# Risk Fact or Fiction: The Information Content of Risk Factor Disclosures

Maclean Gaulin

RICE | BUSINESS
Jones Graduate School of Business

Thesis: Doctor of Philosophy
Accounting, Jones Graduate School of Business
Rice University, Houston, Texas (March 2017)

RICE UNIVERSITY

# Risk Fact or Fiction: The Information Content of Risk Factor Disclosures

by

**Maclean Gaulin**

A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree

**Doctor of Philosophy**

Approved, Thesis Committee:

K. Ramesh, Chair
Herbert S. Autrey Professor of Accounting

Brian Akins
Assistant Professor of Accounting

Brian Rountree
Associate Professor of Accounting

Robin Sickles
Reginald Henry Hargrove Chair of Economics

Shiva Sivaramakrishnan
Henry Gardiner Symonds Professor of Accounting

Houston, Texas
March, 2017

ABSTRACT

**Risk Fact or Fiction: The Information Content of Risk Factor Disclosures**

by

**Maclean Gaulin**

Inconsistent with concerns of uninformative boilerplate or 'copy and paste' disclosure, I find that managers time their identification of new risk factors and removal of previously identified ones to align with the expected occurrence of future adverse outcomes. By using individual risk factors as the unit of disclosure, I am able to provide novel evidence that managers remove stale disclosures on a timely basis. After controlling for firm-specific heterogeneity, I find that the count of individual risk factors disclosed, rather than an aggregate word count, explains time-series variation in managerial disclosure decisions, consistent with the regulatory intent. To shed light on what shapes the disclosure equilibrium, I study the managerial response to demand 'shocks' from public and private enforcement actions. The results show that firms respond to investor demand in a manner consistent with the litigation shield hypothesis, and that this effect persists for multiple years. Consistent with the regulatory cost-benefit function, public enforcement does not result in a net increase in disclosed risk factors, but does evoke more definitive disclosures through more specific language and an increased use of numbers.

# Acknowledgments

I would like to sincerely thank everyone who helped me over the course of this adventure. I am honored to have completed my PhD dissertation under the mentoring and tutelage of Dr. K. Ramesh. I benefited immensely from Dr. Ramesh's insightful economic approach to research, and owe much of my success and productivity to his shared wisdom. I am also greatful for the support and guidance from my committee members Drs. Brian Akins, Brian Rountree, Robin Sickles, and Shiva Sivaramakrishnan.

I thank Dr. Thomas Hemmer for his advice and all-too-infrequent lunch meetings. I thank my co-authors Brian Akins and David De Angelis for their collaborative work and valuable experience, and Karen Nelson for her RA supervision. I also thank the research assistants from the Jones School, specifically Yamil Kaba, Sarah Diez, James Lenz, Dejan Suskavcevic, and Richard Swartz. I am grateful for financial support from the Jones Graduate School of Business.

I thank my friends and family for their emotional support and all the paper drafts they suffered through. I also thank my office mates Edwin Hu, Alberto Teguila, Yang Wang, Alex Chaudhry, Lizzy Berger, Morad Zekhnini, and Ioannis Spyridopoulos for the countless discussions about research, and for all the coffee, food, and scotch they helped enjoy.

Lastly, I thank Lacey Pyle for making work life.

# Contents

# Illustrations

# Tables

# — Chapter 1 —
# Introduction

Do managers disclose *risk factors* consistent with the regulatory requirement to warn of *future adverse outcomes*? This question is of renewed importance in light of the SEC's current *Disclosure Effectiveness Review* which seeks to modernize and 'improve the [disclosure] requirements for the benefit of investors and registrants.'[1] Surprisingly however, when addressing the informativeness of risk factors, academic literature has focused on neither *risk factors* (as the unit of disclosure) nor specific *future adverse outcomes*. Instead, the focus has been primarily on the relation between aggregate measure of risk disclosure (such as total word count, as opposed to individual risk factors) and market outcomes (e.g. beta, return volatility) in the cross section.[2] These studies provide important evidence that high-risk firms have longer risk disclosures, but they provide little evidence as to whether this association describes individual firm's disclosure behavior over time. Motivated in part by the SEC's review of disclosure regulations, this study seeks to address this gap through two channels: first, by testing whether managers disclose risk factors in a timely fashion to warn of future adverse outcomes, and second, by examining how the demands for risk factors from various stakeholders shape the managerial disclosure decision.

I develop two novel approaches to address the question of whether managers warn of future adverse outcomes. First, I develop a set of measures that track individual

---

[1]Concept release File No. S7-06-16, p.147: www.sec.gov/rules/concept/2016/33-10064.pdf

[2]For example Kravet and Muslu (2013), Campbell, Chen, Dhaliwal, Lu, and Steele (2014), Nelson and Pritchard (2016), Bao and Datta (2014), Chiu, Guan, and Kim (2015), Israelsen (2014), or Filzen, McBrayer, and Shannon (2016) among others.

risk factors over time, allowing for direct observation of the managerial decision to add, maintain, and remove distinct risk factors across reporting periods. Guided by regulatory language and its practitioner interpretation, I focus on the individual risk factor as the unit of measure to capture each risk factor as a self-contained, unique risk. Second, for similar reasons, I test the informativeness of my measures on distinct adverse outcomes rather than general market risk, which presents a more clean identification of the relation between the time-series evolution of risk factor disclosures and the potential outcomes about which they are intended to warn. Together, these two approaches allow me to test directly whether managers, on average, add new, retain existing, and remove obsolete risk factor disclosures in advance of specific adverse events, including realized negative reporting outcomes (net loss and net operating loss) as well as adverse real outcomes (significant sales decline, general business lawsuits, and securities litigations). Inconsistent with concerns of uninformative boilerplate or 'copy and paste' disclosure, my results suggest that managers time their identification of new risks and the removal of previously identified risks to align with the expected occurrence of future adverse outcomes. These results support the hypothesis that managers disclose risk factors in accordance with the SEC regulation that requires distinct risk factors addressing specific risks.

The regulation —and court precedence regarding the use of risk factors as legal protection— also requires risk factors be detailed and specific. To test whether managers use improved descriptive precision in their disclosures to warn of future adverse outcomes, I measure the 'definitiveness' of disclosures using three proxies: specificity (effectively proper nouns, as used in Hope, Hu, and Lu, 2016), numeric intensity (percentage of numbers used, as used in Bozanic, Dietrich, and Johnson, 2015), and verbosity (average words per risk factor, capturing the ontology of Bloomfield, 2008). I find that managers generally increase the level of definitiveness ahead of adverse outcomes, consistent with managers gaining and disclosing improved information as adverse events become more probable.

My findings regarding both risk factor evolution and definitiveness are robust to controlling for the firm's ex-ante risk and performance, which suggests that managers' disclosure of risk factors is *incrementally* informative about underlying risks. I then compare my measures of risk factor evolution and definitiveness to previous literature by contrasting them with an aggregate measure of risk disclosure. I find that across most specifications, managers convey information through the evolution of individual risk factors, rather than the total length of risk factor disclosure (by word count), especially after controlling for average between-firm differences (using firm fixed effects). The greater predictive ability of my measures over total word count extends to market outcomes as well, specifically abnormal returns (over both short and long horizons). My results also suggest that previous findings regarding the role that risk disclosures play in reducing information asymmetry is more likely to represent firm-specific attributes rather than improvements stemming from risk factor disclosure evolving over time. These findings support the claim in this paper that treating risk factors as distinct units may be a more faithful representation of the information being conveyed by managers.

Given this evidence that managers provide risk factor disclosures which are informative about future adverse outcomes, I then study the demand for these disclosures from two salient sources: investors and regulators. Specifically, I test whether and how managers provide risk factor disclosure in response to direct demand evidenced by investor and regulator enforcement actions: private enforcement through securities litigation, and public enforcement through SEC comment letter process. Consistent with predictions that disclosure of risk factors can act as a litigation shield (Skinner 1994, 1997; Robbins and Rothenberg, 2005), I find that firms respond to private enforcement by improving their risk disclosures via an increase in the number of new risk factors they identify. Also consistent with the large cost of securities litigation (Rose, 2008), I find that this penchant for increased disclosure persists for multiple years. Together, these results suggest that private investors' demand for risk factors is in line with the regulation requiring disclosure of a plurality of distinct risk factors, but suggests that investors are informed by *which* risks managers deem 'most significant,' rather than through descriptions of those risks. In response to public enforcement actions however, I find that firms do not increase the total number of risk factors they identify, but do significantly increase the definitiveness of their risk factors. This suggests that the demand for risk factor disclosures from regulators is consistent with their abstaining from dictating *which* risks managers identify, but requiring those already identified factors to be definitive.

My findings are important in light of the fact that risk factor disclosures have become an increasingly large portion of annual reports, growing from 11.3% to 16.1% of the length (by word count) over the past decade. This growth has raised questions among practitioners and academics alike as to whether the increase in disclosure is actually informative or merely boilerplate.[3] Previous literature addressing this concern has typically focused on the length of risk disclosure using a word count or "bag of words" approach. However, to address the topical issue of whether managers disclose risk factors per the regulation, these previously used measures may not fully capture the information managers attempt to convey. If managers adhere to the regulation and

[3]For example SEC Chair Mary Jo White (2013), Brown and Tucker (2011), Lukomnik (2016), Dyer, Lang, and Stice-Lawrence (2016).

convey the *most significant factors that make the offering speculative or risky*, they may be focusing on *which* risks they believe to be significant, rather than concerning themselves with the risk factor section in aggregate. Therefore, to the extent that the breadth of risks factors identified by managers improves information 'fineness' (e.g. Blackwell, 1951, D'Souza, Ramesh, and Shen, 2009, Cheng, Huang, and Li, 2016), the length of disclosure could be a byproduct of increased informativeness. Similarly, studying these disclosures in the cross section may fail to fully capture an individual firm's disclosure choices because it assumes an absolute relation between risk and disclosure length, rather than allowing for firm specific differences. In other words, while finding that generally riskier firms have longer disclosures is an important first step, my study adds to the literature by studying a potential underlying mechanism through which these disclosures convey information. By demonstrating that firms update their disclosures in an informative fashion over time, I hope to provide valuable evidence to the current regulatory initiative seeking to modernize the disclosure framework.

My measures are consistent with the principal outlined in Bloomfield, Nelson, and Soltes (2016), which suggests that researchers construct "variables [...] in a way that is most likely to capture the constructs specified in the theory they wish to test." Given that my paper studies the information content in mandatory disclosures, I construct my measures in a way that faithfully represents the regulatory intent of the disclosure requirements. The regulation requires firms to include a *discussion of the most significant factors that make the offering speculative or risky* and that the factors be 'organized logically,' with each risk factor summarized by a caption. Therefore, I focus on the individual risk factor itself as the unit of disclosure. By narrowing the unit of measure to individual risk factors, I observe specific risks being addressed through time, and capture managers' decision to add, remove, or continue to disclose about individual risks, thereby signaling whether they consider a given risk factor to be *significant*. By testing whether disclosing these individual risks predicts adverse outcomes, I find that the manager's decision to signal that a specific risk is *significant* is timely and informative.

The prior literature, which examines the informativeness of risk disclosures by largely focusing on the aggregate level, has drawn mixed conclusions. Some studies of voluntary risk factor disclosures find that the length of risk disclosure reflects both ex-ante litigation risk and ex-post market risk measures (e.g. Kravet and Muslu, 2013). Other findings suggest that the risk disclosure correlates with subsequent systematic risk only for firms facing high ex-ante litigation risk (Nelson and Pritchard, 2016). In 2005, the SEC required all firms to start including a risk factor section in their annual and quarterly reports. Studies regarding these regulated risk factors have also reach varied conclusions. Campbell et al. (2014) find that risk disclosures reflect subsequent market measures of risk, mirroring the results of Kravet and Muslu (2013). Other studies call into question the efficacy of regulation on the value of risk factors. Nelson and Pritchard (2016) and Beatty, Cheng, and Zhang (2015) both find that regulatory influence (mandatory disclosure and enforcement, respectively) decrease the association between risk disclosures and measures of the underlying risk. My study complements this literature by studying risk factor disclosures at the firm level, and suggests a potential reason for the disagreement may be that the information in risk factors is not reflected primarily through aggregate measures, but rather through the time-series evolution of individual risk factors. Additionally, by finding that the regulators and investors demand different dimensions of information from the disclosures, my results may suggest a reason for previous findings that regulatory impact weakens the association between disclosure length and underlying risk.

My study contributes to the literature in two ways. First, my paper contributes to the growing literature on the informativeness of risk factor disclosures. While previous work by Nelson and Pritchard (2016) and Beatty et al. (2015) conclude that regulatory intervention from the SEC leads to less informative disclosures, my findings suggest an alternative perspective. Specifically, I find that the SEC elicits more detailed disclosures, both in specificity and in length of descriptions, for each risk factor. While one outcome of this regulatory focus may be that the length of disclosure is less correlated with underlying economic activity, my findings suggest that regulatory intervention may lead to more definitive or quantitative disclosures. Additionally, my results point to an interesting role that risk factor disclosures may play in the information milieu. Rogers and Van Buskirk (2009) find that firms provide less voluntary disclosure in response to securities litigation, and conclude that managers view *voluntary* disclosure as a legal liability.[4] My finding that risk factor disclosures are *expanded* after litigation suggests that these *mandated* disclosures may serve as a mitigating factor for litigation risk, potentially substituting for voluntary information when the cost of the latter increases.

Second, this study contributes to the developing literature on information overload and boilerplate disclosure. I demonstrate that managers update their disclosures in a timely fashion, suggesting that they are not merely 'copy and pasting.' Instead, my results suggest that firms are informatively choosing when to add, and potentially more importantly, when to remove information. In contrast to previous studies of

[4] While Naughton, Rusticus, Wang, and Yeung (2015) suggest this result may not be driven by lower litigation costs, my findings of increased *mandatory* risk factor disclosures are still in contrast to the reduction of *voluntary* disclosure.

boilerplate disclosure, which take the position that increasing similarity over time connotes uninformative disclosure (e.g. Brown and Tucker, 2011; Nelson and Pritchard, 2016; Dyer et al., 2016), my finding that the removal of disclosures is informative suggests that persistent disclosures may not be inherently uninformative. I also show that managers supply disclosure to meet information demand differentially based on the source of the demand. This is in line with the findings of Bird and Karolyi (2016) and Boone and White (2015) regarding the provision of voluntary information to meet institutional investor demand. Together my results suggest caution in interpreting the increasing length of disclosure as a sufficient signal of information overload. This study demonstrates that in some cases a more nuanced measure of information may be warranted, potentially one derived from the regulation or intent of managers.

# — Chapter 2 —
# Risk Factor Disclosure Information Content

This dissertation studies the information content of mandatory risk factor disclosures filed in annual reports by public firms in the United States. To shed light on the dimensions and timliness of the information in these disclosures, I present evidence on both the supply of, and the demand for risk factor disclosures. In this chapter, to study the supply of risk factor disclosures I investigate whether managers forecast adverse outcomes through their choice of disclosure timing and content. My findings are consistent with the hypothesis that managers compound their expectations of future cash flows and other real adverse events into their disclosure decisions.

## 2.1 Hypothesis Development

Previous literature has focused primarily on the decision value of risk factor disclosures by studying the capital market effects of risk factor disclosures in the cross section. However the evidence in this extant research on the informativeness of risk factor disclosures is mixed.[1] Earlier studies on risk disclosures addressed the voluntary disclosure of risk factors. Using the entire 10-K report, Kravet and Muslu (2013) measure risk disclosure as the change in the number of sentences with at least one risk-related word. They find that industry-level risk-related disclosures, rather than their idiosyncratic variations, are correlated with stock-based risk measures and analyst perceptions of risk. Using a hand collected sample of disclosures from 293 firms, Nelson

[1] Again 'informativeness' in these studies is typically defined as 'being decision useful,' and is often demonstrated by correlating risk disclosure with future returns or other market risk measures.

and Pritchard (2016) finds a similar association between the 'unexpected' length of risk disclosure and market risk measures, but only for firms facing high ex-ante litigation risk. They conclude that only high-risk firms have incentives to disclose meaningful warnings of future risk. Despite disagreeing on whether risk disclosures are informative generally or just conditionally so in some cross-sections, both studies do highlight that the length of these disclosures has increased significantly over time.

Motivated in part by this increase in voluntary disclosure, in 2005 the SEC mandated that firms include a risk factor section in their annual and quarterly reports.[2] Studies regarding this mandatory disclosure are mixed on the efficacy of the regulation and the informativeness of the disclosures themselves. In the high litigation risk firms, Nelson and Pritchard (2016) find that the association between the length of risk disclosure and ex-ante market risk vanishes after the regulation, and this weakened relation is seen in the association with ex-post market risk as well. They conclude that litigation risk still largely drives the informativeness of these disclosures, and the mandate did not fully substitute for the value of previous voluntary risk disclosures. Campbell et al. (2014) focus on the total number of words in the item 1A section (post SEC mandate) as well as author defined sub-categories of risk types (idiosyncratic, systematic, tax, financial, and legal risk). Unlike Nelson and Pritchard (2016), they find significant associations between risk factor disclosures and both ex-ante and ex-post measures of market risk (as well as ex-post information asymmetry).

While the disparate evidence in prior research could be driven by differences in research method choices (for example section 2.3.3 and evidenced in Table 2.12), I focus on how the extant literature has measured risk disclosures themselves. Prior

[2]Securities Offering Reform, SEC File No. S7-38-04 p. 259: www.sec.gov/rules/final/33-8591.pdf

research has largely treated the entire risk disclosure as atomic, using the 'size' of the risk factor section as the disclosure proxy. For example, past studies have measured risk disclosure using: whether any disclosure is present (Filzen, 2015), change in the number of sentences containing risk words (Kravet and Muslu, 2013), total word count (Campbell et al., 2014), absolute value of the change in number of words (Brown, Tian, and Tucker, 2015), a key-word proportion based on risk categories chosen by the researchers[3] (Campbell et al., 2014), a topic-based measure from a labeled classification model (Huang and Li, 2008), a topic-based measure from a computational word-clustering (Bao and Datta, 2014), and a 'similarity' score between disclosures (Nelson and Pritchard, 2016; Brown et al., 2015). Kravet and Muslu argue for their sentence level focus, claiming that "by using sentences instead of words, we avoid multiple counting of the same risk-related information" (2013, p. 1095). This last point is notable because it highlights that the proxy should capture the information purportedly being measured.

This issue of matching proxy to construct is what Bloomfield et al. (2016) refer to as *distillation*. A key input to this distillation process is the construct the researcher is attempting to capture. Research in other areas of accounting have solved this problem by utilizing specialized and sometimes novel proxies developed in other literature. For example, Li (2008) studies whether the complexity of information affects its consumption, and uses the Gunning Fog measure of reading ease developed in linguistics to proxy for information complexity. Loughran and McDonald (2011) address the issue of measuring tone in narrative disclosures, but argue that a customized proxy is required due to the specialized nature of business communication. Previous

[3] Campbell et al. (2014) compile their word lists from non-overlapping subsets of words derived from the largest weightings of an LDA model.

research on risk factors has primarily focused on aggregate *risk disclosure*, and selected proxies accordingly as outlined above. This study instead focuses on risk factors as defined by the regulation for two reasons: doing so provides a concrete foundation from which to derive measures, and because the regulation is consistent with practitioner guidance and court precedence (discussed below).

The requirement to disclose risk factors is set forth in Item 503(c) of Regulation S-K:

> **§229.503 (Item 503) (c) Risk factors**. Where appropriate, provide under the caption "Risk Factors" a discussion of the most significant factors that make the offering speculative or risky. This discussion must be concise and organized logically. Do not present risks that could apply to any issuer or any offering. Explain how the risk affects the issuer or the securities being offered. Set forth each risk factor under a subcaption that adequately describes the risk.

The regulation requires that managers disclose 'the most significant factors,' 'organized logically' with each risk factor summarized by a caption. This suggests that risk factors should be treated as a) distinct, and b) a signal of managers' belief about specific adverse outcomes. The regulation also requires a concise explanation of the risk that is specific to the firm. Because of the presence of the SEC review process (discussed further below), firms must adhere to the verifiable portion of these requirements, or justify otherwise to the SEC (Bozanic et al., 2015). However, investor demand for information can also motivate managers to disclose per the regulation.

Investors can potentially influence managers to disclose risk factors consistent with the regulation because by doing so, managers potentially can protect against securities litigation. This 'litigation shield' hypothesis (Skinner 1994, 1997) posits that warning investors of negative outcomes in advance can reduce the expected cost of class-action securities litigation. By disclosing in advance, managers can potentially reduce the stock impact of the 'corrective disclosure' and shorten the class period.[4] Additionally, by disclosing information in a timely manner, firms can potentially refute a plaintiff's claim that the firm did not adequately provide investors with information (Robbins and Rothenberg, 2005). However, to invoke the litigation shield, firms must disclose specific and detailed risk factors (consistent with the regulation). This is because the value of risk factors as legal protection stems from the 'bespeaks caution' doctrine and the subsequent Private Securities Litigation Reform Act (PSLRA). The PSLRA, enacted in 1995, provided a 'safe harbor' against legal liability when firms disclose forward-looking information, as long as they include cautionary language regarding the uncertainty of the forecasts.

Anecdotally, firms do use risk factor disclosures to gain legal protection,[5] when the disclosures adhere to quality standards set forth by court precedent:

> Cautionary language must be extensive and specific. A vague or blanket (boilerplate) disclaimer which merely warns the reader that the investment has risks will ordinarily be inadequate to prevent misinformation. To suffice, the cautionary statements must be substantive and tailored to the specific future projections, estimates or opinions in the prospectus which the plaintiffs challenge. (Inst. Investors Group v. Avaya, Inc. 564 F.3d 242, 256; 3d Cir. 2009)

In Slayton vs American Express (604 F.3d 758, 762; 2d Cir. 2010) the court concluded that the above standards were not met, stating: "the defendants' [risk factor disclosure] verges on the mere boilerplate [...] Our conclusion is bolstered by the fact that the defendants' cautionary language remained the same even while the problem changed." This legal precedent is consistent with the regulatory requirement that risk factor

[4]To bring a securities litigation, plaintiffs must define both an initial misleading disclosure and a subsequent corrective disclosure (and demonstrate scienter among other factors).

[5]See e.g. In re Convergent Technologies Security Litigation, 948 F.2d 507,515(9th Cir. 1991), In re Worlds of Wonder Securities Litigation, 35 F.3d 1407 (9th Cir. 1994).

disclosures be detailed and address specific outcomes. The latter court opinion also suggests that to gain legal protection, firms must update their risk factor disclosures in a timely fashion to reflect new information.

Together, the demand from investors and regulatory enforcement suggests that managers have strong incentives to disclose risk factors consistent with the regulation. As these disclosures potentially afford protection from both litigation and regulatory actions, firms may be incentivized to 'overload' stakeholders and disclose of every conceivable risk factor. This information overload concern is voiced by both legal opinions and regulators. Justice Marshall raised this concern, cautioning that "management's fear of exposing itself to substantial liability may cause it simply to bury the shareholders in an avalanche of trivial information – a result that is hardly conducive to informed decision-making"[6] SEC Chair Mary Jo White (2013) echoed this concern, stating "I am raising the question here and internally at the SEC as to whether investors need and are optimally served by the detailed and lengthy disclosures about all of the topics that companies currently provide in the reports they are required to prepare and file with us."" As these suggest, it is unclear ex-ante whether disclosures will convey an informative signal, especially one with a subjective materiality threshold such as the risk factor disclosures. Thus in the absence of disclosures costs, manager's choice of risk factor disclosures may convey no information about the underlying probabilities of the adverse outcomes they address, resulting in the 'information overload' Chair White was concerned of.

To test whether managers do disclose risk factors consistent with the regulation and in a timely fashion, I study two distinct dimensions through which managers potentially convey information. The first dimension focuses on the individual risk

[6]TSC Industries, Inc. v. Northway, 426 U.S. 438, 448-449 (1976))

factor as the unit of measure, and captures the evolution of each risk factor over time. To do so I label each risk factor a firm discloses in their annual report as belonging to one of three groups: new risk factors that were not present in the previous year, old risk factors that were present in the previous year and persist in the current year, and removed risk factors, that were present in the previous year but no longer included in the current year. This allows me to observe how many new risk factors firms are identifying, as well as how many firms are removing. The latter is indicative of risks managers deem no longer applicable to the firm, and this ability to capture managers disclosing information through the *removal* of risk factors is unique to my approach. The second dimension I study captures the level of specific detail firms disclose by measuring the specificity and succinctness. I measure specificity with two proxies that capture the definitiveness, or reference to specific entities and quantities. I capture succinctness by measuring the average number of words used in each risk factor.[7]

As discussed above, the regulation suggests that risk factor disclosures be treated as distinct, and as signals of managers' belief about specific adverse outcomes. Rather than test whether risk factor disclosures reflect these specific adverse outcomes, previous literature primarily studies the impact of risk disclosures on general stock market measures to test whether risk factors provide information on firm risk in general.[8] However, a manager's risk factor discloses are not necessarily conditioned on the market's expectations. In other words, testing whether risk disclosures are associated with market measures demonstrates that managers provide *novel* information, but

[7] I omit stop words from these counts such as *the*, *and*, *is*, etc. I use the default list of English stop words from the NLTK Python library.

[8] See, for example, Campbell et al. (2014), Nelson and Pritchard (2016), Huang and Li (2008), Hope et al. (2016), Kravet and Muslu (2013), Chiu et al. (2015). One very recent contrary example is Campbell, Cecchini, Cianci, Ehinger, and Werner (2016), who study the relation between tax-related risk disclosures and taxes paid. However, they still focus on aggregate disclosure (through word count) and the relationship between firms, omitting firm level fixed effects.

does not address whether their disclosures warn of specific events, which is the intent of my study.

Instead of focusing on general risk, I employ a novel test of informativeness by studying whether risk factor disclosures warn of adverse outcomes themselves. By studying whether risk factor disclosures forecast discrete outcomes which a manager can plausibly forecast, my tests provide potentially stronger inferences about the information contained in risk factor disclosures because the discreteness of the outcomes being tested mitigates the issue of reverse causality. This leads to my first set of hypotheses that managers provide information in risk factor disclosures to signal future adverse outcomes. I predict that they do so in two ways: updating *which* risks they identify, and updating the definitiveness with which they disclose the risks. If managers identify risks they deem 'most significant,' and intend to fulfill both regulatory requirement and investor demand, then they will add a risk factor outlining this risk. As the probability of an adverse outcome increases, managers will plausibly gain more information about the event, as well as have increased incentives to disclose the risk and mitigate disclosure costs associated with not warning of the event. Consistent with this, Heinle and Smith (2015) argue that firms receiving bad information will expend resources to gather more information. If managers disclose consistent with their private information, this will translate into managers forecasting, or forewarning of these adverse outcomes. Conversely, when the probability of an adverse outcome decreases, if managers are continually updating their risk factors to reflect only those 'most significant factors,' then they will remove factors when adverse outcomes are less likely.

**H1:** Firms disclose more and remove fewer risk factors in advance of adverse economic outcomes.

The second dimension through which I predict that managers signal future adverse outcomes is the level of detail provided about these events. A recent practitioner review of risk factor disclosures from 50 of the largest 10-K filers demonstrates a wide variety in the level of specificity provided by managers.[9] Hope et al. (2016) study this heterogeneity in specificity across firms, and find that more specific risk factor disclosures are associated with stronger market and analyst responses, suggesting that writing more definitive risk factor disclosures conveys more information to investors. More definitive risk factors being more informative to investors is also consistent with court precedence requiring risk factors to be detailed in order to benefit from the legal protection of the 'bespeaks caution doctrine.'[10] Additionally, as discussed above, managers may obtain more precise information about adverse outcomes as they become more probable. If managers are disclosing risk factors per the regulation and consistent with the legal requirements, together these forces will result in an increase in definitiveness of the risk factor disclosures in advance of adverse outcomes.

> **H2:** Firms provide more definitive risk factors in advance of adverse economic outcomes.

## 2.2 Sample Construction and Data Collection

Risk Factor disclosures have been required in annual and quarterly disclosures under Item 1A since the SEC regulation took effect in 2005. My sample starts with these earliest filings, and extends through fiscal year 2015. I start gathering the sample from the Compustat database between fiscal year end 2005 and 2015 (112,402 firm years),

[9]Investor Responsibility Research Center Institute 2016 report (Lukomnik, 2016)

[10]e.g. In re Worlds of Wonder Securities Litigation, 35 F.3d 1407 (9th Cir. 1994), Inst. Investors Group v. Avaya, Inc., 564 F.3d 242, 256 (3d Cir. 2009)

removing those firms with missing historical Central Index Key (CIK) identifiers (22,715 missing). I further require firm-years have non-missing total assets (10,654 missing) and a valid match to the CRSP securities database (27,784 missing), resulting in a sample of 51,249 firm years. With the 9,632 unique CIKs from this sample of firms, I extract the list of all matching 10-K and 10-KSB filings from the EDGAR index available from the SEC. This results in a sample of 67,648 annual reports filed for fiscal year 2005 or later, from which I then attempt to extract the Item 1A Risk Factor section.[11]

I extract the risk factor section following the methodology described in Campbell et al. (2014). Appendix B describes this process in detail, resulting in a final sample of 31,549 firm years with non-missing total assets, previous year returns from CRSP, and non-missing risk factors.[12] The risk factors are extracted using contextual clues provided in the HTML document, such as bold, italic, underline, and paragraph demarcations. I require that each risk factor be defined by an accentuated heading (bold, underlined, or italic) which is isolated on its own line or located at the beginning of a paragraph. The practice of captioning risk factors is pervasive because it is required in the wording of the regulation, and is enforced by SEC reviewers through comment letters. I use these extracted risk factor heading to define a unique risk factor.

The methodology proposed in this paper tracks risk factor evolution over time, using three proxies to do so: the total number of risk factors, the number of new risk

[11] The number of EDGAR filings I search is larger than the number of CRSP-Compustat merged firm years because I do not filter out firm-years with missing data. I do this to avoid unnecessarily dropping observations in the year-over-year comparison of risk factor statements, for example when a firm has a valid annual report in the previous year but no listed CRSP identifier.

[12] The sample becomes 26,547 firm years when previous year non-missing risk factor disclosure is required for the regressions.

factors, and the number of dropped risk factors which were included in the previous filing but do not appear in the current filing. To determine whether a risk factor was included in a previous year, I compare the risk factor heading from the text with all the previous year's set of risk factors without replacement. To compare the text, I use the Ratcliff and Metzener Gestalt Pattern Match algorithm (RMGPM).[13] The RMGPM is a flexible string match algorithm that allows for a parameterized amount of flexibility for determining if two strings are the same. It is the algorithm that is used in the Python language to compare source code for differences.

To match risk factors across years, I start with the set of risk factors comprising $N$ text strings in year $t$, sorted from longest to shortest number of letters. I then iterate through the $N$ strings, comparing each to the set of $M$ risk factors from year $t$-$1$, also sorted by decreasing length. For each $n$ of the $N$ risk factors, I search for an exact match, and failing that, I iteratively reduce the parameter restricting the 'exactness' of the match. If no match is found after allowing for a minimum of 50% character level match, the risk factor is counted as 'new.' When a match is found, the matching string from the set of $t$-$1$ risk factors is also removed. After searching for all $N$ risk factors from year $t$, the remaining risk factors from year $t$-$1$ are counted as 'dropped.'

I measure the definitiveness of risk factors using three proxies: specificity, numeric intensity, and average words per risk factor. I calculate the specificity of risk factor disclosures using the Stanford Named Entity Recognition (NER) algorithm, similar to that employed in Hope et al. (2016). The NER algorithm extracts specific entities (proper nouns), and is intended to capture whether disclosure uses general language or names a specific entity or location (e.g. our competitor vs. Microsoft, our supplier vs. Foxconn). I use three of the seven entity categories from the pre-trained classifier

---

[13]Included in the *difflib* package in the Python standard library.

provided with the algorithm: location, person, and organization.[14] Unlike Hope et al. (2016), I omit the remaining four categories to avoid overlap with the other proxy of definitiveness, numeric intensity. Consistent with Brown and Tucker (2011), I measure numeric intensity as the percentage of words that are numeric. The numeric intensity is intended to capture the level of detail provided in disclosures, through instances of numbers, currencies, percentages, or dates. These two measures are included as changes to capture the aggregate difference in definitiveness of risk factors, including new risk factors but also changes in the definitiveness of the existing risk factor disclosures. To capture the brevity of the descriptions used in the risk factors, I include the change in average words per risk factor. This proxy is consistent with what Bloomfield (2008) describes as 'ontology,' whereby negative outcomes are difficult to describe and require longer, more detailed explanations. Additionally, as adverse events become more likely, managers potentially gather more information (Heinle and Smith, 2015) which could translate into more information available to disclose.

To test whether managers warn of adverse outcomes directly, I focus on four outcomes: negative net income, negative operating income, sales decline (greater than 10% of previous year's sales),[15] and business, or non-securities, litigation. These outcomes are chosen to reflect significant adverse events which are ubiquitously negative.[16] The business lawsuit data are from the CapitalIQ Key Developments database. I define a business lawsuit as an event in a fiscal year with event code 25,

---

[14]Using the pre-trained, seven-class classifier (Location, Person, Organization, Money, Percent, Date, Time from *english.muc.7class.distsim.crf.ser.gz*) downloaded from nlp.stanford.edu/software/CRF-NER.shtml#Models

[15]The results are generally robust to using a 5% threshold, with some weakening in the significance of the definitiveness results.

[16]To the extent that these outcomes may not be conditionally negative signals, I later perform various robustness tests on other outcomes to provide evidence supporting the conclusion that risk factor disclosures are informative of future real economic outcomes.

corresponding to Litigation events. I form the variable *Lawsuit Intensity* as the log of the number of lawsuits that occur in a given fiscal year (plus one). I gather SEC comment letters from the Audit Analytics Comment Letter database, and classify the comment letter as relating to risk factors if the risk factor column in the database is non-empty (consistent with Brown et al., 2015). I gather securities litigation events from the Stanford Law School's Securities Class Action Clearinghouse.[17] The data gathered includes class start and end dates, as well as filing date and outcome (settled, dismissed, or ongoing). A firm is said to have a securities litigation event if a filing date occurs within the fiscal year. To differentiate them in the tables, I refer to securities litigation as *Litigation* or *Securities Litigation*, and business lawsuits as *Lawsuit*.

Table 2.1 presents the descriptive statistics for the final sample. Table 2.2 presents the pairwise contemporary correlations of the main variables used in the regressions. Of note is the significant negative correlation between the specificity and numeric intensity of risk factors and the FOG score (-0.05 and -0.16 respectively). This suggests that these measures of definitiveness are not provided concomitantly with obfuscatory, more complex language. This is consistent with managers providing definitive disclosures to inform stakeholders.

Figure 2.1 shows the average number of new and dropped risk factors over time. The significant increase in new risk factors in fiscal year 2008 corresponds to the financial crisis.

The average firm in my sample has 29.5 risk factors, and adds 3.7 new risks and removes 2.5 obsolete risks per year. I find that younger firms (below the median firm age of 13 years) disclose 34.4 risk factors on average, and identify 4.4 new factors per year and remove 3.2 factors. Older firms (above median age) disclose significantly fewer

---

[17]Scraped from securities.stanford.edu/filings.html using code from github.com/gaulinmp/lit_scrape.

**Table 2.1**: Summary Stats

Table 2.1 reports the summary statistics for the variables used in the regressions as defined in Appendix **??**.

| | Mean | Std. Dev | Min | 25% | 50% | 75% | Max | N |
|---|---|---|---|---|---|---|---|---|
| $Log(Assets)_t$ | 6.70 | 2.01 | 2.12 | 5.34 | 6.74 | 8.05 | 11.56 | 26,547 |
| $Log(Market\ Equity)_t$ | 6.33 | 1.96 | 1.84 | 4.99 | 6.35 | 7.70 | 10.56 | 26,537 |
| $Book-to-Market_t$ | 0.73 | 0.90 | -0.71 | 0.31 | 0.58 | 0.96 | 4.60 | 26,518 |
| $Net\ Income/AT_t$ | -0.02 | 0.32 | -1.09 | -0.02 | 0.03 | 0.07 | 0.40 | 26,536 |
| $Opr.\ Inc./AT_t$ | 0.02 | 0.28 | -1.01 | 0.01 | 0.06 | 0.12 | 0.48 | 25,276 |
| $Sales/AT_t$ | 0.90 | 0.85 | 0.00 | 0.26 | 0.70 | 1.27 | 4.04 | 26,536 |
| $Leverage_t$ | 0.23 | 0.25 | 0.00 | 0.02 | 0.17 | 0.35 | 0.96 | 26,547 |
| $Tangibility_t$ | 0.21 | 0.24 | 0.00 | 0.03 | 0.12 | 0.32 | 0.89 | 25,479 |
| $Turnover_t$ | 0.87 | 0.78 | 0.04 | 0.33 | 0.66 | 1.13 | 4.15 | 26,538 |
| $Beta_{t-1}$ | 1.04 | 0.54 | -0.10 | 0.69 | 1.05 | 1.40 | 2.31 | 26,547 |
| $Excess\ Ret._{t-1}$ | 0.02 | 0.48 | -0.84 | -0.24 | -0.04 | 0.19 | 1.92 | 26,547 |
| $Ex.\ Ret.\ Std_{t-1}$ | 0.03 | 0.02 | 0.01 | 0.02 | 0.02 | 0.04 | 0.10 | 26,547 |
| $Min.\ Excess\ Ret._{t-1}$ | -0.12 | 0.09 | -0.47 | -0.15 | -0.09 | -0.06 | -0.02 | 26,547 |
| $Ex.\ Ret.\ Skew_{t-1}$ | 0.44 | 1.45 | -3.91 | -0.16 | 0.34 | 0.95 | 6.18 | 26,545 |
| $CAR_{+3\ days}$ | -0.00 | 0.06 | -0.21 | -0.02 | -0.00 | 0.02 | 0.21 | 26,375 |
| $CAR_{+3\ months}$ | 0.00 | 0.25 | -0.72 | -0.12 | -0.00 | 0.11 | 0.86 | 26,382 |
| $Bid-Ask\ Spread_{t+1}$ | 0.75 | 1.57 | 0.02 | 0.07 | 0.16 | 0.58 | 8.40 | 26,386 |
| **Indicator and Negative Outome Variables** | | | | | | | | |
| $Negative\ NI_{t+1}$ | 0.31 | 0.46 | 0 | 0 | 0 | 1 | 1 | 22,026 |
| $Negative\ OperatingInc._{t+1}$ | 0.23 | 0.42 | 0 | 0 | 0 | 0 | 1 | 20,970 |
| $Sales\ Decline_{t+1}$ | 0.21 | 0.41 | 0 | 0 | 0 | 0 | 1 | 22,025 |
| $Sec.\ Litigation_{t-1}$ | 0.02 | 0.14 | 0 | 0 | 0 | 0 | 1 | 26,546 |
| $Lawsuit\ Intensity_{t-1}$ | 0.19 | 0.44 | 0 | 0 | 0 | 0 | 2.08 | 26,547 |
| $RF\ Comment_{t-1}$ | 0.07 | 0.25 | 0 | 0 | 0 | 0 | 1 | 26,224 |
| $Comment\ (Any)_{t-1}$ | 0.53 | 0.50 | 0 | 0 | 1 | 1 | 1 | 26,223 |
| **Textual Variables** | | | | | | | | |
| $\#\ Risk\ Factors_t$ | 29.52 | 14.05 | 6 | 19 | 27 | 37 | 69 | 26,547 |
| $\#\ New\ RF_t$ | 3.67 | 5.03 | 0 | 1 | 2 | 5 | 25 | 26,547 |
| $\#\ Dropped\ RF_t$ | 2.45 | 3.73 | 0 | 0 | 1 | 3 | 21 | 26,547 |
| $\Delta\#\ RF_t$ | 1.22 | 4.56 | -11 | 0 | 1 | 2 | 18 | 26,547 |
| $\#\ of\ Words_t$ | 2731.17 | 1707.36 | 396 | 1478 | 2364 | 3570 | 8847 | 26,547 |
| $\#\ of\ Sentences_t$ | 257.39 | 150.96 | 43 | 147 | 228 | 335 | 729 | 26,547 |
| $\#\ of\ Specific\ Words_t$ | 210.10 | 182.34 | 12 | 89 | 162 | 273 | 887 | 26,547 |
| $\#\ of\ Numerics_t$ | 58.24 | 47.43 | 3 | 25 | 46 | 77 | 225 | 26,547 |
| $\#\ of\ Words/RF_t$ | 88.95 | 25.77 | 38.71 | 71.30 | 86.32 | 102.93 | 176.54 | 26,547 |
| $Specificity_t$ | 5.45 | 3.25 | 0.60 | 3.21 | 4.76 | 6.90 | 17.88 | 26,547 |
| $Numeric\ Intensity_t$ | 2.11 | 1.05 | 0.49 | 1.37 | 1.92 | 2.63 | 5.86 | 26,547 |
| $FOG\ Index_t$ | 20.68 | 1.23 | 17.64 | 19.91 | 20.69 | 21.45 | 23.89 | 26,547 |

risk factors, only 25.8, and identify new factors and remove old factors at significantly lower rates as well (3.4 and 2.2, respectively). This potentially suggests life cycle effects in risk factor disclosures, with older, more established, and less volatile firms disclosing fewer risks and experiencing less risk 'turnover.' However, when looking at

**Table 2.2**: Summary Stats: Correlation

Table 2.2 reports the pairwise correlations for the variables used in the regressions as defined in Appendix **??**. Coefficients in bold are significant at the 5% level.

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| (1) $Negative\ NI_{t+1}$ | 1 | | | | | | | | |
| (2) $Sales\ Decline_{t+1}$ | **0.24** | 1 | | | | | | | |
| (3) $Sec.\ Litigation_{t-1}$ | **0.04** | 0.01 | 1 | | | | | | |
| (4) $Lawsuit\ Intensity_{t-1}$ | 0.00 | **0.02** | **0.11** | 1 | | | | | |
| (5) $RF\ Comment_{t-1}$ | **-0.01** | **0.02** | **0.03** | **0.07** | 1 | | | | |
| (6) $Comment\ (Any)_{t-1}$ | -0.01 | 0.00 | **0.04** | **0.10** | **0.25** | 1 | | | |
| (7) $Log(Assets)_t$ | **-0.34** | 0.00 | **0.03** | **0.27** | **0.09** | **0.17** | 1 | | |
| (8) $Book-to-Market_t$ | **0.10** | **0.16** | -0.00 | **-0.05** | **0.03** | 0.01 | **0.07** | 1 | |
| (9) $Net\ Income/AT_t$ | **-0.48** | -0.01 | **-0.02** | **0.04** | **0.02** | 0.00 | **0.37** | **0.04** | 1 |
| (10) $Beta_{t-1}$ | -0.01 | **0.04** | **0.04** | **0.05** | **0.06** | **0.08** | **0.26** | **-0.08** | **0.06** |
| (11) $Excess\ Ret._{t-1}$ | **-0.19** | **-0.16** | **-0.07** | **0.02** | **0.03** | **0.02** | **0.06** | **-0.19** | **0.15** |
| (12) $Ex.\ Ret.\ Std_{t-1}$ | **0.41** | **0.16** | **0.06** | **-0.04** | **0.02** | -0.01 | **-0.51** | **0.14** | **-0.44** |
| (13) $\#\ Risk\ Factors_t$ | **0.16** | 0.00 | **0.06** | -0.01 | **0.02** | **0.06** | -0.00 | **0.03** | **-0.17** |
| (14) $\#\ New\ RF_t$ | **0.11** | **0.05** | **0.05** | **0.05** | **0.04** | **0.06** | **0.03** | **0.06** | **-0.11** |
| (15) $\#\ Dropped\ RF_t$ | **0.10** | **0.04** | **0.03** | **0.05** | **0.05** | **0.05** | -0.00 | **0.01** | **-0.12** |
| (16) $Log(\#\ of\ Words)_t$ | **0.19** | -0.01 | **0.08** | **0.02** | **0.03** | **0.07** | **0.02** | 0.00 | **-0.22** |
| (17) $Specificity_t$ | **-0.04** | **0.03** | **0.01** | **0.05** | **0.04** | **0.02** | **0.13** | **0.08** | **0.03** |
| (18) $Numeric\ Intensity_t$ | **0.08** | **0.06** | 0.00 | -0.01 | **0.01** | 0.00 | **-0.07** | **0.06** | **-0.08** |
| (19) $FOG\ Index_t$ | **-0.06** | **-0.02** | **0.03** | **0.07** | **0.04** | **0.06** | **0.32** | **0.03** | 0.01 |

*Continued...*

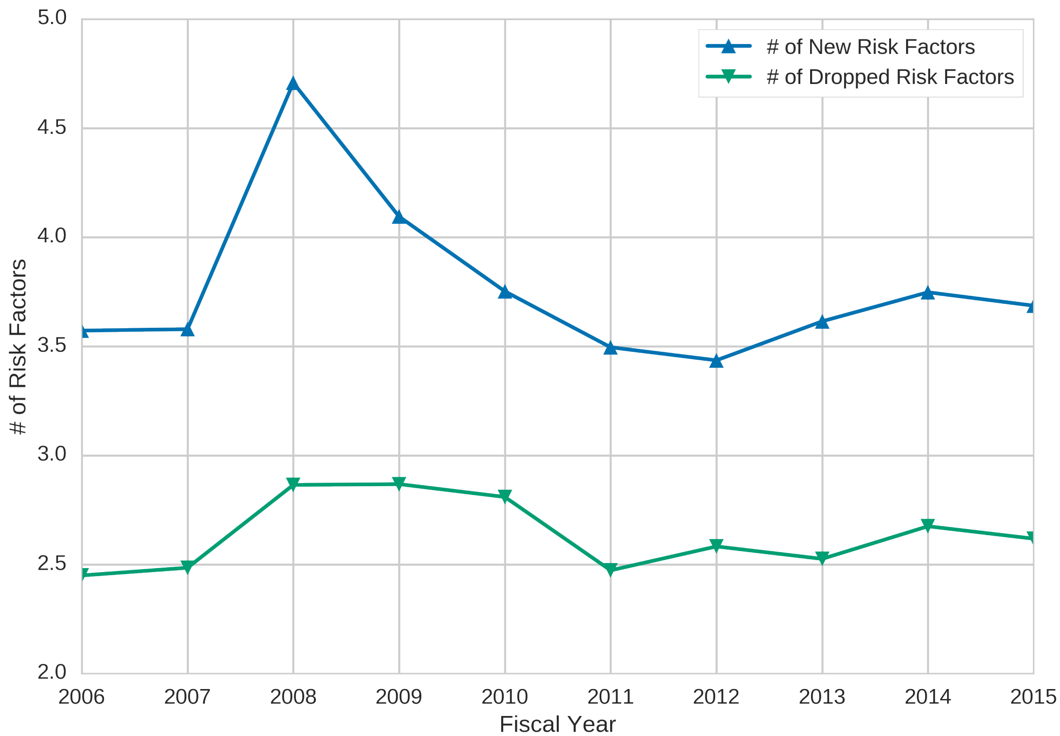
| | (10) | (11) | (12) | (13) | (14) | (15) | (16) | (17) | (18) |
|---|---|---|---|---|---|---|---|---|---|
| (11) $Excess\ Ret._{t-1}$ | **0.04** | 1 | | | | | | | |
| (12) $Ex.\ Ret.\ Std_{t-1}$ | **-0.11** | **-0.05** | 1 | | | | | | |
| (13) $\#\ Risk\ Factors_t$ | **0.06** | **-0.02** | **0.08** | 1 | | | | | |
| (14) $\#\ New\ RF_t$ | -0.00 | **-0.02** | **0.14** | **0.40** | 1 | | | | |
| (15) $\#\ Dropped\ RF_t$ | 0.01 | -0.01 | **0.14** | **0.25** | **0.49** | 1 | | | |
| (16) $Log(\#\ of\ Words)_t$ | **0.11** | **-0.02** | **0.10** | **0.86** | **0.32** | **0.26** | 1 | | |
| (17) $Specificity_t$ | **-0.05** | **-0.03** | 0.01 | -0.00 | **0.07** | **0.04** | **0.02** | 1 | |
| (18) $Numeric\ Intensity_t$ | **-0.02** | **-0.04** | **0.17** | **-0.03** | **0.05** | **0.05** | **0.03** | **0.42** | 1 |
| (19) $FOG\ Index_t$ | **0.05** | -0.00 | **-0.14** | **0.20** | **0.10** | **0.09** | **0.33** | **-0.05** | **-0.16** |

the change in word count of the risk factors between these two groups, there is no significant difference.[18] This suggests that the time series evolution of risk factors may capture some underlying economic differences that a word count approach does not. A graphical depiction of this trend is presented in Figure 2.2.

[18]The change in word counts are 135.5 and 136.4 for young and old firms respectively (t-stat = -0.16). The t-stat for difference in new, dropped, and total risk factors are all significant at <0.001 level.

**Figure 2.1**: New and Removed Risk Factors Over Time

Figure 2.1 plots the average number of new and removed risk factors by fiscal year. The sample comprises firm-years with a non-missing risk factor section in the previous year.

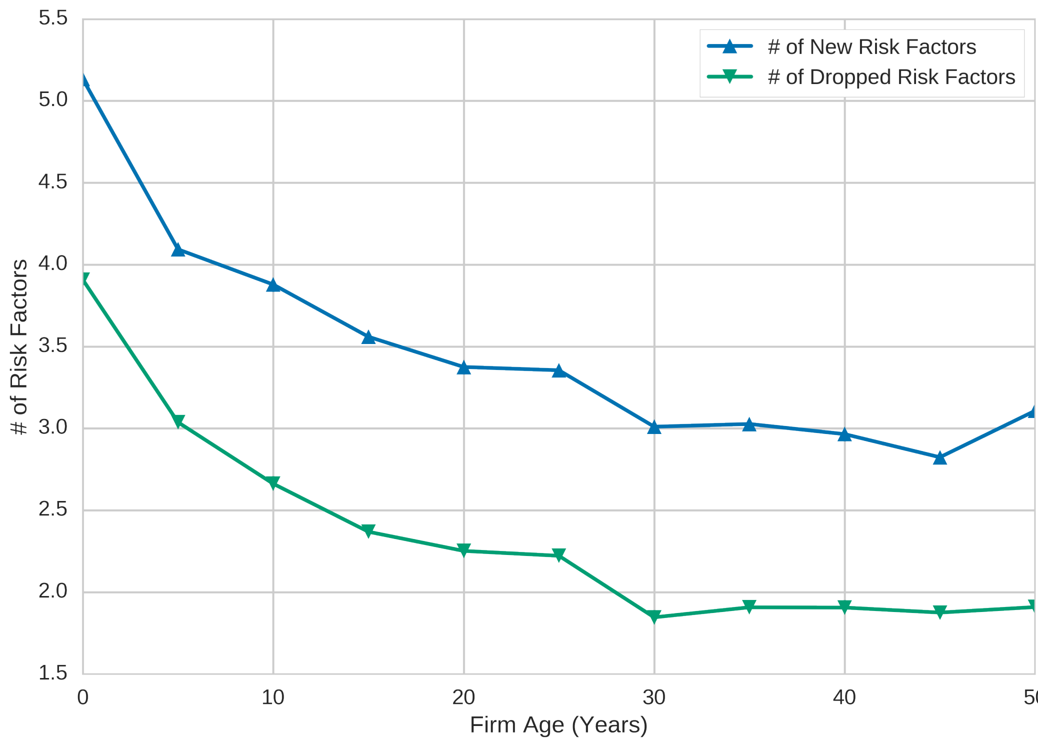



## 2.3 Results

The first two hypotheses posit that managers will increase their identification of risk factors in advance of adverse events. To capture the time-series evolution of both the underlying economics and firm disclosure, I employ panel regressions and control for firm specific heterogeneity using firm fixed effects. This is in contrast to the extant literature, which has largely focused on risk factors in the cross-section using a pooled OLS approach. My measure of adverse outcomes is an indicator variable for negative income, negative operating income, and significant sales decline of greater than ten

**Figure 2.2**: New and Removed Risk Factors Over Firm Age

Figure 2.2 plots the average number of new and removed risk factors over the age of the firm. The datapoints represent averages over five year periods. The sample comprises firm-years with a non-missing risk factor section in the previous year.




percent of the previous year's sales.[19] To capture general business lawsuits, I use the natural log of one plus the number of lawsuits brought against the firm in a given year.

To address the first two hypotheses, I employ a predictive framework in which I regress the adverse outcomes (in the following year) on the measures of risk factor disclosures and test whether increases or decreases in risk factors and their definitiveness precede the events. To model the accounting based outcomes, which are coded as

[19]To avoid biasing towards small firms, I require the sales decline to be at least ten million dollars. Thus, firms with sales less than 100 million must experience a decrease in sales of more than ten million.

dichotomous variables, I use a probit specification with correlated random effects to control for firm specific heterogeneity (Wooldridge, 2010). The lawsuit intensity is modeled using OLS with firm fixed effects.[20]

**Table 2.3**: Adverse Outcome Predictability

Table 2.3 reports results from predictive regressions of future adverse outcomes on current risk factor disclosures. Specifications (1)–(3) present probit regressions with correlated random effects to control for average firm effects. Specification (4) presents an OLS regression with firm fixed effects, and the # of Events listed represents the number of firm years with at least one lawsuit. The coefficients in Specifications (1)–(3) represent the average marginal effects, evaluated at the mean of the dependent variables. The *Sign* column denotes the expected sign of the coefficients based on the hypotheses. The variables are defined in Appendix **??**. Robust standard errors are clustered at the firm level and absolute t-statistics are reported in parentheses. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 level, respectively.

| | Sign | Negative NI | Negative Op. Inc. | Sales Decline | Lawsuit Intensity |
|---|---|---|---|---|---|
| | | (1) | (2) | (3) | (4) |
| $Log(\#\ New\ RF)_t$ | + | 0.022*** | 0.014*** | 0.014*** | 0.013*** |
| | | (5.54) | (4.09) | (3.63) | (3.58) |
| $Log(\#\ Dropped\ RF)_t$ | – | -0.024*** | −0.015*** | −0.008* | −0.005 |
| | | (5.62) | (4.16) | (1.90) | (1.31) |
| $Log(\#\ RF)_{t-1}$ | + | 0.092*** | 0.061*** | 0.047*** | 0.044*** |
| | | (5.41) | (3.75) | (3.39) | (3.24) |
| $\Delta Specificity_t$ | + | 0.001 | 0.000 | −0.002 | 0.004** |
| | | (0.79) | (0.22) | (1.45) | (2.12) |
| $\Delta Numeric\ Intensity_t$ | + | 0.018*** | 0.011*** | 0.014*** | −0.001 |
| | | (3.69) | (2.76) | (3.29) | (0.26) |
| $\Delta Log(\#\ of\ Words/RF)_t$ | + | 0.052** | 0.039** | 0.059*** | 0.064*** |
| | | (2.18) | (2.06) | (2.83) | (2.88) |
| *Year F.E.* | | Y | Y | Y | Y |
| *Correlated R.E.* | | Y | Y | Y | |
| *Firm F.E.* | | | | | Y |
| # Observations | | 21,683 | 20,618 | 21,682 | 26,984 |
| # of Events | | 6,774 | 4,793 | 2,944 | 2,306 |

Because the hypotheses predict that managers warn of adverse outcomes, initially I do not include controls in the model beyond year and firm level fixed effects. Effectively, this tests whether managers warn of adverse outcomes, independent of expectations or other potential disclosure/information channels. Consistent with my first hypothesis,

[20]In untabulated results, I also employ Poisson model of the number of lawsuits with firm fixed effects, and find my results are robust to this alternative specification.

Table 2.3 demonstrates that managers do increase the number of risk factors they identify in advance of adverse events. Managers are also less likely to remove old risk factors in advance of negative income and operating income, and marginally so for sales declines. On the other hand, this is consistent with managers removing risk factors when these adverse outcomes are less likely. The lagged total number of risk factors is also significant, suggesting that not only are the contemporary changes in risk factors predictive of adverse outcomes, but lagged increases in risk factors are informative as well. Taken together, this evidence supports my first hypothesis that managers are informatively adding and removing risk factors in a timely fashion.

Consistent with my second hypothesis, the results in Table 2.3 suggest that managers disclose more definitive risk factors in advance of adverse outcomes. For the three accounting measure outcomes, managers increase both the frequency of numbers used in the disclosures and the verbosity with which they describe them. Because this increase in length of risk factors is concomitant with increases in the other measures of definitiveness (specificity or numeric intensity), it is more likely that the verbosity of risk factors provided are in line with Bloomfield's (2008) ontology explanation rather than obfuscation. Interestingly managers do not use more numbers to describe the external risk of lawsuits, but they do use more specificity. This is consistent with managers reflecting the numerical nature of the internal accounting information they possess, while still providing qualitative detail about potentially less quantifiable external risks.

One limitation of testing the unconditional predictive ability of risk factors is that managers may be merely reacting to realized adverse outcomes through ex-post disclosure. To address this, I further include controls for ex-ante risk as identified in previous studies of risk and disclosure (e.g., Campbell et al., 2014; Nelson and

Pritchard, 2016). I also include a control for the lagged dependent variable using the contemporary continuous (not dichotomous) value measured at the time of the risk factor disclosure release, specifically: net income, operating income, and sales (scaled by total assets) for specifications (1)–(3), respectively, and the log number of suits for specification (4).

Consistent with the unconditional results, managers disclose the new risk factors in advance of adverse outcomes, incrementally to what is explained by observable ex-ante risk.[21] This suggests that managers are not merely reacting to previous adverse news when updating their risk factor disclosures.

### 2.3.1 Alternative Specifications

In any empirical setting, it is often important to test whether the results found are driven primarily by the researcher's specification decisions. To address this concern, I further include tests using alternate specifications to the modeling decisions hitherto employed.

#### Continuous Predictability

One concern stemming from my approach to studying the predictive ability over extreme adverse outcomes is that these tail events may be considered too extreme to connote meaningfully informative disclosures. If managers *do* forecast adverse events but the typical adverse event forecast does not result in as extreme outcomes as those I test, my results would miss this relationship, and thereby under-estimate the extent of information in managers' disclosure decisions. If, as I hypothesize, managers'

[21] Note the number of observations is reduced from the unconditional table due to both data availability of the control variables and the further omission of perfectly predicted outcomes.

**Table 2.4**: Adverse Outcome Incremental Predictability

Table 2.4 reports results from predictive regressions of future adverse outcomes on current risk factor disclosures, and controls for ex-ante risk. Specifications (1)–(3) present average marginal effects from probit regressions with correlated random effects to control for average firm effects. Specifications (4) presents an OLS regression with firm fixed effects. The $Dependent\ Var.\ (Level)_t$ variable is different for each specification. In specification order, it is: net income / total assets, operating income / total assets, sales / total assets, and lawsuit intensity. The variables are defined in Appendix **??**. Robust standard errors are clustered at the firm level and absolute t-statistics are reported in parentheses. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 level, respectively.

| | Negative NI | Negative Op. Inc. | Sales Decline | Log(# Suits) |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| $Log(\#\ New\ RF)_t$ | 0.017*** | 0.009*** | 0.009** | 0.012*** |
| | (4.14) | (2.81) | (2.33) | (3.13) |
| $Log(\#\ Dropped\ RF)_t$ | −0.018*** | −0.011*** | −0.007* | −0.005 |
| | (4.14) | (2.93) | (1.71) | (1.24) |
| $Log(\#\ RF)_{t-1}$ | 0.036** | 0.018 | 0.042*** | 0.045*** |
| | (2.40) | (1.37) | (3.01) | (3.09) |
| $\Delta Specificity_t$ | 0.001 | 0.001 | −0.002 | 0.004** |
| | (0.86) | (0.41) | (1.15) | (2.29) |
| $\Delta Numeric\ Intensity_t$ | 0.010* | 0.006 | 0.010** | −0.001 |
| | (1.92) | (1.39) | (2.31) | (0.15) |
| $\Delta Log(\#\ of\ Words/RF)_t$ | 0.061** | 0.042** | 0.042** | 0.050** |
| | (2.48) | (2.19) | (1.98) | (2.25) |
| $Dependent\ Var.\ (Level)_{t-1}$ | −0.209*** | −0.113*** | 0.089*** | −0.027** |
| | (5.60) | (7.29) | (9.12) | (2.47) |
| $Log(Market\ Equity)_t$ | 0.007 | −0.000 | −0.004 | 0.046*** |
| | (0.76) | (0.02) | (0.64) | (6.45) |
| $Big\ N_t$ | −0.000 | −0.007 | 0.022 | −0.031* |
| | (0.02) | (0.43) | (1.30) | (1.75) |
| $Book-to-Market_t$ | 0.052*** | 0.027*** | 0.011** | 0.001 |
| | (7.92) | (5.65) | (2.17) | (0.25) |
| $Tangibility_t$ | 0.168** | 0.237*** | 0.071 | 0.132** |
| | (2.22) | (4.03) | (1.24) | (2.36) |
| $Leverage_t$ | −0.116*** | −0.079** | −0.011 | −0.001 |
| | (2.98) | (2.50) | (0.51) | (0.04) |
| $Turnover_t$ | 0.014* | 0.019*** | 0.032*** | 0.040*** |
| | (1.89) | (3.09) | (6.17) | (6.10) |
| $Beta_{t-1}$ | 0.002 | −0.004 | −0.007 | −0.005 |
| | (0.25) | (0.71) | (1.00) | (0.61) |
| $Excess\ Ret._{t-1}$ | −0.059*** | −0.028*** | −0.063*** | −0.022*** |
| | (8.60) | (5.03) | (8.77) | (3.30) |
| $Ex.\ Ret.\ Std_{t-1}$ | −0.194 | −0.015 | 0.107 | 0.352 |
| | (0.57) | (0.06) | (0.39) | (1.21) |
| $Ex.\ Ret.\ Skew_{t-1}$ | −0.000 | 0.002 | 0.001 | −0.005** |
| | (0.11) | (1.03) | (0.75) | (2.17) |
| *Year F.E.* | Y | Y | Y | Y |
| *Correlated R.E.* | Y | Y | Y | |
| *Firm F.E.* | | | | Y |
| # Observations | 20,303 | 19,945 | 20,302 | 25,252 |
| # of Events | 6,372 | 4,604 | 2,780 | 2,235 |

disclosure decisions are driven by their subjective assessment of cash flow distributions (or other outcomes, such as general business litigation), it may equally reasonable to assume that the disclosure decision is related to the continuous underlying stochastic time-series economic process, rather than just discrete adverse events I primarily focus on. To test whether this is the case, and risk factor disclosure decisions are related to the continuous economic evolution of the firm, I regress continuous measures of accounting performance on the risk factor disclosure measures.[22] The accounting measures I choose are (all scaled by total assets): *Net Income*, *Operating Income*, *Sales*, and *ROA* to capture profitability changes from either revenue declines or cost increases. I also control for the contemporary value of these measures, measured at the time of the risk factor disclosure. The results are presented in Table 2.5, and includes both the cross sectional predictability (using pooled OLS regressions) and time-series predictability (using a panel regression with fixed effects) of the risk factor disclosures.

The results in Table 2.5 suggest that disclosing new risk factors is strongly correlated with declines in all four outcomes, consistent with expectations of risk disclosures containing adverse information. This relationship holds in both the cross section and the time series. Similarly, removing risk factor disclosures is associated with increases in the income measures (but not so for *Sales*), and this relationship holds primarily in the time series, suggesting firm specific heterogeneity confounds the information about earnings in risk factor disclosures. There is a significant relationship between the *Specificity* of risk factor disclosures and future declines the accounting outcomes (and to a lesser extent the verbosity), but not so for the *Numeric Intensity*. One

[22]The sample size for these tests is larger than that of the corresponding tests in Table 2.4 because the latter omits observations in the presence of perfectly predicted outcomes.

**Table 2.5**: Risk Factor Disclosures and Continuous Accounting Outcomes

Table 2.5 reports results from predictive regressions of future continuous accounting outcomes on the risk factor disclosure measures, and controls for ex-ante risk and accounting performance. Odd and even specifications correspond to cross sectional and firm fixed effects models respectively, and all dependent variables are scaled by total assets. Each specification includes $Dependent\ Var.\ (Level)_{t-1}$, which is the contemporary value of the continuous dependent variable. The results are robust to the omission of this variable. The variables are defined in Appendix **??**. Robust standard errors are clustered at the firm level and absolute t-statistics are reported in parentheses. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 level, respectively.

| | Net Income$_{t+1}$ | | Op. Inc.$_{t+1}$ | | Sales$_{t+1}$ | | ROA$_{t+1}$ | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $Log(\#\ New\ RF)_t$ | -0.015*** | -0.015*** | -0.013*** | -0.012*** | -0.040*** | -0.028*** | -0.015*** | -0.015*** |
| | (5.89) | (5.60) | (4.94) | (4.46) | (6.57) | (4.21) | (5.94) | (5.63) |
| $Log(\#\ Dropped\ RF)_t$ | 0.006* | 0.012*** | 0.003 | 0.008*** | 0.003 | 0.007 | 0.006** | 0.012*** |
| | (1.94) | (4.01) | (0.99) | (2.59) | (0.49) | (0.99) | (1.99) | (3.93) |
| $Log(\#\ RF)_{t-1}$ | -0.017*** | -0.004 | -0.011*** | 0.012 | -0.009 | 0.041 | -0.016*** | -0.003 |
| | (4.14) | (0.37) | (2.72) | (1.09) | (0.95) | (1.51) | (4.04) | (0.29) |
| $\Delta Specificity_t$ | -0.003** | -0.002* | -0.003** | -0.002* | -0.007** | -0.004 | -0.003** | -0.002** |
| | (2.52) | (1.85) | (2.49) | (1.88) | (2.35) | (1.51) | (2.56) | (1.99) |
| $\Delta Numeric\ Intensity_t$ | -0.002 | -0.001 | -0.003 | -0.001 | -0.009 | -0.006 | -0.002 | -0.001 |
| | (0.63) | (0.32) | (0.89) | (0.23) | (0.99) | (0.78) | (0.62) | (0.22) |
| $\Delta Log(\#\ of\ Words/RF)_t$ | -0.011 | -0.029* | -0.007 | -0.018 | -0.097** | -0.068* | -0.012 | -0.028* |
| | (0.62) | (1.83) | (0.38) | (1.14) | (2.43) | (1.89) | (0.71) | (1.81) |
| $Dependent\ Var.\ (Level)_{t-1}$ | 0.514*** | 0.037* | 0.672*** | 0.174*** | 0.725*** | 0.266*** | 0.523*** | 0.041** |
| | (36.07) | (1.83) | (60.00) | (7.17) | (75.10) | (10.40) | (36.71) | (1.99) |
| $Log(Market\ Equity)_t$ | 0.017*** | 0.034*** | 0.014*** | 0.033*** | 0.011*** | -0.033*** | 0.017*** | 0.034*** |
| | (11.71) | (6.73) | (9.83) | (6.51) | (3.59) | (2.76) | (11.73) | (6.87) |
| $Big\ N_t$ | -0.007 | -0.026* | -0.008* | -0.026* | 0.023** | -0.003 | -0.007 | -0.025 |
| | (1.30) | (1.70) | (1.66) | (1.70) | (2.11) | (0.09) | (1.39) | (1.62) |
| $Book-to-Market_t$ | 0.000 | -0.009** | 0.002 | -0.002 | -0.013** | -0.036*** | 0.001 | -0.008** |
| | (0.15) | (2.31) | (0.74) | (0.51) | (2.06) | (3.61) | (0.51) | (1.98) |
| $Tangibility_t$ | 0.031*** | -0.035 | 0.029*** | -0.035 | 0.068*** | 0.385*** | 0.032*** | -0.035 |
| | (3.15) | (0.85) | (2.95) | (0.85) | (3.18) | (3.78) | (3.22) | (0.84) |
| $Leverage_t$ | -0.006 | 0.034 | -0.024** | -0.006 | -0.091*** | -0.185*** | -0.008 | 0.025 |
| | (0.63) | (1.53) | (2.53) | (0.26) | (3.78) | (3.17) | (0.89) | (1.17) |
| $Turnover_t$ | -0.001 | 0.006 | 0.000 | 0.010** | -0.016** | -0.008 | -0.001 | 0.006 |
| | (0.50) | (1.23) | (0.02) | (2.29) | (2.57) | (0.85) | (0.33) | (1.32) |
| $Beta_{t-1}$ | 0.007* | 0.003 | 0.007* | 0.002 | 0.023*** | -0.004 | 0.007* | 0.004 |
| | (1.76) | (0.62) | (1.88) | (0.32) | (2.77) | (0.30) | (1.87) | (0.71) |
| $Excess\ Ret._{t-1}$ | 0.036*** | 0.017*** | 0.035*** | 0.015*** | 0.048*** | 0.066*** | 0.037*** | 0.017*** |
| | (8.12) | (3.72) | (7.90) | (3.29) | (4.79) | (6.46) | (8.31) | (3.74) |
| $Ex.\ Ret.\ Std_{t-1}$ | -1.120*** | -0.317 | -0.872*** | -0.721*** | -0.232 | -0.923* | -1.120*** | -0.346 |
| | (5.74) | (1.35) | (4.63) | (3.06) | (0.56) | (1.71) | (5.79) | (1.49) |
| $Ex.\ Ret.\ Skew_{t-1}$ | -0.010*** | -0.009*** | -0.009*** | -0.009*** | -0.017*** | -0.012*** | -0.010*** | -0.009*** |
| | (6.27) | (5.52) | (5.84) | (5.41) | (5.10) | (3.64) | (6.21) | (5.56) |
| $Year\ F.E.$ | Y | Y | Y | Y | Y | Y | Y | Y |
| $Industry\ F.E.$ | Y | | Y | | Y | | Y | |
| $Firm\ F.E.$ | | Y | | Y | | Y | | Y |
| # Observations | 25,252 | 25,252 | 25,252 | 25,252 | 25,252 | 25,252 | 25,252 | 25,252 |

explanation may be that firms update their qualitative information across a broader spectrum of adverse outcomes, and provide more quantitative information in advance of significantly negative adverse outcomes, consistent with Heinle and Smith (2015). This supports the idea underlying the hypotheses that managers are making risk factor disclosure decisions based on their expectations of future cash flows.

### Omitting Repeated Adverse Events

One concern of studying the *predictive* ability of risk factor disclosures is that managers' disclosure decisions may be reacting to observed adverse outcomes, but these outcomes are correlated in the time series. What may appear to be predictive ability is in actuality an endogenous reaction to repeated adverse events. For example a firm suffering a net loss may contemporaneously disclose potential causes in their risk factor section, but if the loss persists into the next year, these risk factor disclosure changes may appear to 'predict' the loss. To mitigate this concern, I repeat the same predictive tests but reduce the sample of firm-year observations by omitting repeated observations. This effectively reduces the sample to firm years leading up to an adverse outcome, but does not continue to include firm years if the adverse outcomes persist.

The results of this analysis are presented in Table 2.6. Consistent with the results in Table 2.4, the managers disclose more new risk factors ahead of adverse outcomes. Managers also decrease their removal of disclosures, which suggests they remove risk factor disclosures when these adverse outcomes are less likely. The significance of the lagged number of risk factors declines sharply, which supports the conclusion that risk factor disclosures have short-term predictive ability of adverse outcomes. This is because as adverse outcomes are repeated, the predictive risk factor disclosures immediately preceding the adverse outcome move into the 'lagged' risk

**Table 2.6**: Adverse Outcome Predictability Omitting Repeated Outcomes

Table 2.6 reports results from predictive regressions of future adverse outcomes on current risk factor disclosures, and controls for ex-ante risk. The sample consists of firm-years where the contemporaneous value of the adverse outcome is zero. Specifications (1)-(3) present average marginal effects from probit regressions with correlated random effects to control for average firm effects. Specification (4) presents coefficients from an OLS regression with firm fixed effects. The $Dependent\ Var.\ (Level)_t$ variable is different for each specification, and is equal to the continuous dependent variable scaled by total assets. Specification (4) does not have a lagged dependent variable by definition of the removal of repeated adverse outcomes. The variables are defined in Appendix **??**. Robust standard errors are clustered at the firm level and absolute t-statistics are reported in parentheses. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 level, respectively.

| | Negative NI<br>(1) | Negative Op. Inc.<br>(2) | Sales Decline<br>(3) | Log(# Suits)<br>(4) |
|---|---|---|---|---|
| $Log(\#\ New\ RF)_t$ | 0.019*** | 0.010*** | 0.012*** | 0.008** |
| | (4.27) | (3.41) | (3.18) | (2.29) |
| $Log(\#\ Dropped\ RF)_t$ | -0.016*** | -0.010*** | -0.010** | 0.000 |
| | (3.28) | (2.87) | (2.47) | (0.07) |
| $Log(\#\ RF)_{t-1}$ | 0.018 | 0.022* | 0.060*** | 0.021 |
| | (1.15) | (1.92) | (4.18) | (1.61) |
| $\Delta Specificity_t$ | 0.003* | 0.001 | -0.001 | 0.004** |
| | (1.82) | (0.94) | (0.42) | (2.53) |
| $\Delta Numeric\ Intensity_t$ | 0.003 | 0.004 | 0.009** | -0.001 |
| | (0.48) | (0.89) | (2.05) | (0.29) |
| $\Delta Log(\#\ of\ Words/RF)_t$ | 0.058** | 0.028* | 0.032 | 0.057** |
| | (2.29) | (1.65) | (1.45) | (2.56) |
| $Dependent\ Var.\ (Level)_{t-1}$ | -0.361*** | -0.220*** | 0.056*** | |
| | (5.50) | (5.36) | (5.77) | |
| $Log(Market\ Equity)_t$ | 0.047*** | 0.004 | -0.007 | 0.022*** |
| | (4.44) | (0.64) | (0.97) | (3.31) |
| $Big\ N_t$ | 0.015 | 0.006 | -0.000 | -0.019 |
| | (0.66) | (0.38) | (0.01) | (1.24) |
| $Book-to-Market_t$ | 0.115*** | 0.027*** | 0.014** | 0.001 |
| | (8.84) | (5.44) | (2.47) | (0.15) |
| $Tangibility_t$ | 0.107 | 0.185*** | 0.059 | 0.016 |
| | (1.51) | (3.75) | (1.03) | (0.33) |
| $Leverage_t$ | 0.011 | -0.032 | -0.015 | 0.032 |
| | (0.31) | (1.09) | (0.60) | (1.05) |
| $Turnover_t$ | 0.005 | 0.013*** | 0.033*** | 0.043*** |
| | (0.62) | (2.61) | (6.10) | (6.26) |
| $Beta_{t-1}$ | -0.014 | -0.010* | -0.008 | -0.009 |
| | (1.57) | (1.72) | (1.15) | (1.30) |
| $Excess\ Ret._{t-1}$ | -0.063*** | -0.029*** | -0.059*** | -0.006 |
| | (7.29) | (5.17) | (8.06) | (0.91) |
| $Ex.\ Ret.\ Std_{t-1}$ | 0.939** | 0.326 | 0.743** | 0.102 |
| | (2.18) | (1.39) | (2.40) | (0.36) |
| $Ex.\ Ret.\ Skew_{t-1}$ | -0.001 | 0.001 | 0.000 | -0.004** |
| | (0.37) | (0.82) | (0.10) | (2.06) |
| $Year\ F.E.$ | Y | Y | Y | Y |
| $Correlated\ R.E.$ | Y | Y | Y | |
| $Firm\ F.E.$ | | | | Y |
| # Observations | 13,837 | 15,334 | 17,628 | 20,130 |
| # of Events | 1,649 | 939 | 2,038 | 817 |

factor disclosures, but are still predictive of the repeated outcomes. The significance of the definitiveness of risk factor disclosures is reduced from that in Table 2.4, but remains marginally significant in some cases. This suggests that firms may not exhibit a strong increase in the definitiveness of their disclosures immediately, but continue to add more detail as adverse events persist. This is again consistent with the theory of Heinle and Smith (2015) which suggests managers expend resources to gather more information when they receive bad-news signals.

### Net Change in Disclosed Risk Factors

One potential concern of including the dual measures of new and removed risk factor disclosures simultaneously is that they could jointly result in no predictability of the net change measure. For example, in Specification (1) of Table 2.3 the coefficients for new and dropped risk factor disclosures have opposite signs but are very similar in absolute value. While Table 2.2 shows the correlation between these variables is only 0.49, it could be the case that empirically the net combination of new and dropped risk factors do not predict adverse outcome. To determine whether the net change in disclosed risk factors predicts adverse outcomes, I combine the number of new and dropped risk factors in to one variable, and call this variable $\Delta\#\ RF$. If the results are mechanically driven by offsetting new and removed risk factors, then the coefficient on the net change should be zero.

Table 2.7 presents these regressions with the inclusion of the control variables employed in Table 2.4, but the findings are robust to their exclusion. The results suggest that the net change in disclosed risk factors also forecast these adverse events. The coefficient on $\Delta\#\ RF$ is significantly positively related to the adverse outcomes, suggesting that managers increase the total number of risk factors they disclose in

**Table 2.7**: Net Change in Risk Factors Predicting Adverse Outcomes

Table 2.7 reports results from predictive regressions of future adverse outcomes on the net change in number of risk factor disclosures, and controls for ex-ante risk. The variable $Log(\Delta\#\ RF)_t$ is calculated as the natural logarithm of the $\Delta\#\ RF$ plus one, less the minimum net change in risk factors (-22). Specifications (1) and (2) present average marginal effects from probit regressions with correlated random effects to control for average firm effects. Specifications (3)–(5) present an OLS regression with firm fixed effects. The $Dependent\ Var.\ (Level)_t$ variable is different for each specification. In specification order, it is: net income / total assets, sales / total assets, lawsuit intensity, and the change in net income / total assets for both specifications (4) and (5). The variables are defined in Appendix **??**. Robust standard errors are clustered at the firm level and absolute t-statistics are reported in parentheses. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 level, respectively.

| | Negative NI | Negative Op. Inc. | Sales Decline | Log(# Suits) |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| $Log(\Delta\#\ RF)_t$ | 0.081*** | 0.035** | 0.045*** | 0.041** |
| | (4.50) | (2.43) | (2.63) | (2.23) |
| $Log(\#\ RF)_{t-1}$ | 0.037** | 0.016 | 0.047*** | 0.047*** |
| | (2.41) | (1.13) | (3.23) | (3.22) |
| $\Delta Specificity_t$ | 0.002 | 0.001 | −0.002 | 0.004** |
| | (1.03) | (0.58) | (1.18) | (2.25) |
| $\Delta Numeric\ Intensity_t$ | 0.010** | 0.006 | 0.010** | −0.001 |
| | (2.07) | (1.57) | (2.42) | (0.16) |
| $\Delta Log(\#\ of\ Words/RF)_t$ | 0.059** | 0.038** | 0.045** | 0.051** |
| | (2.35) | (1.97) | (2.04) | (2.31) |
| $Dependent\ Var.\ (Level)_{t-1}$ | −0.209*** | −0.113*** | 0.090*** | −0.026** |
| | (5.53) | (7.29) | (9.23) | (2.44) |
| $Log(Market\ Equity)_t$ | 0.007 | 0.000 | −0.004 | 0.046*** |
| | (0.80) | (0.01) | (0.65) | (6.45) |
| $Big\ N_t$ | −0.000 | −0.006 | 0.023 | −0.031* |
| | (0.01) | (0.38) | (1.33) | (1.75) |
| $Book-to-Market_t$ | 0.052*** | 0.027*** | 0.011** | 0.001 |
| | (7.88) | (5.64) | (2.13) | (0.25) |
| $Tangibility_t$ | 0.166** | 0.236*** | 0.075 | 0.131** |
| | (2.16) | (3.97) | (1.31) | (2.35) |
| $Leverage_t$ | −0.115*** | −0.078** | −0.011 | 0.000 |
| | (2.94) | (2.44) | (0.50) | (0.00) |
| $Turnover_t$ | 0.013* | 0.019*** | 0.032*** | 0.040*** |
| | (1.75) | (3.04) | (6.08) | (6.15) |
| $Beta_{t-1}$ | 0.002 | −0.004 | −0.007 | −0.005 |
| | (0.31) | (0.71) | (1.02) | (0.61) |
| $Excess\ Ret._{t-1}$ | −0.060*** | −0.028*** | −0.064*** | −0.022*** |
| | (8.65) | (5.06) | (8.80) | (3.31) |
| $Ex.\ Ret.\ Std_{t-1}$ | −0.204 | −0.020 | 0.127 | 0.376 |
| | (0.59) | (0.08) | (0.46) | (1.30) |
| $Ex.\ Ret.\ Skew_{t-1}$ | −0.000 | 0.002 | 0.001 | −0.005** |
| | (0.13) | (1.01) | (0.76) | (2.17) |
| *Year F.E.* | Y | Y | Y | Y |
| *Correlated R.E.* | Y | Y | Y | |
| *Firm F.E.* | | | | Y |
| # Observations | 20,303 | 19,945 | 20,302 | 25,252 |
| # of Events | 6,372 | 4,604 | 2,780 | 4,534 |

advance of adverse events. This supports the conclusion that my previous results of predictability are not mechanically driven by the dual measures decomposing the net change.

### 2.3.2 Equity Market and Risk Factor Disclosures

The evidence that managers appear to predict adverse outcomes supports the supposition that managers are compounding their expectations of future economic events into their disclosure decisions. The evidence in Table 2.5 lend further credence to this conclusion. However, while this suggests that the evolution of risk factor disclosures contains information, it does not address whether this information is decision useful, i.e. whether it is informative to markets. This latter question is the primary focus of extant literature, as outlined above. However previous studies have generally used aggregated measures of risk disclosure, rather than focusing on the individual risk factor disclosures as the unit of observation. Ex-ante, it is unknown whether the dimensions developed herein to capture the information in risk factor disclosures are consistent with the information being compounded by the equity market.

I therefore reconsider the model of Table 2.4 using market measures as the dependent variables. Consistent with Campbell et al. (2014), I test whether risk factor disclosures are associated with future market risk exposure (*Beta*), as well as short and longer horizon market returns to evaluate whether these risks are already compounded into the firm's stock price, or whether they contain potentially novel information.[23] Similar to the approach in Campbell et al. (2014), I look at a three-day window

[23] I measure the market reaction using raw and idiosyncratic returns. Idiosyncratic returns are calculated as the cumulative abnormal returns using a Fama-French four factor model, calibrated using daily returns over the twelve month period starting three months prior to the filing date of the annual report.

after the filing and control for the earnings news by including the change in earnings (scaled by total assets). For the longer horizon returns, I only calculate the abnormal returns for one quarter (60 business days) to avoid potential confounding inference from subsequent quarterly risk factor updates.

The results in Table 2.8 suggest that in the cross section (Specification (1)), adding new risk factors, and the total number as well, are positively associated with a firm's *Beta*. Removing risk factors is associated with a decrease in the firm's *Beta*. This is an interesting result, because in the regulation for risk factor disclosures, the SEC requires that firms "not present risks that could apply to any issuer or any offering." This suggests that firms should not discuss general risks which affect all (or some significant subset) firms. However it is not necessarily clear that this will preclude risk factors from addressing systematic risks for two reasons. The first is that it is not necessarily the case that managers will follow this prohibition on a systematic basis, or that disclosing a generic risk in a sufficiently descriptive fashion would be a violation of this requirement. As a result, managers may regularly disclose risk factors about systematic risks, which could result in an association with the firm's raw returns. The second is that it is unclear whether managers have the capacity to differentiate between risks which are systematically priced and those which are idiosyncratic.

In light of this potential for both the presence or absence of disclosures about generic, or systematic, risk, the results of Table 2.8 present an interesting juxtaposition to the conclusion in Campbell et al. (2014). The results show a correlation exists in the cross section (Specification (1)) but not so in the time-series, after controlling for firm-specific heterogeneity (Specification (2)). This suggests that the association between risk disclosure and systematic risk exposure may not be driven by the disclosure changes over time, but rather by idiosyncratic firm factors. However, given that the

**Table 2.8**: Risk Factor Disclosures and Market Outcomes

Table 2.8 reports results from OLS regressions of future market outcomes on contemporary risk factor disclosures, and controls for ex-ante risk. Specifications (3) and (5) use raw and idiosyncratic returns, respectively, cumulated over a three trading day window starting on the filing date of the annual report. Specifications (4) and (6) use raw and idiosyncratic returns, respectively, cumulated over a 60 trading day window starting on the filing date of the annual report. Idiosyncratic returns are calculated using the Fama-French four factor model described in Appendix **??**. $\Delta NI_t$ is equal to the change in net income scaled by total assets. The variables are defined in Appendix **??**. Robust standard errors are clustered at the firm level and absolute t-statistics are reported in parentheses. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 level, respectively.

| | Beta$_{t+1}$ | | CAR$_{Raw}$ 3 days | CAR$_{Raw}$ 3 months | CAR$_{Idiosyncratic}$ 3 days | CAR$_{Idiosyncratic}$ 3 months |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $Log(\#\ New\ RF)_t$ | 0.011*** | 0.004 | −0.200*** | −0.557** | −0.176*** | −0.544** |
| | (3.01) | (1.07) | (2.81) | (2.26) | (2.64) | (2.14) |
| $Log(\#\ Dropped\ RF)_t$ | −0.015*** | −0.012*** | 0.050 | 0.404 | 0.031 | 0.779*** |
| | (3.73) | (2.70) | (0.61) | (1.46) | (0.40) | (2.79) |
| $Log(\#\ RF)_{t-1}$ | 0.021*** | 0.023 | −0.208 | 1.467 | −0.324 | 1.379 |
| | (3.34) | (1.40) | (0.71) | (1.60) | (1.24) | (1.46) |
| $\Delta Specificity_t$ | −0.001 | −0.000 | 0.028 | 0.004 | 0.030 | 0.087 |
| | (0.52) | (0.24) | (0.89) | (0.04) | (1.05) | (0.76) |
| $\Delta Numeric\ Intensity_t$ | −0.001 | −0.002 | −0.170* | −0.210 | −0.137 | −0.175 |
| | (0.11) | (0.43) | (1.75) | (0.62) | (1.51) | (0.50) |
| $\Delta Log(\#\ of\ Words/RF)_t$ | 0.066*** | 0.038* | −0.999** | 0.682 | −0.607 | −2.215 |
| | (2.98) | (1.78) | (2.32) | (0.42) | (1.53) | (1.39) |
| $\Delta NI_{t-1}$ | 0.051** | 0.056** | 0.928** | −1.271 | 0.688* | −2.869* |
| | (2.13) | (2.42) | (2.34) | (0.89) | (1.81) | (1.89) |
| $Log(Market\ Equity)_t$ | 0.035*** | 0.093*** | −0.374** | −6.915*** | −0.499*** | −7.222*** |
| | (14.88) | (9.44) | (2.50) | (12.57) | (3.59) | (12.20) |
| $Big\ N_t$ | 0.058*** | −0.017 | 0.521 | −0.330 | 0.289 | 0.753 |
| | (7.20) | (0.67) | (1.47) | (0.23) | (0.83) | (0.52) |
| $Book-to-Market_t$ | −0.007 | 0.001 | 0.197 | 2.398*** | 0.151 | 3.424*** |
| | (1.49) | (0.14) | (1.28) | (4.40) | (1.05) | (5.21) |
| $Tangibility_t$ | 0.011 | 0.022 | −1.493 | −4.861 | −1.130 | −1.788 |
| | (0.71) | (0.32) | (1.27) | (1.10) | (1.04) | (0.41) |
| $Leverage_t$ | 0.028** | 0.072** | 0.656 | −3.345* | 0.752 | −1.056 |
| | (2.32) | (2.13) | (1.19) | (1.74) | (1.49) | (0.54) |
| $Turnover_t$ | 0.056*** | 0.055*** | −0.481*** | −2.565*** | −0.153 | −2.168*** |
| | (12.08) | (6.87) | (3.74) | (5.34) | (1.27) | (4.43) |
| $Beta_{t-1}$ | 0.618*** | 0.212*** | −0.076 | −4.446*** | −0.098 | −3.660*** |
| | (98.11) | (23.28) | (0.48) | (7.89) | (0.65) | (6.50) |
| $Excess\ Ret._{t-1}$ | 0.033*** | −0.020** | 0.015 | −2.188*** | −0.742*** | −19.251*** |
| | (4.50) | (2.51) | (0.11) | (4.39) | (5.71) | (36.78) |
| $Ex.\ Ret.\ Std_{t-1}$ | 4.385*** | 6.624*** | 10.864 | 192.267*** | −4.806 | −42.410 |
| | (13.51) | (15.20) | (1.50) | (7.67) | (0.70) | (1.56) |
| $Ex.\ Ret.\ Skew_{t-1}$ | −0.002 | 0.000 | −0.011 | 0.168 | −0.044 | −0.408*** |
| | (0.88) | (0.03) | (0.32) | (1.36) | (1.33) | (3.22) |
| *Year F.E.* | Y | Y | Y | Y | Y | Y |
| *Industry F.E.* | Y | | | | | |
| *Firm F.E.* | | Y | Y | Y | Y | Y |
| # Observations | 19,859 | 19,859 | 23,991 | 23,997 | 23,991 | 23,997 |

tests fail to reject a null of no association, this evidence could also be merely indicative of a low-power test. It does underline the potential importance of controlling for firm-specific heterogeneity when assessing the informativeness of disclosures which evolve over time. In both specifications, neither the specificity nor numeric intensity are associated with *Beta*, which suggests that managers are not disclosing more definitive information in advance of increased market risk exposure, with the exception of increased verbosity.

The results in Specifications (3)–(6) of Table 2.8 show that adding new risk factor disclosures is associated with more negative returns both immediately and over longer periods.[24] Removing risk factors is only associated with increases in returns over longer horizons, which potentially is due to a slow diffusion process of the news conveyed in (or correlated with) those removed risks managers deem no longer applicable. These results are consistent with those in Tables 2.3 and 2.4, that removing risk factors is an informative signal of future positive cash flow news (or lower risk). As expected, the lagged number of risk factors is not predictive of future returns, consistent with the market already having incorporated the information conveyed in the previous year's disclosures. Interestingly, the risk factor definitiveness does not appear to inform market expectations, however. This suggests that managers convey their private information to markets by signaling which risk factors they believe to be the *most significant*, rather than through their expositional descriptions of those risks.

[24]In untabulated results of increasing return window lengths, the t-statistic for dropped risk factors increases roughly monotonically, up until approximately six months after the filing of the annual report, but is insignificant again at horizons of one year. The coefficient on new risk factors is statistically significant until six months as well, and then becomes insignificant at one year. This is potentially consistent with a systematic diffusion process of the information into prices, of which the tabulated date ranges are representative.

These tests suggest that risk factor disclosures potentially inform, or are correlated with future information revelation to markets about either cash flow news (consistent with the evidence in Tables 2.3 and 2.4) or risk assessments (potentially driven by revised expected returns). It also may be the case that the divulgence of risk factor disclosures as public information affects the information asymmetry in capital markets, either by revealing novel information or confirming existing expectations. Kravet and Muslu (2013) refer to these possibilities as *divergent* and *convergent* outcomes, respectively. To test whether the measures I develop are a channel through which risk disclosure may affect the information asymmetry in markets, I repeat the analysis of Table 2.8, and study the bid-ask-spread and Amihud measure of illiquidity around the filing date.

For both dependent variables, I use two alternative specifications and measures. The two specifications I study are cross-sectional and time series regressions, similar to Specifications (1) and (2) of Table 2.8, again to capture whether these informational effects are due to time-series disclosure variation or firm-specific heterogeneity. The two measures I use are the level of bid-ask-spread and illiquidity, as well as the change in these measures from the two weeks before the filing date. Using a change as a dependent variable measures the incremental effect of the new disclosures, and controls for temporary fluctuations in these measures over time, which if they occur before the filing date are unlikely to be due to the risk-factor disclosures themselves. These regressions are presented in Table 2.9.

The results in Specifications (1) and (5) of Table 2.9 suggest that consistent with the findings regarding bid-ask-spread in Campbell et al. (2014), both the level and addition of new risk factors are associated with reduction in information asymmetry and increases liquidity in capital markets. However in the time-series, after controlling

**Table 2.9**: Market Outcomes: Bid-Ask Spread and Illiquidity

Table 2.9 reports results from OLS regressions of market measures on risk factor measures, and controls of ex-ante risk. The odd specifications represent cross-sectional regressions with industry fixed effects, and the even specifications represent time-series regressions with firm fixed effects. The dependent variables in Specifications (1), (2), (5), and (6) are averaged over the ten trading days starting with the filing date of the 10-K. The dependent variables in the remaining specifications represent changes, calculated as the the average value over the ten trading days starting at the filing date, less the average value over the ten trading days ending the day before the filing date. Industry is defined using the Fama-French 12 industry classification. The variables are defined in Appendix **??**. Robust standard errors are clustered at the firm level and absolute t-statistics are reported in parentheses. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 level, respectively.

| | Bid-Ask Spread | | Δ Bid-Ask Spread | | Illiquidity | | ΔIlliquidity | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $Log(\#\ New\ RF)_t$ | −0.382*** | 0.063 | 0.022 | 0.024 | −0.202*** | −0.101* | 0.023 | −0.023 |
| | (3.26) | (0.62) | (0.49) | (0.47) | (3.52) | (1.89) | (0.78) | (0.71) |
| $Log(\#\ Dropped\ RF)_t$ | 0.251* | 0.022 | 0.038 | −0.014 | 0.083 | 0.010 | 0.023 | 0.007 |
| | (1.88) | (0.19) | (0.75) | (0.24) | (1.27) | (0.19) | (0.67) | (0.19) |
| $Log(\#\ RF)_{t-1}$ | −2.843*** | −0.623 | −0.019 | 0.060 | −0.832*** | −0.487** | 0.112* | 0.055 |
| | (9.49) | (1.41) | (0.24) | (0.31) | (6.45) | (2.26) | (1.84) | (0.41) |
| $\Delta Specificity_t$ | 0.096* | 0.090** | 0.012 | 0.010 | 0.046* | 0.043* | 0.003 | −0.004 |
| | (1.74) | (1.96) | (0.51) | (0.42) | (1.77) | (1.71) | (0.19) | (0.23) |
| $\Delta Numeric\ Intensity_t$ | 0.157 | 0.147 | 0.161** | 0.144* | 0.009 | −0.033 | 0.024 | −0.033 |
| | (0.90) | (0.93) | (2.12) | (1.83) | (0.08) | (0.30) | (0.51) | (0.66) |
| $\Delta Log(\#\ of\ Words/RF)_t$ | 0.737 | 0.777 | 0.454 | 0.755** | 0.286 | 0.303 | 0.117 | 0.222 |
| | (0.89) | (1.05) | (1.41) | (2.11) | (0.65) | (0.76) | (0.56) | (1.01) |
| $\Delta NI_t/AT_{t-1}$ | −1.555** | −1.268** | −0.462 | −0.649** | −0.696** | −0.583** | −0.249* | −0.180 |
| | (2.51) | (2.29) | (1.63) | (2.26) | (2.31) | (2.31) | (1.66) | (1.21) |
| $Log(Market\ Equity)_t$ | −1.401*** | −2.971*** | 0.032 | 0.097 | −0.129*** | −0.442*** | −0.001 | 0.118* |
| | (16.87) | (12.03) | (1.17) | (0.86) | (2.93) | (3.39) | (0.06) | (1.76) |
| $Big\ N_t$ | −2.334*** | 0.371 | 0.119 | −0.071 | −0.516*** | 0.268 | 0.076 | 0.176 |
| | (7.33) | (0.60) | (1.25) | (0.22) | (3.55) | (0.94) | (1.14) | (1.10) |
| $Book-to-Market_t$ | 0.509** | 0.173 | −0.151* | −0.116 | 0.236* | 0.128 | −0.135** | −0.069 |
| | (2.11) | (0.48) | (1.80) | (0.84) | (1.93) | (0.72) | (2.55) | (0.85) |
| $Tangibility_t$ | 0.075 | −2.754** | −0.092 | 0.434 | −0.199 | −1.057* | −0.070 | 0.211 |
| | (0.10) | (1.97) | (0.54) | (0.60) | (0.76) | (1.74) | (0.59) | (0.62) |
| $Leverage_t$ | 1.605*** | 2.324** | −0.297* | −0.845** | 0.188 | 0.015 | −0.125 | −0.353** |
| | (3.06) | (2.57) | (1.73) | (2.32) | (0.54) | (0.04) | (1.57) | (1.98) |
| $Turnover_t$ | −3.286*** | −3.409*** | 0.205*** | 0.333*** | −0.806*** | −0.686*** | 0.196*** | 0.102** |
| | (20.65) | (18.15) | (4.82) | (4.25) | (10.93) | (7.29) | (7.14) | (2.36) |
| $Beta_{t-1}$ | −5.783*** | −0.254 | 0.133* | −0.100 | −1.565*** | 0.011 | 0.362*** | 0.180*** |
| | (27.93) | (1.24) | (1.79) | (0.97) | (14.76) | (0.10) | (7.35) | (2.98) |
| $Excess\ Ret._{t-1}$ | −3.329*** | −1.796*** | 0.054 | 0.103 | −0.665*** | −0.349*** | 0.177*** | 0.120** |
| | (16.95) | (10.10) | (0.66) | (1.11) | (7.02) | (3.74) | (3.97) | (2.28) |
| $Ex.\ Ret.\ Std_{t-1}$ | 32.655*** | 15.011*** | −1.037** | −1.555** | 8.993*** | 3.341*** | −2.178*** | −0.611 |
| | (21.59) | (11.00) | (2.34) | (2.55) | (10.42) | (4.27) | (6.48) | (1.53) |
| $Ex.\ Ret.\ Skew_{t-1}$ | −0.075 | 0.011 | 0.064*** | 0.037 | −0.058** | −0.020 | 0.023* | 0.002 |
| | (1.31) | (0.26) | (2.91) | (1.58) | (2.09) | (0.82) | (1.81) | (0.14) |
| *Year F.E.* | Y | Y | Y | Y | Y | Y | Y | Y |
| *Industry F.E.* | Y | | Y | | Y | | Y | |
| *Firm F.E.* | | Y | | Y | | Y | | Y |
| # Observations | 23,992 | 23,992 | 23,981 | 23,981 | 23,992 | 23,992 | 23,980 | 23,980 |

for firm-specific heterogeneity, these results generally become insignificant. Further, when studying the change specifications, the significance disappears completely in both the cross-section and time-series. Again, with the caveat of not directly testing the null-hypothesis of no information environment effects, this does suggest that potentially the time-series evolution of risk factor disclosure identification does not have a strong influence in reducing information asymmetry or increasing liquidity. Interestingly, Specification (4) provides some evidence that increases in definitiveness may actually increase the bid-ask-spread, which would be consistent with the divergence argument of Kravet and Muslu (2013).

### 2.3.3 Comparison of Risk Disclosures Measures

The results from Table 2.4 support the first and second hypotheses, but it is unclear whether my measures of risk factors capture information differently than proxies used in previous literature. To test this, I compare one of the most used proxies in the literature, word count, to my measures of risk factor evolution. I repeat the analyses in in Table 2.4 and compare the number of words to the risk factors count variables. To separate my measure of risk factors from the word count, I decompose the word count into (*log(word count) – log(number of risk factors)*) and *log(number of risk factors)*.[25] I then compare the difference between a cross-sectional approach and the time-series approach I use in my main tests to further clarify the difference between the measures. The results are presented in Table 2.10. Specifications (1)–(3) present the results of a pooled estimation approach, and Specifications (4)–(6) further include controls for firm specific heterogeneity.

[25] The former being equivalent to *log(number of words / risk factor)*.

**Table 2.10**: Predicting Net Loss: Aggregate vs Risk Factor Measures

Table 2.10 reports results from probit regressions of future negative net income on risk factor measures, and controls of ex-ante risk. Specifications (1)–(3) present probit regressions. Specifications (4)–(6) present probit regressions with correlated random effects to absorb between firm variation. The coefficients represent average marginal effects, evaluated at the mean of the dependent variables. Industry is defined using the Fama-French 12 industry classification. The variables are defined in Appendix **??**. Robust standard errors are clustered at the firm level and absolute t-statistics are reported in parentheses. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 level, respectively.

| | Pooled Probit | | | Correlated Random Effects Probit | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $Log(\#\ of\ Words)_t$ | 0.066*** | | | 0.066*** | | |
| | (11.26) | | | (4.66) | | |
| $Log(\#\ of\ Words/RF)_t$ | | 0.063*** | 0.062*** | | 0.035 | 0.034 |
| | | (5.09) | (5.01) | | (1.37) | (1.34) |
| $Log(\#\ RF)_t$ | | 0.070*** | | | 0.080*** | |
| | | (9.94) | | | (5.04) | |
| $Log(\#\ New\ RF)_t$ | | | 0.026*** | | | 0.017*** |
| | | | (7.14) | | | (4.14) |
| $Log(\#\ Dropped\ RF)_t$ | | | −0.007* | | | −0.020*** |
| | | | (1.73) | | | (4.54) |
| $Log(\#\ RF)_{t-1}$ | | | 0.049*** | | | 0.039*** |
| | | | (6.91) | | | (2.58) |
| $Net\ Income/AT_t$ | −0.834*** | −0.834*** | −0.832*** | −0.202*** | −0.201*** | −0.204*** |
| | (23.70) | (23.70) | (23.75) | (5.29) | (5.27) | (5.37) |
| $Log(Market\ Equity)_t$ | −0.042*** | −0.041*** | −0.042*** | 0.007 | 0.007 | 0.006 |
| | (13.75) | (13.76) | (13.99) | (0.75) | (0.75) | (0.70) |
| $Big\ N_t$ | −0.002 | −0.002 | −0.001 | −0.000 | −0.000 | −0.000 |
| | (0.25) | (0.20) | (0.13) | (0.02) | (0.01) | (0.02) |
| $Book-to-Market_t$ | 0.026*** | 0.026*** | 0.025*** | 0.051*** | 0.051*** | 0.052*** |
| | (6.26) | (6.25) | (6.16) | (7.75) | (7.75) | (7.85) |
| $Tangibility_t$ | −0.015 | −0.015 | −0.016 | 0.166** | 0.170** | 0.169** |
| | (0.74) | (0.75) | (0.82) | (2.16) | (2.22) | (2.22) |
| $Leverage_t$ | 0.001 | 0.001 | 0.001 | −0.125*** | −0.124*** | −0.122*** |
| | (0.06) | (0.06) | (0.07) | (3.03) | (3.03) | (2.97) |
| $Turnover_t$ | 0.030*** | 0.030*** | 0.029*** | 0.013* | 0.013* | 0.014** |
| | (6.32) | (6.33) | (6.10) | (1.83) | (1.84) | (1.99) |
| $Beta_{t-1}$ | 0.046*** | 0.046*** | 0.047*** | 0.001 | 0.000 | 0.001 |
| | (7.63) | (7.66) | (7.76) | (0.08) | (0.05) | (0.18) |
| $Excess\ Ret._{t-1}$ | −0.062*** | −0.062*** | −0.062*** | −0.062*** | −0.062*** | −0.060*** |
| | (9.54) | (9.56) | (9.43) | (8.89) | (8.91) | (8.74) |
| $Ex.\ Ret.\ Std_{t-1}$ | 2.722*** | 2.731*** | 2.680*** | −0.267 | −0.255 | −0.219 |
| | (9.02) | (9.07) | (8.90) | (0.77) | (0.74) | (0.63) |
| $Ex.\ Ret.\ Skew_{t-1}$ | 0.000 | 0.000 | 0.000 | −0.000 | −0.000 | −0.000 |
| | (0.07) | (0.07) | (0.02) | (0.18) | (0.17) | (0.16) |
| *Year F.E.* | Y | Y | Y | Y | Y | Y |
| *Industry F.E.* | Y | Y | Y | | | |
| *Correlated R.E.* | | | | Y | Y | Y |
| # Observations | 20,967 | 20,967 | 20,967 | 20,318 | 20,318 | 20,318 |
| # of Events | 6,695 | 6,695 | 6,695 | 6,379 | 6,379 | 6,379 |

Two features stand out in the analysis. The first is that the results of a pooled cross-sectional approach and a time series approach can differ significantly. For example in specification (2), the words per risk factor is strongly significant, but becomes insignificant in specification (5) when accounting for firm fixed effects. This supports the claim in this study that when evaluating informativeness, it is also important to consider how the disclosure evolves within the firm, rather than only comparing disclosures across firms. Second, the results suggest that managers convey their information primarily through adding and removing distinct risk factors, rather than through the summed word count across all factors. These results are also consistent with the supposition that managers treat risk factors as distinct, where each inclusion provides a separate signal.

I perform the same test on the other adverse outcomes from Table 2.4, omitting tabulation of the controls for brevity. Table 2.11 compares the ability of word count and risk factor number in predicting sales declines and lawsuit intensity. The results are consistent with those from negative net income above, although in the cross section the total word count does not predict either outcome. For both outcomes, the number of risk factors is the dominant predictor, even after controlling for the same factors used in Table 2.4. Interestingly, while word count does not predict either outcome in the cross section, it does so once firm specific heterogeneity is controlled for with firm fixed effects. This further supports the conclusion that the informativeness of risk factors may be best measured in the time series rather than in the cross section.

Table 2.12 compares the ability of word count and risk factor number in predicting two market outcomes: cumulative abnormal returns (measured over a three day window starting on the filing day), and bid ask spread (averaged over the subsequent year after the filing). The latter measure is consistent with the bid-ask spread window

## **Table 2.11**: Comparing Measures: Adverse Outcomes Outcomes

Table 2.11 reports results from regressions of future adverse outcomes on risk factor measures and controls of ex-ante risk. The dependent variables are *Sales Decline*$_{t+1}$ and *Lawsuit Intensity*$_{t+1}$ in panels A and B respectively. The controls included are those found in Table 2.10, where *Net Income / AT* is replaced by *Sales*$_t$ */ AT* and *Lawsuit Intensity*$_t$ in panels A and B respectively. The coefficients in Panel A represent average marginal effects, evaluated at the mean of the dependent variables. Industry is defined using the Fama-French 12 industry classification. The variables are defined in Appendix **??**. Robust standard errors are clustered at the firm level and absolute t-statistics are reported in parentheses. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 level, respectively.

**Panel A**: Sales Decline

| | Pooled Probit | | | Correlated Random Effects Probit | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| *Log(# of Words)*$_t$ | 0.004 | | | 0.073*** | | |
| | (0.72) | | | (5.01) | | |
| *Log(# of Words/RF)*$_t$ | | −0.003 | −0.007 | | 0.050* | 0.037 |
| | | (0.27) | (0.60) | | (1.86) | (1.42) |
| *Log(# RF)*$_t$ | | 0.009 | | | 0.083*** | |
| | | (1.28) | | | (4.90) | |
| *Log(# New RF)*$_t$ | | | 0.008** | | | 0.013*** |
| | | | (2.25) | | | (2.97) |
| *Log(# Dropped RF)*$_t$ | | | 0.013*** | | | −0.008* |
| | | | (3.10) | | | (1.68) |
| *Log(# RF)*$_{t-1}$ | | | −0.005 | | | 0.062*** |
| | | | (0.71) | | | (3.85) |
| Industry F.E. | Y | Y | Y | | | |
| Correlated R.E. | | | | Y | Y | Y |
| # Observations | 20,967 | 20,967 | 20,967 | 20,318 | 20,318 | 20,318 |

**Panel B**: Lawsuit Intensity

| | Pooled OLS | | | Firm Fixed Effects OLS | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| *Log(# of Words)*$_t$ | 0.006 | | | 0.041*** | | |
| | (1.18) | | | (2.84) | | |
| *Log(# of Words/RF)*$_t$ | | 0.003 | −0.000 | | 0.009 | 0.002 |
| | | (0.26) | (0.01) | | (0.33) | (0.09) |
| *Log(# RF)*$_t$ | | 0.008 | | | 0.054*** | |
| | | (1.20) | | | (3.54) | |
| *Log(# New RF)*$_t$ | | | 0.007** | | | 0.012*** |
| | | | (2.10) | | | (3.08) |
| *Log(# Dropped RF)*$_t$ | | | 0.008** | | | −0.006 |
| | | | (2.17) | | | (1.27) |
| *Log(# RF)*$_{t-1}$ | | | −0.001 | | | 0.048*** |
| | | | (0.20) | | | (3.30) |
| Industry F.E. | Y | Y | Y | | | |
| Firm F.E. | | | | Y | Y | Y |
| # Observations | 25,278 | 25,278 | 25,278 | 25,278 | 25,278 | 25,278 |

studied in Campbell et al. (2014). The results in Panel A suggest that the number of risk factors is more predictive of abnormal returns in both the cross section and time series. The results in Panel B specification (1) are consistent with the findings of Campbell et al. (2014), who conclude that risk factors have a convergent effect on market beliefs, because they lower information asymmetry. However, once I add firm fixed effects, this result becomes positive and insignificant. Further, specifications (5) and (6) actually suggest the contrary: increased verbosity in risk factors leads to more uncertainty. As discussed above, this is consistent with the divergent beliefs argument of Kravet and Muslu (2013), whereby risk factors bring new information to markets as signals of increased uncertainty. While the disagreement between Panel B and the findings of Campbell et al. (2014) could be due to potential sample or control variable differences, the stark contrast between specifications (1) and (4) again cautions against attributing the cross sectional market results to individual firm disclosures.

**Table 2.12**: Comparing Measures: Adverse Market Outcomes

Table 2.12 reports results from regressions of future market outcomes on risk factor measures and controls of ex-ante risk. The dependent variables are $CAR_{3days}$ and $Bid - AskSpread_{+1\ year}$ in panels A and B respectively. The controls included are those found in Table 2.10, where *Net Income / AT* is replaced by the change in net income scaled by total assets, in both panels A and B. Industry is defined using the Fama-French 12 industry classification. The variables are defined in Appendix **??**. Robust standard errors are clustered at the firm level and absolute t-statistics are reported in parentheses. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 level, respectively.

**Panel A**: Cumulative Abnormal Returns

| | Pooled OLS | | | Firm Fixed Effects OLS | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $Log(\#\ of\ Words)_t$ | $-0.148^{**}$ | | | $-0.382^{*}$ | | |
| | (2.04) | | | (1.70) | | |
| $Log(\#\ of\ Words/RF)_t$ | | $-0.038$ | $-0.037$ | | $-0.197$ | $-0.116$ |
| | | (0.27) | (0.26) | | (0.53) | (0.31) |
| $Log(\#\ RF)_t$ | | $-0.199^{**}$ | | | $-0.516^{**}$ | |
| | | (2.25) | | | (2.15) | |
| $Log(\#\ New\ RF)_t$ | | | $-0.139^{**}$ | | | $-0.164^{**}$ |
| | | | (2.41) | | | (2.48) |
| $Log(\#\ Dropped\ RF)_t$ | | | 0.038 | | | 0.010 |
| | | | (0.60) | | | (0.13) |
| $Log(\#\ RF)_{t-1}$ | | | $-0.113$ | | | $-0.293$ |
| | | | (1.20) | | | (1.12) |
| Industry F.E. | Y | Y | Y | | | |
| Firm F.E. | | | | Y | Y | Y |
| # Observations | 24,000 | 24,000 | 24,000 | 24,000 | 24,000 | 24,000 |

**Panel B**: Bid Ask Spread

| | Pooled OLS | | | Firm Fixed Effects OLS | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $Log(\#\ of\ Words)_t$ | $-0.226^{***}$ | | | 0.027 | | |
| | (9.18) | | | (0.73) | | |
| $Log(\#\ of\ Words/RF)_t$ | | $-0.050$ | $-0.033$ | | $0.133^{**}$ | $0.136^{***}$ |
| | | (1.19) | (0.77) | | (2.51) | (2.63) |
| $Log(\#\ RF)_t$ | | $-0.296^{***}$ | | | $-0.008$ | |
| | | (8.75) | | | (0.20) | |
| $Log(\#\ New\ RF)_t$ | | | $-0.035^{***}$ | | | 0.007 |
| | | | (2.98) | | | (0.85) |
| $Log(\#\ Dropped\ RF)_t$ | | | $0.022^{*}$ | | | $-0.003$ |
| | | | (1.76) | | | (0.35) |
| $Log(\#\ RF)_{t-1}$ | | | $-0.294^{***}$ | | | $-0.055$ |
| | | | (9.06) | | | (1.30) |
| Industry F.E. | Y | Y | Y | | | |
| Firm F.E. | | | | Y | Y | Y |
| # Observations | 24,000 | 24,000 | 24,000 | 24,000 | 24,000 | 24,000 |

# — Chapter 3 —

# Demand for Risk Factor Disclosure

## 3.1 Hypothesis Development

The motivation for basing my measures and tests on the regulation is that both investors and regulators provide incentive for managers to adhere to the regulation. To capture the validity of this assumption, I test whether these two forces (public regulators and private investors) actually demand timely and definitive risk factors. Observing demand directly is difficult, therefore I employ a methodology similar to that used in Rogers and Van Buskirk (2009) in which I measure manager's disclosure response to a demand 'shock.' By observing how managers respond, I am able to draw conclusions about managers' beliefs as to what disclosures the investors or regulators are demanding. While this assumes that managers are sufficiently incentive-aligned to supply information to meet the demand, the costs associated with not doing so are sufficient to justify this assumption, as discussed below.

As argued in Chapter 2, one driving motivation for risk factor disclosures may be their potential benefit as a litigation shield. However, this same argument drives concern that "Management's fear of exposing itself to substantial liability may cause it simply to bury the shareholders in an avalanche of trivial information" (Justice Marshall, TSC Industries, Inc. v. Northway, 426 U.S. 438, 448-449, 1976). There are two primary reasons why this 'over-disclosure' outcome may not obtain: disclosure costs and regulatory oversight (discussed below). Firms face a plethora of disclosures costs, which they balance against the benefits accruing to disclosure. For risk factor disclosure, these costs could include proprietary costs stemming from competition (e.g. see Verrecchia, 2001), information acquisition or processing costs that increase in the amount of information provided (e.g. Grossman and Stiglitz, 1980; Sims, 2003; Bloomfield, 2002), or market costs associated with increased risk assessments (e.g. Heinle and Smith, 2015; Johnstone, 2016). When making the decision to disclose each risk factor, managers must trade off the benefit of disclosing that specific risk factor against its costs. This suggests that firms are unlikely to disclose completely-uninformative risk factors, or an overabundance of risk factors, because the costs of such disclosures will eventually outweigh the negligible marginal benefits. Consistent with this conclusion, Nelson and Pritchard (2016) find that during the voluntary regime (prior to 2005), risk disclosures were informative and limited in length. Short of disclosure overload, however, it is unclear ex ante whether investors view risk factors as effective litigation shields and thereby demand their disclosure. To test this, I use securities litigation as a direct demand for information from private investors.

Securities litigation as a private enforcement event can be significantly costly to a firm (Rose, 2008), thus managers are likely to attempt to reduce their litigation risk by disclosing information they believe investors demand (Skinner, 1994; White, 2013). If they believe risk factors provide a litigation shield, then a private enforcement event will signal deficient disclosures, and managers will increase their disclosure as a result. On the other hand, if managers believe disclosure causes litigation risk (Francis, Philbrick, and Schipper, 1994; Rogers and Van Buskirk, 2009), then managers will scale back their disclosures after a private enforcement event. However, given the legal practitioner guidance on the litigation value of risk factor disclosure (e.g. Robbins and Rothenberg, 2005), I expect that unlike voluntary disclosures, firms will increase their risk factor disclosures after a private enforcement event. As discussed above, court precedence requires risks factor disclosures to be definitive and specific to the risk they address. Therefore, I expect that in response to a private enforcement event, firms will increase both the number of risk factors they identify and the definitiveness of their disclosure.

> **H3:** Firms respond to private enforcement by identifying more risk factors and increasing their definitiveness.

Regulatory oversight also potentially plays a role in mitigating disclosure overload through the comment letter process, which Cox, Thomas, and Kiku (2003) describe as the "first line of defense against ongoing disclosure violations." Johnston and Petacchi (2017) find that comment letters often result in restatements and improvements in market information environment, but no subsequent changes to voluntary disclosure. Bozanic et al. (2015) find that firms improve both qualitative and quantitative features in their disclosures in response to an SEC comment letter. Brown et al. (2015) find that firms modify their risk factor disclosures after they or an influential peer (competitor or industry leader) receive a comment letter from the SEC pertaining to insufficient risk factor disclosures. Together, these studies provide strong evidence that the SEC comment letter process leads to disclosure improvements. Therefore, I use these SEC comment letters as an indicator of demand for information from regulators.

The incentives of public enforcement differ from those of investors (Cox et al., 2003). The demand for information from the SEC is reflected in their stated goals:[1]

> All investors, whether large institutions or private individuals, should have access to certain basic facts about an investment prior to buying it, and so long as they hold it. To achieve this, the SEC requires public companies to disclose meaningful financial and other information to the public. This

[1]From www.sec.gov/about/whatwedo.shtml

> provides a common pool of knowledge for all investors to use to judge for themselves whether to buy, sell, or hold a particular security. Only through the steady flow of timely, comprehensive, and accurate information can people make sound investment decisions.

However, the SEC is also resource and capacity constrained (Cox et al., 2003; Jackson and Roe, 2009). The information set available to the review process comprises the disclosures provided by the firm.[2] In the absence of manager's private information, it is unclear that the SEC reviewer has the capacity to identify novel omitted, yet significant, risk factors on any systematic basis. This is consistent with the claims made by the SEC: "The Division's review process is not a guarantee that the disclosure is complete and accurate."[3] However, they do have the ability to measure and enforce the definitiveness of disclosures they review, and require adherence to the wording of the regulation Item 503(c), specifically: *This discussion must be concise and organized logically. Do not present risks that could apply to any issuer or any offering. Explain how the risk affects the issuer or the securities being offered.* Because the SEC review process vis-à-vis risk factors is guided by this regulation, I expect the reviews to identify and request improvements in vague or generic disclosures. Thus, I predict that the SEC comment letters will result in improvements in the definitiveness of subsequent disclosures. However, given the resource and information constraints reviewers face, I do not expect that public enforcement will result in firms expanding the set of risk factors they identify.

> **H4:** Firms respond to public enforcement by increasing the definitiveness of risks they identify.

[2] See www.sec.gov/divisions/corpfin/cffilingreview.htm

[3] *Id., referring to the Division of Corporation Finance.*

While these hypotheses do not directly address information overload, the manager's reactions to both public and private enforcement events can provide evidence of their beliefs about stakeholders' views on information overload. Given that the enforcement actions I examine are punitive in nature, managers wishing to avoid future action will correct their disclosures consistent with the incentives driving the enforcement action. If investors or regulators believe risk factor disclosures are boilerplate and uninformative, managers should update and provide more detail in their disclosures to avoid future penalties. Ultimately, this is an empirical question, one that has not been completely addressed in the literature. While Brown et al. (2015) suggest that comment letters impact firms' disclosure, the proxies they use (cosine similarity and absolute value of the change in length) do not indicate directionality nor differentiate between risk factors. The directionality of the response is of primary importance in addressing information overload because it indicates the direction of the demand. Similarly, Beatty et al. (2015) find that firms increase their disclosure lengths after receiving a comment letter from the SEC. However, they also find that the correlation between disclosure of financial constraints and their measure of the underlying financial constraints decreases after SEC comment letter intervention. They conclude that 'less is more,' but do not address the questions of why the SEC would prompt firms for more disclosure if it is uninformative, nor what disclosure changes the SEC may be trying to elicit. The intent of analyzing firm reaction to public and private enforcement, specifically by differentiating between the evolution of risk factors and their definitiveness, is to shed light on this question of whether and to what extent investors and regulators value these disclosures.

## 3.2 Results

Given the results that managers disclose risk factors in a timely and informative manner, I then turn to how firms respond to demand for these disclosures. To test the third hypothesis, I measure private and public disclosure demand with indicator variables representing securities litigation and comment letters, respectively. I separately regress each of the six risk factor disclosure measures (the time series evolution and definitiveness variables) on the indicators equal to one if a public or private enforcement event occurred during the course of the fiscal year. I also include two lagged indicators of both enforcement events to capture the longevity of manager's responses to these events. This will capture the effect that demand for information has on disclosures in the subsequent three filings. In addition to the same controls used in the previous tables including firm fixed effects, I also include lagged values of the dependent variable to control for serial correlation in the variables.

The results in Table 3.1 suggest that managers respond to private enforcement by increasing the new risk factors they identify, which results in an increase in the total number of risk factors disclosed. In subsequent years, firms continue identifying more new risk factors and also increase the number of risk factors they remove from their disclosures, resulting in an overall increase in risk factors that persists for multiple years. Together, these results support my third hypothesis that private enforcement actions evoke the disclosure of more risk factors. Additionally, the subsequent increase in dropped risk factors is consistent with the suggestion in *Slayton vs. American Express* that risk factors should be continually updated. In contrast, Specification (4) suggests that managers do not believe that increasing the verbosity of risk factors is demanded by investors, which is consistent with the results in Table 2.12 Specification (6). Together, the results in Table 3.1 also support the conclusion that managers

**Table 3.1**: Enforcement Response: Risk Factors

Table 3.1 reports results from firm fixed effect regressions of risk factor evolution –total, new, and dropped– on public and private enforcement events. The variables are defined in Appendix **??**. Robust standard errors are clustered at the firm level and absolute t-statistics are reported in parentheses. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 level, respectively.

| | Log(# Total RFs) | Log(# New RFs) | Log(# Dropped RFs) | Log(# Words/RF) |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| *Sec. Litigation*$_{t-1}$ | 0.031*** | 0.178*** | 0.027 | 0.008 |
| | (2.67) | (4.38) | (0.76) | (1.21) |
| —$_{t-2}$ | 0.004 | 0.099** | 0.096** | 0.005 |
| | (0.44) | (2.48) | (2.48) | (0.83) |
| —$_{t-3}$ | 0.023*** | 0.145*** | 0.065* | −0.002 |
| | (3.05) | (4.04) | (1.87) | (0.30) |
| *RF Comment*$_{t-1}$ | 0.004 | 0.072*** | 0.063*** | 0.010*** |
| | (0.88) | (3.11) | (2.88) | (3.14) |
| —$_{t-2}$ | 0.005 | 0.006 | 0.004 | 0.005* |
| | (0.98) | (0.26) | (0.21) | (1.72) |
| —$_{t-3}$ | −0.005 | −0.050** | −0.026 | 0.005 |
| | (1.17) | (2.27) | (1.31) | (1.61) |
| $Log(\# RF)_{t-1}$ | 0.532*** | −0.438*** | 1.069*** | |
| | (33.00) | (10.38) | (27.15) | |
| $Log(\# New RF)_{t-1}$ | | −0.027*** | | |
| | | (3.81) | | |
| $Log(\# Dropped RF)_{t-1}$ | | | −0.091*** | |
| | | | (11.87) | |
| $Log(\# of Words/RF)_{t-1}$ | | | | 0.499*** |
| | | | | (30.52) |
| $Log(Market Equity)_t$ | 0.010*** | 0.010 | −0.050*** | 0.002 |
| | (2.66) | (0.68) | (3.65) | (0.93) |
| $Book-to-Market_t$ | 0.013*** | 0.037*** | −0.001 | 0.005*** |
| | (4.80) | (3.32) | (0.09) | (2.61) |
| *Comment (Any)*$_{t-1}$ | 0.005** | 0.031*** | 0.022** | 0.001 |
| | (2.03) | (2.79) | (2.22) | (0.87) |
| *Sales Growth*$_t$ | 0.018*** | 0.086*** | 0.067*** | −0.001 |
| | (3.42) | (3.15) | (2.68) | (0.28) |
| *Leverage*$_t$ | 0.070*** | 0.285*** | −0.017 | 0.042*** |
| | (4.74) | (4.59) | (0.34) | (4.99) |
| *Turnover*$_t$ | 0.006** | 0.079*** | 0.059*** | 0.003 |
| | (2.13) | (6.18) | (4.94) | (1.44) |
| *Excess Ret.*$_{t-1}$ | −0.003 | −0.018 | 0.039*** | −0.009*** |
| | (1.05) | (1.32) | (3.22) | (4.43) |
| *Ex. Ret. Std*$_{t-1}$ | 0.288* | 4.327*** | 2.706*** | 0.322*** |
| | (1.94) | (6.54) | (4.33) | (3.29) |
| *Ex. Ret. Skew*$_{t-1}$ | −0.002** | −0.000 | 0.004 | −0.000 |
| | (2.11) | (0.12) | (1.02) | (0.60) |
| *Year & Firm F.E.* | Y | Y | Y | Y |
| $R^2$ | 0.497 | 0.056 | 0.113 | 0.375 |
| # Observations | 23,209 | 23,209 | 23,209 | 23,201 |

believe risk factors to be a potential litigation shield. This result is in contrast to the study by Rogers and Van Buskirk (2009), which finds that firms actually decrease their provision of voluntary disclosure in response to shareholder litigation. My novel evidence of disclosure increase is significant because it points to a potentially less risky substitute for the litigation shield disclosures studied in previous literature (e.g., Skinner (1994, 1997; Francis et al., 1994; Field, Lowry, and Shu, 2005; Rogers and Van Buskirk, 2009; Cutler, Davis, and Peterson, 2016).

Table 3.1 also provides evidence consistent with the argument that the SEC comment letter process does not expand the set risk factors identified by firms. This may be due to the limited capacity or information set of the SEC (Cox et al., 2003; Jackson and Roe, 2009), or may indicate that the SEC does not dictate risk factor disclosures on a regular basis.[4] firms do increase the number of new risk factors they identify, but also increase the number they remove. Together with the evidence of no change in overall number, this may be an indicator of the SEC requesting improved risk factor organization or more detailed headings from firms.[5] The influence of the SEC seems to be short lived, as only the filing immediately following the comment letters are materially affected; subsequent filings return to the 'normal' firm level of disclosure updates. Specification (4) demonstrates that firms do increase the verbosity of their risk factor disclosures however. This result contrasts previous findings by Beatty et al. (2015) and Brown et al. (2015), who both argue that risk disclosure increases on average after an SEC comment letter. My results suggest a contrary view, that while the length of disclosure increases, the number of identified factors does not.

My third and fourth hypotheses predict an increase in definitiveness in response to public and private enforcement actions. To test this, I again regress the textual features on indicator variables for public and private enforcement actions for the previous three years. Rather than regressing these features and the controls on the

[4]A notable exception to this practice is the Cyber-security risk factor disclosure guidance issued in 2011. See: www.sec.gov/divisions/corpfin/guidance/cfguidance-topic2.htm.

[5]One example of this is the comment letter to Target referenced in Brown et al. (2015).

change in definitiveness, I allow for more flexible serial correlation by using the level as a dependent variable, with the lagged value as an additional control. These results are presented in Table 3.2.

**Table 3.2**: Enforcement Response: Textual Features

Table 3.2 reports results from firm fixed effect regressions of risk factor textual features on public and private enforcement events. The variables are defined in Appendix **??**. Robust standard errors are clustered at the firm level and absolute t-statistics are reported in parentheses. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 level, respectively.

| | Specificity | Number % | Words / RF |
|---|---|---|---|
| | (1) | (2) | (3) |
| *Sec. Litigation*$_{t-1}$ | 0.154* | 0.049* | 0.666 |
| | (1.87) | (1.82) | (1.03) |
| —$_{t-2}$ | 0.006 | 0.034 | 0.671 |
| | (0.07) | (1.21) | (1.12) |
| —$_{t-3}$ | −0.060 | −0.036 | −0.199 |
| | (0.83) | (1.32) | (0.39) |
| *RF Comment*$_{t-1}$ | 0.125** | 0.046*** | 0.895*** |
| | (2.49) | (2.86) | (2.95) |
| —$_{t-2}$ | −0.039 | 0.040*** | 0.373 |
| | (0.87) | (2.75) | (1.29) |
| —$_{t-3}$ | 0.015 | 0.002 | 0.352 |
| | (0.36) | (0.16) | (1.29) |
| *Specificity*$_{t-1}$ | 0.456*** | | |
| | (36.69) | | |
| *Numeric Intensity*$_{t-1}$ | | 0.448*** | |
| | | (38.94) | |
| *# of Words/RF*$_{t-1}$ | | | 0.501*** |
| | | | (26.98) |
| *Log(Market Equity)*$_t$ | −0.033 | −0.073*** | 0.176 |
| | (1.13) | (6.63) | (0.84) |
| *Book − to − Market*$_t$ | 0.008 | 0.012 | 0.329* |
| | (0.33) | (1.56) | (1.71) |
| *Comment (Any)*$_{t-1}$ | −0.005 | −0.005 | 0.032 |
| | (0.24) | (0.68) | (0.22) |
| *Sales Growth*$_t$ | −0.062 | −0.061*** | −0.204 |
| | (1.35) | (3.58) | (0.65) |
| *Leverage*$_t$ | 0.364*** | 0.178*** | 3.611*** |
| | (2.84) | (4.09) | (4.16) |
| *Turnover*$_t$ | 0.011 | 0.020** | 0.309* |
| | (0.47) | (2.14) | (1.77) |
| *Excess Ret.*$_{t-1}$ | −0.104*** | −0.020** | −0.864*** |
| | (4.01) | (2.18) | (4.63) |
| *Ex. Ret. Std*$_{t-1}$ | 4.471*** | 1.437*** | 30.408*** |
| | (3.47) | (2.84) | (3.38) |
| *Ex. Ret. Skew*$_{t-1}$ | 0.005 | −0.004 | −0.029 |
| | (0.59) | (1.54) | (0.56) |
| *Year & Firm F.E.* | Y | Y | Y |
| $R^2$ | 0.251 | 0.277 | 0.358 |
| # Observations | 23,201 | 23,201 | 23,201 |

The results in Table 3.2 provide only weak evidence in support of the third hypothesis. The marginally significant coefficients on Securities Litigation$_{t-1}$ in specifications (1) and (2) suggest that private enforcement, at most, elicits a weak increase in level of detail with which managers disclose their risk factors. Together with the evidence from Table 3.1, this suggests that managers believe the legal value of risk factors stems from the disclosure of individual risk factors, rather than the level of detail. One potential reason is that firms facing litigation already have a high level of specificity, such that managers interpret the private enforcement as a signal of insufficient identification of risk factors, rather than insufficient definitiveness. Alternatively, firms could be withholding specificity because they believe that providing definitive expositional disclosure opens them up to litigation risk. Consistent with the former explanation, firms facing a securities litigation have on average 8.2% and 5.1% higher levels of specificity and numeric intensity, respectively, when compared to the non-litigation firm years, and both differences are significant at the five percent level. However, in total, the results do not provide strong evidence that firms react to private enforcement by increasing the definitiveness of their risk factor disclosures. These results are in contrast to the regulatory focus and court precedence, which both highlight the importance of detail in cautionary language.

The results in Table 3.2 provide strong evidence consistent with the fourth hypothesis that managers react to public enforcement by increasing the definitiveness of their risk factor disclosures. Firms significantly increase the specificity, numeric intensity, and length of each risk factor in response to SEC comment letters. This effect seems to be temporary, with the increase only occurring in the disclosure immediately succeeding the comment letter, with the exception of numeric intensity, which persists for two years. The regressions control for lagged values of the dependent variables, thus effectively demonstrate an increase in the *change* of definitiveness, not whether these increases in definitive detail persist. To test the persistence, in an untabulated regression I repeat this analysis without controlling for the lagged values of the dependent variables, and find that these increases in definitiveness do persist and are significant for multiple years. This suggests that firms do not undo the improvements in definitiveness that the SEC evokes.

Together with Table 3.1, these results suggest that when managers receive an SEC comment letter about risk factor disclosure inadequacies, they respond by updating their disclosures in the immediately subsequent annual report, but then go back to their *normal* update procedure (i.e. the SEC does not cause firms to fundamentally to alter the underlying data generating process). However it is important to note that while a comment letter only temporarily changes how firms *update* their disclosures, these improvements to increased definitiveness persist: managers do not revert to the pre-comment letter disclosures. This is consistent with the expected costs of the SEC review process, which can evoke immediate changes, but pose few continuing costs once the *no further comment* letter is received. In contrast, private litigation does have a continuing effect on the *updates* that persists for multiple years, consistent with the expectation of future litigation costs being significantly higher (Rose, 2008). The results suggest that managers adjust the how they re-address their risk factors year over year (i.e. changing the underlying data generating process), for example by lowering the threshold of what qualifies as a *significant factor*, leading to an increase in total number of risks they identify. This contrast between the SEC evoking a 'one-off' change to disclosure updates and securities litigation evoking a more lasting change to disclosure updates is potentially of relevance in the ongoing legal literature on whether securities litigation is a useful governance mechanism (Rose, 2008).

### 3.2.1 Alternative Measure of Investor Demand

One potential concern with studying the response to a private enforcement event as a demonstration of demand for information is that managers may be responding to economic changes brought about by the litigation. This issue is less likely to affect the SEC comment letter results, because previous literature has not found significant changes in the economic behavior of firms in response to SEC comment letters (Ryans, 2017).[6] To ameliorate this concern, I therefore study a plausibly exogenous shock to investor demand previously used in the literature by employing the Russell index inclusion discontinuity. While this is not a perfect substitute, as it primarily captures the effect of sophisticated investors through increased institutional ownership, it at least is suggestive of the demand for risk factor disclosures from a significantly influential subset of investors (Crane, Michenaud, and Weston, 2016).

The identification strategy uses a regression discontinuity based on the Russell 1000 and 2000 indexes, which are published by Russell Investment and tracked by a large quantity of investor equity. The Russell 1000 tracks larger firms, but Chang, Hong, and Liskovich (2015) report that as of 2008, $168.6 Billion USD was indexed to the Russell 1000 while $236.7 Billion USD was indexed to the Russell 2000. The Russell indexes are rebalanced yearly by Russell Investments on June 30th. On the last trading day of May each year, the largest 3000 stocks[7] are ordered by market equity and two value weighted indexes are formed, the Russell 1000 containing the largest 1000, and the Russell 2000 containing the next 2000 smaller firms. The sequential

---

[6]Ryans (2017) actually finds higher earnings in the year preceding a comment letter, but no significant reduction in earnings in contemporaneous or subsequent years.

[7] Actual eligibility requires the security be incorporated/headquartered in the US, trade on at least one major US exchange, have a close price greater than $1 USD, market cap greater than $30 Million USD, have greater than 5% shares available for trade, and have an eligible corporate structure. Source and further details at `http://www.russell.com/indexes/documents/Methodology.pdf`.

ordering method of creating the indexes results in firm 1000 in the Russell 1000 and firm 1 in the Russell 2000 being almost identical in size. However the weight assigned to the largest Russell 2000 firm is 91.7 times larger (on average) than the smallest Russell 1000 firm. The result of this is that every institutional investor that tracks the Russell 2000 index holds the two stocks in vastly different proportions (if they hold the small Russell 1000 stock at all, transaction costs considered). This investment discrepancy is purely caused by the Russell indexing weights, thus provides for a setting in which a regression discontinuity design can be applied.

Regression discontinuity (RD) is a methodology which has been used in the Economics literature since Thistlethwaite and Campbell (1960) used it to study whether merit recommendations for schoolchildren led to scholarships and subsequent life performance. The intuition behind the methodology is similar to typical treatment vs control matching. If the cause of the discontinuity is unrelated (or even related but with uncertainty) to the variable being measured, then the assignment into treatment or control can be considered "as-if" random. In this case, the two firms on either side of the index split are almost identical, potentially one trading day away from switching places. However because of an insignificant difference in price on May 30th, one receives a huge increase in institutional investors. This institutional holding—unrelated to the fundamental characteristics of either firm—is arguably the only significant difference between the firms, thus the Russell 1000 firm can be used as a close match for how the Russell 2000 firm would have behaved had it not received investment. This allows for causal inferences that can be attributed specifically to increases in institutional investment.

I estimate the effect of exogenous change in institutional ownership using the RD design established in Calonico, Cattaneo, and Titiunik (2014) to estimate the bandwidth. I derive separate coverage error optimal bandwidths (see Calonico, Cattaneo, and Farrell, 2016), and estimate the difference across the continuity using their bias-adjusted robust methodology for calculating confidence intervals and statistics. The measure the effect of institutional investment on the risk factor measures I develop in this study, and present the results in Table 3.3.

**Table 3.3**: Risk Factor Disclosures and Institutional Investor Demand

Table 3.3 presents the results of the Russell index inclusion regression discontinuity. The results use the optimal bandwidth selection algorithm developed in Calonico et al. (2014). The variables are defined in Appendix **??**. Robust standard errors are clustered at the firm level and absolute t-statistics are reported in parentheses. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 level, respectively.

| Variable | Difference | Std. Err. | Z-statistic | $P(>\lvert Z\rvert)$ |
|---|---|---|---|---|
| New Risk Factors | 2.215** | 0.937 | 2.354 | 0.019 |
| Dropped Risk Factors | 3.364*** | 1.162 | 2.907 | 0.004 |
| Total Risk Factors | 6.359* | 3.553 | 1.764 | 0.078 |
| Specificity | −4.244*** | 1.424 | −2.936 | 0.003 |
| Numerical Intensity | −0.351 | 0.259 | −1.368 | 0.171 |
| Words per Risk Factor | −7.551 | 7.279 | −1.034 | 0.301 |
| Institutional Ownership | 0.336*** | 0.072 | 4.570 | 0.000 |

The results in Table 3.3 show an increase in institutional ownership is associated with an increase in the number of new and removed risk factors disclosed. This is generally consistent with the results in Table 3.1. Firms with high institutional ownership also disclose a larger total number of risk factors, which, while marginally significant, is consistent with the response to private enforcement results being an increase in the risks identified by managers. Somewhat consistent with Table 3.2, the increase in institutional ownership is not associated with quantitative definitiveness or verbosity, but is associated with lower specificity. One potential reason for this may be that firms suddenly adding six new risk factor disclosures (on average) could result in more generic disclosures, or less information being compounded into each risk factor. However another reason may be that the class of investors studied in the regression discontinuity is primarily indexing institutions and liquidity traders,[8] who may derive an informational advantage from private information. Lundholm (1991) suggests that increased precision in a public signal concentrates the demand for private information, potentially reducing the average returns to private information acquisition. As a whole, however, the results in Table 3.3 lend supporting evidence to the conclusions drawn from the private enforcement response tests.

### 3.2.2 Risk Factor Disclosures as a Litigation Shield

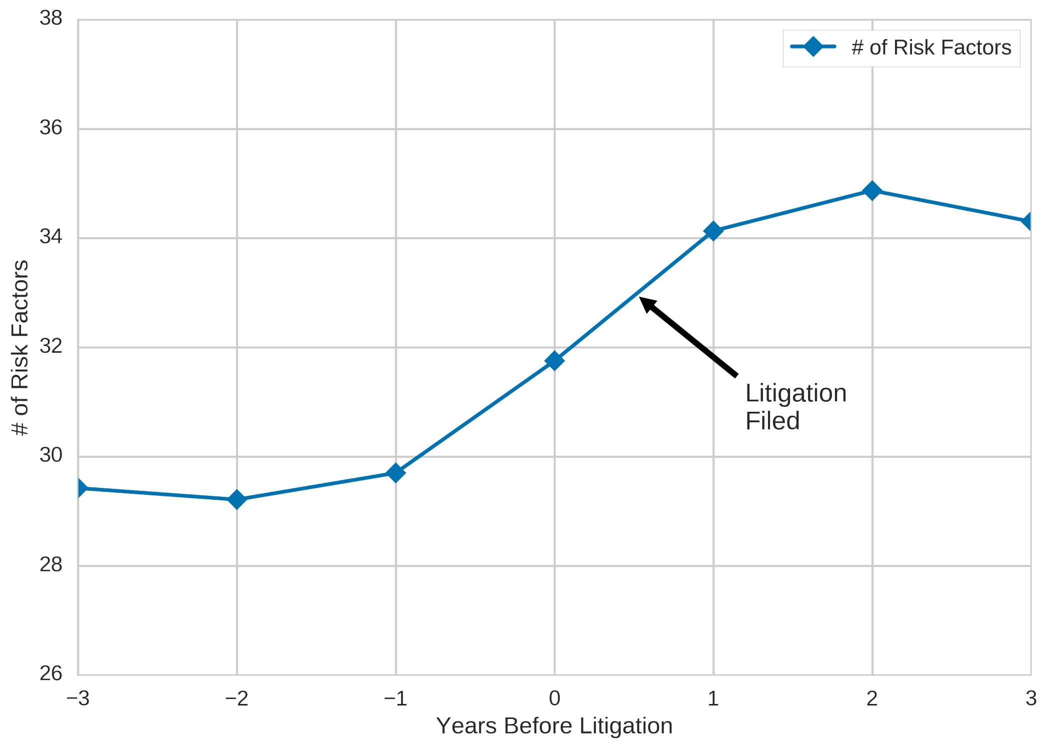
The results in Table 3.1 demonstrate that managers react to private enforcement by providing more disclosures, which suggests they believe risk factors to be an effective litigation shield. However, the extant literature has largely focused on voluntary disclosures when studying the litigation shield hypothesis; it is unclear whether risk factors, as a mandatory disclosure, would have the same legal benefit. To provide suggestive evidence of this, I test whether managers disclose risk factors in advance of litigation events rather than just in reaction to them. Similar to the motivation behind the tests of hypotheses one and two, if managers do believe risk factors provide a litigation shield, they will preemptively increase the disclosure of risk factors as the expectation of litigation increases. I initially provide graphical evidence of this by observing the time-series variation in risk factor disclosure around securities litigation events.

Figure 3.1 demonstrates the average number of risk factors in event time around the filing of a securities litigation. On average, it appears as if firms start increasing the number of risk factors they disclose in the immediately preceding annual report. The increase in risk factors appears to persist for an additional two years after the

[8]Specifically, quasi-indexers and transient investors in the Bushee (1998) classification.

**Figure 3.1**: Number of Risk Factors Around Securities Litigation Filing

Figure 3.1 plots the average number of risk factors in event-time. Year zero corresponds to the last annual report filed before the securities litigation was filed. Year one corresponds to the first annual report filed immediately after the litigation. Observations are averaged across all firms with securities litigations during the sample period with at least one annual report before and after the litigation filing.




litigation filing, suggesting a lasting response consistent with the results in Table 3.1. To bring more statistical rigor to this graphical evidence, I then test whether managers update their risk factor disclosure to preempt the securities litigation by employing a predictive framework similar to that used in Table 2.4. These tests are presented in Table 3.4.

The results in Table 3.4 suggest that managers do, on average, increase their risk factor disclosures in advance of securities litigation. Specification (1) includes the full sample of securities litigation events, including settled, dismissed, and ongoing lawsuits.

**Table 3.4**: Risk Factor Disclosure and Securities Litigation

Table 3.4 reports results from probit and linear regressions of securities litigation outcomes on risk factor disclosures and controls. Specifications (1)–(3) are run on the full sample of firm years; the dependent variable in specification (1) is all litigation events, while the dependent variable in specifications (2) and (3) is settled and dismissed litigation events respectively. Specification (4) is run on the subsample of litigation events, and the dependent variable is equal to 1 if the litigation is settled, and zero otherwise (dismissed or ongoing). Specifications (5) and (6) are performed on only settled litigations, and the dependent variable, class period length, is the number of days between the class start and class end as identified in the litigation. The coefficients in Specifications (1)–(4) represent average marginal effects, evaluated at the mean of the dependent variables. Industry is defined using the Fama-French 12 industry classification. The variables are defined in Appendix **??**. Robust standard errors are clustered at the firm level and absolute t-statistics are reported in parentheses. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 level, respectively.

| | All | Settled | Dismissed | Settled | Class Length | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $Log(\#\ New\ RF)_t$ | 0.003** | 0.002*** | −0.000 | 0.105*** | −94.815** | −94.240 |
| | (2.26) | (2.85) | (0.49) | (5.75) | (2.34) | (1.76) |
| $Log(\#\ Dropped\ RF)_t$ | −0.003** | −0.002** | −0.000 | −0.071*** | 1.585 | 11.034 |
| | (1.98) | (2.34) | (0.03) | (4.31) | (0.03) | (0.17) |
| $Log(\#\ RF)_{t-1}$ | 0.001 | 0.001 | −0.001 | 0.041* | −49.868 | −52.110 |
| | (0.58) | (1.16) | (0.69) | (1.76) | (0.71) | (0.67) |
| $Log(Market\ Equity)_t$ | 0.015*** | 0.006*** | 0.007*** | 0.010 | | 8.729 |
| | (8.16) | (5.49) | (5.51) | (0.89) | | (0.38) |
| $Book-to-Market_t$ | −0.002 | 0.001 | −0.002** | 0.087*** | | −84.053* |
| | (1.32) | (1.04) | (2.50) | (2.80) | | (2.00) |
| $Big\ N_t$ | 0.006 | 0.002 | 0.001 | −0.127** | | 5.367 |
| | (0.89) | (0.51) | (0.19) | (2.01) | | (0.05) |
| $Sales\ Growth_t$ | 0.006 | −0.000 | 0.003 | −0.086 | | −155.945 |
| | (1.56) | (0.17) | (1.42) | (1.18) | | (1.71) |
| $Leverage_t$ | −0.008 | 0.002 | −0.009* | −0.093 | | −258.378 |
| | (1.13) | (0.37) | (1.89) | (1.36) | | (1.36) |
| $Turnover_t$ | 0.004*** | 0.001 | 0.002* | −0.029 | | −102.124* |
| | (2.94) | (1.39) | (1.95) | (0.89) | | (2.13) |
| $Excess\ Ret._{t-1}$ | −0.007*** | −0.003** | −0.004*** | 0.009 | | −91.141 |
| | (3.50) | (2.44) | (2.81) | (0.28) | | (1.38) |
| $Ex.\ Ret.\ Std_{t-1}$ | 0.328** | 0.130* | 0.162* | 1.604 | | −797.140 |
| | (2.54) | (1.78) | (1.77) | (0.90) | | (0.15) |
| $Ex.\ Ret.\ Skew_{t-1}$ | −0.001 | 0.000 | −0.001** | 0.001 | | 29.729 |
| | (1.41) | (0.38) | (2.18) | (0.08) | | (0.86) |
| $Min.\ Excess\ Ret._{t-1}$ | 0.014 | 0.004 | 0.011 | 0.354 | | −613.120 |
| | (0.71) | (0.32) | (0.78) | (0.88) | | (0.70) |
| *Year F.E.* | Y | Y | Y | Y | Y | Y |
| *Industry F.E.* | | | | Y | Y | Y |
| *FPS Industry* | Y | Y | Y | Y | | Y |
| *Correlated R.E.* | Y | Y | Y | | | |
| # Observations | 29,413 | 29,413 | 29,413 | 497 | 160 | 160 |
| # Litigations | 491 | 156 | 233 | 160 | | |

On average, firms increase the number of new risk factors identified, and decrease the number removed before securities litigation filings. Specifications (2) and (3) of

Table use only settled and dismissed litigations, respectively, as a dependent variable. The results show that firms disclose increased risk factors in advance of litigation filings that eventually settle, but have no abnormal forewarning of litigation filings that are dismissed. This is consistent with some of the current academic legal opinion that many securities litigations are inaccurate in identifying fraudulent disclosures (Rose, 2008; Choi and Pritchard, 2016). If dismissed litigations are indicative of a 'scattershot approach' (Choi and Pritchard, 2016), then managers should not be expected to forecast these spurious cases. Specification (4) comprises just the reduced sample of firms that face a securities litigation, and use settled outcomes as the dependent variable, effectively comparing settled to dismissed (or ongoing) litigations. The results confirm those from the full sample, that managers provide significantly more risk factor disclosures in advance of settled cases. Together, these results suggest that managers seem to forecast securities litigation events, but any conclusions suffer from the endogeneity problem stemming from disclosure potentially *causing* litigation (Field et al., 2005).

I attempt to address the endogeneity issue in Specifications (5) and (6) by testing whether the inclusion of risk factors has an effect on the class period length. This is motivated by the method through which disclosure can act as a litigation shield (Skinner, 1994, 1997). Skinner argues that timely disclosures can potentially shorten the class period length because it provides the corrective disclosure, which ends the class period. If this is the case, then disclosing risk factors would be associated with shorter class period lengths.

I test whether timely risk factor disclosures reduce class period length by regressing the number of days between the class start period and end period as defined in the securities litigation on firm's risk factor disclosures. Specification (3) suggests managers may not be capable of (or have incentives to) disclose risk factors warning of dismissed suits, thus I focus on the subsample of 160 lawsuits which are eventually settled.[9] Specification (5) omits controls, effectively testing the correlation between disclosure and class period length, and finds that disclosing more risk factors is significantly associated with a shorter class period. This suggests that more risk factors may cause, or at least are associated with, shorter class periods. However, in Specification (6), which includes the controls used in the previous specifications, the significance on the relation between new risk factors and class period length becomes insignificant (p-value = 0.11).[10] This suggests that the increase in risk factors which cannot be explained by ex-ante risk are not associated with lower class period lengths, or conversely, that the lower class period length is related to correlated omitted factors in Specification (5). Together, this provides suggestive evidence consistent with the litigation shield value of risk factors.

[9]This test of only settled litigations potentially presents a selection problem, therefore I also perform an untabulated robustness test using a Heckman selection model for Specification (5) where the selection for settled versus dismissed is a function of the controls from Specification (6). I find similar results on the coefficient estimate, with lower power (p-value = 0.052).

[10]Because the point estimate is quantitatively similar however, the difference may stem from lack of power given the lowered degrees of freedom.

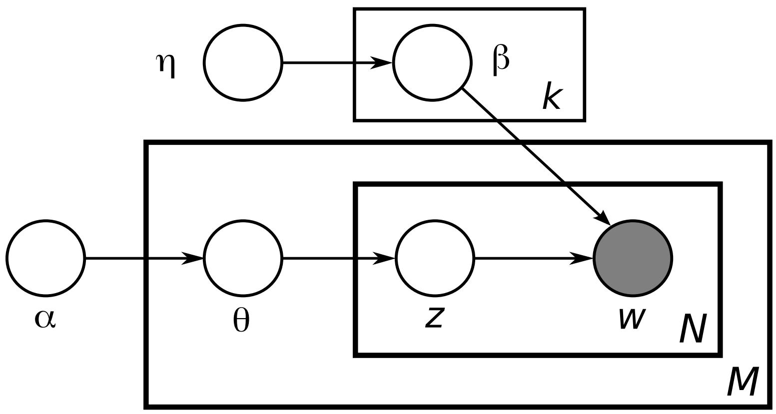
# — Chapter 4 —
# Risk Factor Disclosure Topics

The previous sections treat individual risk factors as fungible, namely that they have equal weighting in their disclosure, as well as their predictive ability of adverse outcomes. This assumption is unlikely to hold, as multiple risk factors commonly address multiple different risks of adverse events which potentially have differing impacts on For example a disclosure about the risk surrounding a new drug approval addresses a dichotomous outcome based on the regulatory decision, while a disclosure about cyber-security risk might warn of detrimental impact to a firm's public image. To investigate to what extent the conclusions of predictive ability of risk factors generalizes to the various different risks being disclosed, I use textual analysis techniques to classify individual risk factor disclosures.

To classify risk factor disclosures and determine to what risks they pertain, I use the latent Dirichlet allocation (LDA) technique to discover and extract the most prominent topics. The approach of using LDA to model topics in textual datasets originated with Blei, Ng, and Jordan (2003). Blei et al. (2003) describes a generative probabilistic model in which documents are composed of a distribution of topics, which themselves are composed of a distribution over the underlying words. The underlying idea is that different topics have a linguistic 'fingerprint' corresponding to the words commonly used within the topic's domain, and a document can be a mixture of topics. The corresponding plate model diagram of LDA is shown in Figure 4.1.

**Figure 4.1**: Graphical plate diagram representation of smoothed LDA model

Figure 4.1 shows a 'plate' diagram representation of the smoothed LDA model developed by Blei et al. (2003). The boxes, or 'plates' represent the objects modeled in the generative LDA approach. In this diagram, the outer plate **M** represents the document level, and the inner plate **N** represents topics (i.e. distributions of words) within the documents.




The model, as developed in Blei et al. (2003) and implemented in this study, assumes the corpus has **M** documents each containing **N** words,[1] which come from a dictionary of **V** total 'known' words. The model has the following steps for each document. First, the distribution of topics, $\theta$ is chosen from a Dirichlet distribution over prior $\alpha$. $\theta$, a $k$-dimension vector, is effectively the true distribution of topics in the document, from which the generation of each word will sample. Second, the distribution of words conditional on the topic, $\beta$ is chosen from a Dirichlet distribution over the prior $\eta$ (a scalar).[2] $\beta$, a $k$ by **V** matrix, is the probability of a given word occurring in a given topic, essentially a look-up table for word occurrence probability conditional on the topic. Lastly, for each word $i$ in the document, a random topic,

[1] **N** can be a random variable which is different for each document, but is independent of the other variables in the model, thus can be considered constant without loss of generality.

[2] The initial model presented assumes that $\beta$ is exogenous, but the smoothed version presented here smooths probabilities across all words in the corpus, including those not seen in training by assuming $\beta$ itself is a random variable to be estimated.

$z_i$, is selected from the prior $\theta$, and the word $w_i$ is selected conditional on this topic. This last step is what differentiates the LDA model from more traditional hierarchical Bayesian models, in that it chooses a topic for *each word*, instead of choosing a topic for *each document*. This feature allows for a document to comprise multiple topics, each with a 'loading' or weight in the document. This results in the joint distribution of latent topic and word loading, as well as observed topics and words:

$$p(\theta, \beta, z, w|\alpha, \eta) = p(\theta|\alpha)p(\beta|\eta)\prod_{n=1}^{N} p(z_n|\theta)p(w_n|z_n, \beta) \tag{4.1}$$

To extract topics from a corpus, which in this study is the set of risk factor disclosures, the researcher must make some discretionary assumptions about $k$, $\alpha$, and $\eta$. The primary choice is the assumption of how many topics contained in the corpus, $k$. This choice is one faced in many taxonomies, specifically how and where to draw divisions between groups. One example of this is the Risk Factor Landscape study from the IRRC Institute (Lukomnik, 2016), which classifies risk factor disclosures into 17 categories. However their classification does not have any domain specific risks. To allow for the potential of domain specific topics to be extracted, I use 25 topics in my LDA analysis.

Another discretionary choice to make in the extraction of textual topics is the definition of document. One approach might be to use the whole risk-factor section as the document. However this would require the algorithm to handle documents containing all the risk factor topics simultaneously, with the majority of differences potentially being domain specific. I take a more sparse approach, and define the document at the individual risk factor level. This allows for a more sparse topic mixture in each document, and potentially allows for more even mixtures of generic risks and domain-specific risks. For example competition is a ubiquitous risk factor, but each industry may discuss it differently, potentially as some mixture between a domain topic and the competition topic.

To estimate the LDA parameters, I use the publicly available gensim software.[3] I scale the bag-of-word vectors by the inverse-document frequency, a method developed by Salton and McGill (1986) commonly adopted and referred to as *tf-idf* (term-frequency inverse document-frequency: word count / number of documents containing the word). I also remove those words which appear in more than half of the risk factors (eliminating frequent words like and, the, etc.), and words occurring in fewer than 200 risk factors, resulting in a final dictionary containing 6,950 words. I use priors for $\alpha$ and $\eta$ empirically based on the data, and perform multiple passes over the corpus. The result of this analysis are displayed in Appendix C, along with a 'name' of the topic based on researcher judgment from the list of most frequent words.[4]

I classify risk factors a pertaining to only one topic, namely that with the highest posterior probability. This may obfuscate the presence of a more weakly discussed risk factor, but it is more consistent with my undergirding assumption in this study that individual risk factor disclosures pertain to specific adverse outcomes, for which one topic plausibly applies. I then compute the same measures of risk factor disclosures, (new, removed, and maintained) for each LDA topic. Table 4.1 presents the average percentage of a firm's disclosed risk factors for each topic, over time. There are some patterns in the distribution of risk factors over topics which evolve over time. For example the *Financing* topic, which discusses equity and real-estate related risks, significantly increases during the financial crisis in 2008. However other factors, for

[3] Available at github.com/RaRe-Technologies/gensim

[4] The name of the topic is for referential convenience, and has no bearing on the analysis.

example the oil price crashes of 2008 and 2014 are not reflected in a material change in the level of discussion of these risks.

**Table 4.1**: Portion of Total Risk Factors by Topic Across Time

Table 4.1 reports the average percentage of risk factors a firm has in a given topic, by fiscal year. Topics are described in Appendix C.

| | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 |
|---|---|---|---|---|---|---|---|---|---|---|
| *Healthcare* | 11.31 | 11.85 | 12.11 | 13.52 | 13.81 | 14.12 | 14.51 | 15.10 | 15.41 | 15.70 |
| *Energy* | 19.26 | 19.30 | 18.86 | 18.85 | 18.92 | 18.85 | 18.65 | 18.11 | 18.25 | 18.03 |
| *Drugs* | 17.65 | 17.64 | 17.13 | 16.51 | 16.33 | 16.51 | 16.15 | 16.08 | 15.95 | 15.60 |
| *Personnel* | 10.27 | 10.27 | 10.20 | 9.62 | 9.29 | 9.18 | 9.03 | 9.04 | 8.83 | 8.67 |
| *Retail* | 4.90 | 4.82 | 4.65 | 4.35 | 4.26 | 4.22 | 4.18 | 4.07 | 4.01 | 3.93 |
| *InformationTechnology* | 2.65 | 2.54 | 2.51 | 2.24 | 2.12 | 2.12 | 2.06 | 2.00 | 2.03 | 1.98 |
| *Claims/Security* | 1.17 | 1.18 | 1.19 | 1.00 | 0.93 | 0.96 | 0.93 | 0.94 | 0.98 | 0.99 |
| *Environ./Regulate.* | 0.74 | 0.76 | 0.81 | 0.74 | 0.73 | 0.74 | 0.76 | 0.79 | 0.80 | 0.89 |
| *Internat. Invest.* | 0.34 | 0.32 | 0.29 | 0.28 | 0.32 | 0.30 | 0.33 | 0.32 | 0.32 | 0.34 |
| *Accounting* | 0.24 | 0.24 | 0.25 | 0.24 | 0.27 | 0.25 | 0.24 | 0.25 | 0.27 | 0.27 |
| *Debt* | 0.22 | 0.23 | 0.22 | 0.31 | 0.24 | 0.24 | 0.24 | 0.29 | 0.28 | 0.30 |
| *Customers* | 0.31 | 0.34 | 0.38 | 0.42 | 0.43 | 0.42 | 0.45 | 0.49 | 0.51 | 0.56 |
| *Financing* | 0.20 | 0.24 | 0.26 | 0.43 | 0.47 | 0.48 | 0.50 | 0.51 | 0.47 | 0.49 |
| *Hedging* | 0.59 | 0.68 | 0.75 | 0.83 | 0.84 | 0.89 | 0.91 | 0.97 | 0.99 | 1.00 |
| *IP* | 0.25 | 0.29 | 0.31 | 0.30 | 0.32 | 0.33 | 0.32 | 0.35 | 0.40 | 0.38 |
| *Internal Control* | 0.26 | 0.29 | 0.31 | 0.29 | 0.30 | 0.31 | 0.33 | 0.34 | 0.33 | 0.36 |
| *Stock Market* | 0.49 | 0.55 | 0.58 | 0.54 | 0.54 | 0.54 | 0.56 | 0.60 | 0.71 | 0.70 |
| *StrategicAlliance* | 0.13 | 0.13 | 0.15 | 0.15 | 0.14 | 0.13 | 0.15 | 0.17 | 0.17 | 0.20 |
| *Competition* | 0.00 | 0.08 | 0.09 | 0.11 | 0.10 | 0.09 | 0.11 | 0.10 | 0.11 | 0.12 |
| *Equity/Dividends* | 0.00 | 0.02 | 0.04 | 0.03 | 0.05 | 0.05 | 0.06 | 0.05 | 0.05 | 0.04 |
| *Governance* | 0.97 | 1.06 | 1.12 | 1.11 | 1.10 | 1.22 | 1.20 | 1.20 | 1.17 | 1.16 |
| *Equity Listing* | 0.00 | 0.02 | 0.06 | 0.04 | 0.04 | 0.04 | 0.04 | 0.03 | 0.05 | 0.08 |
| *Drugs (Generic)* | 0.00 | 0.03 | 0.03 | 0.04 | 0.04 | 0.03 | 0.03 | 0.03 | 0.04 | 0.04 |
| *Supplier* | 0.19 | 0.19 | 0.19 | 0.17 | 0.20 | 0.23 | 0.18 | 0.20 | 0.20 | 0.21 |
| Observations | 2,440 | 3,189 | 3,326 | 3,444 | 3,424 | 3,264 | 3,221 | 3,261 | 3,203 | 3,154 |

The intent of focusing on individual risk factor topics is to lend robustness to the conclusion that firms disclose risk factors to predict adverse outcomes. The tests have focused on firm-specific adverse outcomes, based on the argument that a sufficiently adverse event could detrimentally impact the operations of the firm and result in a the bottom line outcomes I focus on (net loss, etc.). One benefit of focusing on the bottom line is that researcher discretion is not required to relate specific risk disclosures to their appropriate firm accounts. Such discretionary decisions are required, however, when studying the risk *topics* being disclosed, because the channel through which the disclosure forecasts an event becomes specific to the risk disclosure topic. For example a risk factor disclosure under the topic *Competition* could pertain to reduced revenue stemming from increased price competition, or it could pertain to a reduced profit margin stemming from increased upstream prices. Linking the topic of disclosure to firm-specific accounts is potentially confounded by the idiosyncratic nature by which a topic may relate to the firm.

One firm-level adverse outcome which is plausibly related to an extracted topic is the presence of a goodwill impairment. Since the introduction of SFAS 142 in 2001, firm's goodwill has to undergo a yearly impairment test.[5] Failing this test can be significantly detrimental to the firm's net income, thus early warning of potential impairments could provide legal benefit to managers. Hayn and Hughes (2006) suggests that the economic indicators of goodwill impairment lead the accounting write-off, sometimes by multiple years. This suggests that managers may reasonably predict the presence or probability of future adverse impairment test outcomes. This potential for predictability of goodwill impairments presents an interesting setting in which I can test whether managers do systematically warn of these adverse events through the risk factor disclosures, and specifically those topics pertaining to goodwill and impairments: *Strategic Alliance* and *Accounting*. The results of these tests, mirroring those performed in Table 2.4, are presented in Table 4.2.

The results in Table 4.2 suggest that firms increase their disclosure of risks relating to *Strategic Alliance* (e.g. M&A activity) in advance of a goodwill write-off However there is not significant evidence of an increase in the discussion of *Accounting* terms, which includes goodwill and impairments. Given the findings in Hayn and Hughes (2006) that economic indicators of goodwill impairments precede the observed write-off

[5]See www.fasb.org/summary/stsum142.shtml.

**Table 4.2**: Risk Factors and Future Goodwill Impairments

Table 4.2 reports the average marginal effects from a regression of the presence of a goodwill impairment on risk factor disclosures under the *Strategic Alliance* (*SA*) and *Accounting* topics. The topics are described in Appendix C. The sample omits repeated goodwill impairment observations, similar to the filter in Table 2.6. The $Log(\Delta\#SA)$ and $Log(\Delta\#Acct.)$ coefficients are calculated as the log of one plus the difference in new and removed risk factor disclosures of the respective topics, less the minimum difference to maintain a positive value in the logarithm function. The controls included are those used in Table 2.4: *Log(Market Equity)*, *Big N*, *Book-to-Market*, *Tangibility*, *Leverage*, *Share Turnover*, *Beta*, and *Excess Returns*, *Standard Deviation*, and *Skewness*. The variables are defined in Appendix **??**. Robust standard errors are clustered at the firm level and absolute t-statistics are reported in parentheses. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 level, respectively.

| | Goodwill Impairment (1) | Goodwill Impairment (2) | Goodwill Impairment (3) |
|---|---|---|---|
| $Log(\#\ New\ RF)_t$ | 0.025*** | | |
| | (6.92) | | |
| $Log(\#\ New\ SA)_t$ | | 0.074** | |
| | | (2.16) | |
| $Log(\#\ New\ Acct.)_t$ | | 0.031 | |
| | | (1.07) | |
| $Log(\Delta\#SA)$ | | | 0.072** |
| | | | (2.08) |
| $Log(\Delta\#Acct.)$ | | | 0.036 |
| | | | (1.17) |
| $Log(\#\ Dropped\ RF)_t$ | −0.024*** | | |
| | (6.14) | | |
| $Log(\#\ Dropped\ SA)_t$ | | −0.050 | |
| | | (0.95) | |
| $Log(\#\ Dropped\ Acct.)_t$ | | −0.029 | |
| | | (0.82) | |
| $Log(\#\ RF)_{t-1}$ | 0.108*** | | |
| | (7.59) | | |
| $Log(\#\ SA)_{t-1}$ | | 0.014 | 0.003 |
| | | (0.52) | (0.12) |
| $Log(\#\ Acct.)_{t-1}$ | | 0.026 | 0.021 |
| | | (1.37) | (1.18) |
| *Controls* | Y | Y | Y |
| *Year F.E.* | Y | Y | Y |
| *Correlated R.E.* | Y | Y | Y |
| # Observations | 21,969 | 21,617 | 21,617 |

by multiple years, it may be the case that firms disclose of goodwill impairments more than one year in advance. To test this, in untabulated results, I add a second lag of the accounting topic disclosures, and find a significant positive relationship (T-stat=2.19), suggesting managers potentially warn of impending impairments further in advance.

To further provide suggestive evidence that managers warn of adverse outcomes, I test whether managers forecast macro level outcomes. Specifically, I focus the price of oil as proxy for a potentially significant adverse outcomes. Because many industries are affected by the price of oil, the expected cost of significant shifts in the price of oil may be sufficient to warrant a risk factor disclosure. Testing the disclosure of risks relating to the *Energy* topic against future changes in the price of oil lends evidence as to whether managers are incorporating their expectations about future outcomes in their disclosure decisions. The tests of the change of oil price on risk factors are presented in Table 4.3.[6]

The results in Table 4.3 suggest that on average, firms disclose more overall risk factors in advance of relative oil price declines.[7] This relation also exists for the *Energy* topic, suggesting firms disclose more risk factors specifically about energy topics in advance of price declines. Similarly, firms remove more risk factors in advance of subsequent oil price increases, consistent with managers 'informatively' updating their set of disclosed risk factors by actively removing obsolete risk factor disclosures. The net increase in the disclosure of risk factors relating to the *Energy* topic is also negatively associated with oil price changes, suggesting firms disclose overall more disclosure about energy related risk factors ahead of price declines.[8]

In general, the results of robustness tests conducted on the individual topic level are somewhat mixed. Some of the results are consistent with expectations linking topic to outcome, for example Goodwill and Oil prices. However in untabulated

[6]The oil price is measured using the end of month spot price for West Texas Intermediate (WTI), a common benchmark for oil pricing. Data are downloaded from the U.S. Energy Information Administration. Url: www.eia.gov/dnav/pet/pet_pri_spt_s1_d.htm.

[7]Given the nature of a fixed effect regression, declines are relative to the average change in oil price.

[8]These results include the measures used in Table 2.4 to control for ex-ante expected changes in the risk factors, but are robust to the exclusion of these controls, with the exception of the coefficient on new *Energy* risk factor disclosures becoming insignificant in Specification (2).

**Table 4.3**: Risk Factors and Future Oil Price

Table 4.3 reports the results of a regression of the change in oil price on risk factor disclosures under the *Energy* topic. The topic is described in Appendix C. $\Delta$*Oil Price* is measured as the difference between the monthly price of oil 3 months after the fiscal year end, less the price of oil at the fiscal year end. The controls included are those used in Table 2.4: *Log(Market Equity)*, *Big N*, *Book-to-Market*, *Tangibility*, *Leverage*, *Share Turnover*, *Beta*, and *Excess Returns*, *Standard Deviation*, and *Skewness*. The variables are defined in Appendix **??**. Robust standard errors are clustered at the firm level and absolute t-statistics are reported in parentheses. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 level, respectively.

| | $\Delta$*Oil Price* | $\Delta$*Oil Price* | $\Delta$*Oil Price* |
|---|---|---|---|
| | (1) | (2) | (3) |
| *Log(# New RF)$_t$* | $-0.634^{***}$ | | |
| | (5.68) | | |
| *Log(# New Energy)$_t$* | | $-0.349^{**}$ | |
| | | (2.18) | |
| *Log($\Delta$#Energy)* | | | $-0.914^{***}$ |
| | | | (2.88) |
| *Log(# Dropped RF)$_t$* | $0.769^{***}$ | | |
| | (6.00) | | |
| *Log(# Dropped Energy)$_t$* | | $0.425^{**}$ | |
| | | (2.21) | |
| *Log(# RF)$_{t-1}$* | $-5.448^{***}$ | | |
| | (16.52) | | |
| *Log(# Energy)$_{t-1}$* | | $-1.701^{***}$ | $-1.690^{***}$ |
| | | (7.63) | (7.63) |
| *Controls* | Y | Y | Y |
| *Firm F.E.* | Y | Y | Y |
| # Observations | 25,445 | 25,044 | 25,044 |

results, replacing the aggregated risk factor measures in Table 2.3 with the topic-level measures results in only a subset of topics being significantly predictive. This may suggest that only some topics have systematic relationships with the adverse outcomes I test in this study. An alternative interpretation is that some topics are less informative of short-term cash-flow outcomes, and the predictive framework employed herein is inflexible in accounting for differing horizons of forecasting. While future research looking at predictability by topic might find that specific subsets of risk factor disclosures are more predictive of specific outcomes, it could also be the case that aggregate change in risk factors is a more informative measure because it provides a more complete picture of managers assessed (and disclosed) beliefs. However the evidence presented herein suggests that both the aggregate and individual topic level risk factors are disclosed in a manner consistent with the hypothesis that managers disclose individual risk factors to warn of specific adverse outcomes.

# — Chapter 5 —
# Conclusion

This study addresses the question of whether risk factor disclosures are informative by reconsidering both the measure used to capture the risk factors, and the test employed to determine 'informativeness.' My results suggest managers do disclose timely and definitive risk factors to warn of adverse outcomes, in accordance with the regulatory requirement. I find that the addition and removal of risk factors predicts future adverse economic outcomes even after controlling for ex-ante risk and firm performance, and that my measure of risk factors provides superior predictive ability over a popular measure used in extant literature. My approach differs from previous literature in that researchers have typically aggregated the risk disclosure using a bag of words approach, and primarily focused on market outcomes. My results suggest that treating risk factors as distinct units may be a more faithful representation of the information being conveyed by managers.

Given the validation of this time-series approach to measuring the information in risk factors, I use my approach study the demand for these disclosures from two significant sources: investors and regulators. I focus on demand 'shocks' through public enforcement via SEC comment letters, and private enforcement from investors via securities litigation. Consistent with my predictions, I find that firms respond to private enforcement by expanding the set of risk factors they identify, but surprisingly do not increase the definitiveness of those disclosures. In contrast, I find firms respond to public enforcement not by increasing their identified risk factors as suggested in previous literature, but by improving their level of specificity and detail. In line with the cost benefit tradeoff to managers, I find that the response to private enforcement actions persists for multiple years, while public enforcement actions generally only effect changes in the subsequent filing. Together, these results shed light on the beliefs of managers as to the demand for risk factors disclosures from both regulators and investors. These results suggest that managers believe risk factor disclosures to be effective litigation shields for investors, and I provide evidence consistent with this belief.

My study offers three main takeaways. First, my paper suggest that the SEC effectively elicits more detail in disclosures. While this may lead to conclusions in other studies that the length of disclosure is less correlated with underlying economic activity, my findings suggest that this focus on length may be falsely attributing regulatory impact to less informative disclosures. Additionally, my results point to an interesting role that risk factor disclosures may play as litigation shields, potentially substituting for voluntary information when the cost of the latter increases. Second, this study demonstrates that managers update their disclosures in a timely fashion, suggesting that they are not merely ‘copy and pasting.’ I find that firms are informatively choosing when to add and, importantly, remove information, and actively supply disclosure to meet information demand. These results suggest caution against interpreting the increasing length of disclosure as a sufficient signal of information overload; in some cases, a more targeted measure of information may be warranted. Last, my study demonstrate the potential value of considering the time series evolution of disclosures, and carefully constructing textual measures based on the data generating process underlying the disclosure.

# Bibliography

Amihud, Yakov, 2002, Illiquidity and stock returns: Cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.

Bao, Yang, and Anindya Datta, 2014, Simultaneously Discovering and Quantifying Risk Types from Textual Risk Disclosures, *Management Science* 60, 1371–1391.

Beatty, Anne, Lin Cheng, and Haiwen Zhang, 2015, The Information Content of Differences in Financial Constraints Risk Disclosure Requirements, *Working Paper* Ohio State University.

Bird, Andrew, and Stephen A. Karolyi, 2016, Do Institutional Investors Demand Public Disclosure?, *Review of Financial Studies* 29, 3245–3277.

Blackwell, David, 1951, Comparison of experiments, in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, 93–102.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan, 2003, Latent Dirichlet Allocation, *J. Mach. Learn. Res.* 3, 993–1022.

Bloomfield, Robert, 2008, Discussion of "Annual report readability, current earnings, and earnings persistence", *Journal of Accounting and Economics* 45, 248–252.

Bloomfield, Robert, Mark W. Nelson, and Eugene Soltes, 2016, Gathering Data for Archival, Field, Survey, and Experimental Accounting Research, *Journal of Accounting Research* 54, 341–395.

Bloomfield, Robert J., 2002, The Incomplete Revelation Hypothesis and Financial Reporting, *Accounting Horizons* 16, 233–243.

Boone, Audra, and Joshua White, 2015, The effect of institutional ownership on firm transparency and information production, *Journal of Financial Economics* 117, 508–533.

Bozanic, Zahn, J. Richard Dietrich, and Bret A. Johnson, 2015, The SEC Comment Letter Process and Firm Disclosure, *Working Paper* The Ohio State University.

Brown, Stephen V., Xiaoli Shaolee Tian, and Jenny Wu Tucker, 2015, The Spillover Effect of SEC Comment Letters on Qualitative Corporate Disclosure: Evidence from the Risk Factor Disclosure, *Working Paper* The Ohio State University.

Brown, Stephen V., and Jennifer Wu Tucker, 2011, Large-Sample Evidence on Firms' Year-over-Year MD&A Modifications, *Journal of Accounting Research* 49, 309–346.

Bushee, Brian J., 1998, The Influence of Institutional Investors on Myopic R&D Investment Behavior, *The Accounting Review* 73, 305–333.

Calonico, Sebastian, Matias D. Cattaneo, and Max H. Farrell, 2016, Coverage Error Optimal Confidence Intervals for Regression Discontinuity Designs, *Working Paper* University of Michigan.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik, 2014, Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs, *Econometrica* 82, 2295–2326.

Campbell, John L., Mark Cecchini, Anna Cianci, Anne C. Ehinger, and Edward M. Werner, 2016, Do Mandatory Risk Factor Disclosures Predict Future Cash Flows and Stock Returns? Evidence from Tax Risk Factor Disclosures, *Working Paper* University of Georgia.

Campbell, John L., Hsinchun Chen, Dan S. Dhaliwal, Hsin-min Lu, and Logan B. Steele, 2014, The information content of mandatory risk factor disclosures in corporate filings, *Review of Accounting Studies* 19, 396–455.

Chang, Yen-Cheng, Harrison Hong, and Inessa Liskovich, 2015, Regression Discontinuity and the Price Effects of Stock Market Indexing, *The Review of Financial Studies* 28, 212–246.

Cheng, C.S. Agnes, Henry He Huang, and Yinghua Li, 2016, Does Shareholder Litigation Deter Insider Trading?, *Journal of Law, Finance, and Accounting* 1.

Chiu, Tzu-Ting, Yuyan Guan, and Jeong-Bon Kim, 2015, The Effect of Mandatory Risk Factor Disclosures on the Pricing of Credit Default Swaps, *Working Paper* .

Choi, Stephen J., and A. C. Pritchard, 2016, SEC Investigations and Securities Class Actions: An Empirical Comparison, *Journal of Empirical Legal Studies* 13, 27–49.

Cox, James D., Randall S. Thomas, and Dana Kiku, 2003, SEC Enforcement Heuristics: An Empirical Inquiry, *Duke Law Journal* 53, 737–779.

Crane, Alan D., Sébastien Michenaud, and James P. Weston, 2016, The Effect of Institutional Ownership on Payout Policy: Evidence from Index Thresholds, *The Review of Financial Studies* 29, 1377–1408.

Cutler, Joshua, Angela K. Davis, and Kyle Peterson, 2016, Disclosure and the Outcome of Securities Litigation, *Working Paper* University of Houston.

D'Souza, Julia, K. Ramesh, and Min Shen, 2009, Disclosure of GAAP line items in earnings announcements, *Review of Accounting Studies* 15, 179–219.

Dyer, Travis, Mark H. Lang, and Lorien Stice-Lawrence, 2016, The Evolution of 10-K Textual Disclosure: Evidence from Latent Dirichlet Allocation, *Working Paper* University of North Carolina at Chapel Hill.

Field, Laura, Michelle Lowry, and Susan Shu, 2005, Does disclosure deter or trigger litigation?, *Journal of Accounting and Economics* 39, 487–507.

Filzen, Joshua, 2015, The Information Content of Risk Factor Disclosures in Quarterly Reports, *Accounting Horizons* 24, 887–916.

Filzen, Joshua, Garrett McBrayer, and Kyle Shannon, 2016, Risk Factor Disclosures: Do Managers and Markets Speak the Same Language?, *Working Paper* Boise State University.

Francis, Jennifer, Donna Philbrick, and Katherine Schipper, 1994, Shareholder Litigation and Corporate Disclosures, *Journal of Accounting Research* 32, 137–164.

Grossman, Sanford J., and Joseph E. Stiglitz, 1980, On the Impossibility of Informationally Efficient Markets, *The American Economic Review* 70, 393–408.

Hayn, Carla, and Patricia J. Hughes, 2006, Leading Indicators of Goodwill Impairment, *Journal of Accounting, Auditing & Finance* 21, 223–265.

Heinle, Mirko, and Kevin Smith, 2015, A Theory of Risk Disclosure, *Working Paper* The University of Pennsylvania.

Hope, Ole-Kristian, Danqi Hu, and Hai Lu, 2016, The benefits of specific risk-factor disclosures, *Working Paper* University of Toronto.

Huang, Ke-Wei, and Zhuolun Li, 2008, A Multilabel Text Classification Algorithm for Labeling Risk Factors in SEC Form 10-K, *ACM Transactions on Management Information Systems* 2, 18:1–18:19.

Israelsen, Ryan D., 2014, Tell It Like It Is: Disclosed Risks and Factor Portfolios, *Working Paper* .

Jackson, Howell E., and Mark J. Roe, 2009, Public and private enforcement of securities laws: Resource-based evidence, *Journal of Financial Economics* 93, 207–238.

Johnston, Rick, and Reining Petacchi, 2017, Regulatory Oversight of Financial Reporting: Securities and Exchange Commission Comment Letters, *Contemporary Accounting Research* forthcoming.

Johnstone, David, 2016, The Effect of Information on Uncertainty and the Cost of Capital, *Contemporary Accounting Research* 33, 752–774.

Kravet, Todd, and Volkan Muslu, 2013, Textual risk disclosures and investors' risk perceptions, *Review of Accounting Studies* 18, 1088–1122.

Li, Feng, 2008, Annual report readability, current earnings, and earnings persistence, *Journal of Accounting and Economics* 45, 221–247.

Loughran, Tim, and Bill McDonald, 2011, When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks, *The Journal of Finance* 66, 35–65.

Lukomnik, Jon, 2016, Risk Factor Disclosure Landscape, Technical report, Investor Responsibility Research Center Institute.

Lundholm, Russell J., 1991, Public Signals and the Equilibrium Allocation of Private Information, *Journal of Accounting Research* 29, 322–349.

Naughton, James P., Tjomme O. Rusticus, Clare Wang, and Ira Yeung, 2015, Private Litigation Costs and Voluntary Disclosure: Evidence from the Morrison Ruling, *Working Paper* Northwestern University.

Nelson, Karen, and A. C. Pritchard, 2016, Carrot or Stick? The Shift from Voluntary to Mandatory Disclosure of Risk Factors, *Journal of Empirical Legal Studies* 13, 266–297.

Robbins, Robert B., and Philip L. Rothenberg, 2005, Writing Effective Risk Factor Disclosure in Offering Documents and Exchange Act Reports.

Rogers, Jonathan L., and Andrew Van Buskirk, 2009, Shareholder litigation and changes in disclosure behavior, *Journal of Accounting and Economics* 47, 136–156.

Rose, Amanda M., 2008, Reforming Securities Litigation Reform: Restructuring the Relationship Between Public and Private Enforcement of Rule 10b-5, *Columbia Law Review* 108.

Ryans, James P., 2017, Textual Classification of SEC Comment Letters, *Working Paper* London Business School.

Salton, Gerard, and Michael J. McGill, 1986, *Introduction to Modern Information Retrieval* (McGraw-Hill, Inc., New York, NY, USA).

Sims, Christopher A., 2003, Implications of rational inattention, *Journal of Monetary Economics* 50, 665–690.

Skinner, Douglas J., 1994, Why Firms Voluntarily Disclose Bad News, *Journal of Accounting Research* 32, 38–60.

Skinner, Douglas J., 1997, Earnings disclosures and stockholder lawsuits, *Journal of Accounting and Economics* 23, 249–282.

Thistlethwaite, Donald L., and Donald T. Campbell, 1960, Regression-discontinuity analysis: An alternative to the ex post facto experiment, *Journal of Educational Psychology* 51, 309–317.

Verrecchia, Robert E., 2001, Essays on disclosure, *Journal of Accounting and Economics* 32, 97–180.

White, Mary Jo, 2013, The Path Forward on Disclosure.

Wooldridge, Jeffrey, 2010, Correlated random effects models with unbalanced panels, *Working Paper* Michigan State University.

# — Appendix A —
# Variable Descriptions

The following table defines the variables used in this paper. Variable names and calculations provided in brackets correspond to source database. For the regressions presented in the tables, the continuous variables are winsorized at the 1st and 99th percentiles.

| **Variable** | **Description** |
|---|---|
| Assets | Total assets $\{AT\}$ |
| Market Equity | Market value of equity at fiscal year end. $\{CSHPRI * PRCC\_F\}$ |
| Net Income/AT | Net income to lagged assets. $\{NI_t/AT_{t-1}\}$ |
| ROA | Return on assets. $\{IB_t/AT_{t-1}\}$ |
| Operating Inc./AT | Operating income (after depreciation) to lagged assets. $\{OIADP_t/AT_{t-1}\}$ |
| Sales Growth | Ratio of change in sales to lagged assets. $\{(SALE_t - SALE_{t-1})/AT_{t-1}\}$ |
| Sales / AT | Sales to lagged assets. $\{SALE_t/AT_{t-1}\}$ |
| Book Equity | Book value of equity. $\{(TEQ_t \| AT_t - LT_t) + TXDITC_t - PSTK_t\}$ |
| Book-to-Market | Book to market. $\{\text{Book Equity}_t/\text{Market Equity}_t\}$ |
| Leverage | Leverage $\{(DLTT_t + DLC_t)/AT_t\}$ |
| Tangibility | Tangibility $\{PPENT_t/AT_t\}$ |
| Turnover | Ratio of average daily volume (CRSP) to outstanding shares at fiscal year end (Compustat). |
| Beta | Market loading from CAPM model of daily returns on value weighted index, for all available days in the fiscal year. $\{RET = \alpha + \beta \cdot VWRETD + \epsilon\}$ |
| Excess Returns | Cumulative excess daily returns during fiscal year. $\{(\prod RET - VWRETD + 1) - 1\}$ |
| Excess Ret. Std. | Standard deviation of daily excess returns. $\{Std.\ Dev(RET - VWRETD)\}$ |
| Min. Excess Ret. | Minimum daily excess return during fiscal year. $\{Min(RET - VWRETD)\}$ |
| Excess Ret. Skew | Skewness of daily excess returns. $\{Skew(RET - VWRETD)\}$ |

Continued on next page...

...continued from previous page.

| Variable | Description |
|---|---|
| CAR$_{+3\ day}$ | Cumulative abnormal return spanning three business days starting on the filing date of the annual report, using a Fama-French Carhart factor model. Data from Kenneth French's website. $\{RET = MKTRF + SMB + HML + UMD + MOM + \epsilon\}$ |
| CAR$_{+3\ months}$ | Cumulative abnormal return spanning 60 business days starting on the filing date of the annual report, using a Fama-French Carhart factor model. Data from Kenneth French's website. $\{RET = MKTRF + SMB + HML + UMD + MOM + \epsilon\}$ |
| Bid-Ask Spread$_{+1\ year}$ | Average daily bid ask spread at closing for 240 business days starting on the filing date of the annual report, as percentage of average bid and ask. $\{200 * (ASK - BID)/(ASK + BID)\}$ |
| Illiquidity | Amihud Illiquidity (2002), calculated as $\frac{1}{D_{iy}} \sum_{t=1}^{D_{iy}} \lvert R_{iyd} \rvert / Volume_{iyd}$, where $D_{iy}$ is the number of days over which the illiquidity is calculated, and $R_{iyd}$ and $Volume_{iyd}$ are the return and trading volume, respectively, on a given day. |
| Negative NI | Indicator variable equal to 1 if next year's net income is negative. $\{NI_{t+1} < 0\}$ |
| Negative Op. Inc. | Indicator variable equal to 1 if next year's operating income is negative $\{OIADP_{t+1} < 0\}$ |
| Sales Decline | Indicator variable equal to 1 if next year's sales are lower than the current year's sales by 10% or 10 million dollars, whichever is larger. $\{SALE_{t+1} - SALE_t < -10\% * max(100, SALE_t)\}$ |
| Security Litigation | Indicator variable equal to 1 if a securities litigation is filed in the subsequent year. |
| Lawsuit Intensity | Natural log of number of litigation events found in CapitalIQ Key Developments database in a given fiscal year. |
| RF Comment | Indicator variable equal to 1 if an SEC comment letter is received in a given fiscal year that references risk factors. |
| Comment (Any) | Indicator variable equal to 1 if any SEC comment letter is received in a given fiscal year, including those with a reference to risk factors. |
| FPS Industry | Indicator variable equal to 1 if the firm has an SIC code in one of the high litigation risk industries defined in Francis et al. (1994). |
| # Risk Factors | Total number of risk factors disclosed under Item 1A of an annual report. |
| # New RF | Number of new risk factors which were not present in the previous year's annual report. |
| # Dropped RF | Number of risk factors which were in the previous year's annual report, but are no longer included in the present year. |
| # Kept RF | Number of risk factors which were in the previous year's annual report and persist in the current annual report. |

Continued on next page...

...continued from previous page.

| **Variable** | **Description** |
|---|---|
| $\Delta$ # RF | Net change in the number of risk factors, equal to *# New RF - # Dropped RF*. |
| # of Words | Total number of words in Item 1A of an annual report (excluding stop words). |
| # of Sentences | Total number of sentences in Item 1A of an annual report. |
| # of Specific Words | Total number of words identified by the Stanford Named Entity Recognition algorithm as being in categories: Location, Person, Organization |
| # of Numerics | Total number of numbers in Item 1A of an annual report. |
| # of Words/RF | Ratio of the total number of words to the total number of risk factors in Item 1A of an annual report (excluding stop words). |
| Specificity | Ratio of the total number of specific words to the total number of risk factors in Item 1A of an annual report (excluding stop words). |
| Numeric Intensity | Ratio of the total number of numbers to the total number of risk factors in Item 1A of an annual report (excluding stop words). |
| FOG Index | Gunning Fog score for the text in Item 1A of an annual report (excluding stop words). Calculated as $\left\{0.4\left(\frac{\#\ of\ words}{\#\ of\ sentences}+100\frac{\#\ of\ complex\ words}{\#\ of\ words}\right)\right\}$ |

# — Appendix B —

# Risk Factor Extraction

To derive an initial list of filings, I extract the gvkey and historical CIKs from the Compustat annual file, starting from 89,687 firm years. I then merge the non-missing CIKs from Compustat with the EDGAR filings index file provided by the SEC.[1] I filter the index files to include only form 10-Ks, excluding amended 10-K/As, leaving 67,648 filings from 9,632 firms. Form 10-KSB is also included, but are no longer filed after 2009.

I then search through these filings to extract *Item 1A: Risk Factor* section. Filings, as they are submitted to the EDGAR system, comprise an SGML header (with information about the filing such as company identifier, name, SIC code, and period with which the filing is associated), and multiple 'documents' that are typically the main form and exhibits. I extract the first document in a filing, which is the 10-K report, but ignore the accompanying exhibits. This potentially biases against extracting the risk factor section if it is included in Exhibit 13, but is in line with the other studies of risk factors that omit sections included by reference. Similar to the approach in Campbell et al. (2014), I extract sections by assuming visual prominence of Item headings through font and whitespace delineation. Thus, I also exclude filings that are not submitted in HTML format because they omit the visual features required to reliably identify the sections in annual reports.

[1] Located, for example, at ftp.sec.gov/edgar/full-index/2005/QTR1/master.idx. Downloaded using script from github.com/gaulinmp/pyedgar

The general algorithm I use for extracting a specific section in an annual or quarterly report is as follows. First, I identify the Item headings, which should occur in order throughout the document (omitting them when they occur in the table of contents, whether it is at the beginning or end of the document). I identify headings as being the only text on a 'line' of text, in the format Item 1A: Risk Factors (with flexibility in punctuation), and emphasized with either bold or underlined font. Once I identify Item 1A and either Item 1B or Item 2 (whichever is found and comes first), I extract the text between the first instance of Item 1A to first instance of the next non 1A Item header. Some firms repeat the Item 1A header at the top of each page of the section, thus I include all the text until the next item number.

This method is based on visual identification of features that stand out to a human reader of the rendered document. The HTML format of annual reports allows for programatic identification of visual elements through the Document Object Model, or DOM. HTML DOM is a platform independent graph based method for describing content, primarily textual, in the case of HTML (but also arbitrary data in XML, one example of which being XBRL). Each element in a page is a node in the DOM graph, with parents that fully contain it, and potentially children it fully contains. Nodes in the graph, called HTML tags, have multiple types that each have different features. For example, the paragraph tag `<p>` defines a paragraph which is separated/isolated visually from text above and below (children nodes or parent nodes), while the font tag `<font>` defines a subset of words (potentialy children of a parent paragraph tag) that have a specific font, but are not isolated with whitespace from the surrounding text. These features of the different tag types allow for a researcher to programatically extract data based on what a human reader sees as whitespace or emphasis when viewing the rendered document in a web browser. Specifically, I use the beautifulsoup library to handle the HTML DOM which allows for traversing tag nodes in a graph-like format.

For each HTML annual report, I iterate through all instance of an HTML tag containing the text *Item 1A*, *1B*, or *2*, using the case insensitive regular expression (`item(?:[^ a-z0-9]| )*(1[AB]|2)`). The ` ` is an HTML encoded representation of a non-breaking space, which is necessary to include when searching through raw HTML encoded documents. By only searching for Items 1A, 1B, and 2, I impose the assumption that items are presented in order. The result of this search is then checked for emphasis, based on either HTML emphasis tags or CSS styles.[2] Those items that are emphasized are then checked for whitespace separation. Whitespace separation is defined as the tag being at the beginning of a visual line, or as described in Campbell et al. (2014) as segmentation.

To identify whitespace separation, I iterate up the HTML DOM to find the first parent node that is visually separated. This includes the tags: `h1`, `h2`, `h3`, `h4`, `h5`, `h6`, `p`, `div`, `ul`, `ol`, `tr`, or `table`. While it is possible that `div` tags are not visually isolated (using `float` CSS styling), in a non-exaustive search of HTML documents I did not find any instances of floats. Additionally, while some documents use paragraphs and CSS as mentioned above to lay out their annual report, other filings employ tables to do so. The difference is akin to using tab-stops in Microsoft word to put text on both the left and right side of a document, or alternatively creating a table with two columns. For firms that use the latter table method, merely checking within the immediate parent table cell may result in false conclusions as to the visual layout. Therefore, when an item is included in a table, I include the entire row as the parent

[2] HTML emphasis tags: b, em, strong, h1, h2, h3, h4, h5, h6, u. CSS Styles *bold* or *underline* within HTML tags: p, font, div, span, li.

element, rather than just the cell. This separating parent element thus contains the 'block' of text in which the word Item occurs and is emphasized, be it just the header *Item 1A: Risk Factors*, or a paragraph containing a reference to the Item 1A section. To filter out the latter false positives (including table of contents matches), I require that the full (plain) text in the separating element comprise solely the Item number and description using regular expressions.[3]

The result of this step is the elements containing the Item headers for 1A, and 1B or 2. I then keep all of the HTML code between the beginning of the first Item 1A header and the beginning of the first Item 1B or 2 header, whichever occurs first. These extracted sections represent the Item 1A for each given filing. To match the filings to their associated Compustat fiscal years, I extract the period assigned with the filing from the heading of the EDGAR document. To do so, I search in the first 5,000 bytes of the SGML header provided in the raw daily feed file for the `PERIOD` header. I keep the filing if the `PERIOD` header contains a string which conforms to an eight character date format (YYYYMMDD). I match the EDGAR filings to the Compustat data based on CIK (exact match) and period date (within a date window). I require the period listed in the filing header be within five business days of the `datadate` variable in Compustat. While greater than 90% of the filings have an exact match between the reported period and `datadate`, there are some discrepancies because Compustat always sets the `datadate` variable equal to the last day of a month. To verify this assumption, I manually selected a random sample of 50 reports with a date match 5 days apart, and found that all were matched to the appropriate fiscal year in Compustat.

[3]Item 1A: `/^\s*item[^a-z0-9]*1A[^a-z]*risk[^a-z]*factors?\s*$/`,
Item 1B: `/^\s*item[^a-z0-9]*1B[^a-z]*((?:unresolved|sec|staff|comments)[^a-z]*)+$/`,
Item 2: `/^\s*item[^a-z0-9]*2.{,20}propert(y|ies)\D{,35}$/`

# — Appendix C —
# Extracted LDA Topics

This appendix presents 24 topics extracted from the risk factor disclosures, along with the 20 most heavily weighted words in each topic. The LDA was calibrated on 25 topics, but I omit the 'preamble' topic, which consists of general legal language pertaining to the PSLRA safe harbor disclaimer. Topics are presented in order of their weight in the corpus of risk factor disclosures.

| **Topic Name** | **Top Weighted Words** |
|---|---|
| Health/Pension | healthcare, care, insurance, medicare, health, federal, pension, reimbursement, services, agreement, its, certain, medicaid, government, programs, state, act, claims, legislation, tenants |
| Energy | gas, oil, natural, energy, drilling, production, properties, fuel, exploration, costs, weather, prices, construction, coal, reserves, mining, transportation, facilities, aircraft, equipment |
| Drugs (Production) | fda, products, product, clinical, drug, approval, regulatory, candidates, patients, trials, medical, drugs, reimbursement, approvals, manufacturing, approved, pharmaceutical, marketing, physicians, market |
| Personnel | personnel, key, retain, employees, attract, management, qualified, executive, growth, ability, senior, success, chief, officer, manage, team, skilled, officers, services, loss |
| Retail | products, stores, distributors, merchandise, retail, brand, restaurants, customers, brands, retailers, store, consumer, new, distribution, restaurant, food, product, channel, apparel, automotive |

Continued on next page...

...continued from previous page.

| Topic Name | Top Weighted Words |
|---|---|
| Information Technology | internet, services, content, gaming, advertising, software, wireless, products, online, mobile, digital, customers, network, fcc, television, solutions, media, video, satellite, users |
| Claims/Security | systems, liability, insurance, security, claims, product, products, data, customers, coverage, reputation, computer, damage, defects, software, errors, result, failures, technology, breaches |
| Environmental/Regulation | regulations, laws, environmental, subject, hazardous, federal, state, comply, compliance, government, costs, regulatory, regulation, materials, safety, substances, penalties, emissions, fines, local |
| International Investment | foreign, international, china, countries, political, prc, united, export, labor, states, economic, currency, products, chinese, instability, outside, laws, terrorist, including, import |
| Accounting | goodwill, impairment, losses, assets, million, intangible, net, value, estimates, assumptions, fair, history, profitability, accounting, lived, charges, carrying, profitable, significant, continue |
| Debt | debt, indebtedness, capital, credit, financing, covenants, cash, facility, notes, ability, terms, funds, raise, obligations, interest, subsidiaries, default, revolving, senior, available |
| Customers | customers, economic, products, conditions, quarter, spending, revenue, demand, revenues, clients, contracts, global, markets, consumer, customer, services, result, credit, orders, timing |
| Financing | loans, loan, estate, real, mortgage, interest, portfolio, investments, credit, deposits, investment, rates, market, bank, institutions, value, losses, borrowers, banks, conditions |

Continued on next page...

...continued from previous page.

| **Topic Name** | **Top Weighted Words** |
|---|---|
| Hedging | tax, currency, foreign, currencies, dollar, rates, reinsurance, exchange, rate, hedging, ratings, income, taxes, fluctuations, rating, hedge, dollars, denominated, jurisdictions, counterparties |
| Intellectual Property (IP) | intellectual, patent, patents, rights, property, proprietary, litigation, claims, products, protect, technology, infringement, license, third, parties, secrets, licenses, against, proceedings, trade |
| Internal Control | internal, reporting, controls, accounting, sarbanes, oxley, over, control, act, public, weaknesses, disclosure, rules, governance, section, standards, effective, management, procedures, required |
| Stock Market | stock, price, common, market, trading, analysts, fluctuations, volatile, volatility, announcements, volume, shares, companies, has, performance, class, unrelated, litigation, fluctuate, competitors |
| Strategic Alliance | acquisitions, acquired, acquisition, businesses, merger, integration, integrate, joint, benefits, strategic, integrating, realize, venture, technologies, successfully, acquire, management, anticipated, ventures, strategy |
| Competition | products, new, competitors, competitive, market, product, technologies, services, compete, competition, technological, develop, customers, development, industry, technology, resources, markets, greater, companies |
| Equity/Dividends | stock, common, shares, dividends, price, warrants, preferred, convertible, stockholders, market, series, conversion, notes, options, dividend, foreseeable, outstanding, pay, cash, holders |
| Governance | directors, stockholders, stock, board, provisions, stockholder, shares, common, control, voting, delaware, shareholders, incorporation, takeover, preferred, bylaws, certificate, approval, rights, interests |

Continued on next page...

...continued from previous page.

| **Topic Name** | **Top Weighted Words** |
|---|---|
| Equity Listing | reit, tax, unitholders, income, partnership, units, taxable, partner, distributions, stock, common, penny, nasdaq, irs, carryforwards, broker, qualify, listing, code, ownership |
| Drugs (Clinical) | clinical, candidates, trials, product, development, commercialization, drug, products, research, commercialize, trial, collaboration, collaborators, pharmaceutical, marketing, collaborative, collaborations, preclinical, candidate, regulatory |
| Supplier | products, manufacturing, customers, suppliers, components, materials, raw, supply, inventory, customer, manufacturers, production, product, manufacture, orders, supplier, capacity, third, demand, significant |