Understanding Risk Disclosures and Exposures: Insights from a Novel Measure of Information Content

Bruce D. Grundy[a]
University of Melbourne

Stefan Petry[b]
University of Manchester

15 May 2022

Abstract

We use machine-learning to investigate the information content of 10-K Item 1A Risk Factors disclosures by identifying and quantifying 30 risk-factors. Sentence-counts of salient risk-factors are strongly positively related to 14 familiar proxies for firm risk. The average of the adjusted R^2 values of the 14 risk-proxy, risk-factors relations estimated each year is not decreasing through time and the hypothesis that Item 1A's informativeness is declining is rejected. Indices of operating and financing risk constructed from risk-factor sentence-counts help explain asset and equity volatility. Annual changes in the operating and financing risk indices are significantly related to subsequent firm performance.

JEL classification: G10, G38, K22, M41

Key words: risk-factors, textual analysis, 10-K, risk indices

[a] Research School of Finance, Actuarial Science and Statistics, Australian National University. Bruce.Grundy@anu.edu.au

[b] Alliance Manchester Business School, University of Manchester, Booth Street West, Manchester M15 6PB, United Kingdom. +44 (0)161 306 2069. stefan.petry@manchester.ac.uk

We thank Jaewon Choi, Gerard Hoberg, Aneel Iqbal, Zacharias Sautner, Catherine Schrand, Stefano Zedda, and participants at the Alliance Manchester Business School, Oslo Business School, Monash University, the 2021 European Financial Management Association Conference, the 2021 French Finance Association Conference, and the 2021 British Accounting and Finance Association Conference for comments and advice. We thank Qi Deng and Zhuo Wei for valuable research assistance.

Firms face differing exposures to a multitude of risks. Some face high risks associated with product defects and the ensuing litigation, while for others the greatest risk is the loss of a major customer. Understanding the entirety of these risks is important for investors, analysts, regulators, and policymakers. Firms are required to disclose the risks they face in Item 1A of their annual 10-K filings. The information is provided in a textual format. Because of the complexity and impossibility of manually interpreting and comparing the risk disclosures of thousands of firms, many questions remain unanswered. First and foremost, is the information contained in Item 1A accurate? Answering this question requires a method for evaluating the accuracy of textual disclosures. Further questions concern whether the accuracy of Item 1A disclosures has changed over time? Do changes in disclosed risks lead or lag changes in observed metrics of risk exposures? And finally, can the risk disclosures in Item 1A be used to improve our understanding of firms' asset volatility and future operating performance?

To address these questions, we must first overcome two challenges. First, how to identify risks being disclosed textually, and second, how to quantify those risks, in a large panel of annual reports. Our sample consists of 13,470 firm-year observations on 1,708 S&P 1500 firms between 2005, when the disclosure of risks in 10-Ks was made mandatory, and 2015. We employ a variant of the well-established Latent Dirichlet Allocation (LDA) machine learning algorithm of Blei, Ng and Jordan (2003) to identify 30 risk topics, which we refer to as risk-factors, in the sample firms' Item 1A risk disclosures. We quantify the identified risk-factors in a simple and intuitive way, which serves as a measure of firm-level risk exposures; The more sentences a company uses to describe a particular risk-factor, the larger the importance of that risk for the firm. Theoretical support for this approach is provided in Lopez-Lira (2021) which establishes sufficient conditions for firms to optimally trade off benefits and costs of disclosing risk accurately, with more lengthy disclosures of greater risks.

Regulation S-K requires the reporting of "the most significant factors that make [an investment in the security] speculative or risky." While firms may have incentives, for example for competition reasons, to withhold the disclosure of certain types of risks, this incentive is tempered by the risk of shareholder litigation. Nelson and Pritchard (2016) show that after the introduction of the mandatory disclosure of Item 1A, even firms with ex-ante low-litigation-risk started to provide a similar level of disclosure (measured by word count, readability, and statement similarity across years) as ex-ante high-litigation-risk firms.

Having identified and quantified the 30 risk factors in Item 1A via LDA, we manually assign a label to each of the 30 LDA risk factors for expositional purposes only. For example, the top 5 risk factors, based on sentence counts, mentioned by the S&P1500 firms in our sample are *Product Innovation* risk, *Revenue Volatility* risk, *Supply Chain* risk, *Net Cash Flow Volatility* risk, and *Energy Sector* risk. It is important to recognize that our choice of labels has no impact on the outcome of our analysis, as we explain in Section II.

We first confirm via univariate tests that using the sentence-counts of the risk-factors as risk exposure quantifications provides intuitive outcomes. For example, the *Energy Sector* risk-factor is the most frequently mentioned risk-factor in industries associated with energy, such as coal mining, metal mining, oil & gas extraction, and petroleum & coal products. The *Real Estate* risk-factor is the most frequently mentioned factor in the SIC code-defined General Building Contractors and the Hotels & Other Lodging Places industries.

We address our first research question by examining whether the risk disclosures in Item 1A are informative in the sense of accurately describing a firm's risks. We use the terms 'accuracy' and 'informativeness' interchangeably. While theoretical relations between information disclosure and market quality, information production, real investment efficiency, and investor welfare are well-established (Goldstein and Yang (2017)), there is little empirical evidence on the accuracy of textual risk disclosures. Rather than asking how Item 1A disclosures relate to prices and trading volumes, we examine their relation to proxies for firm-level risk. We consider 14 familiar risk-proxies that measure various uncertainties, including customer concentration risk, the firm's sensitivity to economic conditions, competitive pressures associated with product innovation, financing-related risks, intangible asset risk, corporate governance risk, and uncertainties associated with the costs of goods sold, net cash flows, revenues, and the value of the firm. However, our approach is not limited to the 14 risk proxies and applies in a straightforward manner to alternate risk-proxies of interest.

We measure the accuracy of Item 1A as the adjusted R^2 of linear regressions of the 14 risk-proxies on the sentence-counts of salient risk-factors. We use backward-selection and a random subset of 40% of the sample to identify parsimonious sets of the LDA risk-factors with *p*-values associated with a null of no multivariate relation between the risk-proxy under examination and the risk-factor that are less than 1%. We term this set the salient risk-factors for the risk-proxy. We eliminate risk-factors with *p*-values associated with a null of no relation with the risk-proxy that exceed 1%. Once the salient risk-factors for each risk-proxy are identified, we use a hold-out sample of the remaining 60% of the observations to estimate regressions of risk-proxies on their salient risk-factors' sentence-counts. The adjusted R^2 values of these relations provides a natural metric of the information content of Item 1A and can be viewed as indicative of the accuracy of 10-K statements. The approach naturally accommodates both overlap in risk-factors as explanators of a risk-proxy as well as a multiplicity of risk-proxies.

Our findings of the contemporaneous relation between risk-proxies and salient risk-factor sentence-counts show that each of the 14 risk-proxies is related to numerous risk-factors, and several risk-factors are significant explanators of a number of risk-proxies. This highlights the multi-dimensional nature of risk. After removing the effects of industry membership, we find that, on average, 45% of the remaining variation in the risk-proxies is explained by firm-level variation in the salient risk-factors.

When we examine the relation between contemporaneous *changes* in risk-factor sentence-counts and *changes* in risk-proxies, we find the adjusted R^2 are much lower than in the levels regression. However, the *p*-values of the null of no relation is still significant at the 5% (1%) level for ten (seven) of the 14 risk-proxies. We also examine whether annual changes in risk-factor sentence-counts contain forward-looking looking information about future changes in risk-proxies and seven (six) of the *p*-values of the predictive relations are significant at the 5% (1%) level. The reduced R^2 values of the changes regressions relative to the levels regressins is consistent with measurement error in the differenced regressors being higher than in the levels of the regressors (Angrist and Pischke, 2009). The result suggests that the Item 1A risk disclosures are more informative in stable environments and lose accuracy when risk is changing.

We further investigate time-variation in the informativeness of Item 1A by examining changes through time in the adjusted R^2 of the risk-proxy, risk-factor relations. We find that the average adjusted R^2 is consistently around 28% between 2005 and 2015 and we reject the hypothesis that risk disclosures have become less informative through time. The results provide an opposing view to the notion that lengthier annual reports are associated with lower information quality. For example, Dyer, Lang and Stice-Lawrence (2017) document a post-1996 increase in the length of 10-K statements and Item 1A, which is accompanied by an increase in boilerplate content, stickiness, and redundancy of the descriptions in annual reports. An increase in these textual attributes is typically interpreted as a decrease in informativeness. Similarly, based on textual analyses of Management Discussion and Analysis (MD&A) disclosures, Brown and Tucker (2011) suggest a decline in MD&A usefulness over time. In 2019 the SEC called for public comment on proposals to discourage both repetition and the disclosure of immaterial information in Item 1A.[1] But, as Miller (2017) postulates, even if annual reports have become longer, less readable, and more boilerplate over time, they need not have become less informative. For example, simply repeating every sentence in a 10-K may make statements less readable, but their information content will not be reduced. What has been missing to date, and what our methodology provides, is an empirical measure of the accuracy of Item 1A at a point in time and through time for a large sample of firms and over any extended period.

Finally, we demonstrate the utility of the sentence-counts of LDA-identified risk-factors by using them to construct meaningful indices of operating risk and financial risk. We employ these risk indices in two quite different settings. First, we show that cross-sectional differences in firms' asset volatilities reflect differences in the operating and financial risk indices in a manner consistent with the relation between asset volatility and leverage documented in Choi and Richardson (2016). In addition, the operating and financing risk indices are shown to be significant positive explanators of equity volatility. Second, we show that larger increases in either the operating risk index or the financing risk index are associated with significantly lower operating income, net income, and sales performance in the second, third and fourth quarters of the following year. Increases in the financing risk index are associated with significantly lower performance in the first quarter of the following year, though this is not the case for increases in the operating risk index. A one-sentence increase in the sentence-count associated with financing risk has a significantly more negative effect on subsequent performance in every quarter of the following year, than a one-sentence increase in the discussion of operating risk. This result complements the existing literature, which has demonstrated the negative effect of a decrease in the similarity of 10-

[1] RIN 3235-AL78 Modernization of Regulation S-K Items 101, 103, and 105.

K statements in consecutive years on operating income, net income, and sales in the second quarter of the following year (Cohen, Malloy and Nguyen, 2020).

Our findings are important for academics, regulators, and investors. First, our research is among the first to assess the accuracy of textual information in annual reports in a quantifiable and reproduceable way. A methodology that enables a quantitative assessment of the accuracy of Item 1A was not well developed in 2019 when the SEC called for comments on proposed amendments to Item 1A designed to discourage repetition and the disclosure of immaterial information.[2] Second, our findings show that, while Item 1A may have become longer over time, its accuracy has not declined. Third, our approach will be particularly useful in measuring exposures to risks for which risk-proxies are not readily available. One way to determine a firm's exposure to catastrophe risk would be to investigate its Item 1A similarity to that of firms that have experienced a natural disaster in the past After determining the set of risk-factors that have high sentence-counts for firms that previously experienced a catastrophe, presumably this set includes the *Catastrophe* risk-factor, firms with similarly high sentence-counts for these risk-factors could be classified as being highly exposed to catastrophe risk.[3] Finally, our approach can be used to develop quantitative metrics of operational risk for firms in the financial industry, measures of which are still rare (Brown, 2012; Brown, Goetzmann, Liang and Schwarz, 2012). Constructing such a metric would also fulfil part of the Basel requirement of an operational risk, namely that its dimensionality be reduced to a single quantity. From a research perspective, our LDA-based operating and financing risk indices are available for all publicly listed firms, rather than for the small fraction of the public firms in the Compustat database for which loan data is available.

Section I reviews the literature on the textual analysis of firm risk and describes the LDA identification of risk factors. Section II describes the data and presents summary statistics. Section III documents the R^2 of the contemporaneous relations between LDA-identified risk-factors and various proxies for firm risk. Section III also investigates the incremental information content of Item 1A Risk Factor section given knowledge of a firm's industry membership. Section IV

[2] RIN 3235-AL78 Modernization of Regulation S-K Items 101, 103, and 105.

[3] Florackis, Louca, Michaely and Weber (2021) use this style of approach and compare the language in Item 1A of firms that had previously experienced a cyber-attack with that of other firms. Firms with high measured textual similarity to that of the attacked firms are predicted to have high future cybersecurity risk.

examines the changing informativeness of Item 1A through time. Section V examines the predictive relation between changes in risk-factors and changes in risk-proxies. Section VI constructs firm-level operating and financial risk indices from the LDA-identified risk-factors and examines the indices' relations with asset and equity volatility. Section VI also shows that changes in the indices are negatively related to future levels of operating income, net income, and sales. Section VII concludes.

I. Textual Analysis of Risk Reporting

Li (2010) and Loughran and McDonald (2016) survey the growing finance literature that uses textual analysis. Textual analysis in finance and accounting research has primarily been dictionary-based. Only recently, has unsupervised learning approaches have been employed, most of which still rely on a bag-of-words assumption that ignores the sequence of words when identifying the thematic structure in documents. Researchers have applied textual analysis to investigate financial constraints, innovation, the competitive position of firms and their industry membership, the role of financial analysts, and communication with regulators. Bodnaruk, Loughran and McDonald (2016), Hoberg and Maksimovic (2015) and Law and Mills (2015) use textual analysis of 10-K statements to develop measures of financial constraints that have incremental power relative to the Kaplan and Zingales (1997), Hadlock and Pierce (2010), and Whited and Wu (2006) financial constraint indices. Hoberg, Phillips and Prabhala (2014) develop the fluidity measure of the product market threat faced by a company by examining the overlap of the words in a firm's product description in its 10-K statement in a given year with a normalized measure of the aggregate change in the words describing other firms' products. High-fluidity firms are shown to be less likely to distribute dividends or repurchase their stock and to hold more cash than do low-fluidity firms. Ball, Hoberg, and Maskimovic (2015) apply LDA to the Management, Discussion and Analysis section of firms' 10-K statements and conclude that textual information is a better explanator of the valuation of firms undergoing business change than is the content of financial statements. Hoberg and Phillips (2016) use the similarity of firms' 10-K product descriptions to determine a novel measure of industry membership. Bellstam, Bhagat and Cookson (2020) identify topics within analyst reports and conclude that firm performance and growth opportunities are positively related to firm-level similarity of the innovation topic with textbook innovation language. Huang, Lehavy, Zang and Zheng (2018) conclude that analysts discover information beyond that disclosed in conference calls by comparing analyst reports issued soon after earnings conference calls to the content of the calls themselves. Lowry, Michaely and Volkova (2020) analyze communications between the SEC and issuing firms prior to IPOs and show that increased SEC concern about revenue recognition is associated with a higher probability of withdrawal of the IPO.

Particularly relevant for our study are papers that employ textual analysis to investigate the risk-related information in 10-K Statements. Kravet and Muslu (2013), Filzen (2015), and Hope, Hu and Lu (2016) examine the ability of risk-related information to move prices and/or induce trade. Kravet and Muslu (2013) find that an increase in the number of sentences containing at least one pre-defined risk-related word in a firm's 10-K is associated with an increase in the firm's stock return volatility and trading volume after the filing date relative to the level before the filing. Filzen (2015) finds that the abnormal return around the filing of a quarterly update to a firm's risk-factor disclosures is significantly negatively related to whether or not the risk-factor section of a second or third quarter 10Q report is more than 100 words longer than the preceding quarter's risk-factor section and interprets the result as consistent with a firm's preference for withholding bad news. Defining specificity as the occurrence of names of persons, locations, and organizations and quantitative values in percentages and dollars, times, and dates, Hope, Hu and Lu (2016) find that more specific risk-factor disclosures are associated with larger absolute values of three-day abnormal returns and greater trading volume around the 10-K filing date. Brown, Tian and Tucker (2018) find evidence that companies monitor the SEC's response to the risk-factor reporting of peer firms in that SEC comment letters about the risk-factor disclosures of some firms not only lead to changes in the disclosures of those firms, but also in the disclosures of their peers.

Cohen, Malloy and Nguyen (2020) examine the textual similarity of annual 10-K filings and of same-quarter 10-Q filings in years $t-1$ and t and conclude that over the second quarter of year $t+1$ firms with more dissimilar filings experience lower returns than firms with more similar filings and that the predictability is strongest when changes occur in the Risk Factors section. Although returns at the time of the filing are not related to the degree of change in the Risk Factors section, subsequent returns are impacted, and thus Cohen, Malloy and Nguyen (2020) conclude that textual changes in Item 1A are informative with the information impounded into prices with a "lazy" delay. Considering firm operations over the second quarter of year $t+1$, firms with less textual similarity of their financial statements from one year to the next are shown to have significantly lower operating income, net income, and sales performance than firms with more similar financial statements.

Brown and Tucker (2011) examine whether changes made in the Management Discussion and Analysis (MD&A) section of 10-K reports from one year to the next contain information. They find that firms experiencing larger economic changes modify the MD&A section more and that larger modifications are associated with a larger stock price response to the filing of the respective 10-K report. Campbell, Chen, Dhaliwal, Lu and Steele (2014), Goldsmith-Pinkham, Hirtle and Lucca (2016), Hanley and Hoberg (2019) and Lopez-Lira (2019) use LDA analysis to classify the risks discussed in Item 1A. Campbell et al. (2014) combine a set of key words based on prior literature to a list of additional words that repeatedly appeared in Item 1A identified using Latent Dirichlet Allocation and manually classify keywords as related to financial, litigation, tax, other-systematic, or other-idiosyncratic risks. The keyword count and the percentage of keywords associated with each classification are cross-sectionally related to observable measures of systematic and idiosyncratic risk. Goldsmith-Pinkham, Hirtle and Lucca (2016) use LDA to identify five topics in the matters requiring attention in supervisory demands raised by Federal Reserve examiners for corrective actions by banks. Hanley and Hoberg (2019) show that the application of an LDA analysis of the Item 1A section of bank 10-Ks can detect emerging risks in the financial sector. Lopez-Lira (2019) develops a factor model of returns from an LDA analysis of Item 1A that identifies the four risk topics that affect the largest number of firms and topic-mimicking portfolios of firms that allocate more than 25% of their Item 1A disclosure to a discussion of one of the four principal risk topics are constructed. Rather than focusing on a small number of risks, our LDA analysis of 30 risk-factor topics for S&P 500 firms over an extended period involves a high level of granularity and by backward-selection, sets of the salient topics for each risk-proxy can be identified. Thus, we can undertake a rigorous investigation of the accuracy of the Item 1A risk-factor section.

In recent work, management-analyst discussions in quarterly earnings calls have been used as a data source to analyse firms' exposures to particular types of risk (Hassan, Hollander, van Lent and Tahoun, 2019; Hassan, Hollander, van Lent, Schwedeler and Tahoun, 2021; Sautner, van Lent, Vilkov and Zhang, 2021). However, those studies differ from ours in three important ways. First, and most importantly, this work either relies on a dictionary-based supervised learning algorithm to identify terms or on consecutive words (bigrams) that signify the particular risk in question, whereas we employ an unsupervised learning algorithm. Second, the studies look at one particular type of risk (political risks (Hassan et al., 2019), exposure to epidemic diseases (Hassan, et al., 2021), or climate risks (Sautner et al., 2021)), whereas the objective of our study is to extract and quantify all reported risks. Third, earnings calls are voluntary disclosures, i.e., firms are not required to hold earnings calls and only around 50% of firms do. Firms that select to hold earnings calls decide which topics to discuss and which analysts to invite. In fact, there is evidence that some firms invite favourable analysts to join the discussion in earnings calls in order to keep the tonality of the call positive and avoid negative questions (Cohen, Lou and Malloy, 2020).

II. 10-K Risk Factors

A. Data and Sample

We examine Item 1A risk disclosures by S&P 1500 firms over the period from 2005 when the disclosure of risk became mandatory through 2015. 10-K statements are retrieved from the SEC's Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system. Firms' accounting and stock market data are retrieved from the CRSP-Compustat merged database. Asset betas are obtained from Jaewon Choi's homepage.[4] Following Fracassi, Petry and Tate (2016), we use data from TRACE and Mergent FISD to calculate bond-level credit spreads at fiscal year-ends, or if the bond did not trade on that date, then on the closest preceding bond trade date. Firm-level credit spreads are calculated as a weighted average across each firm's bonds with weights given by the amount of the issue outstanding relative to the total of all available bonds for the same firm. HHI index and fluidity measures based on text-based industry classifications are obtained from the Hoberg-Phillips Data Library. Appendix Table A2 provides a summary of the variables used in the analysis as well as the data sources.

We remove firm-year observations if data on any of total assets, leverage or market capitalization are missing or if the book value of equity is negative. We exclude firms with SIC

[4] https://sites.google.com/site/jaewchoi1203

codes beginning with 6; i.e., finance, insurance, and real estate institutions. The final sample contains 13,470 firm-year observations on 1,708 firms.[5] Table I reports summary statistics for total assets, asset volatility and 14 proxies for firm risk. The average book value of assets is $8.19 billion. The average book debt to assets ratio is 20.4%. On average R&D Expenditures and Intangible Assets comprise 3.1% and 74.3%, respectively, of firm assets. 49.9% of observations involve firms that have a major customer meaning a customer that accounts for at least 10% of sales. The average asset beta is 1.079 and the mean volatilities of quarterly costs, quarterly net cash flows, and quarterly revenues as a percentage of firm assets, are 5.1%, 3.7%, and 7.3%, respectively. The average annualized asset return volatility is 32.3% and the average annualized stock return volatility is 40.4%.

B. Risk Factor Identification and Quantification

To identify and quantify the set of risk factors discussed in Item 1A of firms' 10-K statements, we employ the Bao and Datta (2014) modification of the Blei, Ng, and Jordan (2003) LDA algorithm. LDA is an unsupervised machine learning technique that identifies clusters of words that tend to appear in the same context, referred to as topics. LDA assumes that the words that form the discussion of a topic, and the topics within a document, have a pre-defined distribution. Traditional LDA assumes that the order of words does not matter (a "bag-of-words" approach) when assigning words to topics. Bao and Datta (2014) modify this analysis by assuming that each sentence in Item 1A discusses a single topic and show how to assign the sentences in each document to the various topics so as to maximize the likelihood that the documents examined have the observed word and sentence content. Bao and Datta (2014) show that the sentence-based LDA (*sent*-LDA) method outperforms traditional LDA in its ability to identify more-meaningful topics. Importantly, by associating sentences to identified risk topics, *sent*-LDA provides a quantifiable estimate of the importance of each topic within firms' Item 1A. That is, relative to other firms in the same year, the more sentences that a firm devotes to a given risk-factor, the more important that risk-factor is to the firm. We apply *sent*-LDA to the text corpora of all the retrieved Item 1A after first applying standard text pre-processing techniques that delete words that belong

[5] The sample size is comparable to that of other studies of Item 1A. Campbell et al. (2014) examine 9,076 firm-years over the 2005–2009 period. Hope, Hu and Lu (2016) examine 14,865 firm-years over 2006–2011. Filzen (2015) examines 13,165 firm-quarters over 2006–2010.

to a "stop" list used in computational linguistics (Griffiths and Steyvers (2004)). These are meaningless words for the LDA topic model like "the", "or", "and", "for", etc. In addition, topically similar words are reduced to their root; e.g., acquisition, acquire, acquired, all have the same stem, "acqui". LDA is based on transparent assumptions and produces results that are replicable and independent of human judgement. The only human input needed is the choice of the number of topics to be detected.

Our choice of 30 risk factors is comparable to Huang and Li (2011), Mirakur (2011), and Bao and Datta (2014). The choice of number of topics can be driven by a trade-off between interpretability and topic overlaps. If one chooses too many topics, word clusters will overlap, and topics will be harder to differentiate form one another. If too few topics are chosen, they will be hard to interpret as the words associated with the topic will represent several key themes. Because we use backward-selection in a training dataset to winnow out those factors that do not contribute to explaining cross-sectional variation in the specific risk measures we examine, the interpretability of the assigned sentences is moot and topic overlap is accommodated naturally by the selection of multiple risk-factors as significant explanators of a given risk-proxy. For each empirical risk-proxy we retain only the salient risk-factors and when several topics link to the one risk, then they are all retained.

For expositional purposes we ascribe a label to each risk-factor. Our choice of labels has no impact on our analysis of the linkage between risk-factors and risk-proxies, our conclusions about the accuracy of risk disclosure statements, or our conclusions about the changing informativeness of risk disclosure statements through time. The firm-level operating and financing risk indices we create from the risk-factors are not affected by the labelling of the risk factors. To ensure that the chosen labels are descriptive, we follow three steps. First, we assign an initial label to each LDA risk-factor based on the 30 most common words in the sentences assigned to the risk-factor.[6] Second, we identify firms with a high sentence-count for each factor and read the paragraphs in Item 1A in which the most common words appear. For some risk-factors, this led to a slightly different sense of the risk than was initially suggested by the most common words, and we adjusted the label accordingly. In Appendix Table A.1, we report the risk-factor labels and the

[6] The 30 most common words associated with each risk-factor are set out in Table A1.

most common words in order of frequency.[7] For example, the 30 most common words associated with topic 18 in Appendix Table A.1 might suggest that this risk factor is about litigation risk. After reading the matching paragraphs in Item 1A, however, it becomes clear that the discussion concerns the risk of product defects and ensuing litigation or reimbursement claims without litigation. Accordingly, we label topic 18 *Product Defects* and not litigation risk. Similarly, we initially chose the labels *Credit Market Conditions*, *Debt* risk, and *Financing Uncertainty* risk for topics 7, 8, and 9. However, because of the similarity of many of the sentences assigned to these three risks, we opted not to instead use the labels *Financing I*, *Financing II* and *Financing III*. Readers troubled by our assignment of labels can refer to the risk-factors by the generic labels *Factor 1* through *Factor 30* and they will find the analysis and conclusions unchanged. The labels we chose to assign turned out to be similar to those assigned by Bao and Datta (2014) and Huang and Li (2011). Interestingly, the 30 most common words associated with our set of LDA risk-factors have a strong overlap with the descriptions of the risk subcategories that Mirakur (2011) identified manually, suggesting LDA can yield topic classifications comparable to what a human coder can produce.

Our application of *sent*-LDA produces a panel dataset of observations on the sentence-count associated with each of the 30 risk-factors of S&P 1500 firms in each year from 2005 through 2015. Table II reports summary statistics of the sentence-count for the 30 risk-factors. The five risk-factors with the highest average number of associated sentences are *Product Innovation* risk (15.3 sentences), *Revenue Volatility* (14.02), *Supply Chain* risk (12.2), *Net Cash Flow Volatility* (11.8), and *Energy Sector* risk (9.9). The five risk-factors with the lowest average sentence-counts are *Real Estate* risk (1.6), *Financing I* risk (1.7), *Reporting Compliance* risk (2.9), *Corporate Governance* risk (3.3), and *Catastrophe* risk (3.5). Nine risk-factors are viewed as unimportant by most firms in our sample in that for these risks the median sentence-count is zero. These nine less-common risks are *Corporate Governance*, *Energy Sector*, *Financing I*, *Healthcare Spending*, *Product Approval*, *Real Estate*, *Reporting Compliance*, *Tax Uncertainty,* and *Cost Volatility* risk. Note that *Stock Price Volatility* is identified as a risk factor in Item 1A. However, our reading of

[7] One way to visually present the relative frequency of the 30 most frequent words in the sentences assigned to a risk-factor is via a word cloud. Figure A1 in the Appendix depicts the word lists as word clouds.

sentences allocated to the *Stock Price Volatility* risk-factor leads us to conclude that the discussion is intended to make investors aware of their risk when investing and not as a risk-factor of the firm.

We also examine two normalizations of the sentence-count associated with each risk-factor. The first normalization is the sentence-count as a fraction of the firm's total Item 1A sentence-count. This gives a higher quantification of the risk-factor when a firm devotes a larger fraction of Item 1A to a discussion of that factor and provides a control for verbosity. The second normalization is motivated by the property documented in Table II that many firms do not discuss some of the risk-factors. Thus, we examine the risk-factor sentence-count relative to the average sentence-count of those factors that the firm does report on. For ease of exposition, we describe these factors as the firm's nonzero factors. Suppose firm A were to devote N^A sentences to each of M^A nonzero factors and firm B were to devote N^B sentences to each of M^B nonzero factors. For a risk-factor discussed by both firms, the first normalization would measure the importance of the risk as $N^A / M^A \times N^A = 1/M^A$ and $N^B / M^B \times N^B = 1/M^B$, and a risk faced by both firms would be considered less important for the firm that reports on more risks in total.[8] Under the second normalization, the risk is considered equally important to the two firms with $N^A \Big/ \left(\frac{M^A \times N^A}{M^A}\right) = N^B \Big/ \left(\frac{M^B \times N^B}{M^B}\right) = 1$.

C. Risk Factors and Industries

In this section, we investigate differences in reported risk factor quantifications by industry and whether product market competition impacts the sentence count per risk factor. Table III displays the three (for brevity) most frequently mentioned risk-factors based on the average sentence-counts for firms in each SIC 2-digit industry. The classification confirms that our identification and quantification of risk factors provides meaningful results. For example, the *Energy Sector* risk-factor is the most frequently mentioned factor in industries associated with energy; e.g., Coal Mining, Metal Mining, Oil & Gas Extraction, and Petroleum & Coal Products.

[8] While verbosity will affect N, verbosity does not affect the normalized measure.

The *Real Estate* risk-factor is the most frequently mentioned factor in the General Building Contractors and the Hotels & Other Lodging Places industries.

The voluntary disclosure literature that builds on the seminal Verrecchia (1983) analysis suggests that firms favor less disclosure when they face more competition. But if competition magnifies risk exposures and the costs associated with misreporting place a limit on firms' desire to reduce disclosure, then more intense product market competition may be empirically associated with more risk reporting. In Table IV, we provide summary statistics of the sentence-count of each of the 30 risk-factors for industries classified by a measure of industry competition. We classify firms with Hoberg-Phillips *TNIC HHI* values below the median value as operating in competitive product markets and firms with above median *TNIC HHI* values as operating in concentrated markets. Some risk-factors show very large differences across the two market settings. For example, the average sentence-account associated with the *Energy Sector*, *Healthcare Spending*, and *Products Defects* risk factor is 16.08, 12.56, and 14.47, respectively, in concentrated industries and 3.78, 4.19, and 3.52, respectively, in competitive industries. The last row in Table IV shows that, on average, the Item 1A sections of firms in concentrated industries contain approximately 30% more sentences associated with a risk-factor than Item 1A of firms in competitive industries. Any corporate preference for reduced disclosure in more competitive environments seems to be outweighed by higher risk in more competitive environments and competition appears to have a positive net effect on mandatory textual disclosure.

III. The Relation between LDA-Identified Risk-Factors and Risk-Proxies

A. The Accuracy of Item 1A

We use the terms 'accuracy' and 'informativeness' interchangeably and measure the accuracy of Item 1A by the R^2 of linear regressions of 14 observable risk-proxies on quantifications of the sentence-counts of salient LDA-identified risk-factors. An alternative view of informativeness concerns the link between the release of a financial statement and abnormal trading volume and/or price movements (Hope, Hu and Lu, 2016). Our R^2 measure of the accuracy of financial statements is a complement to this notion of informativeness and recognizes that item 1A can contain information useful to investors in designing their optimal portfolios and to regulators irrespective of whether its release moves prices.

For every risk-proxy, we identify the set of salient risk factors via backward selection by estimating stepwise regressions on a random sample of forty percent of each year's observations ("training sample") to eliminate risk-factors with *p*-values that exceed 1% under a null of no relation between the factor and proxy. Rather than reading the Item 1A Risk Factors sections of thousands of annual reports and subjectively linking risk-factors to risk-proxies with no guidance from a theoretical model of the relation, our use of stepwise regressions in a training sample allows us to identify parsimonious regression relations for each risk-proxy without risking data mining in the holdout sample. We estimate multiple regressions of risk-proxies on their salient risk-factors' sentence-counts in the holdout sample of the remaining 60% of the sample observations. The fraction of the variation in a risk-proxy that is explained by variation in the risk-factor sentence-counts is a natural measure of informativeness.[9] Since we examine only linear regression relations, the calculated R^2 values can be thought of as minimum measures of the informativeness of Item 1A.

The 14 risk-proxies we investigate are diverse measures of risk. Their data sources are set out in Table A2. *Cust. Dummy* is a proxy for customer concentration risk and is equal to 1 if the firm has a single customer which accounts for at least 10% of the firm's sales. *Asset Beta* is a proxy for the firm's sensitivity to economic conditions. *TNIC HHI* is used as a proxy for market power and equal to the HHI index of a firm when industry classifications are determined by the Hoberg-Phillips text-based network of firms offering similar products. Item 1A disclosures of firms with high sentence-counts for the *Product Innovation* risk-factor often contain a discussion of innovations by both the firm and its competitors and so we use the Hoberg-Phillips-Prabhala *Fluidity* measure as a proxy for competitive pressures arising from product innovation.[10] The book

[9] Given the absence of SEC guidance on Item 1A, a firm may choose to report the level of its unhedged risk while the observed hedged risk level may be lower. This possibility biases against finding a relation between reported risks and proxies for the risks to which investors are actually exposed.

[10] For example, Item 1A of the 10-K statement of *Time Warner* states that "The Company's competitive position also may be adversely affected by various timing factors, such as delays in its new product or service offerings or the ability of its competitors to acquire or develop and introduce new technologies, products and services more quickly than the Company." Other examples are *Qualcomm* which reports that "our competitors are aggressively pricing products and services and are offering new value-added products and services, which may impact margins, intensify competition in current and new markets and harm our ability to compete in certain markets" and *Telephone and Data Systems* whose subsidiary *U.S. Cellular*'s

debt to assets ratio, *Book D/A*, and the *Credit Spread* between the yield on a firm's bonds and contemporaneous Treasury yields are both proxies for financing-related risks. (*TA – Net PP&E*)/*TA*, being the fraction of a firm's assets not accounted for by property, plant and equipment, is used as a proxy for intangible asset risks. The *R&D Exp/TA* ratio is a proxy for risks associated with intangible assets, intellectual property, product development, and product approval. The Hoberg and Maksimovic (2015) *delaycon* variable is a measure of the tightness of a firm's financial constraints and a proxy for the risk of *Investment Delay* in the event of a negative shock. *Indep. Directors* denotes the fraction of independent directors on the firm's board and is a proxy for corporate governance risk with a higher fraction being associated with lower governance risk (Knyazeva, Knyazeva and Masulis, 2013). σ(*Cost*), σ(*NCF*) and σ(*Revenues*) are the standard deviations of the costs of goods sold, net cash flows, and revenues relative to total assets and are used as proxies for the uncertainties associated with costs, cash flows and revenues respectively. σ_E measured as the annualized stock return volatility estimated from daily stock returns observed over the 250 trading days preceding the fiscal year-end is a proxy for the wealth risk faced by the firm's equity holders.

Table V reports multiple regressions using the holdout sample to estimate the relations between the risk-proxies and the set of associated risk-factors identified by backward-selection. Panels A, B and C differ in their quantification of the sentence-count. Panel A examines the total sentence-count. Panel B uses the risk-factor sentence-count as a fraction of the firm's total Item 1A sentence-count. Panel C uses the risk-factor sentence-count as a fraction of the average sentence-count for the firm's non-zero risk-factors. The principal conclusion from Panel A of Table V is that qualitative information in the Item 1A Risk Factors section of 10-K statements identified via an LDA analysis and quantified on the basis of risk-factor sentence-counts is significantly related to each of the 14 risk-proxies, with the *p*-values associated with a null of no relation being less than 0.0001. The average across years of the yearly averages of the adjusted R^2 values of the relations between the 14 risk-proxies and their salient risk-factors is 18.93%.

"'smart follower' strategy may cause consumers that are eager to adopt new technologies more quickly to select U.S. Cellular's competitors as their service provider".

In the training sample, the risk-proxies are significantly related to the sentence-counts of multiple risk-factors, with an average of 6.71 related risk-factors for each of the 14 familiar risk proxies. Consider the *Cust. Dummy* proxy in column 1 of Panel A of Table V, which is significantly related to the sentence-counts of eight risk-factors, namely *Supply Chain*, *Customer Concentration*, *Reporting Compliance*, *Product Approval*, *Healthcare Spending*, *Cost Volatility*, *Real Estate*, and *Regulatory Change* risk. In the holdout sample, seven of the eight *p*-values are below 0.05, the exception being the *Reporting Compliance* risk-factor with a *p*-value of 0.13.

The sets of risk-factors identified by backward-selection include factors that intuitively link to the risk-proxy; e.g., the *Customer Concentration* risk-factor is selected as one explanator of the *Cust. Dummy* risk-proxy. *Product Innovation* and *Product Approval* risk are both selected as risk-factors associated with competitive pressures arising from product innovation risk as measured by *Fluidity*. The *Financing III* risk-factor is linked to both the *Book D/A* risk-proxy and the *Credit Spread* risk-proxy. Both the *Intellectual Property* and *Intangible Assets* risk-factors are linked to the risk-proxy measure (*TA – Net PP&E*)*/TA* of intangible assets risk. The *Intellectual Property* and *Intangible Assets* risk-factors plus the *Product Innovation* and *Product Approval* risk-factors all link to the *R&D Exp/TA* proxy. The *Corporate Governance* risk-factor is related to the *Indep. Directors* risk-proxy. Not surprisingly, the *Revenue Volatility* risk-factor is an explanator of σ(*Revenues*) and the *Stock Price Volatility* risk-factor is an explanator of the σ_E risk-proxy. The only LDA-identified risk-factor that is not significantly related to any of the 14 risk-proxies that we examine is *Product Defects* risk. This does not imply that *Product Defects* are an unimportant risk and simply reflects the fact that none of the risk-proxies we investigate are natural metrics of product defects. In fact, 21 of the 30 risk-factors have a smaller median sentence-count than that of *Products Defects* risk.

The conclusion that the qualitative information in Item 1A is significantly related to each of the 14 risk-proxies examined also applies for the sentence-count as a fraction of the firm's total Item 1A sentence-count measure examined in Table V Panel B and for the sentence-count as a fraction of the average sentence-count of the firm's non-zero risk-factors examined in Table V Panel C. For each of the 14 risk-proxies, the set of risk-factors chosen through backward-selection largely overlap with the set of salient risk-factors when the normalized sentence-count measures are investigated. For both normalized sentence-count measures, the *p*-values associated with a null of no relation between the risk-proxy and the backward-selected risk-factors are not more than 0.0001 for each of the 14 risk-proxies we examine.[11]

B. The Incremental Information Content of Item 1A

In this section, we examine whether Item 1A provides additional information about firm risk exposures over and above what may be provided by knowledge of a firm's industry classification alone. The component of the risk-proxy that can be explained by industry effects is controlled for by using the deviation of a firm-level risk-proxy from the industry mean risk-proxy as the dependent variable in the regression analysis.

Using 2-digit SIC codes to determine industry membership, we use a training sample and backward-selection of risk-factors to identify the set of salient risk-factors as explanators of risk-proxy deviations from industry means. Panel A of Table VI reports the results of regressing firm-level deviations from industry-average risk-proxy values on the sentence-counts of the related set of salient risk-factors selected in the training sample. The significance of the *p*-value of the adjusted R^2 of this holdout-sample regression is a measure of whether Item 1A has incremental explanatory power conditional on knowledge of a firm's industry membership. The average of the adjusted R^2 values across the relations between the 14 risk-proxies and their salient risk-factors after controlling for industry fixed effects is 7.44%.

We can view the adjusted R^2 values in Table V as an upper bound on the portion of the variation in a risk-proxy that is related to variation in the salient risk-factors' sentence-counts. The adjusted R^2 values in Panel A of Table VI represent a lower bound after removing the portion of the variation in the risk-proxy explained by industry effects. The difference in adjusted R^2 values between Tables V and VI reflects the potential overlap of the explanatory power of industry effects and salient risk- factors. At its extreme, suppose all firms in an industry had identical sentence-

[11] Easton (1998) observes that the R^2 of a regression of share prices on either per-share accounting earnings or per-share book value measures can overestimate the value relevance of accounting disclosures because the estimated relation will reflect the scaling of the variables. Brown, Lo and Lys (1999) expand this observation on the effect of scaling and observe that if the coefficient of variation of the scale factor increases (decreases) through time, then the R^2 of the relation between per share market values and per share earnings and/or book values will increase (decrease) through time. The 14 risk-proxy measures that we examine do not involve issues of scaling. Sentence-counts are, for example, unaffected by stock splits.

counts for each risk-factor and identical values for each risk-proxy. While differences in risk-proxy values would be perfectly explained by differences in risk-factor sentence-counts, the result could be characterized as simply an industry fixed effect. Removing the potential overlap produces the adjusted R^2 values of Table VI as a conservative lower bound on the information content of Item 1A.

Compared with Table V, the adjusted R^2 values in Table VI are lower by an average of 55%. While industry membership can explain up to 55% of the variation in the risk proxies, variation in the reported risk-factors is also an important information source in that risk-factor exposure explains the remaining 45%. After removing what might be industry effects, the *p*-values associated with the null of no relation between each one of the 14 the risk-proxies and the backward-selected risk-factors are less than 0.004. Panel B of Table VI repeats the analysis using the full sample and all 30 risk-factors. Again, we conclude that after removing industry effects, the null of no relation between risk-proxies and risk-factor sentence-counts can be rejected.

In brief, interpreting textual information necessarily involves error. Risk-factor sentence-counts are then noisy measures of the information contained in the Risk Factors sections of financial statements and errors-in-variables will manifest as a downward bias in the R^2 of regressions of risk-proxies on risk-factor sentence-counts. Tables V and VI report that despite this bias and even after controlling for the explanatory effect of industry membership, the R^2 values of the estimated relations between proxies and factors are significantly greater than zero even when the relation is constrained to be linear.; i.e., qualitative risk disclosures are informative.

IV. Predictability of Changes in Risk Proxies

Having demonstrated the significant contemporaneous relation between risk-proxy levels and risk-factor sentence-count levels, we now investigate the relation between changes in proxies and factors. If risk-proxies are Markovian and risk-factor sentence-counts are noisy measures of risk, then changes in risk-factors will be noisy measures of contemporaneous changes in risk-proxies. If risk-factor sentence-counts are forward-looking estimates of future risk, then changes in risk-factors will be noisy predictors of future risk-changes. A number of our risk-proxy measures are estimated using data over a period of more than one year; e.g., the risk-proxy σ(NCF) is measured as the annualized standard deviation of the quarterly cash flows relative to quarter-end total assets estimated over a three-year period. A positive relation between the annual changes in such a risk-proxy measure and past lagged changes in risk-factor sentence-counts can arise even when Item 1A is not forward-looking. If risk follows a martingale and a risk-proxy is estimated using data from the current and past years, then the expectation of the estimated value of the risk-proxy next year will be greater than the current year's estimate of the risk-proxy whenever realized risk during the current year is higher than its value estimated using data from the current and past years. Thus, we do not investigate the predictability of changes in the risk-proxies σ(NCF), σ(Cost), and σ(Revenue) since each is an annualized realized standard deviation of quarterly values observed over the three years preceding the fiscal year-end.

In examining the relation between changes in risk proxies and changes in risk-factor sentence-counts, we use the same approach as that of the Section III analysis of their contemporaneous relation. We use lagged changes in those risk-factors identified as salient in the levels analysis as the explanators of changes in risk-proxies. The adjusted R^2 values of the regressions of changes in the risk-proxies on lagged changes in the sentence-counts of risk-factors is determined in the holdout sample.

The contemporaneous relation between changes in the salient risk-factors and changes in risk-proxies is reported in Panel A of Table VII. Although the contemporaneous relation is significant at the 1% (5%) level for seven (nine) of the 14 risk-proxies, the relation explains only a trivial faction of the variability of the proxies. For only three of the 14 risk-proxies does the adjusted R^2 exceed 0.01 with the maximal adjusted R^2 being 0.057.

The predictive relation between lagged changes in the salient risk-factors and changes in risk-proxies is examined in Panel B. A would-be trader will require that the set of salient risk-factor changes be determined using only past observations. An economist interested in investigating predictability can use observations over the entire sample period in constructing the training sample used to determine the set of salient factors and this is the approach we take. As explained, we do not analyse the predictability of σ(NCF), σ(Cost), and σ(Revenue). For the 11 risk-proxies investigated in Panel B, three (four) are significant at the 1% (5%) level. Lagged changes in salient risk-factor sentence-counts are though unable to explain a non-trivial amount of the variability of subsequent changes in risk-proxies with the maximal adjusted R^2

being 0.005. The low R^2 values do not though imply that changes in risk-factor sentence-counts lack predictive ability for all purposes. In fact, Section VII will show that changes in risk-factor sentence-counts can be used to predict several measures of firm performance in subsequent quarters.

V. Has the Informativeness of Item 1A Changed Through Time?

Researchers have examined whether risk disclosures have become less informative over time.[12] Dyer, Lang and Stice-Lawrence (2017) consider whether the readability of qualitative information has declined over time and attribute the increase in the length of 10-K statements over time to, in part, new disclosure requirements such as Item 1A risk disclosure. They document that 10-K statements have increased in their boilerplate nature, stickiness, and redundancy. However, the paper's conclusion notes that its tests do not measure informativeness. Miller (2017) observes that even if annual reports have become longer, less readable, and more boilerplate, this does not mean that they have become less informative.[13] As explained before, we use the terms accuracy and informativeness interchangeably.

In this section, we test whether the Item 1A Risk Factors section has become less accurate over time by examining the relation between risk-proxies and risk-factor sentence-counts on a yearly basis. The year t sentence-count of risk-factor j for firm i, S_{it}^{j}, can be thought of as a noisy measure of the fundamental information in firm i's disclosure concerning risk-factor j, f_i^j. Assume for the moment that the risk-factor themselves are constant through time.[14] Suppose risk-factor reporting has become increasingly boilerplate and/or redundant through time and, as a result, the annual sentence-count is time-dependent and linked to fundamental information as

$$S_{it}^{j} = \alpha(t) + \beta(t) f_{it}^{j} + \beta(t) \varepsilon_{it}^{j},$$

[12] Francis and Schipper (1999) examine whether the ability of earnings to explain annual market-adjusted returns has altered through time. Beaver, McNichols and Rhie (2005) conclude that the ability of financial statement data to predict bankruptcy is little changed over the 40 years from 1962 to 2002.

[13] We find that the number of sentences in Item 1A in our sample firms has increased from an average of 170 in 2005 to 288 in 2015, a near 70% increase.

[14] Hanley and Hoberg (2019) investigate a setting where risk changes through time and show that LDA analysis can be used to predict heightened risk exposures in the financial sector well in advance of the 2008 financial crisis.

where t denotes time and $\beta(t)\varepsilon_{it}^{j}$ is the noise in the sentence-count as a measure of fundamental information with this noise being independent across firms. The coefficient $\beta(t)$ reflects the time-varying level of repetition of the same information across multiple sentences, while $\alpha(t)$ reflects the time-varying use of boilerplate language.

The year t value of fundamental risk k for firm i is f_{it}^{k}. Let X_{it}^{k} denote the year t value of a single proxy for f_{it}^{k} with $X_{it}^{k} = f_{it}^{k} + \omega_{it}^{k}$. The measurement error ω_{it}^{k} is independent across firms and independent of the noise in the sentence-count. Let $\rho^{t}\left(S_{it}^{j}, X_{it}^{k}\right)$ denote the year t cross-firm correlation between risk-factor sentence-counts and values of the risk-proxy.

$$\rho^{t}\left(S_{it}^{j}, X_{it}^{k}\right) = \frac{\beta(t)\sigma^{2}\left(f_{it}^{j}\right)}{\beta(t)\sigma\left(f_{it}^{j} + \varepsilon_{it}^{j}\right)\sigma\left(f_{it}^{k} + \omega_{it}^{k}\right)} = \frac{\sigma^{2}\left(f_{it}^{j}\right)}{\sigma\left(f_{it}^{j} + \varepsilon_{it}^{j}\right)\sigma\left(f_{it}^{k} + \omega_{it}^{k}\right)}.$$

An increasing use of boilerplate language and increasing redundancy will not affect the expected R^2 value of the regression of the risk-proxy on the risk-factor sentence-count. If the variability of noise in the sentence-count measure and measurement error in the risk-proxy are not time-dependant, then the year t cross-sectional correlation between sentence-counts and risk-proxies is unaffected by time variation in either the repetition measure $\beta(t)$ or the boilerplate measure $\alpha(t)$. Thus, the R^2 measure of information content will not be expected to change through time.

Panel A of Table VIII reports the adjusted R^2 and ρ values of multivariate regressions of the 14 risk-proxies on the sentence-counts of salient risk-factors for each of the years 2005 to 2015.[15] One hundred and forty-one of the 151 annual risk-proxy, risk-factor relations are significant at the 1% level. Eight combinations are significant at the 5% in a given year. The remaining two annual combinations are significant at the 10% level. Averaged across the 14 risk-proxies, the adjusted R^2 values of the annual relations between risk-proxies and salient risk-factors are quite stable and range from 0.1664 in 2005 to 0.1922 in 2015. Panel B reports results when the factor sentence-count as a fraction of the total Item 1A sentence-count is used as the measure of a risk-factor. Averaged across the 14 risk-proxy, fractional sentence-count relations, the mean

[15] Values of the asset beta risk-proxy are unavailable during the years 2013 through 2015.

annual adjusted R^2 value ranges from 0.1994 in 2005 to 0.2041 in 2015. Panel C examines the annual relations when the sentence-count relative to the average sentence-count of the firm's non-zero risk-factors is used as the risk-factor measure. Averaged across the 14 risk-proxy, relative risk-factor relations, the average annual adjusted R^2 value is 0.1983 in 2005 and 0.2023 in 2015.

To determine the statistical significance of the relation between the time-series of R^2 values of the annual risk-proxy, risk-factor sentence-count relation and the passage of we use unreported regressions of the relation to investigate whether the informativeness of Item 1A has declined through time. For eight of the 14 risk-proxies, the estimated sensitivity of the adjusted R^2 value to the passage of time is *positive*. The relation is significant at the 1% level in three of these cases. Only one of the estimated sensitivities is significantly negative, with significance at the 5% level. The time-series of the average across the 14 risk-proxies is significantly increasing with the passage of time, with the relation is significant at the 1% level.

To conclude, the results from the analysis of Table VIII do not support the hypothesis that that annual estimates of the minimal information content of Item 1A given by the adjusted R^2 values of multivariate linear regressions of risk-proxies on salient risk-factors' sentence-counts declined between 2005 and 2015.

VI. LDA-Identified Indices of Operating Risk and Financing Risk

In this section, we demonstrate the utility of the sentence-counts of LDA-identified risk-factors by using them to construct meaningful indices of operating risk and financial risk. We show that these operating and financial risk indices are related to asset and equity volatility as predicted by theory, and that they can help predict firm performance.

A. Construction of the Operating Risk and Financing Risk Indices

Following the pioneering separation of a firm's leverage-related financial risk and its operating risk in Hamada (1972), we create firm-year indices of *Operating Risk* and *Financing Risk* from the sentence-counts of the LDA identified risk-factors. The *Financing Risk* index is constructed as the equal-weighted average of the sentence-counts of the *Financing I*, *Finance II*, and *Financing III* risk-factors. The *Stock Price Volatility* risk-factor reflects both operating and financing and is therefore not included in either of the indices. The *Operating Risk* index is constructed as the equal-weighted average of the sentence-counts of the remaining 26 risk-factors. The *Financing* (*Operating*) *Risk* index has an average sentence count of 4.05 (7.05) and a standard deviation of 4.41 (3.98).

B. Relation of Risk Indices to Equity and Asset Volatility

Table IX explores the relation between asset volatility and the risk indices and leverage. Asset volatility is measured as the average over the 12 months preceding the firm's fiscal year-end of the asset volatility values compiled by Choi and Richardson (2016) for the years 2005 through 2012. Year fixed effects are included in the regressions as a control for variation in the length of Item 1A through time. Column (1) reports that the positive univariate relation between asset volatility and operating risk is significant at the 1% level. The *Operating Risk* index and year fixed effects together explain 11.9% of the cross-firm variation in asset volatility. Column (2) examines the relation between asset volatility and the two risk indices. Asset volatility is significantly positively related to the *Operating Risk* index and significantly negatively related to the *Financing Risk* index. The relation is such that a one-standard deviation increase in the *Operating Risk* index (*Financing Risk* index) is associated with on average a seven percent point increase (a four percent point decrease) in asset volatility. This is consistent with a firm's financing decision being informative about its asset volatility and the *Operating Risk* index measure being a noisy measure of asset volatility. A negative relation between asset volatility and *Financing Risk* index accords with the documented negative relations between firms' leverage choices and asset volatility (Choi and Richardson, 2016) and between firms' leverage choices and the economic risks faced by a firm (Bartram, Brown and Waller, 2015).

Columns (3) and (4) show that asset volatility is significantly positively related to the *Operating Risk* index and significantly negatively related to both the *Financing Risk* index and *Market Leverage*. That the relations with *Financing Risk* and *Market Leverage* are both significant suggests that the *Financing Risk* index contains additional information about financial risk that is not captured by a simple leverage measure.

That the relations with *Financing Risk* and *Market Leverage* are both significant suggests that the *Financing Risk* index and the leverage measure capture different elements of the relation between firms' financing choices and asset volatility,

Table X examines the relation between equity volatility and the *Operating* and *Financing Risk* indices.[16] Choi and Richardson (2016) document that equity volatility is increasing in both asset volatility and leverage. Choi and Richardson (2016) avoid common measurement error in the explanatory and dependent variables by using option-implied volatility to estimate the volatility of the return on equity, σ_E, and an EGARCH estimation of the volatility of the return on a weighted-average of the firm's debt and equity to estimate asset volatility, σ_A. The relation investigated in Table X also avoids common measurement error in that measurement error in σ_E is independent of measurement error in the risk indices.

Columns (1) reports that *Operating Risk* and *Market Leverage* are significant positive explanators of equity volatility at the 1% level. Column (2) shows that *Financing Risk* is also a significant positive explanator of equity volatility once industry effects are controlled for. In Panel A of Table V, we showed that the most significant of the multivariate risk-factor explanators of σ_E is the *Stock Price Volatility* risk-factor. Therefore, in column (3) we add *Stock Price Volatility* as an explanator. Perhaps not surprisingly, the *Operating Risk* index is now no longer a significant determinant of σ_E; i.e., its explanatory power is subsumed by the *Stock Price Volatility* risk-factor. Interestingly, however, *Stock Price Volatility* does not subsume the explanatory power of either the *Financing Risk* index or *Market Leverage*, both of which remain significant at the 1% level. Together Tables IX and X show that LDA-identified *Operating Risk* and *Financing Risk* indices are useful predictors of the volatility of firms' asset and equity values.

C. Predictability of Firm Performance

Are changes in our constructed risk indices indicative of future firm operating performance? Cohen, Malloy and Nguyen (2020) use textual analysis of entire 10-K and 10-Q statements to show that changes in the similarity of a report with the prior year's report predict firm performance in the second quarter of the following year. Operating performance in the second

[16] The Table X analysis of equity volatility involves a larger number of observations than the Table IX analysis of asset volatility. The asset volatility values of Table IX are taken from Jaewon Choi's website which reports asset volatility estimates constructed by delevering equity volatility estimates at month-ends prior to 2013 for the set of firms for which the required loan data is available. Equity volatility can be estimated over the full 2005-2015 period and without regard to the availability of loan data.

quarter is investigated because Cohen, Malloy and Nguyen (2020) find that the similarity metrics are predictive of second-quarter stock returns.

We build upon their analysis and test whether annual changes in indices of risk-factor sentence-counts, i.e., changes in the *Operating Risk* and *Financing Risk* indices, predict firms' future quarterly operating performance. We examine all four quarters of the year following the change, because, although prices may well cease being 'lazy' after the public release of a 10-K typically by the end of the first-quarter, real effects underlying a price change may occur in quarters other than the quarter following a 10-K's release. Table XI reports the results.[17]

Panels A, B and C report the respective relations of quarterly *Operating Income*, quarterly *Net Income*, and quarterly *Sales* and past annual changes in the *Operating Risk* and *Financing Risk* indices. *Operating Income* is measured as quarterly operating income before depreciation divided by beginning-of-quarter total assets. *Net Income* and *Sales* are measured analogously. The adjusted R^2 values are the percent of the variation in firm performance that is explained by variation in the operating and financing risk index changes and by variation across years. The second quarter results are consistent with those of the Cohen, Malloy and Nguyen (2020) analysis, namely that the more dissimilar are past 10-K statements, the lower is firm performance in the second quarter of the subsequent year. In each panel of Table XI, larger changes in the *Operating Risk* and *Financing Risk* indices are associated with lower firm performance in the second quarter of the subsequent year. The negativity of the relation is not confined to the second quarter. In fact, the estimated relation is negative in all four quarters of the following year. For financing risk, we see that the relation between quarterly *Operating Income* and past annual changes in the *Financing Risk* index is significant at the 1% level in all four quarters, and the relation with *Net Income*

[17] Consider a document in consecutive years and assume the document differs by one sentence across the years. Similarity metrics will not distinguish between the addition of a sentence to and the subtraction of a sentence from the first year's document. In contrast, our sentence-count change measure is positive if the discussion of a risk-factor grows across the years and negative if the risk-factor discussion is shortened between years. Distinguishing between sentence-count changes and the absolute value of sentence-count changes is less important when the fraction of negative changes is lower and when negative changes are small relative to positive changes. 28.7% (25.6%) of changes in the *Operating Risk* (*Financing Risk*) index are negative, with the median index change being +0.5 for increases and –0.15 for decreases (+1 for increases and -0.66 for decreases). In unreported results we find our conclusions from an analysis of the relation investigated in Table XI between lagged changes risk indices and firm performance are unchanged if instead the lagged absolute changes in risk indices are considered as explanatory variables.

(*Sales*) is significant at the 1% level in all quarters except the fourth (first), the negative relation in that quarter being significant at the 5% level.

For operating risk, we see that the relation between quarterly *Operating Income* and past annual changes in the *Operating Risk* index is significant at the 1% level in the second, third and fourth quarters of the following year and insignificant in the first quarter. The relation between quarterly *Net Income* and past changes in the *Operating Risk* index is insignificant in the first quarter, and significant at the 10%, 5%, and 1% levels in the second, third, and fourth quarters respectively. The relation between quarterly *Sales* and past changes in the *Operating Risk* index is insignificant in the first quarter, significant at the 5% level in the second and third quarters, and significant at the 1% level in the fourth quarter. Thus, the relations between annual changes in the risk indices and quarterly firm performance over the following year can be summarized as being negative, becoming more negative and becoming significant as the year progresses.

Is distinguishing between changes in operating risks and changes in financing risks important when predicting future firm performance? If document similarity alone, rather than the similarity of specific components of a document, is important, then distinguishing between changes in operating risks and changes in financing risks will be unimportant when predicting firm performance. Consistent with the importance of considering risk types, the relation between future firm performance and a one-sentence increase in a financing risk and is more significant than that with a one-sentence increase in an operating risk. As an illustration, take the fourth quarter results in Panel A of Table XI as an exemplar. The estimated coefficient on the past annual change in the *Operating Risk* index is −0.0651. The estimated coefficient on changes in the *Financing Risk* index is −0.0827. A one-sentence increase in the sentence-count of any of the 26 (three) risks reflected in the equal-weighted *Operating Risk* (*Financing Risk*) index will increase the index by 1/26 (1/3). The predicted effect on *Operating Income* of a one-sentence increase in an operating risk will then be −0.0651/26; i.e., a decrease of 0.25 of a percent point. The predicted effect on *Operating Income* of a one-sentence increase in a financing risk is larger in absolute value and equal to −0.0827/3; i.e., to a 2.76 percent point decrease in operating income as a fraction of total assets. The reported *F*-statistics associated with equality of the relation of quarterly firm performance with a one-sentence increase in an operating risk-factor versus with a one-sentence increase in a financing risk-factor show that the relation differs significantly for all three firm performance measures and in all four quarters. One potential explanation of the difference in significance is that disclosed increases in operating risks may be associated with a management perception of higher future performance volatility. The perceived higher volatility may or may not be accompanied by a lower perceived level of performance. In contrast, increases in disclosed financing-related risks, such as increases in the likelihood of default, higher credit spreads, and rollover difficulties, may reflect a managerial expectation of lower future firm performance.

A consideration of the relation between firm performance and lagged changes in the individual LDA-identified risk-factors may allow a better understanding of the economics of the relation between document similarity and firm performance. The predictive power of specific risk-factor changes could be investigated using a training-sample to determine the set of risk-factor changes potentially salient for a specific performance measure, and a hold-out sample with which to investigate the relation between changes in particular risk-factors and future firm performance.

VII. Conclusions

We implement a machine learning-based quantitative assessment of the accuracy of the textual risk disclosures in Item 1A of firms' 10-K reports to identify risk topics whose importance is measured by the number of sentences associated with the topic. Using training and hold-out samples, we relate 14 familiar proxies for firm risk to a set of salient LDA-identified risk-factors and take the adjusted R^2 of those relations as a measure of the accuracy of Item 1A disclosure. The average of adjusted R^2 has not declined through time. The average was 16.6% in 2005 and 19.2% in 2015. This suggests that the accuracy of Item 1A disclosure has not declined between 2005 and 2015 despite the increasing boilerplate content, stickiness, and redundancy of Item 1A that the extant literature has documented.

LDA-identified risk-factor sentence-counts are approximately 30% higher for firms in competitive industries than for firms in concentrated industries, a result which is contrary to models of voluntary disclosure where firms favor less disclosure when they face more competition (e.g., Verrechia (1983)) which suggests that any preference for concealment is outweighed by the increased overt risk of operating in a competitive environment.

Sentence-counts of LDA-identified risk-factors can be used to construct indices of operating risk and financial risk that are significantly related to firms' asset and equity volatilities.

Quantitative measures of operational risk may be particularly useful for regulators in the financial industry. We further show that larger increases in either the operating or financing risk indices are associated with significantly lower operating income, net income, and sales performance in the second, third and fourth quarters of the following year. In every quarter of the year following a risk-index change, a one-sentence increase in the sentence-count associated with financing risk has a significantly more negative effect on subsequent performance than a one-sentence increase in operating risk.

Our approach will be particularly useful in measuring exposures to risks for which risk-proxies are not readily available. One way to determine a firm's exposure to, say, catastrophe risk would be to investigate its Item 1A similarity to that of firms that have previously experienced a natural disaster. After determining the set of risk-factors that have high sentence-counts for firms that have previously experienced a catastrophe, firms with similarly high sentence-counts for the same risk-factors could be classified as more exposed to catastrophe risk. A further application would be to investigate inter-firm differences in risk-factor sentence-counts as a measure of differences in the characteristics of a characteristics-based model of return predictability.

References

Angrist, Joshua D. and Jörn-Steffen Pischke, 2009, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, Princeton.

Ball, Christopher, Gerard Hoberg and Vojislav Maksimovic, 2015. Disclosure, business change and earnings quality, SSRN: http://dx.doi.org/10.2139/ssrn.2260371

Bao, Yang and Anindya Datta, 2014, Simultaneously discovering and quantifying risk types from textual risk disclosures, *Management Science* 60(6), 1371–1391.

Bartram, Shönke M., Gregory W. Brown and William Waller, 2015. How important is financial risk? *Journal of Financial and Quantitative Analysis* 50(4), 801–824.

Beaver, William H., Maureen F. McNichols and Jung-Wu Rhie, 2005, Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy, *Review of Accounting Studies* 10(1), 93–122.

Bellstam, Gustaf, Sanjai Bhagat, Sanjai and J. Anthony Cookson, 2020, A text-based analysis of corporate innovation, Management Science 67(7), 3985-4642.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan, 2003, Latent dirichlet allocation, *Journal of Machine Learning Research* 3, 993–1022.

Bodnaruk, Andriy, Tim Loughran and Bill McDonald, 2016, Using 10-K text to gauge financial constraints, *Journal of Financial and Quantitative Analysis* 50(4), 623−626.

Brown, Stephen J., 2012, Quantitative measures of operational risk: an application to funds management, *Accounting and Finance* 52, 1001−1011.

Brown, Stephen, William Goetzmann, Bing Liang and Christopher Schwarz, 2012, Trust and delegation, *Journal of Financial Economics* 103(2), 221−234.

Brown, Stephen, Kin Lo and Thomas Lys, 1999, Use of R^2 in accounting research: measuring changes in value relevance over the last four decades, *Journal of Accounting and Economics* 28(2), 83−115.

Brown, Stephen V. and Jennifer Wu Tucker, 2011, Large-sample evidence on firms' year-over-year MD&A modifications, *Journal of Accounting Research* 49(2), 309–346.

Brown, Stephen V., Xiaoli Tian and Jennifer Wu Tucker, 2018, The spillover effect of SEC comment letters on qualitative corporate disclosure: Evidence from the risk factor disclosure, *Contemporary Accounting Research* 35(2), 622–656.

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(1), 396–455.

Choi, Jaewon and Matthew Richardson, 2016, The volatility of a firm's assets and the leverage effect, *Journal of Financial Economics* 121(2), 254–277.

Cohen, Lauren, Dong Lou and Christopher J. Malloy, 2020, Casting conference calls, *Management Science* 66(11), 51015–5039.

Cohen, Lauren, Christopher Malloy and Quoc Nguyen, 2020, Lazy prices, *Journal of Finance* 75(3), 1371-1415.

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

Easton, Peter D., 1998. Discussion of revalued financial, tangible, and intangible assets: Association with share prices and non-market-based value estimates, *Journal of Accounting Research* 36 (Supplement), 235-247.

Filzen, Joshua J., 2015, The information content of risk factor disclosures in quarterly reports. *Accounting Horizons* 29(4), 887–916.

Florackis, Chris, Christodoulos Louca, Roni Michaely and Michael Weber, 2021, Cybersecurity risk, Chicago Booth Research Paper No. 20-47, Available at SSRN: https://ssrn.com/abstract=3725130

Fracassi, Cesare, Stefan Petry and Geoffrey Tate, 2016, Does rating analyst subjectivity affect corporate debt pricing? *Journal of Financial Economics* 120(3), 514-538.

Francis, Jennifer and Katherine Schipper, 1999, Have financial statements lost their relevance? *Journal of Accounting Research* 37(2), 319–352.

Goldsmith-Pinkham, Paul, Beverly Hirtle and David Lucca, 2016, Parsing the content of bank supervision, Federal Reserve Bank of New York Staff Report 770.

Goldstein, Itay, and Liyan Yang, 2017, Information disclosure in financial markets, *Annual Review of Financial Economics* 9, 101–125.

Griffiths, Thomas L. and Mark Steyvers, 2004, Finding scientific topics, *Proceedings of the National Academy of Sciences* 101, 5228–5235.

Hadlock, Charles J. and Joshua R. Pierce, 2010, New evidence on measuring financial constraints: Moving beyond the KZ index, *Review of Financial Studies* 23(5), 1909–1940.

Hamada, Robert S., 1972, The effect of the firm's capital structure on the systematic risk of common stocks, *Journal of Finance* 27(2), 435–452.

Hanley, Kathleen Weiss and Gerard Hoberg, 2019, Dynamic interpretation of emerging risks in the financial sector, *Review of Financial Studies* 32(12), 4543-4603.

Hassan, Tarek A., Stephan Hollander, Laurence van Lent and Ahmed Tahoun, 2019, Firm-level political risk: Measurement and effects, *Quarterly Journal of Economics* 134(4), 2135–2202.

Hassan, Tarek A., Stephan Hollander, Laurence van Lent, Markus Schwedeler and Ahmed Tahoun, 2021, Firm-level exposure to epidemic diseases: COVID-19, SARS, and H1N1, NBER Working Paper 26971.

Hoberg, Gerard, Gordon Phillips and Nagpurnanand Prabhala, 2014, Product market threats, payouts, and financial flexibility, *Journal of Finance* 69(1), 293-324.

Hoberg, Gerard and Vojislav Maksimovic, 2015, Redefining financial constraints: A text-based analysis, *Review of Financial Studies* 28 (5), 1312-1352

Hoberg, Gerard and Gordon Phillips, 2016, Text-based network industries and endogenous product differentiation, *Journal of Political Economy* 124(5), 1423-1465.

Hope, Ole-Kristian, Dangi Hu and Hai Lu, 2016, The benefits of specific risk-factor disclosures, *Review of Accounting Studies* 21(4), 1005–1045.

Huang, Allen H., Reuven Lehavy, Amy Y. Zang and Rong Zheng, 2018, Analyst information discovery and interpretation roles: A topic modeling approach, *Management Science* 64(6), 2833-2855.

Huang, Ke-Wei and Zhuolin Li, 2011, A multilabel text classification algorithm for labeling risk factors in SEC Form 10-K, *ACM Transactions on Management Information Systems* 2(3), 18.1-18.19.

Kaplan, Steven N. and Luigi Zingales, 1997, Do financing constraints explain why investment is correlated with cash flow? *Quarterly Journal of Economics* 112(1), 169–216.

Kravet, Todd and Volkan Muslu, 2013, Textual risk disclosures and investors' risk perceptions, *Review of Accounting Studies* 18(4), 1088–1122.

Knyazeva, Anzhela, Diana Knyazeva and Ronald W. Masulis, 2013, The supply of corporate directors and board independence, *Review of Financial Studies* 26(6), 1561-1605.

Law, Kelvin K.F. and Lilian F. Mills, 2015, Taxes and financial constraints: Evidence from linguistic cues, *Journal of Accounting Research* 53(4), 777-819.

Li, Feng, 2010, Textual analysis of corporate disclosures: A survey of the literature, *Journal of Accounting Literature* 29, 143–165.

Lopez-Lira, Alejandro, 2019, Risk factors that matter: Textual analysis of risk disclosures for the cross-section of returns, SSRN: http://dx.doi.org/10.2139/ssrn.3313663.

Lopez-Lira, Alejandro, 2021, Why do managers disclose risks accurately? Textual analysis, disclosures, and risk exposures. *Economics Letters*, *204*, 109896.

Loughran, Tim, and Bill McDonald, 2016, Textual Analysis in Accounting and Finance: A Survey, *Journal of Accounting Research* 54(4), 1187–1230.

Lowry, Michelle, Roni Michaely and Ekaterina Volkova, 2020, Information revealed through the regulatory process: Interactions between the SEC and companies ahead of their IPO, *Review of Financial Stu*dies 33(12), 5510-5554.

MacKay, Peter and Gordon M. Phillips, 2005, How does industry affect firm financial structure? *Review of Financial Studies* 18(4), 1433–1466.

Miller, Gregory S., 2017, Discussion of "The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation", *Journal of Accounting and Economics* 64(2-3), 247–252.

Mirakur, Yatin, 2011, Risk disclosure in SEC corporate filings, Wharton Research Scholars Journal, Paper 85.

Nelson, Karen K. and A. C. Pritchard, 2016, Carrot or stick? The shift from voluntary to mandatory disclosure of risk factors, *Journal of Empirical Legal Studies* 13(2), 266-297.

Sautner, Zacharias, Laurence van Lent, Grigory Vilkov and Ruishen Zhang, 2021, Firm-level climate change exposure, ECGI Finance Working Paper N° 686/2020.

Verrecchia, Robert E, 1983, Discretionary disclosure, *Journal of Accounting and Economics* 5, 179–194.

Whited, Toni and Guojun Wu, 2006, Financial constraints risk, *Review of Financial Studies* 19(2), 531–559.

Table I: Descriptive Statistics

Sample summary statistics of total assets and asset volatility (σ_A) in Panel A and the 14 risk-proxies in Panel B. Variable definitions are contained in Appendix Table A2. *Total Assets* are in million USD. *Book D/A* is the book value of long-term plus short-term debt over the book value of total assets. *Credit Spread* is the fiscal year-end percentage point difference between firms' bond yields and a benchmark Treasury yield as in Fracassi, Petry and Tate (2016). *R&D Exp/TA* is a proxy for research and development intensity. (*TA–Net PP&E*)/*TA* is a measure of Asset Intangibility calculated as 1 − Net property plant and equipment/total assets. The *Cust. Dummy* variable is equal to one if a single corporate customer is responsible for at least 10% of firm sales. *TNIC HHI* is a measure of industry concentration based on the text-based network industry classifications in Hoberg and Phillips (2016). *Fluidity* is the measure of the intensity of product market change due to Hoberg, Phillips and Prabhala (2014). *Investment Delay* is the delaycon variable from Hoberg and Maksimovic (2014), which measures the firms' risks of delaying their investments due to issues with liquidity. *Indep. Directors* is the fraction of independent directors. *Asset Beta* is the beta of monthly returns with respect to value-weighted market estimated over a fiscal year. σ_A is the 12-month average of the Choi-Richardson monthly asset volatilities during a fiscal year. σ(*Cost*) is the annualized standard deviation of quarterly costs of goods sold relative to quarter-end total assets over the 12 quarters preceding the fiscal year-end. σ(*NCF*) is the annualized standard deviation of quarterly cash flow relative to total assets over the 12 quarters preceding the fiscal year-end. σ(Revenue) is the annualized standard deviation of quarterly revenue relative to quarter-end total assets over the 12 quarters preceding the fiscal year-end. σ_E is the annualized stock return volatility estimated using daily stock returns over the 250 trading days preceding the fiscal year-end.

	N	Mean	p25	Median	p75	Std Dev	Min	Max
Panel A: Asset Values and Asset Volatility								
Total Assets	13,470	8,186.7	584.9	1,652.4	5,255.2	3,0391.6	10.012	797,769
σ_A	5,105	0.323	0.211	0.292	0.410	0.158	0.060	1.670
Panel B: Risk Proxies								
Cust. Dummy	10,451	0.499	0	0	1	0.500	0	1
Asset Beta	5,153	1.079	0.503	0.947	1.481	0.976	−16.708	13.653
TNIC HHI	13,375	0.285	0.100	0.183	0.385	0.261	0.017	1
Fluidity	13,281	6.085	3.821	5.479	7.638	3.141	0.416	22.371
Book D/A	13,470	0.204	0.036	0.192	0.320	0.171	0	0.852
Credit Spread	4,513	2.976	1.288	2.231	4.068	2.243	0.033	9.995
(TA−Net PP&E)/TA	13,464	0.743	0.634	0.822	0.916	0.226	0.017	1
R&D Exp/TA	13,470	0.031	0	0.001	0.041	0.058	0	0.887
Investment Delay	10,368	−0.023	−0.081	−0.023	0.035	0.082	−0.262	0.303
Independ. Directors	10,947	0.778	0.714	0.800	0.875	0.119	0.100	1
σ (Cost)	13,031	0.051	0.016	0.033	0.064	0.058	0	1.082
σ (NCF)	11,614	0.037	0.011	0.019	0.037	0.076	0	4.187
σ (Revenue)	13,035	0.073	0.030	0.051	0.091	0.071	0.002	1.340
σ_E	13,464	0.400	0.262	0.354	0.480	0.206	0.010	3.246

Table II: Risk-Factors Sentence Counts

Summary statistics of the firm-year sentence counts associated with each of the 30 LDA risk factors obtained via a *sent*-LDA analysis of the text corpora of Item 1A of the annual reports of S&P1500 firms for the years 2005 through 2015. A word cloud representation can be found in Appendix Figure A1.

	Mean	p25	Median	p75	St.Dev.	min	max
Catastrophe	3.54	1	2	5	4.08	0	34
Corporate Governance	3.27	0	0	6	4.97	0	64
Country-Related	4.59	0	3	7	5.90	0	93
Customer Concentration	9.33	3	7	12	11.55	0	272
Economic Conditions	6.62	2	5	9	6.15	0	62
Energy Sector	9.91	0	0	3	28.70	0	539
Financing I	1.68	0	0	2	4.12	0	81
Financing II	4.39	0	1	6	7.57	0	230
Financing III	6.29	1	4	9	7.11	0	66
Growth and Restructuring	8.50	3	7	12	7.79	0	80
Healthcare Spending	8.37	0	0	4	29.54	0	672
Human Capital	7.47	2	6	10	7.79	0	78
Incomplete Contracts	4.87	0	1	4	12.46	0	568
Information Systems	6.44	0	3	9	10.45	0	209
Intangible Assets	5.33	0	3	8	8.52	0	279
Intellectual Property	8.06	0	4	13	10.89	0	99
Product Approval	8.95	0	0	1	37.18	0	909
Product Defects	6.96	2	6	10	6.36	0	58
Product Innovation	15.26	2	7	18	23.09	0	574
Real Estate	1.60	0	0	1	7.25	0	115
Regulatory Change	8.55	3	6	12	8.55	0	182
Regulatory Compliance	4.65	0	3	6	6.64	0	127
Reporting Accuracy	4.29	2	3	6	3.97	0	71
Reporting Compliance	2.87	0	0	4	6.28	0	253
Supply Chain	12.17	2	7	16	15.64	0	203
Tax Uncertainty	4.05	0	0	6	8.61	0	169
Cost Volatility	3.56	0	0	3	11.19	0	872
Net Cash Flow Volatility	11.75	5	10	16	8.36	0	76
Revenue Volatility	14.02	5	11	20	12.13	0	92
Stock Price Volatility	4.77	0	2	7	6.51	0	60

Table III: Top 3 Risk-Factors Sentence Counts by Industry

Top 3 LDA risk-factors with the highest average firm-year sentence counts by SEC 2-digit industry. The LDA risk factors are obtained via a *sent*-LDA analysis of the text corpora of Item 1A of the annual reports of S&P1500 firms for the years 2005 through 2015. The column headed # Firms shows the number of firms per industry.

		Average Sentence-Count of Top 3 risks					
Industry (SIC 2)	# Firms	1st		2nd		3rd	
Agricultural Production – Crops	2	*Product approv.*	23.91	*Revenue vol.*	11.73	*Intellectual prop.*	9.82
Agricultural Production – Stock	1	*Energy sector*	20.70	*Product approv.*	11.50	*Corporate gov.*	10.80
Agricultural Services	1	*Revenue vol.*	22.27	*Growth and restruct.*	19.00	*Product approv.*	17.09
Amusement & Rec. Services	10	*Customer concent.*	21.47	*Incomplete cont.*	19.37	*Financing III*	14.57
Apparel & Accessory Stores	36	*Revenue vol.*	22.88	*Human capital*	14.98	*Customer concent.*	12.26
Apparel & Other Textile Products	24	*Revenue vol.*	22.51	*Customer concent.*	16.71	*NCF volatility*	16.17
Auto Repair, Services & Parking	5	*Revenue vol.*	16.13	*Customer concent.*	13.78	*Human capital*	9.40
Auto Dealers & Service Stations	11	*Revenue vol.*	20.46	*Customer concent.*	20.08	*NCF volatility*	13.67
Building Materials & Garden Supplies	5	*Revenue vol.*	18.93	*Info. sys.*	13.54	*Human capital*	13.07
Business Services	220	*Product innov.*	36.30	*Info. sys.*	19.26	*Revenue vol.*	16.32
Chemical & Allied Products	149	*Product approv.*	66.83	*Intellectual prop.*	15.74	*NCF volatility*	11.38
Coal Mining	6	*Energy sector*	137.91	*Cost volatility*	27.40	*Financing III*	23.32
Communications	48	*Product innov.*	49.80	*Incomplete cont.*	14.77	*Health. spend.*	14.41
Eating & Drinking Places	33	*Human capital*	29.57	*Revenue vol.*	25.52	*Regulatory compl.*	10.52
Educational Services	11	*Health. spend.*	192.56	*Human capital*	17.28	*Regulatory Δ*	15.90
Electric, Gas, & Sanitary Services	85	*Energy sector*	66.22	*Cost volatility*	21.60	*Financing III*	12.86
Electronic & Other Elect. Equipment	146	*Supply chain*	37.40	*Product innov.*	29.12	*Revenue vol.*	21.45
Engineering & Man. Services	30	*Health. spend.*	17.69	*Product innov.*	16.51	*Human capital*	14.13
Fabricated Metal Products	25	*Supply chain*	18.38	*Revenue vol.*	15.54	*NCF volatility*	11.74
Food & Kindred Products	51	*Revenue vol.*	16.30	*NCF volatility*	11.51	*Supply chain*	11.49
Food Stores	9	*Revenue vol.*	14.88	*Customer concent.*	9.46	*Human capital*	7.91

Table III continued

Industry (SIC 2)	# Firms	Average Sentence-Count of Top 3 risks					
		1st		2nd		3rd	
Furniture & Fixtures	12	*Revenue vol.*	15.62	*Supply chain*	13.59	*Customer concent.*	11.65
Furniture & Home Stores	8	*Revenue vol.*	16.20	*Product innov.*	14.19	*Human capital*	10.64
General Building Contractors	15	*Real estate*	60.12	*Financing III*	14.03	*Revenue vol.*	12.37
General Merchandise Stores	18	*Revenue vol.*	17.64	*Human capital*	10.30	*Info. sys.*	10.09
Health Services	40	*Health. spend.*	93.48	*Regulatory Δ*	16.02	*Product defects*	12.98
Heavy Construction excl. Building	9	*Supply chain*	24.93	*Intangible assets*	24.89	*NCF volatility*	19.34
Hotels & Other Lodging Places	5	*Real estate*	41.63	*Tax uncertainty*	18.78	*Regulatory Δ*	15.93
Industrial Mach. & Equipment	115	*Product innov.*	22.29	*Supply chain*	21.15	*Revenue vol.*	16.95
Instruments & Related Products	130	*Product approv.*	29.79	*Product innov.*	18.20	*Supply chain*	17.52
Leather & Leather Products	10	*Revenue vol.*	25.84	*Supply chain*	17.13	*Customer concent.*	13.14
Local & Interurban Passenger Transit	1	*Health. spend.*	121.27	*Regulatory Δ*	21.18	*Human capital*	17.91
Lumber & Wood Products	8	*Revenue vol.*	13.96	*Tax uncertainty*	10.38	*Supply chain*	6.92
Metal, Mining	2	*Energy sector*	75.08	*Customer concent.*	29.92	*Cost volatility*	25.46
Miscellaneous Manufacturing Ind.	17	*Product innov.*	22.80	*Revenue vol.*	19.30	*Customer concent.*	14.65
Miscellaneous Retail	35	*Revenue vol.*	21.76	*Info. sys.*	12.51	*Human capital*	11.61
Motion Pictures	4	*Product innov.*	64.43	*Revenue vol.*	15.29	*Customer concent.*	13.17
Non-Classifiable Establishments	4	*Health. spend.*	39.23	*Regulatory Δ*	8.08	*Intangible assets*	7.36
Nonmetallic Minerals excl Fuels	6	*Energy sector*	55.61	*Revenue vol.*	19.64	*Regulatory compl.*	15.11
Oil & Gas Extraction	64	*Energy sector*	79.26	*Regulatory compl.*	11.90	*Product defects*	11.12
Paper & Allied Products	24	*Revenue vol.*	12.57	*Supply chain*	11.77	*NCF volatility*	9.08
Personal Services	6	*NCF volatility*	10.71	*Report. accuracy*	9.27	*Financing I*	8.71
Petroleum & Coal Products	9	*Energy sector*	41.35	*NCF volatility*	10.44	*Revenue vol.*	8.37
Primary Metal Industries	31	*Supply chain*	20.17	*Revenue vol.*	15.26	*NCF volatility*	14.36
Printing & Publishing	22	*Product innov.*	19.05	*Revenue vol.*	13.25	*Growth and restruct.*	8.60

Table III continued

Industry (SIC 2)	# Firms	Average Sentence-Count of Top 3 Risks					
		1st		2nd		3rd	
Railroad Transportation	5	*NCF volatility*	18.43	*Customer concent.*	18.36	*Energy sector*	17.15
Rubber & Misc. Plastics Products	14	*Supply chain*	13.79	*Revenue vol.*	13.77	*NCF volatility*	11.46
Social Services	2	*Health. spend.*	36.00	*Financing II*	34.90	*Corporate gov.*	27.30
Special Trade Contractors	4	*Intangible assets*	29.87	*Product defects*	22.40	Supply chain	21.50
Stone, Clay, & Glass Products	9	*Energy sector*	26.88	*Supply chain*	16.09	*Revenue vol.*	11.37
Textile Mill Products	5	*Customer concent.*	13.61	*Revenue vol.*	13.32	*Supply chain*	12.75
Tobacco Products	5	*Incomplete cont.*	47.47	*Product approv.*	21.10	*Revenue vol.*	16.60
Transportation Equipment	47	*Supply chain*	17.75	*Revenue vol.*	12.51	*Customer concent.*	11.34
Transportation Services	8	*Revenue vol.*	13.76	*Customer concent.*	13.75	*NCF volatility*	11.43
Transportation by Air	16	*Customer concent.*	47.69	*Revenue vol.*	20.54	*Catastrophe*	16.25
Trucking & Warehousing	15	*Revenue vol.*	15.59	*Energy sector*	12.16	*NCF volatility*	10.21
Water Transportation	9	*Energy sector*	45.77	*Regulatory compl.*	17.37	*NCF volatility*	13.49
Wholesale Trade – Durable Goods	42	*Revenue vol.*	17.61	*Supply chain*	11.73	*NCF volatility*	11.48
Wholesale Trade – Nondurable Goods	24	*Revenue vol.*	16.17	*NCF volatility*	12.44	*Customer concent.*	10.85
Instruments & Related Products	2	*Product approv.*	29.79	*Product innov.*	18.20	*Supply chain*	17.52

Table IV: Risk-Factors Sentence Counts by Competitiveness of the Product Market

Summary statistics of the firm-year sentence counts associated with each of the 30 LDA risk factors for high and low levels of competitiveness of the product market. Firms operating in competitive *(*concentrated) product markets have a TNIC3 HHI value below (above) the median of its distribution. The last column (ΔMeans) reports the differences in means of the sentence counts in competitive and concentrated product markets. The significance levels are based on two-tailed, two-sample t-tests of the difference in means. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Competitive (below-median TNIC3 HHI)				Concentrated (above median-TNIC3 HHI)				
	Mean	Median	St.Dev.	Max	Mean	Median	St.Dev.	Max	ΔMeans
Catastrophe	4.15	3	4.47	34	2.91	2	3.54	33	1.24***
Corp. Governance	3.72	1	5.30	64	2.79	0	4.52	36	0.92***
Country-Related	4.30	2	5.80	93	4.86	3	5.94	52	–0.57***
Customer Concent.	9.11	6	12.03	272	9.50	7	10.94	217	–0.40**
Economic Conditions	6.13	5	5.91	50	7.08	6	6.28	62	–0.95***
Energy Sector	16.08	0	37.59	539	3.78	0	12.96	401	12.30***
Financing I	2.08	0	4.75	81	1.28	0	3.33	67	0.79***
Financing II	4.92	2	8.32	230	3.87	1	6.75	166	1.05***
Financing III	7.22	5	7.79	65	5.35	4	6.23	66	1.87***
Growth and Restruct.	8.57	7	8.08	74	8.40	7	7.48	80	0.17
Healthcare Spending	12.56	1	38.32	604	4.19	0	15.82	672	8.37***
Human Capital	8.30	6	8.31	78	6.62	5	7.09	66	1.68***
Incomplete Contracts	5.75	1	13.98	568	3.92	0	10.46	215	1.83***
Information Systems	6.36	3	9.66	114	6.52	3	11.19	209	–0.17
Intangible Assets	5.80	3	9.72	279	4.81	2	6.75	217	0.99***
Intellectual Property	9.37	5	12.29	99	6.73	3	9.04	65	2.64***
Product Approval	14.47	0	50.83	909	3.52	0	11.78	212	10.95***
Product Defects	7.61	6	6.75	58	6.32	5	5.87	56	1.29***
Product Innovation	15.76	7	24.79	574	14.70	8	21.06	318	1.07***
Real Estate	2.19	0	9.32	113	1.01	0	4.25	115	1.18***
Regulatory Change	9.68	8	8.55	121	7.41	5	8.40	182	2.26***
Regulatory Compl.	5.26	3	7.73	127	4.04	2	5.29	44	1.22***
Reporting Accuracy	4.48	3	4.23	53	4.11	3	3.66	71	0.37***
Reporting Compliance	3.30	1	7.38	253	2.40	0	4.81	126	0.90***
Supply Chain	12.98	6	17.40	125	11.31	8	13.53	203	1.67***
Tax Uncertainty	4.32	0	8.38	150	3.71	0	8.50	169	0.61***
Cost Volatility	4.16	0	10.49	144	2.96	0	11.87	872	1.20***
NCF Volatility	12.32	11	8.61	76	11.17	10	8.07	75	1.15***
Revenue Volatility	13.86	10	13.29	92	14.16	12	10.81	92	–0.30*
Stock Price Volatility	5.51	3	6.97	60	4.00	1	5.87	47	1.51***
Column average	7.68	3.62	12.57	203	5.78	3.23	8.40	173.53	

Table V: Analysis of Relation Between Risk Proxies and Selected LDA Risk Factors

The relation between risk proxies (dependent variable) and the set of LDA risk factors (independent variables) identified via backward elimination. The stepwise, backward estimation is carried out on a random sample that contains 40% of the observations in every year. The results shown in this table are based on the remaining 60% of the sample observations. The reported coefficients in Panel A are ordered by size. Robust t-statistics are in parentheses. The standard errors are clustered at the 2-digit SIC industry level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Risk-factor disclosure quantified by the risk-factor sentence-count

Risk Proxy	Cust. Dummy	Asset Beta	–TNIC HHI	Fluidity	Book D/A	Credit Spread	(TA–Net PP&E)/TA
	Supply Chain	*Supply Chain*	*Catastrophe*	*Stock Price Vol.*	*Financing III*	*Financing III*	*Intellectual Prop.*
	0.0068***	0.0142***	0.0062***	0.0469***	0.0077***	0.0969***	0.0052***
	(8.02)	(5.58)	(4.50)	(4.83)	(13.22)	(5.06)	(6.50)
	Customer Concent.	*Info. Sys.*	*Human Capital*	*Energy Sector*	*Financing II*	*Corp. Govern.*	*Growth and Restruct.*
	0.0029**	0.0081***	0.0035***	0.0370***	0.0044***	0.0805***	0.0040***
	(2.26)	(2.82)	(3.01)	(11.65)	(4.75)	(3.48)	(3.35)
	Reporting Compl.	*Real Estate*	*Stock Price Vol.*	*Regulatory Δ*	*Real Estate*	*Human Capital*	*Intangible Assets*
	0.0022	0.0050**	0.0033**	0.0350**	0.0016***	0.0388***	0.0021**
	(1.51)	(2.29)	(2.51)	(2.15)	(3.33)	(2.74)	(2.59)
	Product Approv.	*Report. Accuracy*	*Financing I*	*Financing I*	*Supply Chain*	*Product Approv.*	*Info. Sys.*
	0.0021***	–0.0097*	0.0031***	0.0341***	–0.0017***	–0.0095***	0.0011**
	(5.12)	(–1.85)	(3.45)	(3.15)	(–5.06)	(–5.03)	(2.29)
	Health. Spend.	*Financing I*	*Energy Sect.*	*Product Approv.*	*Human Capital*	*Cost Vol.*	*Energy Sector*
	–0.0017***	–0.0128*	0.0019***	0.0313***	–0.0036***	–0.0178***	–0.0029***
	(–2.79)	(–1.97)	(7.31)	(24.71)	(–3.14)	(–3.08)	(–6.49)
	Cost Vol.	*Regulatory Δ*	*Health Spend.*	*Health Spend.*	*Stock Price Vol.*	*Info. Sys.*	*Regulatory Compl.*
	–0.0030***	–0.0163***	0.0010***	0.0277***	–0.0037***	–0.0273***	–0.0042**
	(–4.97)	(–4.99)	(3.04)	(4.30)	(–5.58)	(–3.19)	(–2.38)
	Real Estate		*Product Approv.*	*Product Innov.*		*Catastrophe*	*Catastrophe*
	–0.0049**		0.0008***	0.0264***		–0.0354	–0.0101***
	(–2.41)		(5.65)	(6.70)		(–1.36)	(–6.17)

Table V continued

Risk Proxy	Cust. Dummy	Asset Beta	−TNIC HHI	Fluidity	Book D/A	Credit Spread	(TA−Net PP&E)/TA
	Regulatory Δ		*Country-Related*	*NCF Volatility*			
	−0.0050**		−0.0025**	−0.0242**			
	(−2.50)		(−2.16)	(−2.51)			
				Information Systems			
				−0.0350***			
				(−3.71)			
				Economic Conditions			
				−0.0473***			
				(−2.94)			
	Constant	Constant	Constant	Constant	Constant	Constant	Constant
	0.4244***	1.0502***	−0.3809***	5.0487***	0.2001***	2.3105***	0.7331***
	(7.72)	(11.17)	(−15.72)	(17.32)	(15.04)	(16.31)	(25.89)
Observations	6,261	3,079	8,035	7,963	8,090	2,728	8,087
Adjusted R^2	0.119	0.068	0.114	0.375	0.296	0.188	0.442
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table V continued

Panel A: Risk-factor disclosure quantified by the risk-factor sentence-count							
Risk Proxy	R&D Exp/TA	Invest. Delay	Indep. Directors	σ(Cost)	σ(NCF)	σ(Revenue)	σ_E
	Intellectual Prop.	*Financing III*	*Regulatory Δ*	*Revenue Vol.*	*Stock Price Vol.*	*Revenue Vol.*	*Stock Price Vol.*
	0.0018***	0.0011***	0.0010***	0.0007***	0.0012***	0.0012***	0.0066***
	(10.01)	(4.42)	(2.77)	(3.26)	(4.56)	(3.78)	(5.95)
	Supply Chain	*Corporate Gov.*	*Product Approv.*	*Real Estate*	*Intellectual Prop.*	*Product Innov.*	*Financing III*
	0.0007***	0.0011**	0.0002**	0.0005***	0.0007***	−0.0004***	0.0047***
	(4.38)	(2.44)	(2.54)	(3.99)	(4.67)	(−4.12)	(3.43)
	Stock Price Vol.	*Real Estate*	*Cost Volatility*	*Intel. Prop.*	*Financing III*	*Country-Rel.*	*Revenue Volatility*
	0.0006**	0.0006***	0.0007**	−0.0011***	0.0005***	−0.0012***	0.0021***
	(2.55)	(2.78)	(2.50)	(−3.37)	(3.47)	(−3.02)	(3.45)
	Product Innov.	*Product Approv.*	*NCF Vol.*		*Country-Related*		*Real Estate*
	0.0005***	0.0003***	0.0012***		−0.0005*		0.0010**
	(3.07)	(7.97)	(4.14)		(−1.83)		(2.28)
	Product Approv.	*Energy Sector*	*Corporate Gov.*		*Regulatory Δ*		*Country-Rel.*
	0.0004***	0.0003***	−0.0057***		−0.0006***		−0.0011**
	(7.08)	(5.51)	(−7.27)		(−3.39)		(−2.04)
	Human Capital	*Product Innov.*	*Tax Uncertainty*				*Cost Volatility*
	−0.0005***	0.0002**	0.0013***				−0.0012***
	(−3.99)	(2.30)	(3.65)				(−6.41)
	Customer Concent.		*Incomplete Cont.*				*Growth and Restruct.*
	−0.0006***		−0.0005***				−0.0023***
	(−3.93)		(−2.93)				(−2.81)
	Financing III						*Regulatory Δ*
	−0.0007***						−0.0029***
	(−3.04)						(−5.09)
	Regulatory Δ						
	−0.0010***						
	(−4.79)						

Table V continued

Risk Proxy	R&D Exp/TA	Invest. Delay	Indep. Directors	σ(Cost)	σ(NCF)	σ(Revenue)	σ_E
	Constant	Constant	Constant	Constant	Constant	Constant	Constant
	0.0174***	–0.0433***	0.7668***	0.0487***	0.0310***	0.0669***	0.3600***
	(4.00)	(–10.91)	(92.18)	(14.37)	(16.45)	(17.51)	(25.90)
Observations	8,090	6,244	6,547	7,837	6,980	7,839	8,087
Adjusted R^2	0.437	0.051	0.072	0.049	0.022	0.045	0.102
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0003	0.0000

Panel B: Risk-factor disclosure quantified by the risk-factor sentence-count as a fraction of the total Item 1A sentence-count

	Cust. Dummy	Asset Beta	–TNIC HHI	Fluidity	Book D/A	Credit Spread	(TA–Net PP&E)/TA
Adjusted R^2	0.099	0.091	0.129	0.446	0.314	0.150	0.544
p-value	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

	R&D Exp/TA	Investment Delay	Indep. Directors	σ(Cost)	σ(NCF)	σ(Revenue)	σ_E
Adjusted R^2	0.439	0.024	0.100	0.069	0.023	0.060	0.141
p-value	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000

Panel C: Risk-factor disclosure quantified by the risk-factor sentence-count as a fraction of the average sentence-count for the firm's non-zero risk-factors

	Cust. Dummy	Asset Beta	–TNIC HHI	Fluidity	Book D/A	Credit Spread	(TA–Net PP&E)/TA
Adjusted R^2	0.120	0.085	0.137	0.453	0.327	0.183	0.533
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

	R&D Exp/TA	Investment Delay	Indep. Directors	σ(Cost)	σ(NCF)	σ(Revenue)	σ_E
Adjusted R^2	0.451	0.042	0.106	0.073	0.026	0.064	0.082
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table VI: The Incremental (Net of Industry) Information Content of Item 1A

This table estimates the incremental information content of the LDA risk-factors over and above the information implicit in knowledge of a firm's industry. Industry membership is determined at the 2-digit SIC level. The stepwise, backward estimation to identify salient the risk-factors investigated in Panel A is carried out on a random sample that contains 40% of the observations in every year to identify the salient risk-factors for each risk-proxy. The Adjusted R^2 and p-values shown in Panel A are based on using the remaining 60% of the sample observations to regress deviations from industry-average risk-proxy values on the sentence-counts of the related salient risk-factors. In Panel B, all 30 risk factors are included as independent variables in the analysis of risk-factor deviations from industry means in the full sample. Standard errors are clustered at the 2-digit SIC industry level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Risk-factor disclosure quantified by the risk-factor sentence-count

Panel A: Analysis of holdout sample using the salient risk factors identified via backward selection regressions as independent variables

	Cust. Dummy	Asset Beta	–TNIC HHI	Fluidity	Book D/A	Credit Spread	(TA–Net PP&E)/TA
Adjusted R^2	0.048	0.011	0.050	0.166	0.176	0.140	0.056
p-value	0.0000	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000

	R&D Exp/TA	Investment Delay	Indep. Directors	σ(Cost)	σ(NCF)	σ(Revenue)	σ_E
Adjusted R^2	0.195	0.034	0.041	0.005	0.018	0.006	0.095
p-value	0.0000	0.0000	0.0000	0.0020	0.0000	0.0039	0.0000

Panel B: Analysis of full sample using all 30 risk factors as independent variables

	Cust. Dummy	Asset Beta	–TNIC HHI	Fluidity	Book D/A	Credit Spread	(TA–Net PP&E)/TA
Adjusted R^2	0.069	0.033	0.074	0.194	0.208	0.175	0.066
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

	R&D Exp/TA	Investment Delay	Indep. Directors	σ(Cost)	σ(NCF)	σ(Revenue)	σ_E
Adjusted R^2	0.224	0.044	0.064	0.022	0.031	0.029	0.106
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table VII: Contemporaneous and Lagged Changes in Risk-Factors

Panel A reports the results of panel regressions of changes (Δ) in risk-proxies on contemporaneous changes in the set of salient LDA risk factors identified via backward elimination as per Table V. Panel B reports the results of panel regressions of changes (Δ) in risk-proxies on lagged changes in the set of salient LDA risk factors identified via backward elimination as per Table V. The results shown in the table are based on the 60% of the sample observations not used in the stepwise, backward determination of the salient risk-factors. Regression constants and the independent variables are not reported.

Panel A: Contemporaneous Relation Between Changes in Risk-Proxies and Changes in Risk-Factors

Dependent variable	Observations	adj. R^2	Model p-value
Δ Cust. dummy	5,249	0.000	0.122
Δ Asset Beta	2,273	–0.001	0.291
Δ (−TNIC HHI)	6,861	0.000	0.072
Δ Fluidity	6,766	0.006	0.000
Δ Leverage	6,924	0.057	0.000
Δ Credit spread	2,236	0.033	0.000
Δ (TA-Net PP&E)/TA	6,922	0.001	0.058
Δ R&D Exp/TA	6,924	0.004	0.000
Δ Investment Delay	5,062	0.000	0.004
Δ Independent Directors	5,421	0.000	0.019
Δ σ(Cost)	6,745	0.000	0.029
Δ σ(NCF)	5,401	0.000	0.001
Δ σ(Revenue)	6,746	0.000	0.321
Δ σ_E	6,915	0.044	0.000

Table VII continued

Panel B: Relation Between Changes in Risk-Proxies and Lagged Changes in Risk-Factors

Dependent variable	Observations	adj. R^2	Model p-value
Δ Cust. dummy	5,078	–0.001	0.030
Δ Asset Beta	1,881	–0.002	0.515
Δ (−TNIC HHI)	6,620	0.000	0.391
Δ Fluidity	6,493	0.001	0.638
Δ Leverage	6,622	0.000	0.131
Δ Credit spread	2,213	0.005	0.000
Δ (TA-Net PP&E)/TA	6,631	0.000	0.001
Δ R&D Exp/TA	6,633	0.001	0.000
Δ Investment Delay	4,442	0.001	0.000
Δ Independent Directors	5,480	0.000	0.013
Δ σ_E	6,686	0.003	0.003

Table VIII: Annual Cross-Sectional Regressions of Risk Proxies on Selected Risk-Factor Sentence-Counts

Each pair of values in Panel A shows the adjusted R^2 and the p-value of the F-test of the annual regression of the associated risk proxy on the sentence-counts of the salient risk-factors identified through backward elimination. Salient risk-factors are determined using a training sample of 40% of each year's observations. Standard errors are clustered at the 2-digit SIC industry level. The adjusted R^2 values are calculated using the remaining 60% of the sample observations being those that were not used in the stepwise, backward determination of the salient risk-factors.

Panel A: Risk-factor disclosure quantified by the risk-factor sentence-count

Risk Proxy	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Cust. Dummy	0.1109	0.1099	0.1172	0.1203	0.1079	0.1570	0.1160	0.1071	0.1327	0.1326	0.1279
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Asset Beta	0.1022	0.0974	0.0238	0.0646	0.1036	0.1379	0.0707	0.0555			
	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000			
−TNIC HHI	0.1004	0.1091	0.0851	0.0892	0.0940	0.1384	0.1467	0.1376	0.1272	0.1409	0.1458
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Fluidity	0.3963	0.4847	0.4582	0.3498	0.4856	0.4657	0.4548	0.4730	0.4596	0.3919	0.3748
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Book D/A	0.2850	0.2202	0.2723	0.3304	0.3535	0.3088	0.2868	0.3316	0.2917	0.2947	0.2999
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Credit Spread	0.1118	0.1875	0.2201	0.1779	0.2450	0.2053	0.3004	0.2531	0.1847	0.2692	0.2012
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
(TA−Net PP&E)/TA	0.3855	0.3815	0.3988	0.4002	0.4090	0.4921	0.4769	0.4478	0.5203	0.5138	0.5452
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
R&D Exp/TA	0.4536	0.4929	0.5294	0.4165	0.4729	0.4978	0.4429	0.4916	0.3806	0.4205	0.3917
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Investment Delay	0.0296	0.0406	0.0863	0.0861	0.0784	0.0592	0.0568	0.0494	0.0319	0.0337	0.0074
	0.0000	0.0004	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0230
Indep. Directors	0.0518	0.0542	0.0860	0.0611	0.0499	0.0975	0.0725	0.0716	0.0980	0.0770	0.0544
	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
σ (Cost)	0.0321	0.0681	0.0608	0.0845	0.0706	0.0701	0.0548	0.0470	0.0196	0.0286	0.0271
	0.0415	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0014	0.0047	0.0835	0.0706

Table VIII Panel A continued

Risk Proxy	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
σ(NCF)	0.0432	0.0157	-0.0040	0.0016	0.0825	0.0444	0.0709	0.0165	0.0519	0.0692	0.0490
	0.0002	0.0011	0.0148	0.0042	0.0000	0.0012	0.0000	0.0159	0.0001	0.0417	0.0034
σ(Revenue)	0.0597	0.0354	0.0420	0.0580	0.0504	0.0342	0.0572	0.0460	0.0330	0.0368	0.0480
	0.0054	0.0071	0.0112	0.0006	0.0027	0.0126	0.0030	0.0002	0.0007	0.0032	0.0201
σ_E	0.1679	0.1491	0.1546	0.1928	0.1108	0.1142	0.1203	0.1542	0.1422	0.1767	0.2263
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Average adj. R^2	0.1664	0.1747	0.1808	0.1738	0.1939	0.2016	0.1948	0.1916	0.1903	0.1989	0.1922

Panel B: Risk-factor disclosure quantified by the risk-factor sentence-count as a fraction of the total Item 1A sentence-count

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Average adj. R^2	0.1992	0.1810	0.1958	0.1968	0.2071	0.2075	0.2079	0.2200	0.1967	0.2025	0.2042

Panel C: Risk-factor disclosure quantified by risk-factor sentence-count as a fraction of the average sentence-count of the firm's non-zero risk-factors

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Average adj. R^2	0.1983	0.1862	0.2020	0.1969	0.2099	0.2032	0.2082	0.2141	0.1998	0.2076	0.2028

Table IX: Relation of Asset Volatility to Indices of *Operating Risk* and *Financing Risk*

Regressions of Asset Volatility on LDA-identified *Operating Risk* and *Financing Risk* indices. *Asset Volatility* is the average over the 12 months preceding the fiscal year-end of the monthly asset volatility values available from Jaewon Choi's homepage. The *Financing Risk* index is constructed using sentence-counts of the *Financing I*, *Finance II*, and *Financing III* risk-factors. The *Operating Risk* index is constructed using sentence-counts of all risk-factors other than the financing risk-factors and the *Stock Volatility* risk-factor. The firm-level *Operating* (*Financing*) index in a year is the equal-weighted average of the firm-level sentence-counts for all *Operating* (*Financing*) risk-factors that year. The risk indices are winsorized at the 1% level. *Market Leverage* is the sum of the book value of debt and market value of equity divided by the market value of equity as at fiscal year-end. Year and industry fixed effects are included as indicated. Robust *t*-statistics based on clustered standard errors at the SIC 2-digit industry level are shown in parentheses. ** and *** denote significance at the 5% and 1% level respectively.

	(1)	(2)	(3)	(4)
Operating Risk Index	0.0135***	0.0165***	0.0159***	0.0114***
	(7.32)	(10.39)	(10.23)	(7.24)
Financing Risk Index		−0.0092***	−0.0077***	−0.0034***
		(−3.48)	(−3.00)	(−3.04)
Market Leverage			−0.0151***	−0.0076**
			(−2.74)	(−2.14)
Year FE	Yes	Yes	Yes	Yes
Industry FE	No	No	No	Yes
Observations	5,105	5,105	5,105	5,105
Adjusted R^2	0.119	0.176	0.186	0.396

Table X: Relation of Equity Volatility to Indices of *Operating Risk* and *Financing Risk*

Regressions of Equity Volatility, σ_E, on LDA-identified *Operating Risk* and *Financing Risk* indices and the *Stock Price Volatility* risk-factor. The dependant variable σ_E is estimated as the annualised daily return volatility over the fiscal year. The *Risk Indices* are as defined in Table IX and are winsorized at the 1% level. *Market Leverage* is the sum of the book value of debt and market value of equity divided by the market value of equity as at fiscal year-end. Year and industry fixed effects are included as indicated. Robust *t*-statistics based on clustered standard errors at the SIC 2-digit industry level are shown in parentheses. *** denotes significance at the 1% level.

	(1)	(2)	(3)
Operating Risk Index	0.0069***	0.0050***	0.0007
	(5.59)	(5.76)	(0.98)
Financing Risk Index	0.0024	0.0046***	0.0039***
	(0.88)	(4.24)	(3.72)
Market Leverage	0.0236***	0.0242***	0.0244***
	(2.88)	(2.77)	(2.72)
Stock Price Volatility risk-factor			0.0056***
			(8.48)
Year FE	Yes	Yes	Yes
Industry FE	No	Yes	Yes
Observations	13,464	13,464	13,464
Adjusted R^2	0.366	0.443	0.464

Table XI: Prediction of Operating Income, Net Income, and Sales Performance

Panel regressions of operating income, net income, and sales in each quarter of year $t + 1$ on annual changes (Δ) in operating and financing risk indices from years $t - 1$ to t. *Operating Income* is measured as quarterly operating income before depreciation divided by beginning-of-quarter total assets. *Net Income* is measured as quarterly net income divided by beginning-of-quarter total assets. *Sales* are measured as quarterly sales divided by beginning-of-quarter total assets. The dependent variables are measured over quarters one (q+1), two (q+2), three (q+3), and four (q+4) of year $t + 1$. The *Financing Risk Index* is the equal-weighted average of the firm-level sentence-counts of the *Financing I*, *Finance II*, and *Financing III* risk-factors. The *Operating Risk Index* is the equal-weighted average of the firm-level sentence-counts of all risk-factors other than the *Stock Volatility* risk-factor and the three financing risk-factors. Regressions are performed after multiplying the dependent variable by 100. All variables are winsorized at the 1% level. All regressions include firm and year fixed effects. Robust t-statistics based on clustered standard errors at the yearly level are shown in parentheses. The F-statistic is that associated with the restriction that the predicted effect on the dependant variable of a one-sentence increase in an operating risk-factor is equal to that of a one-sentence increase in a financing risk-factors. *, **, and *** denote significance of estimated coefficients or F-statistics at the 10%, 5%, and 1% levels, respectively.

Panel A: Relation between quarterly *Operating Income* and past annual changes in the *Operating Risk* and *Financing Risk* indices

	q+1	q+2	q+3	q+4
Past annual change in *Operating Risk* index	–0.0243	–0.0440***	–0.0651***	–0.0651***
	(–1.14)	(–3.30)	(–3.77)	(–4.03)
Past annual change in *Financing Risk* index	–0.0664***	–0.0700***	–0.0721***	–0.0827***
	(–3.94)	(–4.81)	(–4.13)	(–4.00)
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	10,931	10,824	10,755	10,628
Adjusted R^2	0.036	0.031	0.023	0.029
F-statistic	11.59***	18.15***	11.36***	12.17***

Table XI continued

Panel B: Relation between quarterly *Net Income* and past annual changes in the *Operating Risk* and *Financing Risk* indices

	q+1	q+2	q+3	q+4
Past annual change in *Operating Risk* index	–0.0626	–0.0506*	–0.0756**	–0.1310***
	(–1.62)	(–2.02)	(–2.35)	(–3.89)
Past annual change in *Financing Risk* index	–0.0716***	–0.0744***	–0.0763***	–0.1044**
	(–3.33)	(–4.85)	(–4.33)	(–2.95)
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	10,928	10,822	10,755	10,628
Adjusted R^2	0.032	0.030	0.016	0.046
F-statistic	6.51**	16.66***	10.86***	6.52**

Panel C: Relation between quarterly *Sale*s and past annual changes the *Operating Risk* and *Financing Risk* indices

	q+1	q+2	q+3	q+4
Past annual change in *Operating Risk* index	–0.0923	–0.1749**	–0.1716**	–0.2134***
	(–1.31)	(–2.93)	(–2.48)	(–4.51)
Past annual change in *Financing Risk* index	–0.1610**	–0.2156***	–0.2290***	–0.2495***
	(–2.78)	(–3.52)	(–3.90)	(–4.79)
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	10,931	10,824	10,755	10,628
Adjusted R^2	0.063	0.067	0.061	0.056
F-statistic	5.55**	9.00**	10.42**	17.96**

Appendix Figure A1

Risk-Factor Labels – Word Cloud Representation

The figure displays word clouds of the 30 most common words in each risk-factor topic. The scaling of the size of the words reflects the frequency of the words in a cloud.

Appendix Table A1

LDA Risk-Factor Word lists

This Table presents the 30 LDA-identified risk-factors, their assigned labels, and the 30 words most frequently associated with each risk-factor across all firm-years. The words are listed in order of decreasing count frequency.

Topic Counter	LDA Risk Factors	Topic Word Composition
1	*Catastrophe*	operation, event, disaster, business, natural, result, facility, loss, damage, disruption, system, weather, failure, condition, including, adversely, terrorist, affect, attack, risk, act, interruption, earthquake, impact, adverse, hurricane, significant, fire, financial, company.
2	*Corporate Governance*	stock, stockholder, director, share, common, board, provision, control, shareholder, right, change, company, law, voting, preferred, incorporation, bylaw, certain, prevent, class, approval, outstanding, interest, transaction, price, business, vote, ownership, holder, certificate.
3	*Country-Related*	rate, currency, interest, foreign, risk, exchange, result, dollar, fluctuation, change, financial, operation, market, value, increase, affect, exposure, income, asset, company, impact, investment, revenue, net, hedging, adversely, hedge, contract, price, loss.
4	*Customer Concentration*	sale, revenue, customer, approximately, company, product, year, agreement, united, state, service, operation, business, significant, portion, fiscal, market, facility, net, store, december, total, international, contract, outside, including, number, accounted, located, ended.
5	*Economic Conditions*	economic, business, foreign, risk, operation, condition, market, change, including, country, political, international, subject, state, law, product, result, financial, rate, united, regulation, factor, currency, control, cost, tax, affect, government, adversely, trade.
6	*Energy Sector*	gas, price, natural, oil, cost, operation, energy, result, production, market, demand, power, company, fuel, customer, drilling, including, increase, facility, service, regulation, emission, change, coal, risk, supply, condition, new, business, future.
7	*Financing I*	company, financial, insurance, investment, market, risk, credit, capital, rating, bank, business, security, fund, institution, including, asset, certain, service, subsidiary, result, loss, regulation, subject, rate, regulatory, requirement, federal, ability, state, liquidity.
8	*Financing II*	credit, facility, debt, note, agreement, million, senior, indebtedness, covenant, certain, default, interest, asset, amount, subsidiary, outstanding, financial, ability, payment, secured, december, event, loan, term, company, stock, revolving, obligation, including, make.
9	*Financing III*	capital, debt, ability, cash, business, credit, financing, additional, operation, future, fund, financial, market, indebtedness, term, condition, flow, obligation, available, affect, obtain, result, adversely, make, company, interest, need, acquisition, able, service.
10	*Growth and Restructuring*	business, acquisition, operation, company, acquired, risk, result, cost, management, financial, future, growth, including, product, benefit, investment, new, significant, successfully, strategy, system, difficulty, anticipated, strategic, venture, resource, technology, time, integration, able.

Table A1 continued

11	*Healthcare Spending*	program, state, service, health, government, federal, care, contract, medicare, payment, law, healthcare, regulation, cost, act, change, including, provider, result, business, certain, medicaid, reimbursement, rate, revenue, patient, hospital, year, subject, insurance.
12	*Human Capital*	personnel, business, retain, ability, key, employee, attract, management, qualified, success, service, new, executive, operation, depends, future, company, adversely, officer, growth, result, senior, loss, able, upon, sale, unable, store, highly, competition.
13	*Incomplete Contracts*	agreement, company, director, certain, court, officer, action, state, stock, board, filed, law, executive, subject, interest, merger, business, including, share, inc, party, shareholder, district, right, claim, security, term, transaction, addition, time.
14	*Information Systems*	system, information, business, security, customer, service, data, result, operation, breach, technology, failure, party, loss, financial, risk, reputation, including, third, company, network, disruption, damage, product, computer, significant, ability, employee, client, adversely.
15	*Intangible Assets*	asset, value, impairment, result, goodwill, future, intangible, charge, estimate, fair, million, loss, change, financial, net, cost, significant, amount, contract, operation, investment, period, cash, carrying, assumption, market, required, company, revenue, december.
16	*Intellectual Property*	right, property, intellectual, patent, product, technology, party, protect, proprietary, license, third, claim, trademark, business, infringement, use, agreement, trade, others, litigation, secret, protection, company, result, law, certain, obtain, future, service, addition.
17	*Product Approval*	product, approval, fda, regulatory, clinical, market, development, drug, trial, sale, result, new, including, candidate, company, marketing, manufacturing, use, device, regulation, subject, medical, state, process, obtain, certain, patient, pharmaceutical, requirement, business.
18	*Product Defects*	claim, insurance, liability, result, business, litigation, coverage, product, cost, loss, financial, operation, damage, significant, future, risk, time, company, legal, subject, proceeding, amount, adverse, management, material, action, substantial, matter, lawsuit, property.
19	*Product Innovation*	product, new, service, market, customer, technology, competitor, business, company, ability, competitive, industry, develop, development, resource, change, sale, result, financial, compete, revenue, existing, competition, system, marketing, future, continue, greater, software, including.
20	*Real estate*	loan, property, real, estate, loss, market, mortgage, condition, result, rate, lease, increase, risk, value, tenant, interest, economic, ability, adversely, credit, portfolio, cost, business, affect, financial, investment, company, home, including, change.
21	*Regulatory Change*	regulation, law, business, result, operation, subject, change, regulatory, cost, comply, compliance, state, financial, requirement, government, affect, future, adversely, penalty, new, federal, company, adverse, effect, impact, including, applicable, environmental, fine, failure.
22	*Regulatory Compliance*	law, regulation, environmental, state, subject, property, hazardous, federal, liability, operation, product, material, cost, substance, including, safety, local, certain, waste, act, use, disposal, health, facility, company, damage, requirement, various, site, contamination.

Table A1 continued

23	*Reporting Accuracy*	risk, statement, result, financial, factor, report, forward, looking, information, business, operation, uncertainty, see, condition, item, form, discussion, future, note, management, annual, company, described, materially, additional, following, actual, currently, consolidated, analysis.
24	*Reporting Compliance*	financial, control, internal, reporting, accounting, result, statement, material, report, company, standard, management, change, effective, weakness, system, procedure, act, required, public, stock, future, business, sarbanesoxley, operation, section, sec, disclosure, maintain, requirement.
25	*Supply Chain*	product, customer, cost, material, supplier, result, supply, manufacturing, component, contract, price, increase, raw, delay, production, business, ability, time, order, demand, service, purchase, company, manufacturer, certain, operation, adversely, significant, new, sale.
26	*Tax Uncertainty*	tax, income, reit, change, rate, subject, federal, taxable, law, would, result, certain, state, distribution, year, asset, liability, company, future, jurisdiction, qualify, net, foreign, amount, provision, transaction, revenue, authority, dividend, effective.
27	*Cost Volatility*	plan, cost, rate, pension, benefit, increase, asset, change, result, future, employee, company, expense, obligation, million, return, assumption, funding, liability, contribution, requirement, significant, market, including, required, interest, operation, certain, year, financial.
28	*NCF Volatility*	result, financial, operation, condition, business, adverse, adversely, effect, material, affect, cash, impact, flow, materially, operating, customer, affected, company, market, significant, product, risk, loss, future, cost, failure, economic, position, revenue, would.
29	*Revenue Volatility*	result, product, customer, sale, revenue, price, market, operating, business, demand, cost, industry, adversely, increase, condition, economic, change, affect, margin, future, quarter, impact, significant, consumer, service, level, factor, decline, period, operation.
30	*Stock Price Volatility*	stock, common, price, market, share, security, result, company, future, operating, trading, sale, decline, factor, fluctuation, change, affect, financial, analyst, equity, adversely, dividend, condition, investor, performance, significant, stockholder, volatility, value, investment.

Appendix Table A2

Variable Definitions

This Table defines variables and lists the data sources.

Variable name	Definition	Source
Book D/A	Sum of book value of long- and short-term debt relative to book value of total assets	CRSP Compustat Merged (CCM)
Cash Flow	Earnings before extraordinary items plus depreciation divided by total assets	CCM
Credit Spread	Percentage point difference between the last available yield to maturity on a firm's bonds before the fiscal year-end and a contemporaneous benchmark Treasury yield as in Fracassi, Petry and Tate (2016)	Trace, Mergent FISD, CRSP
Customer Dummy	Dummy variable equal to one if a corporate customer makes up at least 10% of the firm's sales	Compustat Segments
Fluidity	Intensity of product market change as in Hoberg, Phillips and Prabhala (2014)	Hoberg-Phillips Data Library
Independent Directors	Fraction of independent directors on firm's board	IRRC
Investment Delay	The delaycon variable from Hoberg and Maksimovic (2014), measuring the firms' risks of delaying their investments due to liquidity-related issues	Hoberg-Phillips Data Library
Market Leverage	Book value of debt plus market value of equity relative to market value of equity	CCM
RD Expense/TA	Research & development (R&D) expense divided by total assets. Set to zero if R&D expense is missing	CCM
TNIC HHI	HHI value in the TNIC3HHIdata.txt file	Hoberg-Phillips Data Library https://hobergphillips.tuck.dartmouth.edu/
(TA−Net PP&E)/TA	(Total assets – Net property plant and equipment) /Total assets	CCM
σ(Asset Return)	Average of monthly stacked EGARCH estimates of asset volatility over 12 months ending at the fiscal year-end	Jaewon Choi website https://sites.google.com/site/jaewchoi1203

Table A2 continued

Variable name	Definition	Source
σ(Cost)	Annualized standard deviation of quarterly costs of goods sold divided by total assets estimated over the 12 quarters preceding the fiscal year-end	CCM
σ(NCF)	Annualized standard deviation of quarterly cash flow (EBITDA) divided by total assets estimated over the 12 quarters preceding the fiscal year-end	CCM
σ(Revenue)	Annualized standard deviation of quarterly revenue divided by total assets estimated over the 12 quarters preceding the fiscal year-end	CCM
σ_E	Annualized stock return volatility estimated using daily stock returns over the 250 trading days preceding the fiscal year-end	CRSP

Appendix Table A3

30 Most Common Words in the sentences assigned to *Corporate Governance* Risk

stock, stockholder, director, share, common, board, provision, control, shareholder, right, change, company, law, voting, preferred, incorporation, bylaw, certain, prevent, class, approval, outstanding, interest, transaction, price, business, vote, ownership, holder, certificate.

Selected paragraphs from Facebook's 2012 10-K Item 1A: Words in bold are *Corporate Governance* common words

Our **Class** B **common stock** has ten **votes** per **share**, and our **Class** A **common stock** has one **vote** per **share**. Because of the ten-to-one voting ratio between our **Class** B and **Class** A **common stock**, the **holders** of our Class B **common stock** collectively **control** a majority of the combined **voting** power of our **common stock** and therefore are able to **control** all matters submitted to our **stockholders** for **approval** so long as the **shares** of **Class** B **common stock** represent at least 9.1% of all **outstanding shares** of our **Class** A and **Class** B **common stock**. This concentrated **control** will limit or preclude your ability to influence corporate matters for the foreseeable future.

Because we qualify as a "**controlled company**" under the corporate governance rules for NASDAQ-listed companies, we are not required to have a majority of our **board** of **directors** be independent, nor are we required to have a compensation committee or an independent nominating function. […] Accordingly, should the **interests** of our controlling **stockholder** differ from those of other **stockholders**, the other **stockholders** may not have the same protections afforded to **stockholders** of companies that are subject to all of the corporate governance rules for NASDAQ-listed companies. Our status as a **controlled company** could make our **Class** A **common stock** less attractive to some investors or otherwise harm our **stock price**.

Our status as a Delaware corporation and the anti-takeover provisions of the Delaware General Corporation **Law** may discourage, delay, or **prevent** a **change** in **control** by prohibiting us from engaging in a **business** combination with an **interested stockholder** for a period of three years after the person becomes an **interested stockholder**, even if a change of **control** would be beneficial to our existing **stockholders**. In addition, our restated **certificate** of **incorporation** and **bylaws** contain provisions that may make the acquisition of our **company** more difficult.

Source: https://www.sec.gov/Archives/edgar/data/1326801/000132680113000003/fb-12312012x10k.htm.