulatory assessments of risk factor disclosures and motivate our hypotheses, we also identify the reasons

⁵ Large auditing firms provide disclosure consulting services which include template disclosures based on peer firm disclosures for their clients. We spoke with a director of audit and assurance at one of the big four accounting firms who stated his firm tags and databases disclosures from Item 1A risk factors as well as other sections of the 10-K to use as a reference for aiding client firms to prepare their own disclosures. He indicates his firm also tags and databases the comments in SEC comment letters, which allows the firm to model the likelihood their client will receive an SEC comment letter based on the client's disclosures.

cited for why the risk factor disclosure was or was not adequate. We find significant variation in the rationale judges and the SEC offer to support their assessments of risk factor disclosure adequacy, suggesting this determination is far from straightforward.

Our multivariate regression analyses suggest that the probability a federal judge rules that a firm's risk factor disclosures are inadequate for safe harbor protection is decreasing in disclosure length, specificity, and standardization. Including both disclosure specificity and standardization in the model reveals that only length and specificity remain statistically significant. Together, our results suggest that, conditional on a firm being sued, risk factor disclosures are less likely to be assessed as inadequate under judicial review if they are longer and less specific. These results are consistent with judges interpreting less specific language as covering a broader set of adverse potential outcomes and using length as a heuristic for adequate disclosure. Results are robust to controlling for determinants of firm risk, industry affiliation, fiscal year, the U.S. circuit in which the firm is headquartered, and the topics discussed in the firm's risk factor disclosures.

Turning to the SEC comment letter sample, we find a strong negative association between our measure of standardized disclosure language and the probability of receiving an SEC comment letter related to risk factor disclosures. The finding that standardized risk disclosures are negatively associated with the likelihood of being flagged as inadequate by the SEC is consistent with the current structure of the SEC, which uses industry specialization of regional offices, leading SEC employees to develop a benchmark of adequate disclosure based on industry peers. Overall, our findings suggest that one reason firms may be willing to bear the negative capital market consequences of lengthy, non-specific, and standardized risk factor disclosures is that these attributes are associated with a lower likelihood of judicial and

regulatory assessments of disclosure inadequacy. We run various robustness tests that support these main findings. Our main results are robust to alternative approaches to measuring disclosure length and standardization and to controlling for the specific topics disclosed in the risk factor section.

Our paper contributes directly to the stream of research examining the quality of firms' risk factor disclosures. Whereas recent research focuses on whether risk factor disclosures contain *any* information content, our study is among the first to propose and test an explanation for why lengthy, non-specific, and standardized risk factor disclosure continues to persist despite the potential for adverse capital market consequences. Our results suggest that such disclosure characteristics may benefit the firm by reducing expected legal and regulatory costs arising from judicial and regulatory enforcement.

These findings also have direct implications for regulators and policy makers seeking to understand the forces that shape current risk factor disclosure practices. The evidence we find suggests that one reason firms continue to provide lengthy, non-specific, and standardized risk factor language is that regulatory and legal outcomes generally do not reward disclosures that are concise, specific, and unique-to-the-firm. Our results also suggest that judges and the SEC may assess the adequacy of risk factor disclosures in a manner that is inconsistent with official guidance.

2. Institutional Background and Hypothesis Development

2.1 CAPITAL MARKET CONSEQUENCES OF RISK FACTOR DISCLOSURE LENGTH AND BOILERPLATE

Several recent studies have examined capital market outcomes associated with firms' risk factor disclosures. Kravet and Muslu (2013) find that increases in risk disclosure length is associated with subsequent increases in stock return volatility and trading volume and

conclude that longer risk disclosures increase investors' perception of firm risk. Campbell et al. (2014) find that increases in risk factor disclosures increase firms' cost of capital, and suggest that this relation may provide strong incentives for managers to resist disclosure. Hope et al. (2016) find that more specific disclosures are associated with stronger market responses and with greater analyst ability to assess fundamental risk, and concurrent research finds that risk factor disclosures can predict firm-specific adverse events (Gaulin 2017). Overall, these studies suggest that risk factor disclosures that are less specific (i.e., more boilerplate) and lengthier can subject firms to negative capital market consequences.

Despite the benefits of concise and specific risk disclosure documented by prior research, critics allege that risk factor disclosures continue to be excessively boilerplate and lengthy (e.g. Johnson 2010; IRRC 2016; SEC 2016; Berkman 2018). Consistent with critics' concerns about the informativeness of risk factor disclosures, Beatty, Cheng, and Zhang (2018) report that the market's response to unexpected 10-K risk factor disclosures has decreased in recent years. However, existing research sheds little light on the benefits firms derive from lengthy and boilerplate risk factor disclosures that could explain the persistence of these attributes. We contribute to this literature by examining how risk factor disclosure length, non-specificity, and standardization correlate with the likelihood regulators and federal judges will flag those risk factor disclosures as inadequate.

2.2 CAUTIONARY LANGUAGE AND THE PSLRA'S SAFE HARBOR

Risk factor disclosures play an important role in providing legal protections to firms making forward-looking statements. In response to concerns that securities lawsuits were discouraging managers from providing useful forward-looking information to investors, Congress enacted a safe harbor provision to protect forward-looking statements as part of the

Private Securities Litigation Reform Act of 1995 (PSLRA) (15 U.S.C. § 78u5(c)).⁶ In order to receive the full protection of the PSLRA safe harbor, firms must accompany forward looking statements with “meaningful cautionary language” or risk factor disclosure. This safe harbor protection often results in the preliminarily dismissal of claims brought by shareholders over allegedly false or misleading forward looking statements.⁷ Conditional on a firm being sued for

Congress did not explicitly define what makes cautionary language “meaningful”, though the statute indicates that cautionary language must identify important risks that could cause actual results to vary from projections. The Congressional Conference Report accompanying the PSLRA asserts that risk factor disclosures should focus on specific risks, that boilerplate cautionary language would not suffice for purposes of the statute’s safe harbor, and that only important factors should be listed (Cong. Rec. 28 Nov. 1995 13703). After the SEC began mandating the disclosure of firms’ principal risks as Item 1A in their periodic filings starting in 2005, Item 1A became the primary location for much of the cautionary language firms use to invoke safe harbor protection for their forward-looking statements (Nelson and Pritchard 2016). Ultimately, whether a firm’s risk factor disclosures are deemed as adequate cautionary language depends on judicial interpretation and application of the statutory safe harbor provisions (e.g., Olazábal 2000; De Simone, Ingber, and Creutz 2004; Pritchard and Sale 2005; Cornerstone

⁶ While the PSLRA codified a safe harbor protecting forward looking statements, even prior to the PSLRA a common law rule called the Bespeaks Caution Doctrine protected projections and estimates accompanied by meaningful cautionary language because misrepresented statements that were accompanied by such warnings were deemed to be immaterial. In re Trump Casino Sec. Litig., 7 F.3d 357, 371 (3d Cir. 1993); Harris v. IVAX Corp., 998 F. Supp. 1449, 1454 (S.D. Fla. 1998).

⁷ The safe harbor of the PSLRA has two prongs and the first protects forward looking statements that are immaterial and statements accompanied by “meaningful cautionary statements identifying important factors that could cause actual results to differ materially” (15 U.S.C. § 78u-5(c)(1)(A)). Claims based on statements covered by the first prong are typically dismissed by courts on preliminary motion without inquiry into the state of mind of the firm that made the statement. In the absence of such cautionary statements, lesser protections may apply under the second prong that permits an inquiry into the state of mind of the firm making the statement (15 U.S.C. § 78u-5(c)(1)(B)). The second prong protections are more similar to the weaker general PSLRA protections for non-forward looking statements and potentially opens the firm up to costly discovery proceedings.

2016). Thus, we focus on explicit judicial assessments of firms' risk factor disclosures to provide direct insight into how risk factor disclosure properties relate to safe harbor protections.

In order to gain institutional insight into the factors that judges claim influence their assessments of risk factor disclosure adequacy and to guide the development of our hypotheses, we read over five hundred judicial opinions published between 1996 and 2015 relating to defendants' request to dismiss securities lawsuits that allege false or misleading forward-looking statements. We find that in over eighty percent of these decisions, the judge makes an explicit ruling over the firms' cautionary language as either adequate or inadequate to avail the firm's forward-looking statements of safe harbor protection. Next, we categorize the judges' rationale and tabulate in Table 1 the reasons cited by judges to support their rulings.

Panel A indicates that the extent to which risk factor language is boilerplate versus firm-specific is commonly cited as a factor influencing judicial assessments of risk factor disclosure adequacy. However, consistent with similar views expressed by legal scholars (Bloomenthal and Wolff 2012), our reading of the underlying judicial opinions suggests many of these assessments are largely ad hoc and are not based on any standard, objective benchmark of what constitutes sufficient specificity. For instance, Panel A of Table 1 indicates that several judges explicitly consider how similar a firm's risk factor disclosures are to risk disclosures that were considered adequate by prior courts. This practice is consistent with the legal doctrine of *stare decisis*, which binds judges to uphold precedents set in prior cases. In one case, a federal court granted a favorable ruling of risk factor disclosure adequacy because the firm's disclosure was "virtually identical to language approved by the Ninth Circuit in instances in which forward-looking

statements were immunized by the PSLRA Safe Harbor.”⁸ Firms’ adoption of risk factor language that has been vetted by prior courts could lead to more standardized risk factor language that is less likely to be considered inadequate under subsequent judicial review. We formalize the association between standardized risk factor disclosures and judicial assessments of inadequacy as our first hypothesis, stated in null form:

H1a: There is no association between risk factor disclosure standardization and the likelihood risk factor disclosures will be deemed inadequate for purposes of the PSLRA’s safe harbor.

The most common factor cited by judges ruling that risk factors are adequate is that the disclosure simply fulfils the requirement to warn investors of risks that could cause actual results to vary. The most commonly cited reason for why risk factor language is inadequate is the firm’s omission of a material risk factor. Because specific risk factor disclosures inherently cover a narrower subset of potential adverse outcomes, it is possible that less specific risk factor disclosures provide superior legal protection. These considerations suggest that even generic risk factor disclosures may be adequate to avail firms of safe harbor protection.⁹

Consistent with this notion, one partner at a national law firm recently asserted that all 10-Ks contain boilerplate risk factor language to make such factors “as all-encompassing as possible” (see Berkman 2018). This incentive may help explain Hope et al.’s (2016) finding that Item 1A risk factor disclosures are less specific than other sections of the 10-K. Thus, despite legislative guidance that risk factors should be firm-specific, the actual association between risk

⁸ In *Re Fusion-IO, Inc. Securities litigation*, 2015 U.S. Dist. LEXIS 18304; see also *Plumbers & Pipefitters Local Union No. 630 Pension-Annuity Trust Fund v. Allscripts-Misys Healthcare Solutions, Inc.*, 778 F. Supp. 2d 858 (2011).

⁹ Bloomenthal and Wolff (2012) questioned whether the 7th circuit court correctly identified boilerplate language in determining the adequacy of risk factor disclosures in the case *Asher v. Baxter International, Inc.*, 377 F.3d 717 (7th Cir. 2004) because such disclosures, although deemed adequate by the court, appeared to “cover as many conceivable generic factors that could relate to the company’s business as possible.”

factor specificity and judicial assessments of adequacy is unclear. We formalize the tension in the relationship between risk factor specificity and the likelihood risk factors are flagged as inadequate for purposes of the PSLRA's safe harbor as our next hypothesis, stated in null form as follows:

H1b: There is no association between risk factor disclosure specificity and the likelihood risk factor disclosures will be deemed inadequate for purposes of the PSLRA's safe harbor.

The PSLRA and its supporting legislative guidance indicate that only important risk factor disclosures entitle firms to safe harbor protections. However, there are a number of reasons that longer risk factor disclosures may provide superior legal protection. First, judges may view lengthier risk factor disclosures as better fulfilling the PSLRA's requirement that risk factor disclosures convey "substantive" information about the risks that could cause actual results to vary. Panel A of Table 1 indicates that judges sometimes cite the length of a firm's risk factor disclosures as supporting evidence for their assertion that the firm's cautionary language was adequate. Second, uncertainty regarding which risks are most important may lead assessors of risk disclosure adequacy to view lengthier risk factor disclosures as representing a good faith effort to warn of all potentially material risks the firm may face. Third, the inclusion of a longer list of risk factor disclosures helps ensure no obvious omissions have been made. Given that the congressional record and some evidence from our reading of judicial opinions may be at odds with one another, we formalize our next hypothesis in null form as follows:

H1c: There is no association between risk factor disclosure length and the likelihood risk factor disclosures will be deemed inadequate for purposes of the PSLRA's safe harbor.

2.3 THE SEC'S ITEM 1A RISK FACTOR DISCLOSURE MANDATE

In 2005, the SEC began mandating disclosure of firms' most significant risk factors in Item 1A of firms' periodic filings to help investors understand the nature of the risk inherent in the company. Prior to 2005, any cautionary language that firms provided for purposes of the PSLRA's safe harbor was not consistently reported in any particular section of firms' periodic filings. SEC reporting requirements stipulate that Item 1A risk factor disclosures should avoid vague boilerplate (17 C.F.R. 230.421(b)) and be clear, concise, and tailored to the specific risks facing the reporting firm (e.g., SEC 2004; SEC 2005).

The SEC reviews Item 1A risk factor disclosures, along with firms' other periodic filings, as part of its filing review process. Firms whose risk factor disclosures are non-compliant with disclosure guidelines may receive SEC comment letters asking them to rectify the perceived deficiency. The SEC states it "concentrates its review resources on disclosures that appear to be inconsistent with Commission rules or applicable accounting standards, or that appear to be materially deficient in their rationale or clarity." The SEC Division of Corporate Finance's 11 offices possess industry-specific accounting and disclosure expertise and manage the review of filings of firms in these industries.¹⁰

The SEC's review of Item 1A risk factor disclosures varies in important respects from judicial review of risk factor disclosures to determine disclosure adequacy. Perhaps most importantly, the SEC is under no obligation to explicitly rule on the adequacy of risk factor disclosures that it reviews. Rather, the SEC only sends comment letters relating to risk factor disclosures if it identifies a perceived deficiency. Thus, no observable documentation exists regarding the SEC's assessments when risk factor disclosures are deemed to be adequate for

¹⁰ See <https://www.sec.gov/divisions/corpfin/cffilingreview.htm> (accessed on March 20, 2018).

regulatory compliance. To identify the specific disclosure deficiencies the SEC asks firms to rectify and to guide the development of our hypotheses, we read a randomly selected sample of two hundred SEC comment letters identified by Audit Analytics as relating to firms' risk factor disclosures. Based on our reading, we identified six separate categories of comments and tabulate them as Panel B in Table 1. Because the SEC may have multiple comments relating to risk factor disclosures within a single comment letter, the total number of comments tabulated in Panel B exceeds two hundred. Panel B indicates that the SEC often requests firms to clarify, expand, and occasionally reduce risk factor disclosures.

Although SEC comment letters often solicit additional information about the firm's own risks, it is possible that risk factor disclosures that more closely mirror industry peers' risks are less likely to attract regulatory scrutiny in the first place incident to the structure of the SEC's process. The SEC's filing review process is conducted by staff who review filings for a large number of firms within the same industry, and as such the reviewing staff may be less likely to challenge risk disclosures that look similar to those of industry peers. Recent research suggests that firms preparing their own disclosures often look to peer firm disclosures to form their own expectation of what constitutes adequate disclosure, and that borrowing language from industry peers' disclosures is common (e.g., McMullin 2016, Berkman 2018; see also Brown, Tian, and Tucker 2017). This process is likely facilitated by large audit firms' statutory reporting and disclosure assistance services, which provide financial reporting templates for their clients based on their knowledge of the types of disclosures that were flagged as inadequate by regulators in the past.

Thus, despite the SEC's formal guidance that risk factor disclosures should be tailored to the firm, it is possible that more standardized risk factor language is less likely to attract

regulatory attention. We formalize the tension in the relationship between standardized risk factor disclosures and the likelihood of an SEC comment letter as our next hypothesis, stated in null form as follows:

H2a: There is no association between risk factor disclosure standardization and the likelihood risk factor disclosures will be deemed inadequate under the SEC's filing review process.

Panel B of Table 1 indicates that lack of specificity is the most frequently addressed issue in SEC comment letters relating to risk factor disclosures, suggesting that firms with more specific risk factor language may avoid receiving a comment letter from the SEC. However, there are at least two reasons that a lack of disclosure specificity may not lead to a greater likelihood of receiving an SEC comment letter in the first place. First, the SEC may use other filtering mechanisms, such as the extent to which a firm's risk disclosures vary from those of industry peers, to select risk disclosures for further investigation. In that case, the SEC's solicitation for greater specificity in risk disclosures is only conditional on, and not a driver of, the decision to issue the comment letter.

Second, the SEC's formal disclosure guidance requires firms to be concise while still disclosing the most significant risks that make their offering speculative or risky (SEC 2005). Because less specific risk factor language inherently covers a broader set of adverse potential outcomes concisely than does more specific language, disclosures that describe risks in less specific terms may be less likely to initially attract the SEC's attention. We consider the association between risk factor specificity and the likelihood risk factor disclosures are targeted by an SEC comment letter in our next hypothesis, stated in null form:

H2b: There is no association between risk factor disclosure specificity and the likelihood risk factor disclosures will be deemed inadequate under the SEC’s filing review process.

Although SEC guidance cautions firms to make risk factors concise, lengthier risk factor disclosure may decrease the likelihood regulators will observe an obvious omission that could spawn further scrutiny. The descriptive evidence in Panel B of Table 1 suggests the SEC is more likely to request a firm to lengthen its risk factor disclosures rather than remove an existing factor. In 57 of the comment letters, the firm was asked to add an additional risk factor. In 40 comment letters the firm was asked to remove “qualifying language” and in only 7 cases did the SEC ask the firm to remove an existing risk factor. Thus, how risk factor disclosure length relates to the probability risk factor disclosures are targeted by the SEC is an open empirical question. We formulate our final hypotheses in null form as follows:

H2c: There is no association between risk factor disclosure length and the likelihood risk factor disclosures will be deemed inadequate under the SEC’s filing review process.

3. Sample Construction

Our analysis of judicial and regulatory assessments of firms’ risk factor disclosures requires the construction of two distinct samples. The first sample consists of firms whose risk factor disclosures were assessed for adequacy by judges making safe harbor determinations during the dismissal phase of securities class action lawsuits. We search the Lexis Advance legal database for all published judicial opinions relating to the motion to dismiss securities litigation alleging false or misleading forward-looking statements between 2006 and 2015. Keyword

searches retrieved an initial judicial opinion sample for 328 securities lawsuits that we were able to match with lawsuit data from Stanford's Securities Class Action Clearinghouse database.¹¹

We eliminate several cases due to missing Compustat or CRSP data and cases where either the judge determined the statements were not forward-looking or failed to explicitly rule on the adequacy of the cautionary language. Our procedure for selecting this sample is detailed in Panel A of Table 2. Ultimately, our regression analyses examining judicial assessments of risk factor disclosure are based on a sample of 144 distinct lawsuits in which the judge explicitly rules regarding the adequacy of the firm's cautionary language.

We measure the judicial assessment of risk factor disclosure adequacy for each lawsuit and the boilerplate for the Item 1A Risk Factor disclosures in the firm's 10-K filed during the lawsuit's class period.¹² Because the class period of some lawsuits includes more than one year, our final sample for our main analysis of judicial assessments includes 231 unique firm-years corresponding to 231 unique 10-K filings made during the class period.

Our second sample consists of firm-years in which a firm received an SEC comment letter relating to at least one Item 1A risk factor disclosure between 2005 and 2015. We use Audit Analytics' SEC Comment Letter database to construct this sample. As illustrated in Panel B of Table 2, we begin with an initial sample of 4,450 firm-years for which Audit Analytics indicates the corresponding SEC filings were the target of a comment letter relating to risk factors. After eliminating firm-years for which Item 1A Risk Factor disclosures were not retrievable from the SEC website, firm-years for which textual analysis variables could not be

¹¹ Our initial search includes securities cases between 1996 and 2015 with judicial opinions that use the following phrases: "risk factor," "cautionary language," and "safe harbor." This search returned over 500 cases. This sample is reduced to 328 align the time period with the presence of Item 1A (2006-2015).

¹² In some cases, the class period did not include a 10-K filing date. In these cases, we selected the 10-K filed immediately prior to the start of the class period.

computed from the extracted Item 1A disclosure, and firm-years missing necessary CRSP or Compustat data, we are left with a final sample of 1,607 firm-years for which risk factor disclosures were the target of an SEC comment letter in the following year. We use this same dataset to identify firm-years in which a firm did not receive an SEC comment letter relating to risk factor disclosures to form a control sample. After identifying these firms, our final sample for the SEC comment letter analysis includes a total of 33,725 firm-years.

4. Variable Measurement

4.1 MEASURING JUDICIAL AND REGULATORY ASSESSMENT OF RISK FACTOR INADEQUACY

Hypotheses 1a, 1b, and 1c center on the likelihood risk factor disclosures will be ruled inadequate under judicial review as a function of risk language standardization, specificity, and length. To code our dependent variable for these analyses, we identify risk factor disclosures deemed inadequate under judicial review by reading each opinion in our sample of lawsuits and determining whether the judge assesses the firm's cautionary language as sufficient to avail itself of the PSLRA's safe harbor. We set the indicator variable *RF_INADEQUATE_JG* equal to one for firm-years in class periods in which the cautionary language was ultimately deemed inadequate by the judge, and zero for firm-years in class periods for which the judge ruled the cautionary language adequate.

Hypotheses 2a, 2b, and 2c consider the probability risk factor disclosures will be flagged as inadequate under regulatory review as a function of risk language standardization, specificity, and length. We use the SEC's decision to issue a risk factor-related comment letter as a proxy for the SEC's assessment that the firm's risk factor disclosure was inadequate. We set the indicator variable *RF_INADEQUATE_SEC* equal to one if a firm-year's risk factor disclosures were targeted by an SEC comment letter, and zero otherwise. We use Audit Analytics' Comment

Letter database to identify firms that received comment letters relating to their risk factor disclosures in the 10-K filing for any given year.

Our objective in testing our hypotheses is to capture the association between risk factor disclosure characteristics and judicial or regulatory assessments of risk factor disclosure adequacy, holding constant the firm's underlying risks, industry affiliation, headquarter location, and fiscal year. We next discuss our measure of disclosure standardization, specificity, and length, as well as the construction of each of our control variables.

4.2 MEASURES OF DISCLOSURE LENGTH AND BOILERPLATE

Prior literature has examined two measures of “boilerplate” that reflect distinct disclosure practices that cause disclosures to convey less firm-specific information. The first measure captures the extent to which disclosure language is so standardized that it is unlikely to be informative (e.g., Lang and Stice-Lawrence 2015). To capture firms' use of standardized language in their risk factor disclosure, we first identify commonly used trigrams (3-word phrases) in risk factor disclosure sentences for all firms in the same two-digit SIC industry. Next, we identify standardized sentences as those that either use 10 or more of these commonly used trigrams or for which 10 percent or more of the trigrams in the sentence are commonly used. We then count the number of words in these standardized sentences and divide this number by the total number of words in the risk factor section. Our approach to measuring this variable, which we label *STANDARDIZATION*, is similar to the approach used by Lang and Stice-Lawrence (2015) to measure boilerplate in annual reports of non-US firms.

The second measure of “boilerplate” is based on the Stanford Named Entity Recognition (NER) tool that determines the extent to which specific entities are discussed in the disclosure (e.g., Hope et al. 2016). This measure captures the extent to which disclosures are more or less

specific based on the inherent precision of the language. Following Hope et al. (2016), we use a computing algorithm to identify words that specifically describe an entity (name or numeric expression), and count the number of specific entities described in a firm's risk factor section. We scale this word count by the total number of words in the risk factor section to adjust for differences in the length of this disclosure across firms. We label this measure *NON_SPECIFICITY*.¹³ Because we are interested in examining the effects of specificity after controlling for determinants of the underlying disclosure, we benchmark this measure against that of each firm's most similar industry peer. Each firm's most similar industry peer is based on the similarity of the two firms' 10-K Item 1 Business Description disclosure. Details of this matching procedure can be found in the Appendix.

We measure Item 1A risk factor disclosure length based on the total number of words, consistent with how disclosure volume is measured by a host of prior studies (e.g., Li 2008; You and Zhang 2009; Miller 2010; Lawrence 2013; Campbell et al 2014). *LENGTH* is the natural logarithm of the number of words in the Item 1A risk factor disclosure in the firm-year's 10-K. Because we are interested in examining the association between risk factor disclosure length and judicial and regulatory assessments of adequacy after holding other determinants of disclosure constant, we also benchmark this variable against the most similar industry peer.

4.3 CONTROL VARIABLES

We include several control variables that may confound the relationship between risk factor disclosure attributes and adjudicators' assessments of risk factor disclosure adequacy. First, we include the correlation of firms' stock returns and industry peer firm stock returns to

¹³ Untabulated analyses indicate that *NON_SPECIFICITY* and *STANDARDIZATION* have only a 22 percent correlation. Thus, although these two variables capture related constructs, the underlying disclosure behaviors are distinct.

control for the fact that firms with similar underlying economic events may adopt similar language to describe for those events. *RET_COR* is the mean correlation between each firm's 24-month return ending at the end of the current fiscal year and that of its 2-digit SIC industry peers. We also construct several measures of firm risk that may influence firms' risk factor disclosures as well as adjudicators' assessments of disclosure adequacy. *RET_VOL* is the standard deviation of daily returns for the year ending two days prior to the 10-K filing date. *BETA* is the firm's market beta measured using daily returns for the one year ending two trading days prior to the 10-K filing date. *SIZE* is the natural logarithm of the market value of equity. *LEV* is total liabilities scaled by total assets, measured as of the end of the fiscal year. *BIG_N* is an indicator variable equal to one if the firm is audited by one of the largest auditors (the Big N), and zero otherwise. *SKEW* is measured as the negative coefficient of skewness for daily returns over the year ending two days prior to the 10-K filing date. *TURNOVER* is the mean number of shares traded during the fiscal year scaled by the number of common shares outstanding. *ETR* is the firm's effective tax rate, computed as tax expense scaled by pre-tax income. *ROE* is income before extraordinary items scaled by the market value of equity. We include various fixed effects to control for industry affiliation, fiscal year, and the U.S. circuit in which each firm is headquartered because research indicates judicial practices and litigation risk sometimes vary by circuit (Hopkins 2018; Cazier, Christensen, Merkley, Treu 2018).¹⁴ In some specifications we also include controls for the actual topics discussed in the firm's risk factor disclosures (See

¹⁴ We measure industry affiliation at the 2-digit SIC level for the SEC comment letter analysis. Because of the relatively small sample size in the lawsuit sample, we measure industry affiliation at the Fama-French 12 industry classification level for those tests to maintain a practical number of degrees of freedom (http://mba.tuck.dartmouth.edu/pages/faculty/ken_french/Data_Library/det_12_ind_port.html).

Appendix B for construction of these risk topic controls) however, these controls are excluded from the main specification to avoid overfitting concerns

Equation 1 below shows the form of our primary regression model:

$$\Pr(\text{Risk Factors Assessed as Inadequate}_{i,t}) = \beta_0 + \beta_1 \text{STANDARDIZED}_{i,t} + \beta_2 \text{NON_SPECIFICITY}_{i,t} + \beta_3 \text{LENGTH}_{i,t} + \beta_4 \text{RET_COR}_{i,t} + \beta_5 \text{RET_VOL}_{i,t} + \beta_6 \text{BETA}_{i,t} + \beta_7 \text{SIZE}_{i,t} + \beta_8 \text{LEV}_{i,t} + \beta_9 \text{BIG_N}_{i,t} + \beta_{10} \text{SKEWNEWS}_{i,t} + \beta_{11} \text{TURNOVER}_{i,t} + \beta_{12} \text{ETR}_{i,t} + \beta_{13} \text{ROE}_{i,t} + \text{Industry controls} + \text{Circuit controls} + \text{Year controls} + \text{Topic controls} + \varepsilon_{i,t} \quad (1)$$

4.4 DESCRIPTIVE STATISTICS

Table 3 presents sample means for each of our independent variables of interest. Panel A displays the descriptive statistics for our sample of firms sued for false or misleading forward-looking statements. Panel B displays the descriptive statistics for firm-years in our comment letter analysis. We note that our variables do not appear to exhibit extreme skewness, as the mean for each continuous variable is within the inner-quartile range in nearly every case. To mitigate the effects of extreme outliers, all continuous variables are winsorized at the 1st and 99th percentiles.

5. Empirical Results

5.1 JUDICIAL ASSESSMENT OF RISK FACTOR DISCLOSURE ADEQUACY

Our first set of hypotheses focuses on judicial assessments of risk factor disclosure inadequacy. To substantiate the notion that judicial assessments of risk factor disclosure inadequacy are associated with costly legal outcomes and to validate our measurement of *RF_INADEQUATE_JG*, we first estimate a logistic regression of the likelihood of lawsuit dismissal as a function *RF_INADEQUATE_JG* and control variables. Column 1 of Table 4 displays results from estimating this regression over the entire sample period 1996 to 2015. Column 2 of Table 4 re-estimates this regression only on post-2005 observations for which we

have Item 1A data. Both columns report a significantly negative coefficient on *RF_INADEQUATE_JG* ($p < 0.01$), confirming that a ruling of risk factor disclosure inadequacy for purposes of the safe harbor imposes significant costs for defendant firms.

Results from modeling judicial assessments of risk factor disclosure adequacy as a function of Item 1A standardization, specificity, and length are reported in Table 5. In the first column of Table 5, we present results from including *LENGTH* as the only risk factor disclosure attribute and find a significantly negative ($p < 0.05$) relationship between risk factor disclosure length and the likelihood risk factor disclosures will be ruled as inadequate for purposes of the safe harbor. In the second column of Table 5, we replace *LENGTH* with *NON_SPECIFICITY* and find it also has a significantly negative coefficient estimate ($p < 0.05$), suggesting that risk factor disclosures containing less specific language are *less* likely to be assessed as inadequate by judges for purposes of safe harbor protections.¹⁵ Re-estimating this model after replacing *NON_SPECIFICITY* with *STANDARDIZATION* produces the results displayed in the third column of Table 5, which show no significant association between the extent to which risk factor disclosures use standardized language and the likelihood risk factor disclosures are flagged as inadequate under judicial review.

The last column of Table 5 presents results after including all three risk disclosure attributes together, along with indicator variables controlling for the topics discussed in the firm's risk factor disclosure. Results indicate that even after controlling for disclosure content, *NON_SPECIFICITY* has a strong negative association ($p < 0.01$) with the likelihood risk factor

¹⁵ Stanford's NER tool identifies entities in seven categories, three of which refer to proper nouns (persons, organizations, and locations), and four of which refer to numeric values (time, money, percent, and date). In untabulated analyses, we split the *NON_SPECIFICITY* variable into the portion made of proper nouns versus numeric values and find that the results in Table 5 appear to be driven primarily by a lack of specificity relating to proper nouns.

disclosures are deemed inadequate for safe harbor purposes. *LENGTH* continues to be negatively associated with the likelihood risk disclosures are ruled inadequate, though the inclusion of the specific topic indicators attenuates the statistical significance of this relationship. We note that when estimating this regression without topic indicators, the coefficient estimate on *LENGTH* is negative and strongly significant ($p < 0.01$) and the coefficient estimate on *NON_SPECIFICITY* is significant at the 5% level.

Results in Table 5 suggest that, despite the caution expressed in the congressional record and by courts that risk factor disclosure should be specific and concise, less specific and lengthy risk factor disclosure correlates with more favorable judicial outcomes. These results are consistent with the view that lengthy and non-specific risk factor disclosure may provide greater legal protection by covering a broader set of potentially adverse events and, as occasionally expressed in judicial opinions, that relatively generic language can meet the requirements of the statute.

To provide a sense of the economic magnitude of our results, we compute the mean marginal effects for the two explanatory variables of interest that load significantly in the last column of Table 5. Untabulated marginal effects suggest that going from the first quartile to the third quartile of *LENGTH* decreases the probability risk factor disclosures are ruled as inadequate under judicial review by 7.6 percent. The probability a firm's risk factor disclosures are ruled inadequate under judicial review decreases by 13.7 percent as those disclosures move from the first to the third quartile of *NON_SPECIFICITY*. This evidence suggests one reason firms may continue to issue lengthy and non-specific risk factor disclosures is to benefit from an increased likelihood of being granted safe harbor protection.

5.2 REGULATORY ASSESSMENT OF RISK FACTOR DISCLOSURE ADEQUACY

Panel A of Table 6 presents results from estimating equation 1 on our sample of firms for our SEC comment letter analysis. We replace the dependent variable with *RF_INADEQUATE_SEC*, which is an indicator variable equal to one in firm-years if the firm receives an SEC comment letter relating to their Item 1A risk factor disclosures. The ordering of results in Table 6 parallels that in Table 5. The first column of Panel A indicates that *LENGTH* is not significantly associated with the likelihood of receiving an SEC comment letter relating to risk factor disclosures. The second column shows a negative and marginally significant ($p < 0.10$) coefficient estimate on *NON_SPECIFICITY*, providing modest support for the notion that SEC comment letters are less likely to target risk factor language that is less specific. The third column of Panel A reveals a strongly negative ($p < 0.01$) coefficient estimate on *STANDARDIZATION*, suggesting the SEC is significantly less likely to issue comment letters relating to risk factor disclosures when those disclosures use language that is pervasive in industry peers' disclosures. The final column of Panel A displays results from estimating this model with all three risk factor disclosure variables included, along with indicator variables capturing the topics discussed in the risk disclosure. Results indicate only *STANDARDIZATION* is significantly associated with the likelihood of receiving an SEC comment letter in this specification.

An analysis of marginal effects suggests that the association between standardized language (*STANDARDIZATION*) and the probability of an SEC comment letter targeting risk factor disclosures is relatively modest. Going from the first quartile to the third quartile of *STANDARDIZATION* from Model 4 of Panel A in Table 6 decreases the probability of an SEC letter by only approximately 0.5 percent. However, results from Panel A do suggest that, despite

the SEC's frequent admonition to tailor risk factor disclosures to the specific risk factors faced by the firm, risk factor language that more closely resembles that of industry peers is less likely to be flagged as inadequate under regulatory review.

5.3 ROBUSTNESS TESTS

One potential limitation of the sample examined in Panel A of Table 6 is that not all firms undergo the SEC filing review process each year. The SEC reviews filings of every publicly traded firm at least every three years, with some firms' filings being reviewed more frequently. However, the SEC does not reveal its formula for determining which firms to review more frequently, and in the absence of a comment letter, we cannot discern which firms were under review each year. To test the sensitivity of the results in Panel A of Table 6, we rerun our estimation only on firms that received at least one SEC comment letter during the year to ensure our sample includes only firms whose filings were actually under review. We tabulate results from this sample, which includes 13,419 firm-year observations, in Panel B of Table 6. We find the association between *STANDARDIZATION* and *RF_INADEQUATE_JG* to be qualitatively and quantitatively similar to those reported in Panel A, providing further support for the view that the likelihood risk factor disclosures are targeted by an SEC comment letter is *decreasing* in disclosure boilerplate. We continue to find no association between *LENGTH* and the likelihood of an SEC comment letter.

A second potential limitation of the results in Panel A of Table 6 is that we rely on Audit Analytics' comment letter database to identify comment letters relating to Item 1A risk factor disclosures. Our reading of a significant subsample of the SEC comment letters relating to risk factors as identified by Audit Analytics indicates that the majority (approximately 75%) relate to Item 1A risk factor disclosures. However, we find that a significant minority of SEC comment

letters that Audit Analytics indicates as relating to “risk factors” do not directly relate to Item 1A Risk Factor disclosures, which potentially induces noise or even bias in our main analyses. To examine the impact of this potential bias in our sample, we test the sensitivity of our results to an alternative method of identifying comment letters relating to Item 1A risk factors. We use a regular expression to measure an alternative variable equal to one if the SEC comment letter includes the text “Item 1A,” and zero otherwise. We find this alternative indicator variable to be 45 percent correlated with *RF_INADEQUATE_SEC*. We report results from estimating equation (1) after substituting this variable as the dependent variable in Panel C of Table 6. Results in the second and third columns of Panel C indicate that both *NON_SPECIFICITY* and *STANDARDIZATION* are significantly negatively associated with the likelihood of receiving an SEC comment letter related to Item 1A risk factor disclosures. However, results in the last column of Panel C suggest that when both *STANDARDIZATION* and *NON_SPECIFICITY* are both included in the model, only *STANDARDIZATION* has a significantly negative coefficient estimate ($p < 0.05$). We continue to find no evidence of an association between risk factor disclosure length and the likelihood of an SEC comment letter.

We next examine the sensitivity of our results to alternative cutoffs to identify boilerplate trigrams and sensitives to measure *STANDARDIZATION*. Our original construction of this variable in our main analyses defines boilerplate sentences as those with at least 10% of all trigrams in the sentence being boilerplate trigrams, where boilerplate trigrams are those appearing in at least 10% of all Item 1A disclosures in the same 2-digit SIC industry on average, but not in more than 90% of all industry peers’ Item 1A disclosures. We test the robustness of our results to alternative cutoffs and report the results in Table 7. The first column of Table 7 (Model 1) displays results after changing those cutoffs to 20%, 20%, and 80%, respectively. The

second column of Table 7 (Model 2) displays results after changing those cutoffs to 25%, 25%, and 75%, respectively. Under both alternative sets of cutoffs, *STANDARDIZATION* continues to have a strong negative association ($p < 0.01$) with the likelihood of an SEC comment letter.

The SEC has repeatedly admonished firms to avoid risk factor language that could apply to *any* issuer. We next examine results from the SEC comment letter analysis after broadening our measure of *STANDARDIZATION* to include trigram similarity with all firms in the economy rather than those in just the same industry. The resulting measure of standardized language arguably captures greater reliance on extremely generic risk factor disclosure language. We apply the original cutoffs (10%, 10%, and 90%) to our construction of *STANDARDIZATION* where these percentages apply to trigrams used in all Item 1A risk factors in our sample, rather than those only in the same 2-digit SIC industry and report results in the last column of Table 7. We find this variable does not correlate significantly with the probability of receiving an SEC comment letter. In untabulated analyses, we fail to find any association between market-wide measures of standardized language based on any alternative cutoffs and the likelihood of an SEC comment letter. Overall, our results suggest the SEC does not look favorably or unfavorably upon risk factor language that is so broad that it could apply to *any* issuer, though the SEC appears to reward firms using risk factor language that is pervasively used by firms within the same industry.¹⁶

We emphasize that the negative correlation between industry-level standardized risk factor language and SEC comment letters holds after controlling for the actual underlying risk factor topics disclosed. In untabulated analyses, we also include a control variable that measures the cosine similarity between a firm's risk factor topic probability vector of each firm's risk

¹⁶ We emphasize that our results hold after including controls for the actual underlying risk factor topics being disclosed. In untabulated analyses, we also find the negative relationship between

factors with those of its industry peers. Results are consistent with *STANDARDIZATION* being associated with a lower likelihood of risk-related SEC common letters, even after controlling for the similarity of disclosed risk topics. These results appear to be inconsistent with the SEC's admonition for firms to tailor their risk factor language to reflect the specific risks faced by the firm.

Finally, we test the robustness of our results to an alternative measure of risk factor disclosure length proposed by Campbell et al (2014). This alternative measure of length is based on the natural logarithm of a word count of specific risk-related words identified in their study. To test the robustness of our results to the inclusion of this alternative measure of the length of risk disclosures, we rerun our main analyses using this measure in place of *LENGTH*. We find that this measure also correlates significantly ($p < 0.05$) with the probability risk factor disclosures will be ruled as inadequate under judicial review, and that this variable has a marginally significant negative correlation with the likelihood of receiving an SEC comment letter ($p < 0.10$).

6. Conclusion

Despite official reporting requirements from the SEC that Item 1A risk factor disclosures should be specific and tailored to the reporting firm, practitioners and regulators continue to lament that these disclosures are boilerplate and uninformative. Whereas other academic studies document risk factor disclosures convey *some* information to the market and the use of lengthy boilerplate disclosures result in negative capital market consequences, our study is among the first to propose and test an explanation for why firms continue to provide lengthy and boilerplate disclosures. Specifically, we provide evidence that firms may benefit from using lengthy, non-specific and standardized risk factor disclosures by avoiding unfavorable regulatory or judicial

outcomes. Our results provide insight that may inform policy makers and standard setters regarding how legal standards and SEC mandates are implemented in practice. Our results highlight a potential disconnect between formal regulatory and legislative disclosure guidance and how that guidance is enforced via regulatory and judicial outcomes.

The results from our study are subject to certain caveats. For instance, we note that the correlation we document between boilerplate and the likelihood risk factor disclosures will subsequently be deemed adequate is not necessarily causal. We cannot rule out the possibility that risk factor disclosures that contain more boilerplate happen to be superior along other dimensions not included in our model, and that it is these correlated omitted variables to which adjudicators are responding. However, the link between favorable judicial and regulatory outcomes and disclosure characteristics the SEC asserts are undesirable provides one possible explanation for the persistence of lengthy and boilerplate risk disclosures.

APPENDIX

Description of *LENGTH* measure

First, we calculate the similarity between all firms in the same two-digit SIC industry and identify a firm's peer as the firm with the most similar 10-K Item 1 Business Description disclosure. To measure similarity, we first construct a topic model using Latent Dirichlet Allocation with 200 topics over all firms' Item 1 and calculate the topic probability vector for each firm (Dyer et al. 2017). Our similarity metric is in the same spirit as Hoberg and Phillips (2016), except that we calculate the cosine similarity score between two firms' topic probability vectors rather than between vectors indicating word presence. The motivation for using the LDA topic model relaxes the assumption made by Hoberg and Phillips (2016) that each word is its unique topic. Second, for each firm in our sample, we subtract the natural logarithm of the number of words of the most similar peer to obtain a measure of *abnormal* disclosure length.

Description of *NON_SPECIFICITY*

We use Python's natural language tool kit (NLTK) package (<http://www.nltk.org>) to implement Stanford's Named Entity Recognition algorithm (Lafferty et al. 2001, Finkel et al. 2005). This algorithm identifies specific references to entities in seven categories: person, organization, location, money, date, time, and percent. Hope et al. (2016) explain in detail the implementation and assess the construct validity of this measure.

Description of *STANDARDIZATION*

We define commonly-used trigrams are those that appear in at least 10% and not more than 90% of the risk factor disclosures produced by firms in the same SIC 2-digit industry.

Description of topic model

We estimate a topic model using a Bayesian machine-learning approach proposed by Blei et al. (2003).¹⁷ Specifically, we split each risk factor section into paragraphs, remove stop words, estimate an LDA topic model with 200 topics, and calculate a topic probability vector (TPV) for each risk factor paragraph. We identify the topic for each paragraph as the topic with the highest probability and create 200 indicator variables that equal one if the topic was disclosed in that

¹⁷ Blei (2012) provides an overview of topic models. Also, prior accounting studies have estimated topic models for 10-Ks (Dyer et al. 2017), analyst reports (Huang et al. 2016), and risk factor disclosures (Campbell et al. 2014).

firm's risk factor disclosures, and zero otherwise. To preserve degrees of freedom, we replicate this analysis using only 20 topics when controlling for topic inclusion in the lawsuit analysis.

REFERENCES

- Beatty, A., Cheng, L., & Zhang, H. 2018. Are Risk Factor Disclosures Still Relevant? Evidence from Market Reactions to Risk Factor Disclosures Before and After the Financial Crisis. *Contemporary Accounting Research*, forthcoming.
- Berkman, O. 2018. Disclosure Effectiveness Weakened By Complicated Ownership. *Financial Executives International Daily*.
- Blei, D. M. 2012. Probabilistic topic models. *Communications of the ACM* 55 (4): 77–8.
- Blei, D., Ng, A., & Jordan, M. 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research* 3 (4-5): 993–1022.
- Bloomenthal, H., & Wolff, S. 2016. Pleading Securities Fraud and the PSLRA. *Securities Law Handbook* Vol.2, Ch.29.
- Brown, S., Tian, X., & Tucker, J. 2017. The Spillover Effect of SEC Comment Letters on Qualitative Corporate Disclosure: Evidence from the Risk Factor Disclosure: 1–58.
- Campbell, J., H. Chen, H., Dhaliwal, D., Lu, H., & Steele, L. 2013. The information content of mandatory risk factor disclosures in corporate filings. *Review of Accounting Studies* 19 (1): 396–455.
- Cazier, R., Christensen, T., Merkley, K., and Treu, J. 2017. Litigation Risk and Non-GAAP Reporting. *Working Paper*: 1–55.
- Johnson, S. 2010. SEC Pushes Companies for More Risk Information. *CFO.com*.
- Cornerstone Research. 2017. Securities Class Action Filings 2016 Year in Review: 1–44.
- De Simone, J., Ingber, M. & Creutz, E. 2004. High Courts Should Review Ruling on Securities Fraud “Safe Harbor.” *Legal Opinion Letter - Washington Legal Foundation* 14 (26): 1–2.
- Dyer, T., Lang, M. & Stice-Lawrence, L.. 2016. The Evolution of 10-K Textual Disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting and Economics*: 1–68.
- Finkel, J., Grenager, T., & Manning, C. 2005. Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. In, 363–370. Stroudsburg, PA.
- Gaulin, M. 2017. Risk Fact or Fiction: The information content of risk factor disclosures: 1–51. *Working paper, University of Utah*.
- Hoberg, G., & Phillips, G. 2016. Text-Based Network Industries and Endogenous Product Differentiation Gerard Hoberg. *Journal of Political Economy* 124 (5): 1423–1465.

- Hope, O., Hu, D., & Lu, H. 2016. The benefits of specific risk-factor disclosures. *Review of Accounting Studies* 21 (4): 1005–1045.
- Hopkins, J. 2018. Do Securities Class Actions Deter Misreporting? *Contemporary Accounting Research*.
- Huang, A., Lehavy, R., Zang, A., & Zheng, R.. 2017. Analyst Information Discovery and Interpretation Roles: A Topic Modeling Approach. *Management Science*: mns.2017.2751–24.
- Investor Responsibility Research Center Institute (IRRC). 2016. The Corporate Risk Factor Disclosure Landscape.
- Kahan, M., & Klausner, M. 1997. Standardization and Innovation in Corporate Contracting (Or “The Economics of Boilerplate”). *Virginia Law Review* 83 (4): 713–770.
- Kravet, T., & Muslu, V. 2013. Textual risk disclosures and investors’ risk perceptions. *Review of Accounting Studies* 18 (4): 1088–1122.
- Lafferty, J., McCallum, A., & Pereira, F. 2001. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proceedings of International Conference on Machine Learning (ICML-2001)*, 282–289.
- Lang, M., Stice-Lawrence, L. 2015. Textual analysis and international financial reporting: Large sample evidence. *Journal of Accounting and Economics* 60 (2-3): 110–135.
- Lawrence, A. 2013. Individual investors and financial disclosure. *Journal of Accounting and Economics* 56 (1): 130-147.
- Li, F. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45 (2-3): 221-247.
- McMullin, J. 2016. Can I Borrow Your Footnotes? Learning and Network Benefits of Footnote Similarity. *Working Paper, Kelley School of Business*.
- Miller, B. 2010. The Effects of Reporting Complexity on Small and Large Investor Trading. *The Accounting Review* 85 (6): 2107-2143.
- Nelson, K., & Pritchard, A. 2016. Carrot or Stick? The Shift from Voluntary to Mandatory Disclosure of Risk Factors. *Journal of Empirical Legal Studies* 13 (2): 266–297.
- Olazábal, A. 2000. Safe Harbor for Forward-Looking Statements Under the Private Securities Litigation Reform Act of 1995. *Dickson Law Review* 105 (1): 1–30.
- Pritchard, A., & Sale, H. 2005. What Counts as Fraud? An Empirical Study of Motions to Dismiss Under the Private Securities Litigation Reform Act. *Journal of Empirical Legal Studies* 2 (1): 125– 149.

Securities and Exchange Commission (SEC). 2004. Proposed rule: securities offering reform, release nos. 33-8501; 34-50624; ic-26649, international series release no. 1282, file no. s7-38-04. Available at: <https://www.sec.gov/rules/proposed/33-8501.htm#VII>.

Securities and Exchange Commission (SEC). 2005. Securities and exchange commission final rule, release no. 33-8591: 1-468. Available at: <https://www.sec.gov/rules/final/33-8591.pdf>

Securities and Exchange Commission (SEC). 2016. Business and financial disclosure required by regulation S-K, release no. 33-10064; 34-77599; File No. S7-06-16: 1-341. Available at: <https://www.sec.gov/rules/concept/2016/33-10064.pdf>

You, H. & Zhang, X. 2009. Financial reporting complexity and investor underreaction to 10-K information. *Review of Accounting Studies* 14 (4): 559-586.

TABLE 1**Factors Cited in Judicial and Regulatory Assessments of Risk Factor Disclosure Adequacy**

Panel A: Judicial Assessments of Risk Factor Disclosure Adequacy for Safe Harbor Protections

<i>Supporting Reasons Cited When Risk Factor Disclosure is Ruled Adequate</i>	
Risk factor language fulfils the statutory requirement to warn investors of risks that could cause actual results to vary.	94
Risk factor language is specific/detailed/tailored to the firm	57
Risk factor language included the risk that actually transpired.	56
Assertion that risk factor language is adequate with no clear supporting explanation.	47
Risk factor language is extensive/lengthy/numerous	30
Risk factor language is similar to that found to be adequate by prior courts.	13
Other	12
Total	309
<i>Supporting Reasons Cited when Risk Factor Disclosure is Ruled Inadequate</i>	
Failure to disclose a material risk factor	51
Failure to disclose important facts made risk factors disclosures misleading	46
Failure to provide any risk language with the forward-looking statements	35
Risk factor disclosure was too vague/boilerplate/non-specific	35
Adverse events disclosed as "risk factors" were actually happening	20
Other	12
Total	199
Panel B: SEC Comments Relating to Risk Factor Disclosures	
Increase specificity relating to information already disclosed	84
Explain further or clarify a risk factor	83
Add an additional risk factor	57
Remove qualifying language from risk factors	40
Change heading or formatting to make content of risk factor more clear	9
Remove a risk factor	7
Total	280

Panel A of Table 1 categorizes the factors cited by judges in explaining why a given firm's risk factor disclosures are deemed adequate or inadequate to provide legal protection to a defendant's forward-looking statements under the safe harbor of the PSLRA. Panel A is based on 409 unique judicial opinions regarding defendants' motion to dismiss lawsuits filed between 1996 and 2015. The total factors cited sum to more than 409 because in some cases the judge cites more than one unique factor. These lawsuits consist of 264 unique cases for which the risk factor disclosures were deemed adequate, and 145 unique cases for which the firm's risk factor disclosures were deemed inadequate. Panel B categorizes the SEC comments relating to firms' risk factor disclosures for a random sample of 200 risk factor related SEC comment letters issued between 2006 and 2015. The total issues cited in the SEC comment letters sums to more than 280 because in some comment letters the SEC requests more than one change to the risk factor disclosure.

TABLE 2
Sample Selection

Panel A: Sample of Lawsuits Alleging False or Misleading Forward-Looking Statements	
Securities lawsuits with published judicial opinions retrieved from keyword searches on Lexis Nexis Legal with a filing date in 2006 or later	328
Eliminate:	
Lawsuits against firms missing necessary Compustat or CRSP data or for which the CIK could not be found	(78)
Lawsuits that do not involve forward-looking statements	(37)
Lawsuits in which the judge does not explicitly rule regarding the adequacy of the firm's cautionary language	(62)
Lawsuits missing Item 1A textual analysis data	(7)
Unique lawsuits examined in post-2005 multivariate regression analyses	144
Panel B: Sample of SEC Comment Letters Relating to Item 1A Risk Factor Disclosures	
SEC Comment Letters relating to Item 1A Risk Factor Disclosures from Audit Analytics Database (June 2017)	4,450
Eliminate:	
Firm-years for which Item 1A textual analysis variables could not be computed	(1,606)
Firm-years missing CRSP and Compustat data necessary for model variables	(1,178)
SEC Comment Letters relating to Item 1A Risk Factors examined in multivariate regression analyses	1,666

TABLE 3
Descriptive Statistics

Panel A: Descriptive Statistics for Firm-Years in Lawsuit Sample (N = 246)

Variable	Mean	Median	Q1	Q3	Std
LENGTH	0.0917	0.0915	-0.3757	0.4911	0.6612
NON_SPECIFICITY	-0.4687	-0.2724	-6.5086	5.8015	11.3285
STANDARDIZATION	0.5460	0.5518	0.4539	0.6421	0.1473
RET_COR	0.2151	0.2159	0.1247	0.2878	0.1222
RET_VOL	0.0343	0.0299	0.0223	0.0424	0.0179
BETA	1.2215	1.1775	0.9057	1.5446	0.4826
SIZE	7.4862	7.2264	6.3059	8.5297	1.8350
LEV	0.2299	0.1500	0.0028	0.3620	0.2478
BIG_N	0.8252	1.0000	1.0000	1.0000	0.3806
SKEW	0.3406	0.2648	-0.3176	0.9282	1.5483
TURNOVER	0.6031	0.4753	0.2665	0.8551	0.4609
ETR	0.1824	0.2888	0.0000	0.3702	0.3679
ROE	-0.0117	0.0430	-0.0326	0.0747	0.1903

Panel B: Descriptive Statistics for Firm-Years in SEC Comment Letter Sample (N = 33,725)

Variable	Mean	Median	Q1	Q3	Std
LENGTH	-0.0267	-0.0067	-0.4747	0.4231	0.7298
NON_SPECIFICITY	-0.3969	-0.1995	-6.2545	5.5401	10.4461
STANDARDIZATION	0.5493	0.5576	0.4725	0.6366	0.1334
RET_COR	0.2259	0.2173	0.1258	0.3203	0.1403
RET_VOL	0.0314	0.0265	0.0184	0.0385	0.0192
BETA	1.0426	1.0491	0.6625	1.4030	0.5758
SIZE	6.2914	6.2730	4.8966	7.6394	1.9964
LEV	0.2215	0.1552	0.0224	0.3346	0.3656
BIG_N	0.7145	1.0000	0.0000	1.0000	0.4517
SKEW	0.4469	0.2731	-0.1261	0.7907	1.5538
TURNOVER	0.5820	0.3822	0.1898	0.6859	1.0851
ETR	0.1171	0.2700	0.0000	0.3610	11.9401
ROE	-0.0739	0.0416	-0.0274	0.0717	24.5087

LENGTH is the natural logarithm of the total words in the 10-K Item 1A risk factor disclosure. *NON_SPECIFICITY* is the number of specific entities identified in the firm's risk factor section using the Stanford Named Entity Recognition tool, scaled by the total number of words in Item 1A and multiplied by -1,000. *LENGTH* and *NON_SPECIFICITY* are benchmarked against the

firm's closest industry peer, where closest industry peers are determined as discussed in Section 4.2. *STANDARDIZATION* is the number of words in boilerplate sentences in Item 1A scaled by the total number of words in Item 1A. Our measure of boilerplate sentences is based on commonly used trigrams in risk factor disclosures of other firms in the same 2-digit SIC industry, and is described in greater detail in Section 4.2 of the text. *RET_COR* is the mean correlation between firm's 24-month return ending at the current fiscal year end and the same window return for each other firm in its same 2-digit industry. *RET_VOL* is the standard deviation of daily returns for the year ending two days prior to the 10-K filing date. *BETA* is the firm's market beta measured using daily returns for the one year ending two trading days prior to the 10-K filing date. *SIZE* is the natural logarithm of the market value of equity. *LEV* is total liabilities scaled by total assets, measured as of the end of the fiscal year. *BIG_N* is an indicator variable equal to one if the firm is audited by one of the largest auditors (the Big N), and zero otherwise. *SKEWNEWSS* is measured as the negative coefficient of skewness for daily returns over the year ending two days prior to the 10-K filing date. *TURNOVER* is the mean number of shares traded during the fiscal year scaled by the number of common shares outstanding. *ETR* is the firm's effective tax rate, computed as tax expense scaled by pre-tax income. *ROE* is income before extraordinary items scaled by the market value of equity.

TABLE 4
Inadequate Risk Factor Disclosures and the Probability of Lawsuit Dismissal

	Dependent Variable	
	DISMISSAL	
	Parameter Estimate	
	Standard Error	
	Model 1	Model 2
RF_INADEQUATE_JG	-1.9931 *** 0.2520	-4.7041 *** 1.0297
<i>Risk control variables:</i>		
RET_COR	-0.3569 0.9688	-0.1630 2.6787
RET_VOL	3.0916 10.1762	-34.0933 22.2919
BETA	0.2138 0.2446	0.9220 0.7265
SIZE	0.2687 *** 0.0740	0.0783 0.1935
LEV	-0.7639 0.5340	-0.5730 1.5982
BIG_N	-0.0386 0.4045	0.1495 0.9497
SKEW	0.1535 * 0.0824	0.5262 *** 0.1822
TURNOVER	0.7542 *** 0.2889	1.9220 ** 0.7679
ETR	0.3259 0.3159	1.3903 1.0701
ROE	0.8553 0.7946	0.3501 2.8759
Industry fixed effects	Yes	Yes
Circuit fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Area under ROC curve	0.857	0.955
N	524	237

Table 4 displays the results of a logistic regression of the probability of a lawsuit dismissal on the indicator variable, RF_INADEQUATE_JG, and control variables. RF_INADEQUATE_JG is equal to one for lawsuits in which the judge explicitly rules the firm's risk factor disclosures are inadequate to avail the firm of the PSLRA's safe harbor, and zero if the judge explicitly rules the risk disclosures are adequate. All control variables are defined in the notes to Table 3. The sample includes all firm-years that have a 10-K filed within the class period of the lawsuit. The first column displays regression results for our entire sample of lawsuits from 1996 to 2015. The second column displays results for the post-2005 sample for which we have Item 1A textual data. *, **, and *** denote statistical significance at the .10, 0.05, and 0.01 significance levels, respectively (two-tailed tests).

TABLE 5
Judicial Assessments of Risk Factor Disclosure Inadequacy

	Dependent Variable			
	RF_INADEQUATE_JG			
	Parameter Estimate Standard Error			
	Model 1	Model 2	Model 3	Model
LENGTH	-0.8383 ** 0.3994			-1.2118 * 0.7298
NON_SPECIFICITY		-0.0464 ** 0.0211		-0.1527 *** 0.0586
STANDARDIZATION			-2.2763 1.905	-1.8447 3.5685
<i>Risk control variables:</i>				
RETURN_CORRELATION	-0.2979 2.2853	-0.317 2.179	-0.3517 2.0981	1.5849 4.6377
RETURN_VOLATILITY	17.4697 20.4671	9.1474 21.803	11.1378 20.6375	13.314 60.828
BETA	-0.3085 0.5523	-0.4518 0.5205	-0.3899 0.4927	-1.8279 * 1.0203
SIZE	-0.0591 0.1846	0.0231 0.2048	0.0643 0.1912	-0.2193 0.4025
LEVERAGE	1.9095 ** 0.9503	1.9987 ** 0.9618	1.5598 0.9625	2.4886 2.2107
BIG_N_AUDITOR	-0.6017 0.7480	-0.5387 0.6909	-0.5556 0.7196	-2.5447 * 1.4052
SKEWNESS	0.1924 0.1227	0.1956 0.134	0.1986 0.1307	0.1341 0.353
TURNOVER	-0.0508 0.4399	-0.2447 0.4308	-0.198 0.4479	0.0292 0.8855
ETR	-0.1413 0.7102	-0.1177 0.6013	-0.1746 0.5894	0.6334 1.0163
ROE	5.1202 *** 1.5186	6.1168 *** 1.8494	4.6931 *** 1.5389	9.5208 ** 4.5805
Industry fixed effects	Yes	Yes	Yes	Yes
Circuit fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Topic fixed effects	No	No	No	Yes
Area under ROC curve	0.884	0.887	0.880	0.967
N	246	246	246	246

Table 5 reports the results from a regression of the probability the sued firm's risk factor disclosures are assessed as inadequate by a judge for purposes safe harbor protection under the PSLRA. These assessments

are based on the published judicial opinion supporting the judicial decision whether to grant the defendant's motion to dismiss the securities lawsuit. Variable definitions are provided in the notes to Table 3. Industry fixed effects are measured at the Fama/French 12-industry classification level. Topic indicators are included based on twenty topics estimated using Latent Dirichlet Allocation over the entire corpus of 10-K Item 1A risk factor disclosures over our sample period. All continuous variables are winsorized at the 1st and 99th percentiles. Robust standard errors are used to assess statistical significance, which is denoted by *, **, and *** for statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

TABLE 6

Regulatory Assessments of Risk Factor Disclosure Inadequacy

Panel A: Full Sample

	Dependent Variable			
	RF_INADEQUATE_SEC			
	Parameter Estimate		Standard Error	
	Model 1	Model 2	Model 3	Model 4
LENGTH	0.039 0.0365			-0.0635 0.0419
NON_SPECIFICITY		-0.0041 * 0.00239		-0.0024 0.00267
STANDARDIZATION			-1.0054 *** 0.2279	-0.8002 *** 0.2767
<i>Risk control variables:</i>				
RET_COR	-0.1075 0.2635	-0.1075 0.2635	-0.0428 0.2641	0.0297 0.276
RET_VOL	15.1548 *** 2.3322	15.1449 *** 2.332	14.9053 *** 2.3405	12.2839 *** 2.4851
BETA	0.0459 0.0579	0.0487 0.0579	0.051 0.0579	0.0341 0.0606
SIZE	0.2121 *** 0.0206	0.2119 *** 0.0206	0.2017 *** 0.0207	0.1998 *** 0.023
LEV	-0.0197 0.1034	-0.0191 0.1031	-0.0035 0.1021	-0.0584 0.1197
BIG_N	-0.2448 *** 0.0722	-0.2418 *** 0.0722	-0.2429 *** 0.0722	-0.1607 ** 0.0755
SKEW	-0.0063 0.0183	-0.0067 0.0182	-0.0062 0.0183	-0.0066 0.0193
TURNOVER	0.004 0.0487	0.007 0.0485	0.0123 0.0484	-0.0003 0.0518
ETR	-0.0688 0.0667	-0.0702 0.0667	-0.0727 0.0664	-0.0624 0.068
ROE	-0.0691 0.0882	-0.0688 0.0882	-0.0631 0.0876	-0.052 0.0884
Industry fixed effects	Yes	Yes	Yes	Yes
Circuit fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Topic fixed effects	No	No	No	Yes
Area under ROC curve	0.713	0.713	0.714	0.741
N	33,407	33,407	33,407	33,407

Table 6 Panel A reports the results from a regression of the probability that a firm's risk factor disclosures are targeted by an SEC comment letter during the SEC's filing review process. Variable definitions are provided in

the notes to Table 3. Industry fixed effects are measured at the two-digit SIC classification level. Topic indicators are included based on two hundred topics estimated using Latent Dirichlet Allocation over the entire corpus of 10-K Item 1A risk factor disclosures over our sample period. All continuous variables are winsorized at the 1st and 99th percentiles. Robust standard errors are used to assess statistical significance, which is denoted by *, **, and *** for statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

TABLE 6 (cont'd)

Panel B: Sample of Firms Receiving Any SEC Comment Letter During the Year

	Dependent Variable			
	RF_INADEQUATE_JG			
	Parameter Estimate			
	Standard Error			
	Model 1	Model 2	Model 3	Model 4
LENGTH	0.0107 0.0387			-0.0654 0.0445
NON_SPECIFICITY		-0.0049 * 0.00253		-0.0044 0.00287
STANDARDIZATION			-0.9947 *** 0.2444	-0.7498 ** 0.2971
<i>Risk control variables:</i>				
RET_COR	-0.2114 0.274	-0.2146 0.274	-0.1506 0.2749	-0.0634 0.2895
RET_VOL	11.0152 *** 2.5918	10.9273 *** 2.5915	10.6329 *** 2.5957	8.7197 *** 2.8041
BETA	0.0045 0.0631	0.0061 0.0631	0.0114 0.0632	0.0005 0.0665
SIZE	0.0669 *** 0.0217	0.0667 *** 0.0217	0.0561 ** 0.0218	0.0539 ** 0.0245
LEV	-0.1773 0.1204	-0.1854 0.1206	-0.1752 0.12	-0.2031 0.1385
BIG_N	-0.1867 ** 0.0752	-0.1856 ** 0.075	-0.1841 ** 0.0751	-0.1139 0.0801
SKEW	0.0151 0.0218	0.015 0.0218	0.0156 0.0218	0.021 0.0232
TURNOVER	-0.004 0.0512	-0.0048 0.0511	-0.0002 0.0511	-0.0026 0.0553
ETR	-0.0261 0.0699	-0.0249 0.0698	-0.0281 0.0695	-0.0391 0.0716
ROE	-0.049 0.0938	-0.0468 0.0938	-0.0452 0.0933	-0.0475 0.0946
Industry fixed effects	Yes	Yes	Yes	Yes
Circuit fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Topic fixed effects	No	No	No	Yes
Area under ROC curve	0.686	0.687	0.688	0.719
N	14,582	14,582	14,582	14,582

Table 6 Panel B reports the results from re-estimating the regression from Panel A on only the subset of firms that received at least one SEC comment letter relating to any disclosure for the fiscal period. Variable

definitions are provided in the notes to Table 3. Industry fixed effects are measured at the two-digit SIC classification level. Topic indicators are included based on two hundred topics estimated using Latent Dirichlet Allocation over the entire corpus of 10-K Item 1A risk factor disclosures over our sample period. All continuous variables are winsorized at the 1st and 99th percentiles. Robust standard errors are used to assess statistical significance, which is denoted by *, **, and *** for statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

TABLE 6 (cont'd)

Panel C: Alternative Measure of SEC Comment Letter Relating to Item 1A Risk Factors

	Dependent Variable			
	RF_INADEQUATE_JG			
	Parameter Estimate			
	Standard Error			
	Model 1	Model 2	Model 3	Model 4
LENGTH	0.038 0.0703			-0.079 0.082
NON_SPECIFICITY		-0.0085 ** 0.00432		-0.0068 0.00496
STANDARDIZATION			-1.4583 *** 0.4312	-1.0633 ** 0.5191
<i>Risk control variables:</i>				
RET_COR	0.3784 0.5097	0.3793 0.5092	0.4786 0.5123	0.3962 0.5459
RET_VOL	14.2611 *** 4.2271	14.066 *** 4.2353	13.8912 ** 4.2503	11.9407 ** 4.6485
BETA	0.0784 0.1121	0.0833 0.112	0.0867 0.1118	0.0675 0.1195
SIZE	0.2243 *** 0.0361	0.2226 *** 0.0362	0.2079 *** 0.0363	0.2072 *** 0.0422
LEV	-0.2928 0.2081	-0.3015 0.208	-0.2765 0.2053	-0.5177 ** 0.2565
BIG_N	0.0706 0.1472	0.0769 0.1469	0.082 0.1466	0.1413 0.1569
SKEW	-0.0182 0.0383	-0.0187 0.0382	-0.0166 0.0384	-0.0279 0.0422
TURNOVER	0.0084 0.0943	0.0124 0.0934	0.0199 0.0925	0.0173 0.0969
ETR	-0.0359 0.1221	-0.0341 0.1224	-0.0406 0.1215	-0.0578 0.1229
ROE	0.2333 0.2157	0.2359 0.215	0.2436 0.2119	0.2398 0.224
Industry fixed effects	Yes	Yes	Yes	Yes
Circuit fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Topic fixed effects	No	No	No	Yes
Area under ROC curve	0.805	0.806	0.809	0.849
N	33,407	33,407	33,407	33,407

Table 6 Panel C reports the results from re-estimating the regression reported in Panel A after replacing the dependent variable with an alternative indicator of whether the firm received an SEC comment letter relating to Item 1A Risk Factors. The alternative dependent variable is equal to one if the firm received an SEC comment letter that contained the phrase “Item 1A” relating to that period’s filings, and zero otherwise. Variable definitions are provided in the notes to Table 3. Industry fixed effects are measured at the two-digit SIC classification level. Topic indicators are included based on two hundred topics estimated using Latent Dirichlet Allocation over the entire corpus of 10-K Item 1A risk factor disclosures over our sample period. All continuous variables are winsorized at the 1st and 99th percentiles. Robust standard errors are used to assess statistical significance, which is denoted by *, **, and *** for statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

TABLE 7
Robustness of Results to Alternative Measures of
Standardized Risk Factor Disclosure Language

	Dependent Variable		
	RF_INADEQUATE_SEC		
	Parameter Estimate		
	Standard Error		
	Model 1	Model 2	Model 3
LENGTH	-0.0604 0.0418	-0.0588 0.0418	-0.0592 0.0422
NON_SPECIFICITY	-0.0028 0.00264	-0.0029 0.00264	-0.004 0.00273
STANDARDIZATION	-1.2935 *** 0.4556	-1.8593 *** 0.5769	0.0177 0.3953
<i>Risk control variables:</i>			
RET_COR	0.0118 0.2761	0.0092 0.2759	-0.0126 0.2752
RET_VOL	12.3061 *** 2.4838	12.3617 *** 2.4839	12.4049 *** 2.4787
BETA	0.0354 0.0607	0.0372 0.0607	0.0345 0.0606
SIZE	0.2003 *** 0.023	0.2012 *** 0.0229	0.2029 *** 0.0229
LEV	-0.0637 0.1201	-0.0639 0.1197	-0.0644 0.1199
BIG_N	-0.1594 ** 0.0756	-0.1614 ** 0.0756	-0.1629 ** 0.0755
SKEW	-0.0065 0.0193	-0.0067 0.0193	-0.0067 0.0193
TURNOVER	0.0007 0.0518	0.0013 0.0518	-0.0061 0.0518
ETR	-0.062 0.0681	-0.0613 0.0681	-0.0634 0.0682
ROE	-0.0533 0.0883	-0.0552 0.0884	-0.0543 0.0885
Industry fixed effects	Yes	Yes	Yes
Circuit fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Topic fixed effects	Yes	Yes	Yes
Area under ROC curve	0.41	0.741	0.740
N	33,407	33,407	33,407

Table 7 displays results from running the regression analysis after remeasuring STANDARDIZATION based on alternative cutoffs for identifying boilerplate sentences. In Table 6, this variable defines boilerplate sentences as those with at least 10% of all trigrams in the sentence being boilerplate trigrams, where

boilerplate trigrams are those appearing in at least 10% of all Item 1A disclosures in the same 2-digit SIC industry on average, but not in more than 90% of all industry-peers' Item 1A disclosures. In Model 1 of Table 7, we change these cutoffs to 20%, 20%, and 80%, respectively. In Model 2, we change these cutoffs to 25%, 25%, and 75%, respectively. In Model 3, we remeasure this variable based on market-wide rather than industry-specific trigrams based on the original 10%, 10%, and 10% cutoffs. Industry fixed effects are measured at the two-digit SIC classification level. Topic indicators are included based on two hundred topics estimated using Latent Dirichlet Allocation over the entire corpus of 10-K Item 1A risk factor disclosures over our sample period. All continuous variables are winsorized at the 1st and 99th percentiles. Robust standard errors are used to assess statistical significance, which is denoted by *, **, and *** for statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.