



Taking Information Seriously: A Firm-side Interpretation of Risk Factor Disclosure

M. J. Histen¹

Accepted: 17 October 2022 / Published online: 22 November 2022
© International Atlantic Economic Society 2022

Abstract Information plays a key role in how markets assess a firm’s value. Disclosure is expensive to comply with and to interpret. Narrative information about risks is particularly challenging to evaluate. Though results on the informativeness of narrative risk disclosure are mixed, previous research focuses on the value to investors. However, risk disclosure does more than provide transparency to investors. It also affects firm behavior through an elevated assessment and increased preparation in response to risk. This study takes this active view of risk disclosure by analyzing firm-side effects through multifactor fixed effects regressions of approximately 13,000 firm-year observations from 2015–2018. The robust results show firm performance is correlated with topic models of narrative risk disclosures. These findings support the argument that risk disclosure corresponds to risk management. A firm disclosing cybersecurity risk responds better in the event of a data breach compared to one that has not, despite the firm itself sourcing the information. Overall, narrative risk disclosure both provides investors with useful information and strengthens a firm’s responsiveness. This provides empirical evidence to policymakers and market participants on the value of forward-looking statements even when the news is pessimistic.

Keywords Narrative information · Topic modeling · Text analysis · Risk disclosures · Econometric analysis

JEL D82 · G18 · G32 · M41 · G14

✉ M. J. Histen
matthew.histen@csun.edu

¹ California State University Northridge, Northridge, CA, USA

Introduction

In its 1998 initial public offering (IPO), eBay disclosed that a decline in the popularity of Beanie Babies could have materially adverse effects on its business. The cultural phenomenon associated with these stuffed toys accounted for over 30,000 simultaneous auctions, a substantial fraction of all eBay trades (eBay 1998). This disclosure about risks was required by the United States (U.S.) Securities and Exchange Commission (SEC) for any firm proceeding with a public offering. To reduce information asymmetries, issuers must elaborate about circumstances that could make a business speculative or risky. Beginning in 2005, the SEC extended the disclosure of critical risk factors as an additional section in all financial statement filings. While eBay's 2020 filing has no mention of Beanie Babies, it does expound upon risks related to coronavirus disease 2019 (COVID-19) and other items that could materially affect its operations.

Information disclosed in public filings about a firm's operations are vital to how markets assess its value. The SEC argues that risk factor disclosures (RFDs) provide investors with important information. However, disclosures are expensive for firms to comply with and expensive for market participants to interpret. Measuring the value of such information is challenging, especially with unstructured narrative data. Moreover, RFDs alone have caused substantial increases to the length and content of annual 10-Ks (Dyer et al., 2017). Attempts to measure their effectiveness rely on natural language processing (NLP) to systematically quantify features of textual disclosures. Much of the published work focuses on readability, where researchers unpack the assimilation of information from text into asset prices (see Loughran and McDonald (2016) for a survey). Such papers typically track a lexicon through documents to detect a sentiment or construct an index to correlate with stock prices. Results are mixed (Isiaka 2018). Many studies (Campbell et al., 2014, 2019; Hope et al., 2016; Li et al., 2018) argue for the relevance of the disclosure for risk assessment in market prices. Practitioners, regulators, and researchers have also criticized their informativeness, remarking that risks are generic, repetitive, or incomplete (Bao & Datta, 2014; Beatty et al., 2019). After all, it is not always in management's best interest to disclose bad news. Nearly all these studies considered disclosure effects on stock market returns after report release dates. That is, they focused on event studies of investor response, representing risk factors as mere informational transfers with volatility consequences. Such answers to whether RFDs contain useful information, then, take a narrow view of the relationship between parties. They fall into a pure agency theory paradigm where the only value generated by disclosure comes from collapsing the distance between parties' information sets, hence the appropriateness of event studies. Does a specific disclosure elicit reaction by investors by reducing informational asymmetry?

However, event studies only focus on the investor side of the information flow, neglecting how disclosure affects firm behavior. This study argues that RFDs influence both sides of the information exchange. In considering risk possibilities, management acts on them, signaling that the firm has well-defined procedures

in place to navigate contingencies even though the firm is itself the source of the information. A firm with a disclosure about cyber security will respond better in the event of a cyber security breach than one that has not. The relevant change was not the wording but the firm's preparedness. The research question in previous studies analyzed whether RFDs were informative. The argument here is stronger: RFDs are about risk management. As disclosure increases, both the market and the firm are affected.

This study expands previous research into qualitative disclosure by analyzing firm-side effects, taking a more active view of management's role in processing and responding to risk. RFD effects are tested through multifactor fixed effects regressions while controlling for firm, market, and industry effects with a sample of approximately 13,000 firm-year observations from 2015–2018. The dependent variable regimes consider year-over-year returns on equity (ROE) and assets (ROA) rather than pricing behavior three days after a report release date. The explanatory variable was generated through an unsupervised topic model that groups text automatically to minimize the researcher's judgements. Further, the analysis hones in on cyber security since this risk can be pinned to precise textual artifacts while broader categories, such as macroeconomic risks, have ambiguous pools of word associations. The results show that risk disclosures are positively correlated with performance and that statistical significance is greater for worse performing firms. This provides evidence that not only are RFDs informative, but management was better prepared to handle risk realizations when they disclosed the RFDs. This sheds some positive light on systemic risks related to cyber security insofar as both the market and the firm itself are incentivized to integrate these effects into asset prices. It also supports unsupervised machine analysis over textual data. Overall, the results and methodology complement research into RFDs and the effects of qualitative disclosures, which provides empirical evidence to both policymakers and market participants on the value of forward-looking statements.

Related Literature

The SEC was founded to ensure an orderly and efficient securities market structure. This hinges on the transparency of information between issuers and investors whereby the SEC requires periodic filings from public and regulated companies. For many contingent parties or policy makers, such reports are important resources about company conditions and potential risks. Information plays a key role in a firm's operations and how markets assess value. Greater specificity reduces the expected cost of capital. The more costly it is to interpret financial reports, the less accurate are market prices (Heinle & Smith, 2017). These filings have transformed over time in response to Congressional acts and procedural changes. The corporate disclosure literature has long analyzed the explanatory power of useful financial information (Francis et al., 2004). Recent requirements about company narratives (particularly future performance) have caused substantial changes to the structure and length of these filings (Dyer et al., 2017). Evaluating their effectiveness has become more challenging.

For example, the Private Securities Litigation Reform Act of 1995 offered liability protection through safe harbor provisions for forward-looking statements. That is, in the S-1 filing for its IPO registration, eBay could deflect potential lawsuits by disclosing market risks related to the Beanie Babies bubble. Beginning in 2005, the SEC updated their requirements to have all companies (not just firms proceeding with public offerings) disclose critical risk factors that made an issuer speculative or risky in Section 1A of their financial filings. With the aim of reducing information asymmetries, these risk disclosures are supposed to be concise, organized logically, furnished in plain English, and updated in quarterly reports (SEC 2005). However, disclosures are expensive to report and expensive to interpret. Narrative and forward-looking statements are not audited externally and have long been criticized for being uninformative (Schrand & Elliott, 1998). Firms have discretion in what they choose to report (Beyer et al., 2010; Hope et al., 2016). As a result, practitioners, regulators, and researchers have all remarked that risk disclosures are boilerplate in nature (Chen et al., 2022). The SEC has responded with legal incentives to provide more firm-specific risk disclosures and updated their guidance with comment letters in 2010 and again in 2015, asking that filers only include specific risk factors related to the individual company (Isiaka 2018). Though many parties may be interested in the information contained, most attention has been aimed at investor usefulness (Isiaka 2018). The research questions boil down to first, whether any useful information is contained in RFDs or whether they are generic and boilerplate. If not, what are the effects of reducing firm-investor asymmetries (Bao & Datta, 2014)?

Approaches to these questions have been greatly enhanced by methods in natural language processing (Loughran & McDonald, 2016). Structured data are divided into standardized and identifiable pieces, such as numbers or dates. Representing unstructured data such as narrative text is far more difficult. NLP techniques aim to organize and extract content from large amounts of unstructured data such that textual artifacts can be represented in quantitative analysis. Prior research has provided evidence that disclosures inform market participants of relevant firm-specific risks (Campbell et al., 2014, 2019; Hope et al., 2016; Li et al., 2018). The most common methodology here attempts to measure textual information through dictionary methods. Text is assigned to categories using keywords or phrases to classify documents (Bao & Datta, 2014). Tracking the lexicon can then detect a sentiment or construct an index to correlate with stock prices (Loughran & McDonald, 2016). In risk disclosures, risks might be classified by words (e.g., “lawsuit”) with aggregate counts for each document regressed against some performance measure. Since Li (2008), many studies have followed this approach to examine tone, sentiment, or readability in public reports or news articles. The most relevant here include seminal studies by Campbell et al. (2014) and Kravet and Muslu (2013), who used predefined dictionaries to taxonomize risk types. These studies counted risk sentences across years to show that certain risk factors affect stock-return volatility. Other papers have focused on qualitative aspects of the narrative, such as readability or specificity measures (Israelsen & Yonker, 2017). Hope et al. (2016) showed that more specific information like proper nouns and precise percentages in RFDs again have economic consequences, supporting their usefulness.

These studies all approached the research question by focusing on value measures to investors. That is, they pursued event studies with a narrow window for stock returns following report releases (e.g., Campbell et al., 2014; Bao & Datta, 2014; Li et al., 2018; Chen et al., 2022). The corporate disclosure literature largely presumes that value stems from an information flow from one party to another, ignoring the role that information plays to structure the firm behavior itself. Consider the work of Chen et al. (2022), who tracked a cyber security lexicon in risk disclosures before and after a data breach. They found that management increased disclosure after a breach occurred, arguing that this reflects better transparency. However, this glosses over the effect of disclosure on the firm side: an elevated assessment of cyber security and stronger protocol for protecting data. The wording is not the relevant change, but the firm's preparedness for the risk is. The process of disclosure influences firm behavior as well as investors. RFDs are as much a risk management disclosure as they are an information transfer. They signal that a firm is actively managing such a risk with well-defined procedures in place. This study contributes to the research on RFDs by examining these effects via expansion of the scope of dependent variables. The seminal studies (Bao & Datta, 2014; Campbell et al., 2014) examined the relationship between three-day abnormal returns around the filing date and the amount of risk information disclosed. In contrast, here the dependent variable regimes took longer measures of firm performance through one-year lenses of the following year's returns on equity and asset.

The hypothesis is that management's assessment of a risk implies procedures are in place to handle it. RFDs signal action. In the event a risk is realized, a firm performs better having disclosed the risk compared to if it had not. Thus, RFDs affect both sides of the information exchange: market participants and the firm itself. Not only are they informative, but they should also positively correlate with performance relative to the absence of disclosure. Boilerplate disclosures by definition are not responsive to changes in circumstances (Chen et al., 2022). If not boilerplate, the empirical results will show differences in performance after controlling for confounding factors. If those coefficients are stronger for worse performing firms, then the argument is further supported for RFDs as risk management rather than merely risk disclosure.

Data and Methodology

Financial narratives contain rich textual information but are hard to interpret. NLP techniques can help represent qualitative aspects of text. A variety of studies attempted to interpret narrative information in broad quantitative analysis. The commonly used dictionary methods count keyword searches in the text body. Though showing promising results, they rely on human coders to develop rules for variables and a predefined set of categories. Identification of risk factors heavily depended on researchers' adjustments and traditional theories in structuring their data (Wei et al., 2019). Minimizing researcher judgements can be improved through an unsupervised machine learning technique called topic modeling. Though practiced less, topic modeling organizes text systematically through emergent structures instead

of top-down assumptions. For example, rather than deciding *a priori* that litigation risks are important, the model develops categories without pouring through textual patterns over word associations. However, previous studies (Bao & Datta, 2014; Wei et al., 2019; Zhu et al., 2016) analyzing risk factors with such models detected far less information than their dictionary counterparts. Often, less than two-thirds of generated topics held any statistical significance. By applying these methods to performance metrics rather than event study windows, the regression results here show statistical significance ($p < 0.10$ or better) across more than 90% of generated topics.

Topic modeling uses generative statistical models that sort text by unobserved groups to explain similarities across documents. That is, rather than the researcher specifying or defining categories, the model attempts to discover abstract topics through hidden semantic structures in the text. The model here uses a Latent Dirichlet Allocation (LDA), which comprehensively discovers risk factors from textual disclosure with no predefined types (Wei et al., 2019). LDA assumes that documents are a mixture of topics whereby words are probabilistically assigned to unobserved categories by an updated Bayesian network. By focusing on the distribution of conjugate relations between terms, the model updates the interaction between observed documents (firm disclosures) and the hidden topic structure (risk factors). Documents are considered random mixtures over these hidden topics, with each topic conceptualized as a distribution of words. To infer these topics, the model repeatedly generates distributional processes until it accurately replicates how documents were created. For example, with a corpus of news articles, an LDA sorts topics according to statistical disparities between words (e.g., {home run, touchdown, goal}, {election, voter, ballot}, {interest rates, unemployment, GDP}) into generated topics (e.g., {sports, politics, economics}). By analogy, documents of RFDs hold different risk types that the model sorts by generated risk topics.

The data for the LDA come from the textual content in each firm's annual 10-K filings. These data are published in the SEC (2022) Financial Statements and Notes Data Sets by accession number (ADSH), the unique firm-year filing key, which was pulled with Python from the SEC (2018) Electronic Data Gathering, Analysis, and Retrieval (EDGAR) website for all fiscal years ending between 2015 and 2018 following the SEC's updated guidance on RFD requirements and before the coronavirus disease (COVID) pandemic. EDGAR maintains these reports in hypertext markup language (html), which were processed using BeautifulSoup4, a Python package to create navigable parsing trees to locate and extract RFDs from Section 1A of the 10-K. After converting html to plain text, the resulting data are unstructured, free-form narratives. To prepare the textual data for analysis, all stop words were removed with Python's Natural Language Toolkit Standard English library (e.g., words such as 'and', 'to', 'with', and 'there' were dropped). The remaining corpus of words were then lemmatized, a process that removes inflection endings by morphological analysis to minimize ambiguity (e.g., {study, studies, studied, studying} are all represented by {studi}). Finally, only parts of speech through nouns and adjectives were analyzed (as is typical in high-capacity NLP analyses). These data were then organized with an LDA model through Python's Gensim module for unsupervised topic modeling using machine learning. Each mentioned library is open source with public documentation.

Though LDA allows for a more emergent generation of risks than dictionary methods, there are tradeoffs. For example, some categories may not seem meaningful in a specific context. Another problem with LDA involves correlations between risk types, which may be related or overlap on word probabilities (Isiaka 2018). Fortunately, other studies (Bao & Datta, 2014; Wei et al., 2019; Zhu et al., 2016) have demonstrated the reliability of LDA categorizations from RFDs by auditing and matching results by hand. Here, the topic model generated 25 categories (the commonly chosen amount since Huang and Li (2011)) which aligned comparably well to previous descriptions. Regarding correlation concerns, this study focused on a well-defined topic, cyber security, with textual artifacts that did not intersect with other categories. This topic's highest probable words include tokens such as security, information, cyber, attack, cybersecurity, cyberattack, breach, computer, and technology. The specificity of this category avoided having to infer a precise meaning behind all generated topics and held less ambiguity than vaguer categorizations like macroeconomic risks. Moreover, cyber risks are particularly salient in the contemporary landscape, having caused about \$1 trillion in damages in 2020 with private firms as major victims (Lewis et al., 2021). In response, the Biden Administration released a memo about the national security threats imposed by cyberattacks following an executive order to strengthen U.S. defenses (Exec. Order No. 14028 2021). Given the severity of such economic harms and their ubiquity across industries, companies must take seriously their private approaches to risks related to cyber security. At the very least, this begins with disclosure about the risks and their corresponding responses.

After the LDA generated a risk vector for each firm weighted by prevalence of topic, ADSH keys were synchronized with Mergent (2022) (generated from the Financial Times Stock Exchange Russell) for numerical data including performance metrics. These data were pulled for all observations with more than \$100 thousand in assets or liabilities. Finally, the top and bottom of the sample was truncated to minimize the effects of outliers. Sectors are organized by letter digit Standard Industrial Classification (SIC) codes, except for manufacturing, which was drilled down into 2-digit classifications for fields with more than 500 firms. The industry ratios track closely to the original after observations were dropped (most of which were shell companies) with no sector variation exceeding a 1% difference.

Empirical Analysis

This paper argues that risk disclosures contain information about firm performance because these statements are not just documentation but signals of management actions. This hypothesis is tested empirically through public company annual data from 2015–2018. Risks as disclosed in Section 1A were categorized with an LDA model and these categories are tested against several measures of firm performance through a fixed effects panel regression by firm and year. These dependent variables were based on the following year's performance metrics. Risk disclosures are forward-looking statements, so the analysis quantified these narratives against the real change in the firm's performance the following year.

The main explanatory variable came from categorizing risks with a topic model. From approximately 13,000 firm-year observations and 4,000 unique firms, 25 risk topics were generated from Section 1A of the annual 10-K forms. Robustness tests with 20 and 30 topics preserved the results. To minimize ambiguity, the results below focus on the topic associated with cyber security based on word probability outcomes. If risk disclosures indicate a firm has taken actions against some risk, in the realization of such a risk, firms making the disclosure should be more prepared and therefore perform better. To test this, the topic model variables were regressed on the following year's performance. The regression took the form

$$Y_{it} = \beta_1 c_{it} + \beta_j RV_{it} + \beta_k X_{it} + \alpha + \varepsilon_{it}$$

where Y_{it} represented the performance metric with i referring to entity and t to time, c_{it} was the cyber (or other) risk topic decimal, RV_{it} was the vector of remaining j risk topics, X_{it} was a vector of controls, α was the intercept, and ε_{it} was the error term. Two different dependent variable regimes were used: return on assets (ROA) and return on equity (ROE). Both ROA and ROE were calculated directly by Mergent. ROE and ROA combine earnings and book value measures and are generally considered key summary measures of financial performance and value creation. Moreover, the risk of a firm is noted to have important effects on these measures. Table 1 above provides descriptive statistics by dependent variable. Some observations were lost due to missing data/panels. The sample was trimmed at the 5th and 95th percentiles to minimize outliers.

The controls vector included firm, market, and industry effects. Firm-specific traits were controlled for with logged assets, revenue, leverage, and age. Market traits like market power were controlled for with a concentration ratio for the top eight firms and Herfindahl–Hirschman Index (both calculated based on revenue data from the population by SIC code) as well as volatility dummies for whether the firm was National Association of Securities Dealers Automated Quotations (NASDAQ) listed or headquartered outside the U.S. Finally, year and industry fixed effects were used, where industry controls used letter digit SIC codes from each SEC filing. The results of the fixed effects models are reported with robust standard errors. In general, the reported values are consistent with a variety of robustness tests, including changes to the LDA topics and parameters, sample size, dropped controls, and other standard econometric tests.

Table 2 shows the summarized results for ROA. In attempting to discover topics in the text, LDA categorizes semantic structures from the text into groups of probabilistic words. Using a grouping with specific word tokens related to cyber security, this study traces the effects of cyber risk on firm performance measures. The remaining 24 generated topic groups were included as a risk vector and treated as controls. Inclusion of

Table 1 Descriptive statistics, all regression regimes, 2015–2018

Dependent variable	No of obs	Median	Min	Max
ROA	13,141	1.26	−68.01	14.90
ROE	13,141	6.03	−124.69	38.48

Table 2 Fixed effects regression results with ROA as dependent variable 2015–2018

	Dependent variable regime			
	(1)	(2)	(3)	(4)
	ROA (no controls)	ROA (w controls)	Sample below Median	Sample above Median
Cyber risk	40.64*** (14.61)	39.87*** (14.54)	59.63** (28.97)	-1.019 (7.11)
Risk vector	yes	yes	yes	yes
Firm controls	no	yes	yes	yes
Market controls	no	yes	yes	yes
Industry controls	no	yes	yes	yes
Time controls	no	yes	yes	yes
Intercept	-30.11** (13.92)	-50.05** (20.96)	-118.6*** (39.80)	23.72** (9.53)
No. of Obs	13,141	13,141	6552	6589
Overall R ²	0.006	0.016	0.025	0.029

Columns (1) and (2) show values without and with controls, respectively. Columns (3) and (4) split the sample at the median, with column (3) containing all observations below the median (indicating poorer performance) and column (4) containing all observations above the median (indicating stronger performance)

***, **, * indicates significance at 0.01, 0.05, 0.1 level, respectively. Robust standard errors in parentheses

topic values increased the R^2 by about 60 percent from 0.010 to 0.016. This provides evidence that RFDs are usefully informative. Firm performance improved with disclosure (p -values < 0.01). After controlling for firm, market, and industry effects, the ROA improved by about half a point for each percentage increase in cyber topic disclosures. Furthermore, results were significant (p -value < 0.05) for worse performing firms, showing higher R^2 and coefficient values (Table 2, column (3)) while being inconclusive for strong performers (Table 2, column (4)).

The results in Table 3 tell a similar story with ROE measures. Including the topic model values nearly doubled the R^2 from 0.010 to 0.018, providing further support that disclosure contains relevant information and disclosure about cyber risks has a material effect on firm performance (p -values < 0.01). After controlling for firm, market, and industry effects, the ROE improved by about one point by each percentage increase in cyber topic disclosures. Moreover, results were again significant (p -value < 0.01) for worse performing firms (column (3)) with a higher coefficient on cyber security and higher R^2 overall, while strong performers demonstrated no statistically significant effect.

Discussion

Transparency through public filings is crucial to well-informed, smoothly functioning markets. Information plays a key role in a firm's operations and how markets assess value. However, disclosures are expensive to comply with and to interpret (Heinle & Smith, 2017). As RFDs are a major contributor to length increases in

Table 3 Fixed effect regression results with ROE as dependent variable, 2015–2018

	Dependent variable regime			
	(1)	(2)	(3)	(4)
	ROE (no controls)	ROE (w controls)	Sample below Median	Sample above Median
Cyber risk	126.5*** (32.26)	122.4*** (32.46)	200.0*** (63.33)	8.512 (15.24)
Risk vector	yes	yes	yes	yes
Firm controls	no	yes	yes	yes
Market controls	no	yes	yes	yes
Industry controls	no	yes	yes	yes
Time controls	no	yes	yes	yes
Intercept	-103.7*** (30.86)	-57.06 (37.62)	-128.8* (67.74)	-0.0101 (16.95)
No. of Obs	13,141	13,141	6567	6574
Overall R ²	0.010	0.018	0.033	0.035

Columns (1) and (2) show values without and with controls, respectively. Columns (3) and (4) split the sample at the median, with column (3) containing all observations below the median (indicating poorer performance) and column (4) containing all observations above the median (indicating stronger performance)

***, **, * indicates significance at 0.01, 0.05, 0.1 level, respectively. Robust standard errors in parentheses

10-K statements, it is vital to assess those costs. However, determining the effect of information is challenging, even more so with unstructured narrative data. Many studies use NLP methods to measure effects on stock return volatility after report release dates. While some provided evidence to doubt their usefulness (Bao & Datta, 2014; Beatty et al., 2019), many (Campbell et al., 2014, 2019; Hope et al., 2016; Li et al., 2018) supported the informativeness of RFDs. However, narrow event studies focused on only one side of the information flow. They assumed value is only generated by reducing information asymmetries between parties. Examining only the investor response neglects how disclosure affects firm behavior. The premise here is stronger: not only are RFDs informative, but they also imply actions taken by management. A disclosure suggests management is better prepared to handle risks should such circumstances arise, even though the firm is itself the source of the information. This argument is supported with multifactor fixed effects regressions over topic models while controlling for confounding effects.

This study contributes to the literature in several ways. First, unsupervised machine learning models were used to analyze the content of risk disclosures through an emergent generation of topics rather than imposing top-down structure through dictionary methods. Previous studies have not found much statistical significance using this methodology. Second, the study considered how information correlated with longer measures of firm performance beyond the narrow window of an event study. The generated categories were regressed on one year ROA and ROE. The data were also generated more recently (fiscal year ends from 2015 to 2018)

and scraped directly from the SEC's EDGAR (2018) website following updated guidance on RFD requirements in 2015. The sample size was approximately 13,000 firm-year observations with additional controls related to firm, market, and industry effects.

The results corroborate the value of RFDs and that information about future performance is captured in forward-looking statements. The data and methodology identified statistical significance across nearly all categorical risks (compared to less than a third in previous studies using unsupervised topic models (Bao and Datta (2014))). From there, the analysis honed in specifically on cybersecurity risk. This topic is pinned to precise textual artifacts versus more general-purpose risk types and represents hazards facing most industries with substantial costs. Results were statistically significant (p -values < 0.01) for both ROA and ROE. Finally, coefficients and p -values across all regression regimes increased substantially when examining the bottom half of firms in the sample. Poor performance indicated that certain risk factors materialized. Worse performing firms showed higher correlation with risk topics, supporting the argument that management had better procedures in place for such outcomes.

In conclusion, these results go beyond arguments from previous studies regarding the value of public company information. Disclosure represents more than a passive information transfer to investors and other parties; it conveys active risk management that transforms firm behavior as well. In considering risk possibilities, management considers the consequences of their contingencies beyond mere sentences in the annual report. For example, after a data breach, firms tended to increase their disclosures about cybersecurity (Hope et al., 2016). It is not better transparency about risks that matters, but an elevated assessment of cybersecurity and stronger protocol by the firm. The relevant change is not the wording but the firm's preparedness. RFDs are as much a risk management disclosure as they are an information transfer. These disclosures are not cheap talk; they imply firm actions to address such possibilities. Boilerplate language would not reveal statistical significance across categories. This counters arguments about disclosure as mere signaling. Going through the motions of the exercise actually exercises the body. Going through the procedures about information security translates into taking those procedures more seriously. As firms increase disclosures about such risks, both the market and firm are changed regarding those issues.

Overall, the analysis showed that disclosures of risk factors provided market participants useful information and strengthens a firm's capacity to respond. This supports theoretical research about public data reducing information asymmetries among investors (Francis et al., 2004) and aligns with previous research on RFD informativeness (Campbell et al., 2014). These findings have practical applications since RFDs create value for investors and managers. The market does not seem to punish pessimistic news, which helps unwind incentives to suppress material risks. In fact, fair risk assessment improved longer term performance especially in materialization of such risks because it forced management to have contingent strategies. If both sides gain from disclosure, there may be concern for RFDs becoming boilerplate by describing all possible contingencies. Other authors have considered the costs and benefits associated with voluntary disclosure (Beyer et al., 2010; Hope

et al., 2016), but constraints on length or specificity might encourage an appropriate prioritization of firm-specific risk factors. Of course, there are limitations to the methods used here, particularly in the clustering methods in the topic model. While the goal was to minimize structures imposed on the modeling space until after it was formed, more robust evaluation methods will surely improve the interpretability of textual artifacts.

References

- Bao, Y., & Datta, A. (2014). Simultaneously discovering and quantifying risk types from textual risk disclosures. *Management Science*, *60*(6), 1371–1391.
- Beatty, A., Cheng, L., & Zhang, H. (2019). Are risk factor disclosures still relevant? Evidence from market reactions to risk factor disclosures before and after the financial crisis. *Contemporary Accounting Research*, *36*, 805–838.
- Beyer, A., Cohen, D. A., Lys, T. Z., & Walther, B. R. (2010). The financial reporting environment: Review of the recent literature. *Journal of Accounting and Economics*, *50*(2), 296–343.
- Campbell, J. L., Chen, H., Dhaliwal, D. S., Lu, H., & Steel, L. B. (2014). The information content of mandatory risk factor disclosures in corporate filings. *Review of Accounting Studies*, *19*(1), 396–455.
- Campbell, J. L., Cecchini, M., Cianci, A., Ehinger, A., & Werner, E. M. (2019). Tax-related mandatory risk factor disclosures. *Review of Accounting Studies*, *24*(1), 264–308.
- Chen, J., Henry, E., & Jiang, X. (2022). Is cybersecurity risk factor disclosure informative? Evidence from disclosures following a data breach. *Journal of Business Ethics*. <https://doi.org/10.1007/s10551-022-05107-z>
- Dyer, T., Lang, M., & Stice-Lawrence, L. (2017). The evolution of 10-K textual disclosure: Evidence from latent Dirichlet allocation. *Journal of Accounting and Economics*, *64*(3), 221–245.
- ebay Inc. (1998). *Form S-1*. <https://sec.report/Form/S-1>. Accessed 1 Dec 2021.
- Exec. Order No. 14028 (2021). *Improving the nation's cyber security*. 86 F.R. 26633–26647. Available at: <https://www.gsa.gov/technology/technology-products-services/it-security/executive-order-14028-improving-the-nations-cybersecurity>
- Francis, J., Lafond, R., Olsson, P. M., & Schipper, K. (2004). Costs of equity and earnings attributes. *The Accounting Review*, *79*(4), 967–1010.
- Heinle, M., & Smith, K. (2017). A theory of risk disclosures. *Review of Accounting Studies*, *22*, 1459–1491.
- Hope, O. K., Hu, D., & Lu, H. (2016). The benefits of specific risk-factor disclosures. *Review of Accounting Studies*, *21*(4), 1005–1045.
- Huang, K. W., & Li, Z. (2011). A multilabel text classification algorithm for labeling risk factors in SEC Form 10-K. *ACM Transactions on Management Information Systems*, *2*(3), 1–19.
- Isiaka, A. (2018). Three essays on risk factor disclosures. *Theses and Dissertations*, 2032. Available at: <https://scholars.wlu.ca/etd/2032/>. Accessed 1 Jun 2021.
- Israelsen, R. D., & Yonker, S. E. (2017). Key human capital. *Journal of Financial and Quantitative Analysis*, *52*(1), 175–214.
- Kravet, T., & Muslu, V. (2013). Textual risk disclosures and investors' risk perceptions. *Review of Accounting Studies*, *18*, 1088–1122.
- Lewis, J., Smith, Z., & Lostri, E. (2021). The hidden costs of cybercrime. *Center for Strategic and International Studies*. www.csis.org/analysis/hidden-costs-cybercrime. Accessed 1 Feb 2022.
- Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, *45*(2), 221–247.
- Li, H., No, W. G., & Wang, T. (2018). SEC's cybersecurity disclosure guidance and disclosed cybersecurity risk factors. *International Journal of Accounting Information Systems*, *30*, 40–55.
- Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, *54*(4), 1187–1230.
- Mergent. (2022). Mergent Online. Available at: <https://www.mergentonline.com/noticescm.php?page=about>. Accessed 1 Sep 2021.

- Schrand, C. M., & Elliott, J. A. (1998). Risk and financial reporting: A summary of the discussion. *Accounting Horizons*, *12*, 271–282.
- SEC. (2005). Securities and Exchange Commission final rule. Release no. 33–8591 (FR-75), www.sec.gov/rules/final/33-8591.pdf. Accessed 1 Aug 2021.
- SEC. (2018). EDGAR database. Available at: <https://www.sec.gov/edgar.shtml>. Accessed 1 Aug 2021.
- SEC. (2022). Financial Statement and Notes Datasets. Available at: <https://www.sec.gov/dera/data/financial-statement-and-notes-data-set.html>. Accessed 1 Aug 2021.
- Wei, L., Li, G., Zhu, X., Sun, X., & Li, J. (2019). Developing a hierarchical system for energy corporate risk factors based on textual risk disclosures. *Energy Economics*, *80*, 452–460.
- Zhu, X., Yang, S. Y., & Moazeni, S. (2016). Firm risk identification through topic analysis of textual financial disclosures. *IEEE Aerospace and Electronic Systems Magazine (SSCI)*, 1–8.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.