10-K Risk-Factors Quantification and the Information Content of Textual Reporting*

Bruce Grundy†
University of Melbourne

Stefan Petry‡
University of Manchester

ABSTRACT

We use machine-learning to determine the information content of the Item 1A Risk Factors section of S&P 1500 10-Ks. We identify and quantify 30 risk-factors and show a strong positive relation between levels of and contemporaneous changes in risk-factors and proxies for the associated risks. Typically, 28% of cross-firm variation in a risk-proxy is explained by cross-firm variation in the associated risk-factor. Risk disclosure is not found to be forward-looking. Item 1A's informativeness has not declined through time despite previously documented increases in boilerplate content, stickiness and redundancy. Indices of operating and financing risk help explain asset and equity volatility.

JEL classification: G3, G38, J53, K22, L51

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

* We thank Qi Deng and Zhuo Wei for valuable research assistance.

† Faculty of Business and Economics, University of Melbourne, 198 Berkeley Street, Carlton 3010, Australia. phone: +61 (0)3 8344 9083, email: bruceg@unimelb.edu.au

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

10-K Risk-Factors Quantification and the Information Content of Textual Reporting

We use machine-learning to determine the information content of the Item 1A Risk Factors section of S&P 1500 10-Ks. We identify and quantify 30 risk-factors and show a strong positive relation between levels of and contemporaneous changes in risk-factors and proxies for the associated risks. Typically, 28% of cross-firm variation in a risk-proxy is explained by cross-firm variation in the associated risk-factor. Risk disclosure is not found to be forward-looking. Item 1A's informativeness has not declined through time despite previously documented increases in boilerplate content, stickiness and redundancy. Indices of operating and financing risk help explain asset and equity volatility.

JEL classification: G3, G38, J53, K22, L51

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

US companies are required to report the risks that they are exposed to in unquantified text format in the Item 1A Risk Factors section of their annual reports (10K-statements). Despite their importance, we know little about the relation between reported risk-factors and measures of the associated risks. Textual disclosure means that understanding, interpreting, and analysing the reported risk-factors presents a challenge for investors, analysts, and researchers who wish to examine a large sample of firms. We employ a machine-learning algorithm, Latent Dirichlet Allocation (LDA), to uncover and quantify topics that represent meaningful risk-factors. The risk-factors are meaningful in that the count of sentences assigned to a topic is significantly cross-sectionally positively correlated with proxies for the risk discussed in the sentences assigned to the topic and changes in risk-factor sentence-counts are significantly positively correlated with contemporaneous changes in the associated risk proxies.

We view the strength of the cross-sectional correlation between risk proxies and LDA-identified risk-factors as a measure of the accuracy of the Item 1A Risk Factor section. That accuracy is a determinant of information content is a complement to the familiar notion that disclosures are informative if they move prices or induce trade. Investors may find the Risk Factors section useful in designing their optimal portfolios even when there are no abnormal returns or trading volume around 10-K filing dates. Our principal contribution is an investigation of the accuracy of the Risk Factors section by analysing the relation between firm-level proxies for risk-factors and the factors' LDA-identified firm-level sentence-count. This investigation answers the following questions. What are the risk-factors disclosed in Item 1A and how well do they describe a firm's risk exposure? Has the information content of Item 1A changed over time? Do changes in risk-factor sentence-counts predict future changes in proxies for the associated risks? And, what is the relation between Item 1A risk-factors and firm operating and financing risk?

The SEC provides little direction as to the risks that need to be disclosed. Item 503(c) of Regulation S-K provides some indirect guidance by requiring the reporting of "the most significant factors that make [an investment in the security] speculative or risky." The risk factor discussion must be "concise and organized logically". Further, the regulation directs firms to explain "how each risk affects the firm" and it "discourages disclosures of risks that could apply to any

registrant".[1] An SEC discussion of the 10-K report states that "In practice, [the risk factor] section focuses on the risks themselves, not how the company addresses those risks. Some risks may be true for the entire economy, some may apply only to the company's industry sector or geographic region, and some may be unique to the company."[2] Thus, even though the SEC's guidance is limited, there is a recognition of firm, industry and market-wide risks.

We examine the set of S&P 1500 firms during the period 2005 to 2015 and use the Bao and Datta (2014) sentence-based Latent Dirichlet Allocation (LDA) modification of the traditional LDA of Blei, Ng and Jordan (2003) to identify a thematic structure within Item 1A. The most frequent words that belong to a common word cluster or topic are extracted and maximum likelihood is used to assign each sentence to a single topic or risk-factor. We attach a descriptive label to each risk-factor based on the most frequent words in the risk-factor's word cluster words and a reading of the sentences assigned to the risk-factor. The importance of a risk-factor for a firm in a given year is quantified by the risk-factor sentence-count and normalizations thereof. Following Bao and Datta (2014), we assume that Item 1A discusses 30 topics or risk-factors. The labels of the five most important risk-factors based on their associated average sentence-counts are *Product Innovation* (15.3 sentences), *Volatile Revenue* risk (14.0), *Supply Chain* risk (12.2), *Volatile Net Cash Flow* risk (11.8), and *Energy Sector* risk (9.9). The five least important risks are *Catastrophe* (3.5), *Corporate Governance* (3.3), *Reporting Compliance* (2.9), *Financing I* (1.7), and *Real Estate* risk (1.6).[3]

To determine how well the LDA-identified risk-factors capture firms' risk exposures, we focus on the 15 risks for which we can determine one of more observable proxies. For example, a potential proxy for the *Volatile Stock Price* risk is the annualized stock return volatility estimated

[1] https://www.govinfo.gov/content/pkg/CFR-2019-title17-vol3/xml/CFR-2019-title17-vol3-sec229-503.xml.

[2] See https://www.sec.gov/fast-answers/answersreada10khtm.html .

[3] The relative importance of the 30 risk factors is in part a reflection of the industry composition of the S&P 1500 index. If pharmaceutical and healthcare companies had constituted a higher proportion of the sample, then the relative importance of *Product Approval* risk (the sixth most important of the 30 risks) would have been higher.

using monthly stock returns over the 60 months preceding the fiscal year-end. Because some risk-factors have more than one associated risk-proxy measure, we are able to analyse 20 risk-factor, risk-proxy pairings. We first examine whether firms with more than the 2005 to 2015 median number of sentences associated with a particular risk-factor have higher average and median values of the associated risk-proxy than firms with a below-median number of associated sentences. We also examine two normalizations: the risk-factor sentence count as a fraction of the firm's total Item 1A sentence-count and the risk-factor sentence-count as a fraction of the average sentence-count for the firm's non-zero risk-factors. An above-median sentence-count is associated with a statistically significantly higher median (mean) value of the risk-proxy for 20 (19) of the risk-factor, risk-proxy pairings. We also examine the cross-sectional relations between risk-factors and associated risk-proxies at an annual level. The estimated relation is positive for 203 of the 217 pairings and significantly so at the 1% level for 122 of the pairings. The relation achieves significance at the 5% level for a further 18 pairings. In sharp contrast, only one pairing is significantly negatively related at the 5% level.

In August of 2019 the SEC called for public comments on proposed amendments to risk factor disclosures.[4] One of the stated goals of these amendments is to discourage repetition and the disclosure of immaterial information. Dyer, Lang and Stice-Lawrence (2017) report that between 1996 and 2013, there has been an increase in the boilerplate content, stickiness and redundancy of 10-K statements, and they identify the introduction of new reporting requirements such as the Item 1A risk disclosure in 2005 as one cause of the increase in the length of 10-Ks. The fraction of the variation in firm-level proxies for the risk-factors explained by variation in the firm-level sentence-counts of the associated risk-factors provides a direct metric of the accuracy of risk-factor disclosures. Using this metric, we find no evidence that risk disclosures have become less informative through time. The average adjusted R^2 values of regressions of risk-proxies on their associated risk-factor sentence-counts has been around 28% in every year from 2005 through 2015.

We explore the possibility that changes in risk-factor sentence-counts for a firm predict changes in the associated risk-proxy value of the firm. That is, do firms adjust their Item 1A descriptions in anticipation of future changes in risk levels or do changes in risk-factor descriptions reflect realized changes in risk. Our results suggest that there is little relation between lagged

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

changes in risk-factor sentence-counts and future changes in risk-proxies. Hence, Item 1A can be regarded as a statement about the levels of firms' risk exposures at fiscal year-end rather than as forward-looking disclosures.

The 30 risk-factors can be divided into 26 measures of firm operating risk, three measures of firm financing risk, and a measure of *Volatile Stock Price* risk. Equity volatility will reflect the volatility of firm assets and firm leverage. We combine risk-factor sentence-counts to create *Operating Risk* and *Financing Risk* indices at the firm-year level. We show that asset volatility is related to our LDA-identified indices of *Operating Risk* and *Financing Risk* in a manner consistent with the findings of Choi and Richardson (2016) and that equity volatility is increasing in both the *Operating Risk* index and leverage.

Section I reviews the literature on the textual analysis of firm risk. Section II describes the LDA identification of risk factors. Section III describes the data and presents summary statistics. Section IV documents the relations between firm-year levels of (changes in) LDA-identified risk factors and firm-year levels of (changes in) the associated risk-proxies. Section V examines the association between financial and operating risk indices and asset and equity volatilities. Section VI concludes.

I. Textual Analysis of Risk Reporting

Li (2010) and Loughran and McDonald (2016) survey the growing literature using textual analysis to address issues in finance. The applications include but are not limited to risk reporting. Bodnaruk, Loughran and McDonald (2015) investigate 10-K Statements and report that the frequency of words like 'required', 'obligations', 'requirements', 'permitted', 'comply', and 'imposed' has incremental power relative to the KZ (Kaplan and Zingales, 1997), SA (Hadlock and Pierce, 2010), and WW (Whited and Wu, 2006) indices of financial constraints as predictors of dividend omissions, equity issues to fund dividends and repurchases, and underfunded pensions indices of financial constraints. Hoberg and Phillips (2016) develop a novel measure of industry membership based on the similarity of firms' 10-K product descriptions. Dyer, Lang and Stice-Lawrence (2017) investigate how 10-K reports have changed over time in terms of length, boilerplate, stickiness, redundancy, specificity, readability, and the relative amount of hard information. Bellstam, Bhagat and Cookson (2019) identify topics within analyst reports and labels

the topic with the lowest divergence from textbook language on innovation as the innovation topic. The likelihood a report is about the innovation topic is positively related to measures of firm performance and growth opportunities. Huang, Lehavy, Zang and Zheng (2019) compare analyst reports issued soon after earnings conference calls to the content of the calls themselves and conclude that analysts discover information beyond that disclosed in conference calls. Lowry, Michaely and Volkova (2019) 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.

Several studies employ textual analysis to investigate risk. 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. The result is interpreted as consistent with a firm's preference for withholding bad news. 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 before. 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. Specificity is measured by the occurrence of names of persons, locations, and organizations and quantitative values in percentages and dollars, times, and dates.

Brown, Tian and Tucker (2018) find evidence that companies monitor the SEC's response to the risk-factor reporting of their peers 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. Hoberg, Phillips and Prabhala (2014) develop the Fluidity measure of the product market threat faced by a company. Fluidity is measured by 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. Larger values of the absolute change in the aggregate number of rival firms that use similar language to describe their products as that used by firm *i* are posited to be indicative of a more uncertain or fluid competitive landscape

for firm *i*. Hoberg, Phillips and Prabhala (2014) show that high-fluidity firms are less likely to distribute dividends or repurchase their stock and hold more cash than low-fluidity firms.

Several text-based investigations of risk have used LDA analysis. Campbell, Chen, Dhaliwal, Lu and Steele (2014) employ LDA analysis to identify risk-related key words that are added to an initial list based on prior literature. The complete list is then manually classified into financial risk, litigation risk, tax risk, and other risks. For each category, the keyword count and the percentage of all keywords associated with a particular classification are shown to be cross-sectionally related to observable measures of systematic and idiosyncratic risk. 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 more useful for business valuation in cases where quantitative information is less relevant, and particularly so during times of business change. Goldsmith-Pinkham, Hirtle and Lucca (2016) use LDA analysis to identify five topics in the matters requiring attention or immediate attention in supervisory demands raised by Federal Reserve examiners for corrective actions by banks. Hanley and Hoberg (2019) show that LDA analysis of item 1A of the 10-K statements of banks can be used to detect emerging risks in the financial sector. Lopez-Lira (2019) develops a factor model of returns based on an LDA identification of the risks disclosed in Item 1A. Factors are associated with the four risk topics that affect the largest number of firms. Factor-mimicking returns are constructed as value-weighted returns on portfolios that contain those firms that allocate more than 25% of their risk disclosure statement to a discussion of one of the four risk topics. A 5-factor asset pricing model constructed from the market and the four risk disclosure measures is found to better explain cross-sectional differences in average returns than does the 5-factor Fama-French model.

II. LDA Identification of Risk-Factors

To identify the set of topics discussed in Item 1A of firms' 10-K statements, we employ the Blei, Ng, and Jordan (2003) LDA machine-learning algorithm as modified by Bao and Datta (2014). LDA is an unsupervised learning technique that identifies clusters of words that represent topics and thereby reduces the dimensionality of the text. We analyse the entirety of Item 1A after applying standard text pre-processing techniques that delete words that belong 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 assumes that the words that form the discussion of a topic have a pre-defined distribution.[5] 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.

We use the terms topic, risk and risk-factor interchangeably. Relative to other firms in the same year, the more sentences that a firm devotes to a given risk, the more important that risk is to the firm. The result is a panel dataset consisting of observations on the sentence-count associated with each of the 30 LDA risk-factors in each year 2005 through 2015. We ascribe a label to each risk-factor based on the 30 most common words in the sentences assigned to the risk-factor and a reading of Item 1A for firms with high sentence-counts for the factor. Our assignment of labels to the topics is necessarily subjective. The labels we assign are similar to those assigned by Bao and Datta (2014) and Huang and Li (2011).[6] Appendix Table A.1 reports the risk-factor labels and the 30 most common words in order of frequency.[7] Because of the similarity of the messages conveyed by many of the sentences assigned to different financing-related risk-factors, we opted not to use our initial label choices of *Credit Market Conditions*, *Debt* risk, and *Financing Uncertainty* risk and instead label them simply as *Financing I*, *Financing II* and *Financing III*.

[5] For details see Blei, Ng, and Jordan (2003).

[6] Huang and Li (2011) consider 25 risk-factors and provides a brief description for each risk while Bao and Datta (2014) considers 30 risk-factors and reports the associated word clouds. Mirakur (2011) examines a random sample of 122 firms and manually assigns the disclosures in Item 1A into 29 risk categories based on an initial assignment into 116 subcategories. The descriptions of the risk subcategories within the separate categories have a strong overlap with the 30 key words associated with our set of LDA risk-factors, suggesting LDA yields results are comparable to what a human coder would produce. The advantage of an LDA approach is that LDA can analyse much more text than a human coder can, the outcome is reproducible, and the underlying assumptions are transparent.

[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 30 word lists as word clouds.

III. Data and Summary Statistics

Disclosure of risk in Item 1A became mandatory in 2005.[8] We examine the Item 1A risk disclosures over the 2005 through 2015 period for the set of firms contained in the S&P 1500. Annual 10-Ks are retrieved from the SEC's Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system and firms' accounting and stock market data from the CRSP-Compustat merged database. Asset betas for the years 2005 through 2012 are obtained from Jaewon Choi's personal homepage.[9] Following Fracassi, Petry and Tate (2016), we use data from TRACE and Mergent FISD to calculate bond-level credit spreads on 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 outstanding bond issues with weights given by the amount outstanding of each issue relative to the total of all available bonds for the same firm. Fluidity measures are obtained from the Hoberg-Phillips Data Library.[10] Appendix Table A2 provides a summary of the variables and 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 missing or negative. We exclude firms with SIC 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.[11] Table I reports summary statistics for the sample. The average book value of assets is $8.19 billion. The average book leverage is 20.4%. Average R&D expenditures and Intangible assets comprise 3.1% and 74.3% of firm assets.

[8] Risk factor disclosure as mandated by Item 503(c) of the SEC Regulation S-K became effective for fiscal periods ending on or after December 1, 2005, with SEC-defined "smaller reporting companies" being exempt from this requirement (Brown, Tian and Tucker (2018)).

[9] We thank Jaewon Choi for making the databases at https://sites.google.com/site/jaewchoi1203 available to researchers.

[10] We thank Gerard Hoberg and Gordon Phillips for providing their databases to researchers via https://hobergphillips.tuck.dartmouth.edu/

[11] The sample size is comparable to that of other studies that examine 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; and Filzen (2015) examine 13,165 firm-quarters over 2006—2010.

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 of the firms in the sample is 1.079. The mean volatilities of quarterly costs, quarterly net cash flows, and quarterly revenues measured 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%.

The result is a panel dataset consisting of observations on the sentence-count associated with each of the 30 LDA risk-factors in each year 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), *Volatile Revenue* (14.02), *Supply Chain* risk (12.2), *Volatile Net Cash Flows* (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-counts over all firm-years is zero. These nine risks are *Corporate Governance*, *Energy Sector*, *Financing I*, *Healthcare Spending*, *Product Approval*, *Real Estate*, *Reporting Compliance*, *Tax Uncertainty*, and *Volatile Costs* risk.

We examine both the firm sentence-counts associated with each risk-factor and two normalizations of the sentence-count. 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 if 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 choose not to discuss all the risk-factors. Thus, we also examine the risk-factor sentence-count relative to average sentence-count for those factors the firm does report on. For ease of exposition, we describe these factors as the firm's nonzero factors. Suppose firm A devotes N^A sentences to each of M^A nonzero sentence-count risk-factors and firm B devotes N^B sentences to each of M^B nonzero sentence-count risk-factors. For a risk-factor discussed by both firms, the first normalization measures the importance of the risk as $N^A \big/ M^A \times N^A = 1 \big/ M^A$ and $N^B \big/ M^B \times N^B = 1 \big/ M^B$. This common risk will be considered less important for the firm that faces

more risks.[12] Considering the same common risk-factor, the second normalization measures the importance of the risk as $N^A \Big/ \frac{M^A \times N^A}{M^A} = 1$ and $N^B \Big/ \frac{M^B \times N^B}{M^B} = 1$. This common risk will be viewed as equally important to the two firms.

IV. The Relation Between LDA-identified Risk-Factors and Risk-Proxies

A. The Contemporaneous Relation between LDA-identified Risk-Factors and Risk Proxies

We wish to examine the relation between firm-level risk disclosures and proxies for the firm-level exposure to the disclosed risks. It is important to recognize the limitations of applying textual analysis of Item 1A to identify "high risk" firms. Given the absence of SEC guidance on Item 1A, a firm may choose to report the level of unhedged risk. The actual (hedged) level may be lower. This possibility biases against finding a relation between the reported risk level and a proxy for the risk to which investors are actually exposed. We examine the cross-sectional relation between firms' reported risk exposure and proxy measures of the actual risk in order to determine how well LDA-identified risk topics capture firms' risk exposures. Columns 1 and 2 of Table III sets out the measurable proxies used for 15 of the 30 risk-factors.[13] Ten of the 15 risks are associated with a single proxy. The other five risks are each associated with two proxies.

The five risk-factors that are associated with two risk proxies are as follows. We pair *Growth and Restructuring* risk with the *negative* of the Herfindahl-Hirschman Index (HHI) and the Hoberg-Phillips-Prabhala Fluidity measure. We use the negative of HHI in order to capture a positive relation between the risk-proxy and the associated risk-factor: The more negative is HHI, the greater the market power of firms in the industry. We posit that firms with more market power face less *Growth and Restructuring* risk. An increase in the negative of HHI implies an increase in competitive pressure and an increase in *Growth and Restructuring* risk. Our second risk-proxy for *Growth and Restructuring* risk is the Fluidity measure gauge of the competitive pressure arising from innovation in the products offered by rival firms. We predict that higher quantifications of

[12] While verbosity will affect *N*, it will not affect the normalized measure.

[13] The proxies and data sources are described in Table A2.

the *Growth and Restructuring* risk-related sentence-count will be associated with higher Fluidity.[14] Leverage and the Credit Spread are both used as proxies for each of the *Financing I*, *Financing II* and *Financing III* risk-factors. Higher sentence-count quantifications of these risk-factors are predicted to be associated with higher leverage and higher credit spreads. The two risk proxies for *Product Innovation* risk are R&D Expenditures/Total Assets and Fluidity.[15] We use Fluidity as a second proxy for *Product Innovation* risk because the Item 1A disclosures of firms with high sentence-counts for the *Product Innovation* risk-factor often discuss innovation by both the firm and its competitors.[16]

The ten risk-factors associated with a single risk-proxy are as follows. For *Customer Concentration* risk, the risk-proxy is a dummy equal to 1 if the firm has a customer that contributes at least 10% to the firm's sales. If our LDA analysis correctly identifies firms facing *Customer Concentration* risk, then the dummy is more likely to equal 1 for firms with high sentence-count quantifications of the risk-factor. Asset beta is used as the proxy for risks related to *Economic Conditions*. Again, if our LDA analysis correctly identify risks with high exposures to Economic Conditions, we predict that asset beta will be high for firms with high risk associated with *Economic Conditions*. We use the fraction of firm assets not accounted for by Net PP&E as a proxy measure for *Human Capital* risk and for *Intangible Assets* risk. We predict that this fraction will

[14] Similarity is determined from the overlap of words in the product description sections of the firms' 10-K statements.

[15] The R&D Expenditures/Total Assets variable is set to zero when the R&D expense is missing in Compustat.

[16] 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 "'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".

be higher when the sentence-count based quantifications of *Human Capital* risk is higher and when the sentence-count based quantifications of *Intangible Assets* risk is higher. We use R&D Expense/Total Assets as the sole proxy for *Intellectual Property* and for *Product Approval* risk and again predict positive relations between the proxy and the risk-factor sentence-count normalizations. The proxy for *Volatile Revenue* is the annualized standard deviation of quarterly revenue relative to the end of quarter book value of total assets with the estimation made using the preceding 12 quarter's data. The proxy for *Volatile Costs* is the annualized standard deviation of quarterly costs of goods sold as a fraction of total assets over the preceding 12 quarters. The proxy for *Volatile Net Cash Flow* is estimated over the 12 quarters preceding the fiscal year as the annualized standard deviation of the quarterly earnings before extraordinary items and depreciation relative to end of quarter total assets. The *Volatile Stock Price* proxy is the annualized stock return volatility estimated over the 60 months preceding the fiscal year-end.

Table III reports the contemporaneous relations between measures of the risk-factor sentence-count and proxies for the risk captured by the factor. Panel A uses the risk-factor sentence-count as the measure of the reported risk. Panels B and C uses the two normalizations of the count discussed in Section III. The Table compares the mean value of the risk-proxy when the risk-factor sentence-count is above the median for the risk-factor (*High*) to the mean value of the risk-proxy when the risk-factor sentence-count is below the median for the risk-factor (*Low*).[17] As

[17] A *High* level for some risk may well be the optimal level of risk. A firm with a reputation for good governance may consider the risk of losing its governance standing to be an important risk and devote considerable discussion to governance risk. Another firm might devote considerable discussion to governance because by popular metrics that firm can be deemed to have bad governance. *Facebook Inc* devotes an annual average (annual minimum) of 25 (23) sentences to a discussion of *Corporate Governance* risk in its Item 1A risk disclosure. The 95th percentile of the distribution across all firms of the annual average number of sentences associated with *Corporate Governance* risk is only 13. Clearly *Facebook* reports a significantly higher level of *Corporate Governance* risk than the typical firm. But this does not imply that *Facebook* has poor governance. Appendix Table A3 sets out several paragraphs from *Facebook*'s Item 1A disclosure that contain many of the 30 most frequent words associated with *Corporate Governance* risk. These paragraphs describe how *Facebook*'s dual class share structure limits outside shareholders' ability to influence decision-making in the firm. Class B shares have ten times the voting rights of Class A

an example, consider the sentence-count for the *Volatile Stock Price* risk-factor and the annualized monthly stock return volatility proxy for this risk. The set of firms with *High* sentence-counts for the *Volatile Stock Price* risk-factor has a mean (median) stock return volatility of 46% (42%) per annum. The set of firms with *Low* sentence-counts has a mean (median) volatility of 36% (32%). The differences in the means and medians are statistically significant at the 1% level.[18]

Nineteen of the 20 risk-factor, risk-proxy pairings are such that firms with *High* risk-factor sentence-counts have significantly larger mean values of the risk-proxy than do firms with *Low* sentence-counts. The one exception is the (*Product Innovation*, Fluidity) pairing. For all 20 pairings, the median value of the risk-proxy is higher for the set of firms with *High* risk-factor sentence-counts than for the set with *Low* risk-factor sentence-counts. The differing result obtained from a comparison of means rather than medians when examining the (*Product Innovation*, Fluidity) pairing is consistent with the Fluidity measure placing more weight on innovation by rival firms than the firm does in its Item 1A disclosure of *Product Innovation* risk.

Panel B of Table III uses a normalized variant of the sentence-count to examine the relation between disclosed risk-factors and proxies for those factors. The normalization in Panel B measures the sentence-count as a fraction of the total Item 1A sentence-count of the firm. The normalization is appropriate if firms devote a higher fraction of their item 1A disclosure to the risks they see as more important. While the results in Panels A and B are similar, the differences

---

shares. The voting rights differential between class A and B shares may be optimal if it the dual class share structure that allows *Facebook* to retain the valuable human capital of its founders who hold the Class B shares. See Cremers, Litov and Sepe (2017) for an analysis of dual class share structures.

[18] Implementation of the above versus below median criterion leads the *High* and *Low* sample sizes to differ, as seen in Table III. Each integer value of the risk-factor sentence-count is associated with many risk-factor, risk-proxy pairings. Since a risk-factor is defined as *High* when its sentence-count is *above* the median, this break never cuts the sample in half and the *High* sample almost always contains fewer observations than the *Low* sample. Only for the (*Financing II*, Credit spread) pairing is the number of observations classified as *High* less than the number classified as *Low*. This occurs because the median sentence-count is determined without conditioning on the availability of the risk-proxy and this is a second reason why the *High* and *Low* sample sizes differ. Credit spreads are the least available of the risk proxies and credit spread data is more frequently available when the sentence-count is high.

are interesting. Fluidity is negatively related to the mean and median of the normalized sentence-count for both the *Growth and Restructuring* risk-factor and the *Product Innovation* risk-factor. Firms that choose to devote a high fraction of their Item 1A risk disclosure to a discussion of these two risks may be firms that face a lesser threat from their competitor's product innovations relative to the other risks they face. Firms with relatively unique products will have low Fluidity and may view the possible failure of their own innovations in their unique product line as risky and a potential cause of a need to restructure the business. The other difference between Panels A and B concerns *Volatile Net Cash Flow* risk. Panel A reports that firms with *High* sentence-counts associated with *Volatile Net Cash Flow* risk have higher values mean and median values of the historical volatility of net cash flows than do firms with *Low* sentence-counts, but Panel B reports an opposite result when the associated sentence-count is measured relative the total Item 1A sentence-count.

Panel C examines the relation between the risk-factor sentence-count measured relative to the average sentence-count for the firm's nonzero risk-factors and the proxy for the risk. A comparison of Panels C and B shows that measuring the risk-factor sentence-count relative to the total sentence-count or relative to the average sentence-count for the firm's nonzero risk-factors does not change the relation with the associated risk-proxy. The principal conclusion from Table III is that an LDA analysis of the Item 1A section of firms' 10-K Statements succeeds in identifying risk-factors that correspond to familiar risk proxies, in that higher sentence-counts and normalized sentence-counts for an LDA-identified risk-factor tend to be associated with higher values of proxy measures of the associated risk. The results using Fluidity as a risk-proxy are not as easy to interpret. It may be that low Fluidity is associated with high sentence-counts arising from risks associated with own-firm *Product Innovation* rather than rival-firm *Product Innovation*. Section C investigates this possibility.

B. The Informativeness of Risk-Factor Disclosure through Time

A general concern with disclosure is whether it has become less informative over time (see, e.g., Francis and Schipper (1999) and Beaver, McNichols, Rhie (2005)). Dyer, Lang and Stice-Lawrence (2017) show that the total 10-K length has increased over time and attribute the increase to in part, to new compliance requirements such as the mandatory disclosure of risks in Item 1A. We find that the number of sentences in Item 1A has increased from an average of 170 in 2005 to

288 in 2015, an increase of close to 70%.[19] Dyer, Lang and Stice-Lawrence (2017) further document that 10-K Statements have increased in their boilerplate nature, stickiness, and redundancy through time. In a discussion of Dyer, Lang and Stice-Lawrence (2017), Miller (2017) observes that even if annual reports have become longer, less readable, and more boilerplate over time, this does not mean they have become less informative. We investigate this question by examining changes through time in the correlation between risk- proxies and the associated risk-factor sentence-counts.

We introduce a novel and natural measure of the informativeness of the Item 1A risk disclosure, namely the fraction of the cross-firm variation in a risk-proxy measure explained by cross-firm variation in the associated risk-factor sentence-count. The higher the correlation between a risk-proxy and the associated risk-factor sentence-count, the more informative is the risk disclosure statement. A common and natural view of the information content of financial statements is the link between their release and abnormal trading volume and/or price movements; i.e., informativeness is defined in terms of new information (Hope, Hu and Lu (2016)). Our measure recognizes that financial statements can contain information useful to investors determining their optimal portfolios and to regulators without there necessarily being a link to abnormal volume or returns. Subsection *D.1* investigates the relation between contemporaneous changes in risk-factor sentence-counts and risk-proxies and subsection D.2 investigates the predictive relation between changes in risk-proxies and lagged changes in risk-factor sentence-counts.

Let S_{it}^{j} denote the year *t* sentence-count for risk-factor *j* for firm *i*. Let s_i^j denote the fundamental information in firm I's Item 1A disclosure concerning risk-factor *j*. For simplicity,

[19] The increase is not driven by changes in the composition of S&P1500 firms through time. For the subset of firms in the S&P1500 in 2005, the number of sentences (words) in Item 1A increased from 170 to 268 (from 4,221 to 7,130) in 2015.

the risk is assumed constant through time.[20] Suppose risk-factor reporting has become increasingly boilerplate and/or redundant through time and the annual sentence-count is time-dependent, with

$$S_{it}^{j} = \alpha(t) + \beta(t) s_i^j .$$

$\alpha(t)$ reflects the time-varying use of boilerplate language and $\beta(t)$ reflects the time-varying repetition of the same information across multiple sentences. Redundancy as measured by the number of words contained in sentences repeated verbatim is a clear example of the repetition of information.

Now consider the information content of Item 1A in this setting of increasing boilerplate language and increasing redundancy. Let X_i^j denote the proxy measure for risk-factor *j*. The year *t* cross-sectional correlation between the risk-factor and the risk-proxy is

$$\frac{\operatorname{cov}\left(S_{it}^{j}, X_i^j\right)}{\sigma\left(S_{it}^{j}\right)\sigma\left(X_i^j\right)} = \frac{\operatorname{cov}\left(\alpha(t)+\beta(t)s_i^j, X_i^j\right)}{\sigma\left(\alpha(t)+\beta(t)s_i^j\right)\sigma\left(X_i^j\right)} = \frac{\operatorname{cov}\left(s_i^j, X_i^j\right)}{\sigma\left(s_i^j\right)\sigma\left(X_i^j\right)} . \quad (1)$$

The correlation does not depend on *t*, despite the increasing use of boilerplate language and increasing redundancy. Thus, consistent with the Miller (2017), increasing verbosity and the increasing use of boilerplate language will not cause Item 1A risk disclosures to become less informative through time.

For each of the 20 (risk-factor, risk-proxy) pairings in Table III, Table IV reports the results of annual regressions of the risk proxies on the associated risk-factor sentence-count. Except for the years 2013 through 2015 when values of the asset beta risk-proxy are unavailable, there are 20 risk-proxy, sentence-count pairings per year for each year 2005 through 2015. Each group of three rows reports annual estimates of the sensitivity of a risk-proxy to its associated risk-factor sentence-count, the t-statistics and the values of the regression adjusted-R^2. Two hundred and three of the 217 annual pairings are positively related with 122 being significant at the 1% level. A

[20] 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.

further 18 of the positive relations achieve a 5% level of significance. Only one of the estimated negative relations is significant and that is at the 5% level.

The regression coefficients in Table IV are estimates of the sensitivity of the risk proxies to their associated risk-factor sentence-counts; i.e., of

$$\frac{\mathrm{cov}\left(S_{it}^{j}, X_{i}^{j}\right)}{\sigma^{2}\left(S_{it}^{j}\right)} = \frac{1}{\beta(t)} \frac{\mathrm{cov}\left(s_{i}^{j}, X_{i}^{j}\right)}{\sigma^{2}\left(s_{i}^{j}\right)}. \tag{2}$$

If verbosity/redundancy as captured by $\beta(t)$ has increased through time, then the expected values of the regression coefficients will have declined through time. If the informativeness of Item 1A risk disclosure has declined through time, then the R^2 values will have tended to decline through time. This is not what is seen for the 20 pairings reported in Table IV. The overall conclusion from Table IV is that strong positive sensitivities and adjusted R^2 values persist throughout the 11-year period.

To formally investigate the relations between the passage of time and both the informativeness of risk disclosure statements and the sensitivity of risk proxies to risk-factor sentence-counts we first calculate annual average measures of the regression R^2 values and the regression coefficient estimates. The third-last and the final row of Table IV report the annual simple-averages and weighted-averages of the R^2 values. The fourth-last and second-last row of Table IV report the annual simple-average and weighted-average of the regression coefficients. To calculate these averages, we first recognize that five of the 15 risk-factors are associated with more than one risk-proxy. For these five risk-factors we equal weight the R^2 values and regression coefficient estimates in each year. This yields annual R^2 and regression coefficient values for each of 15 risk-factors. Simple-averages are straightforward to calculate. A weighted-average is also calculated each year. The weighted-average uses as weights the risk-factor's total sentence-count across all firms that year relative to the total sentence-count of all risk-factors across all firms that year.

Table V reports the results of regressions of the annual average coefficient and annual average R^2 on time with time measured as the excess of the calendar year over 2004. The first and third columns of Table V report regressions of the simple- and weighted-average coefficients on time. An examination of the simple-average leads to the conclusion that there is no evidence to

suggest that there has been an increase in verbosity/redundancy. But the regression of the weighted-average coefficients on time, does allow one to reject a null of no increase in verbosity/redundancy between 2005 and 2015 at the 10% level. This quite weak evidence of increasing redundancy in the risk disclosure section of 10-K Statements is not inconsistent with the Dyer, Lang and Stice-Lawrence (2017) documentation of increasing redundancy in 10-K Statements as a whole. Panel A of Figure 5 of Dyer, Lang and Stice-Lawrence (2017) depicts, inter alia, redundancy within the risk-disclosure section itself and portrays an unchanging level of risk-disclosure redundancy between 1996 and 2013.

The second and fourth columns of Table V report regressions of the simple- and weighted-averages of our novel adjusted-R^2 measure of the informativeness of risk-disclosure statements on time. There is no evidence to suggest that the informativeness of risk-factor disclosures declined between 2005 and 2015. And this is so despite the Dyer, Lang and Stice-Lawrence (2017) documentation of increased boilerplate, stickiness, and redundancy of 10-K Statements as a whole between 1996 and 2013.[21]

C. Fluidity and LDA-identified Risk-Factors

As observed in subsections A and B, the two risk-factors *Growth and Restructuring* and *Production Innovation* do not have a consistently positive relation with Fluidity. Fluidity is a measure of the threat to a firm's profitability as a result of changes in its rival's products. In this subsection, we investigate these two relations in some depth.

A firm's Fluidity is a metric for the changes in rival firms' products determined from a textual analysis of Item 1 for the firm and its competitors. To the extent Fluidity provides a proxy for the competitive threats faced by a firm, the Risk Factors section of firms' 10-K Statements should include a discussion of fluidity-related risks. This though does not imply that there will be a monotonic relation between the Fluidity metric and the sentence-count quantification of either *Growth and Restructuring* risk or *Production Innovation*. For example, *Product Innovation* risk can be high for firms operating in highly fluid environments. *Product Innovation* risk could also

[21] This result is analogous to the Beaver, McNichols and Rhie (2005) finding that the ability of financial ratios to predict bankruptcy did not decline over the forty years from 1962 to 2002.

be high for low-fluidity firms that have few rivals and whose own product innovations could damage their monopolistic profits. And if such potential damage is perceived as a risk to growth or a cause of a need to restructure, then this low-fluidity firm may disclose a high level of *Growth and Restructuring* risk.

In Figure 1, we provide plots of Fluidity against the three *Growth and Restructuring* risk-factor quantifications used in Table III and plots of Fluidity against the three *Product Innovation* risk-factor quantifications used in Table III. Showing such plots using all sample observations would each involve more than 14,000 observations. To show the relation more clearly, the observations are replaced by 16 points (hexadeciles) in fluidity, risk-factor space. The first point corresponds to the set of firms with a zero value for their risk-factor quantification and the simple average of the fluidity of the firms in the set. The other 15 points are determined by first ordering the set of firms which have a nonzero quantification of the risk-factor in increasing size of the risk-factor quantification and forming equal-sized portfolios of firms with increasing quantifications. For portfolios $j = 2, \ldots, 16$, point j corresponds to the portfolio (average risk-factor quantification $_j$, average Fluidity $_j$) pairing. Figure 1 plots the 16 points corresponding to various pairings of average risk-factor quantification and average Fluidity. Panels A, B, and C consider the average risk-factor sentence-count, the average risk-factor sentence-count as a fraction of the firm total Item 1A sentence-count, and the average risk-factor sentence-count as a fraction of the average sentence-count for all the firm's nonzero risk-factors respectively. *Growth and Restructuring* risk is the risk-factor examined in the left-hand section of each panel. *Product Innovation* risk is examined in the right-hand section of each panel. The regression line that best fits the 16 points is displayed in each plot.

The relation between Fluidity and *Growth and Restructuring* risk depicted in the left-hand section of Figure 1 is interesting. The relation switches from positive to negative when we move from the risk-factor sentence-count in Panel A to either of the normalizations in Panels B and C. When the sentence-count for *Growth and Restructuring* risk is high, the product market threat from innovative rivals tends to be high. But firms for which *Growth and Restructuring* risk is high relative to all the risks the firms face tend to operate in environments characterised by low fluidity. In other words, firms that operate in highly-fluid environments tend to disclose a lot about the other risks they face. Once this additional risk reporting is taken into consideration, *Growth and*

Restructuring risk when measured relative to all risk-factors appears to be less important and firms with low relative *Growth and Restructuring* risk tend to have high Fluidity.

The relation between Fluidity and *Product Innovation* risk depicted in the right-hand section of Figure 1 is nonmonotonic. Recall from Table III that the average Fluidity of firms with below-median *Production Innovation* risk sentence-counts at 6.17 which is significantly higher than the 6.00 average Fluidity of firms with above-median sentence-counts. These relative values are consistent with a negative relation between Fluidity and *Product Innovation* risk. Table IV suggests though that the relation is positive. Firm-level Fluidity and *Production Innovation* risk is significantly positively related in each year from 2005 to 2015. An examination of Figure 1 reveals that the relation is actually U-shaped for all three risk-factor quantifications.

It is not surprising that there exist high Fluidity firms that disclose high levels of *Product Innovation* risk. Such firms should see their own and their competitors' innovations as risky. Nor is it surprising that the average *Product Innovation* risk of the portfolio at the turning point of the U-shaped relation is similar to the median level of *Product Innovation* risk. Given a turning point, the turning point portfolio will contain numerous firms with low Fluidity. These firms have few innovative competitors may not view their own product development as being of low risk. What is surprising about the U-shaped relation is that there are a sizeable number of high-fluidity firms that choose to disclose little or nothing about *Product Innovation* risk. Perhaps hubris is the explanation for why these firms appear to underappreciate the risks posed by innovative rivals and to overlook the possibility that their own product development might be as successful as the introduction of "New Coke".[22]

D. Changes in Risk-Factors and Changes in Risk-Proxies

Subsections A and B demonstrate that cross-sectional differences in risk-factor sentence-counts help explain cross-sectional differences in risk-proxies. This subsection examines the time series dimension of the relation by investigating the contemporaneous relation between changes in risk-factor sentence-counts and changes in the associated risk-proxies and whether changes in

[22] Understanding the incentives of product managers, developers and others when providing information to the preparer of a firm's 10K Statement is beyond the scope of this study.

risk-factor sentence-counts predict future changes in the associated risk-proxies. We also introduce an index of the *Financing* risk-factors to recognize the possibility that they capture similar risks.

D.1 Contemporaneous Changes in Risk-Factors and Changes in Risk-Proxies

For each (risk-factor, risk-proxy) pairing, panel regressions of annual changes in risk-proxies on the contemporaneous annual change in their associated risk-factor's sentence-count are reported in Table VI. Sixteen of the 20 contemporaneous relations are positive. Seven (six) of the 16 are significant at the 5% (1%) level. None of the five negative contemporaneous relations are significant.

If changes in risk-factor sentence-counts were uncorrelated with contemporaneous changes in risk-proxies and each of the 20 relations were independent, then the probability of observing 16 positive relations would be only 0.59%. But the pairings are not independent. Some risk-factors pair with more than one risk-proxy and some risk-proxies pair with more than risk-factor. Leverage and Credit Spreads are risk-proxies for each of *Financing I*, *II* and *III*. Changes in the three *Financing* risk-factors are positively related to both changes in Leverage and changes in Credit Spreads. The relations between changes in Leverage and changes in the *Financing II* and *III* risks are significant at the 1% level. The relations between changes in Credit Spreads and changes in the *Financing I* and *III* risks are significant at the 1% level. Interestingly, despite the significant relation between changes in the level of *Financing I* risk and changes in Credit Spreads, Table IV does not document a significant relation between the level of *Financing I* risk and the level of Credit Spreads. That each of the three financing risk-factors is related to both Leverage and Credit Spreads confirms our undifferentiated choice of labels for the financing risks. We therefore combine the financing risks into two index measures. The S*imple Fin Risk Index* is the mean of the sentence-counts of the three risk-factors. The W*eighted Fin Risk Index* is a weighted average with weights given by the fraction of the aggregate financing risk-factor sentence-count over all firm-years due to the particular financing risk. Changes in both *Fin Risk* indices are significantly positively related at the 1% level to contemporaneous changes in Leverage and in Credit Spreads.

Contemporaneous changes in risk-factors are significantly positively related to changes in risk-proxies for *Customer Concentration* risk, *Growth and Restructuring* risk, and *Product Innovation* risk. The relation between changes in *Customer Concentration* risk and changes in a

dummy for whether the firm has a customer that contributes at least 10% to the firm's sales is positive and significant at the 5% level. Changes in *Product Innovation* risk and in *Growth and Restructuring* risk are positively related to changes in Fluidity at the 10% and 1% levels respectively.

In the analysis of changes we use a different proxy for the *Volatile Stock Price* risk-factor than we used in our analysis of levels. In our analysis of levels we used the σ(Stock Return) measure which is annualized monthly volatility of stock returns over the 60 months ending at the fiscal year-end. The annual change in the σ(Stock Return) measure will be driven by the difference in volatilities between the current year and the volatility five years earlier. Therefore, to analyse the relation between annual changes in the *Volatile Stock Price* risk-factor sentence-count and annual changes in a proxy of stock return volatility, we use the annualized stock return volatility estimated using daily stock returns over the 250 trading days preceding the fiscal year-end and term this proxy σ(Equity). Changes in σ(Equity) from one fiscal year to the next are significantly positively related at the 1% level to the contemporaneous change in the *Volatile Stock Price* risk-factor. The significant positive contemporaneous co-movements of changes in risk-factors and risk-proxies coupled with the absence of significant negative contemporaneous co-movements suggest that innovations in risk disclosure statements can be timely reflections of changes in risk.

D.2 Changes in Risk-Proxies and Lagged Changes in Risk-Factors

Table VII reports the results of panel regressions of changes in risk-proxies on lagged changes in the associated risk-factor sentence-counts. These predictive results are mixed. Nine of the 20 predictive relations are positive. The relations between lagged changes in the *Volatile Net Cash Flow*, *Volatile Revenue*, and *Volatile Stock Price* risk-factor sentence-counts and changes in their associated risk-proxies are statistically significant at the 1%, 5%, and 5% levels respectively. The risk-proxy σ(NCF) is the annualized standard deviation of quarterly cash flows relative to quarter-end total assets over the 12 quarters preceding the fiscal year-end. The risk-proxy σ(Revenue) is defined analogously for revenues. The risk-proxy σ(Equity) is the annualized stock return volatility estimated from daily stock returns over the 250 trading days preceding the fiscal year-end.

Given that the σ(NCF), σ(Revenue) and σ(Equity) proxies are measured over an extended period that ends at the fiscal year-end, a positive relation between annual changes in these risk-proxies measures and lagged annual changes in their associated risk-factor sentence-counts does not necessarily imply that risk disclosure statements are forward-looking. To see this, assume that risk-factor disclosures are not forward-looking and are instead a report on the level of risk as at the end of the fiscal year. If risk follows a martingale and has increased during fiscal year *t* and the year *t* risk-proxy is measured over a period preceding the end of fiscal year *t*, then the expected value of the year $t + 1$ extended-period risk-proxy will be greater than the realized value of the year *t* extended-period risk-proxy. The annual change in the risk-factor sentence-count between fiscal year $t - 1$ and fiscal year *t* will reflect the increase in risk that occurred during the year. Thus, the $t - 1$ to *t* change in the sentence-count will predict the *t* to $t + 1$ change in the extended-period risk-proxy even though the Item 1A Risk Disclosure is not forward looking. We can investigate this possibility for the *Volatile Stock Price* risk-factor at least because for this risk-factor we have an alternate proxy that measures risk as at the end of the fiscal year.

We use the fiscal year-end values of Jaewon Choi's stacked EGARCH equity volatility estimates as a fiscal year-end proxy for *Volatile Stock Price* risk and regress annual changes in this risk-proxy on lagged annual changes in the *Volatile Stock Price* risk sentence-count. The estimated regression coefficient of 0.0013 has a *t*-statistic of only 0.95. As seen in Table VII, when annual changes in an equity risk-proxy estimated over the 250 days preceding the fiscal year-end are regressed on lagged annual changes in the *Volatile Stock Price* sentence-count, the estimated regression coefficient of 0.0035 is nearly three times as large and the associated *t*-statistic is 3.31. Comparing the insignificance of the estimated coefficient when changes in an end-of-fiscal year proxy are used to measure changes in equity risk with the significant estimated coefficient when instead changes in a risk proxy measured over an extended period preceding the end of the fiscal year are used, we can conclude that the apparent forward-looking nature of the *Volatile Stock Price* risk-factor in the Table VII analysis is specious.

The relations between lagged changes in the *Financing I* and *III* risk-factor sentence-counts and the current change in credit spreads are significantly negative at the 1% level. This negative relation may reflect a survivorship bias. As seen in Table VI, increases in the *Financing* risk-proxies accompany contemporaneous increases in credit spreads. If a firm successfully responds

to an increase in its *Financing* risk by restructuring its debt, then credit spreads should decline in the year after the initial increase. Any such restructuring does not appear to take the form of a reduction in leverage. The estimated relation between the lagged change in the *Simple Fin Risk Index* and the current change in leverage is insignificantly positive. The relation between the lagged change in the *Weighted Fin Risk Index* and the current change in leverage is also insignificantly positive. We conclude that there is little to suggest that Item 1A Risk Disclosure is forward-looking in that other than for the *Financing* risk and credit spread pairings, lagged changes in risk-factors do not reliably predict future changes in risk-proxies.

V. Asset and Equity Volatility and Operating and Financing Risk Indices

Asset volatility reflects a firm's operating risk. Leverage transforms a firm's asset volatility into the volatility of the firm's equity. Twenty-six of the LDA-identified risk-factors are operating risks. As shown in Tables III and IV, the three *Financing* risk-factors are significantly correlated with leverage. The remaining risk-factor of *Volatile Stock Price* risk is a direct metric of equity and as shown is Tables III and IV is significantly positive related with the σ(Stock Return) risk-proxy. To determine the relation of both asset volatility and equity volatility to operating risk, financing risk, and leverage, we form *Operating Risk* indices in a manner analogous to the construction of the *Fin Risk* indices in Section IV *D.2*. For each firm-year, *Simple* and *Weighted Operating Risk* indices are constructed as averages of the firm's sentence-counts associated with each operating risk-factor. The *Simple Operating Risk Index* is the mean of the operating risk-factor sentence-counts. The *Weighted Operating Risk Index* weights each operating risk-factor sentence-count by the fraction of the aggregate of the operating risk-factor sentence-counts over all firm-years that is due to the particular operating risk.

A. Asset Volatility and Risk Indices

The link between asset volatility and the individual operating risk-factors is complex. As Campbell and Vuolteenaho (2004) observe, asset volatility reflects both cash flow news and discount rate news. A high sentence-count for a risk-factor may be directly related to cash flow risk and to the duration of cash flows. Grundy and Verwijmeren (2019) document that the time until an investment produces positive net cash flows is negatively related to the tangibility of the investment and positively related to its R&D-like nature. Thus, firms with high levels of either

Human Capital or *Intellectual Property* risk will have more distant cash flows and their values will be more sensitive to changes in discount rates. Rather than examining individual risk-factor sentence-counts, we examine the relation between asset volatility and the *Operating Risk* indices. For fiscal-years 2005 through 2012, we measure asset volatility as the average over the 12 months preceding the fiscal year-end of the monthly asset volatility values available from Jaewon Choi's homepage.

Table VIII explores the relation between asset return volatility and indices of *Operating Risk* and *Financing Risk*. Panel A (B) reports panel regressions with (without) industry fixed effects. Calendar year fixed effects are included as a control for any increasing length of the Item 1A Risk Factor disclosure across years. Including a time fixed effect is analogous to replacing the firm-year risk index by the firm-year deviation from the year's mean level of the index. The left-hand (right-hand) sections of Panels A and B report results using *Simple* (*Weighted*) *Operating* and *Financing Risk* indices Columns 1 and 4 report that irrespective of whether industry fixed effects are included, both the *Operating Risk* indices are positively related to asset volatility at the 1% level. Panel B reports that without controlling for industry fixed effects, cross-sectional differences in firms' *Simple* and *Weighted Operating Risk* indices explain 9.7% and 14.4% respectively of cross-sectional differences in asset volatility. That the estimated sensitivity of asset volatility to operating risk is higher in Panel B than in A is natural since the Panel A regression subsumes industry differences in asset volatilities and the *Operating Risk* indices into an industry fixed effect.

The sentence-count-based *Operating Risk* indices will be noisy measures of operating risk. Thus, other measures correlated with asset volatility can also help explain asset volatility even after controlling for *Operating Risk*. Bartram, Brown, Waller (2015) and Choi and Richardson (2016) document a significant negative relation between a firm's asset volatility and the level of leverage chosen by the firm. Columns 2 and 5 of Panels A and B of Table VIII document that asset volatility is significantly positively related to the *Operating Risk* indices while being simultaneously significantly negatively related to *Financing Risk*. The *Financing Risk* indices are themselves noisy measures of leverage and columns 3 and 6 show that, controlling for *Financing Risk*, asset volatility is significantly positively related at the 1% level to the *Operating Risk* indices while being significantly negatively related to leverage at the 1% level. This corroborate the results

in Choi and Richardson (2016), i.e. firms with higher operational risks use less leverage and vice versa. The conclusion from Table VIII is that LDA-identified *Operating Risk* is significantly positively related to asset volatility even after controlling for *Financing Risk*, leverage, and industry fixed-effects.

B. Equity Volatility and Risk Indices

The seminal work of Modigliani and Miller (1958) demonstrates that a firm's equity risk is an increasing function of both its asset risk and leverage: $\sigma_E^2 = \eta^2 \sigma_A^2$, where σ_E^2 is the variance of the return on equity, σ_A^2 is the variance of the return on the firm, and $\eta \equiv \frac{\partial E}{\partial V}\frac{V}{E}$. V denotes the value of the firm and E denotes the value of its equity. Taking logs gives $ln\left(\sigma_E\right) = ln\left(\eta\right) + ln\left(\sigma_A\right)$. Approximating $\frac{\partial E}{\partial V}$ by unity, and hence η by $\frac{V}{E}$, Choi and Richardson (2016) investigate the following regression across firms and time.

$$ln\left(\text{estimate } \sigma_{E\ i,t}\right) = \beta_0 + \beta_1 ln\left(\frac{V}{E}_{i,t-1}\right) + \beta_2 ln\left(\text{estimate } \sigma_{A\ i,t}\right) + \varepsilon_{i,t}. \qquad (3)$$

Choi and Richardson (2016) are careful to avoid the bias that would arise from common measurement error in the explanatory and dependent variables. Option-implied volatilities are used to estimate $\sigma_{E\,i,t}$. A stacked EGARCH estimate of the volatility of a value-weighted average of a firm's debt and equity returns is used to estimate $\sigma_{A\,i,t}$. Measurement error in the two volatility estimates will be independent. Rather than an EGARCH estimate of $\sigma_{A\,i,t}$, we use the sentence-count-based *Operating Risk Index* to estimate asset volatility.[23] Measurement error in our LDA-based measure of asset risk will be independent of measurement error in a stacked EGARCH estimate of equity volatility. We use the stacked EGARCH estimates of equity volatility on Jaewon Choi's website as estimates of $\sigma\left(equity_{it}\right)$. Hence, our investigation of regression relation (4) below also avoids common measurement errors in the explanatory and dependent variables.

[23] As seen in Table VIII, the *Operating Risk Index* is significantly positively correlated at the 1% level with the Choi-Richardson EGARCH estimate of σ_A.

$$ln\left(\text{estimate } \sigma\left(equity\right)_{i,t}\right) = \alpha_0 + \alpha_1 ln\left(\text{Market Lev}_{i,t-1}\right) + \alpha_2 ln\left(Operating\ Risk\ Index_{i,t}\right) + \omega_{i,t}, \quad (4)$$

where beginning of fiscal-year market leverage is used as a measure of $\frac{V}{E}_{t-1}$.

Table IX reports that irrespective of whether a simple- or weighted-average *Operating Risk Index* is considered and whether industry fixed-effects are included, the estimated coefficient on *ln*(Market Lev) is positive and significantly so at the 1% level when industry fixed-effects are included in regression equation (4). Further, in all cases the estimated coefficient on the *ln*(*Operating Risk Index*) measure is significantly positive at the 1% level. The contribution of the analysis in this subsection is not simply a further demonstration that equity volatility is increasing in both leverage and the risk of the underlying assets. Rather, the analysis reported in Table IX demonstrates that an LDA-identified index of a firm's operating risks gives a credible measure of asset volatility in that it can be used to explain equity volatility.

VI. Conclusion

Determining the accuracy of the textual risk disclosures of a large set of firms via a subjective classification of firms' Item 1A Risk Factor disclosures is a mammoth task and difficult to corroborate. We use a machine-learning algorithm to identify and extract the 30 most important risks disclosed in Item 1A of the 10-K Statements of S&P 1500 firms between 2005 and 2015. Risk-factors are quantified for each firm-year by a count of the Item 1A sentences uniquely assigned to each risk. We measure the accuracy of S&P 1500 risk disclosures by determining the fraction of the cross-firm variation in proxy measures of the identified risks that is explained by cross-firm variation in the associated risk-factor sentence-counts. We determine at least one measurable proxy for 15 of LDA-identified risk-factors and investigate 20 risk-factor, risk-proxy pairings through time. Annual values of proxies and their associated risk-factors are positively correlated for the vast majority of risk-factor, risk-proxy pairings. The correlation is positive in 203 of 217 cases. The positive correlation is significant for more than half the pairings. The correlation is significant at the 5% (1%) level for 140 (122) of the pairings.

The average fraction of the cross-firm variation in the risk-proxy measures explained by cross-firm variation in the associated risk-factor sentence-counts has not declined between 2005

when risk-factor reporting became mandatory and the 2015 end of our sample period. This is so despite the documentation of a 1996 through 2003 increase in the boilerplate content, stickiness and redundancy of 10-K statements (Dyer, Lang and Stice-Lawrence (2017)). Risk-disclosures can become more similar without their differences becoming less salient.

Not only are risk-factors and risk-proxies strongly positively correlated in levels, their contemporaneous annual changes are positively related. A panel regression of contemporaneous changes in the values of the 20 risk-factor, risk-proxy pairings shows a positive relation for 16 of the pairings. For six (seven) pairings the positive relation is significant at the 1% (5%) level. None of the negative relations are significant. We also examine whether lagged changes in risk-factors predict subsequent changes in risk-proxies and conclude that there is little to suggest that Item 1A risk disclosures are forward-looking. Other than a negative relation between lagged changes in *Financing* risk sentence-counts and subsequent changes in credit spreads, lagged changes in risk-factors do not reliably predict future changes in risk-proxies. Twenty-six of the 30 risk-factors relate to operating risk. Three relate to financing risk. We combine the 26 and the three to create *Operating Risk* and *Financing Risk* indices. Consistent with Choi and Richardson (2016), firms with higher operational risks as measured by our LDA-identified *Operating Risk* index use less leverage and vice versa. Further, our *Operating Risk* index provides a useful measure of asset volatility in that along with cross-firm differences in leverage, cross-firm differences in the index help explain cross-firm differences equity volatility.

Many potential extensions can be explored by expanding the set of risk proxies considered. For example, one could examine how well LDA-identified *Corporate Governance* risk correlates with director age and experience and how well *Human Capital* risk correlates with employee turnover. While the regression relations between risk-proxies and risk-factors are overwhelmingly positive, the set of negative estimated relations warrants further study as do the relations that are non-monotonic. One interesting non-monotonic relation is that between *Product Innovation* risk and fluidity. Firms with either zero or high sentence-counts for *Product Innovation* risk tend to have high fluidity. Firms with median sentence-counts tend to have lower levels of fluidity. This U-shaped pattern may be driven by some firms reporting own-firm innovation risk and other firms reporting on competitor innovation risk. Understanding the incentives facing those who prepare financial statements may be a prerequisite to understanding their contents.

VII. References

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, 2019, A text-based analysis of corporate innovation, SSRN: http://dx.doi.org/10.2139/ssrn.2803232

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.

Brander, James A. and Tracy R. Lewis, 1986, Oligopoly and financial structure: The limited liability effect, *American Economic Review* 76(5), 956-970.

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.

Campbell, John Y. and Tuomo Vuolteenaho, 2004, Bad beta, good beta, *American Economic Review* 94(5), 1249-1275.

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, L., Malloy, C.J., Nguyen, Q.H., 2018. Lazy prices. Working paper.

Cremers, K.J. Martijn, Lubomir P. Litov and Simone M. Sepe, 2017, Staggered boards and firm value, revisited, *Journal of Financial Economics*, 126(2), 422–444.

Dhaliwal, Dan S., J. Scott Judd, Matthew Serfling and Sarah Shaikh, 2016, Customer concentration risk and the cost of equity capital, *Journal of Accounting and Economics* 61(1), 23–48.

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.

Fama, Eugene F. and James D. MacBeth, 1973. Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81(3), 607–636.

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

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.

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

Grundy, Bruce D. and Patrick Verwijmeren, 2019, The external financing of investment, SSRN: http://dx.doi.org/10.2139/ssrn.2986127

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.

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

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

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

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.

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.

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, 2019, Information revelation through regulatory process: Interactions between the SEC and companies ahead of the IPO, *Review of Financial Studies* forthcoming.

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.

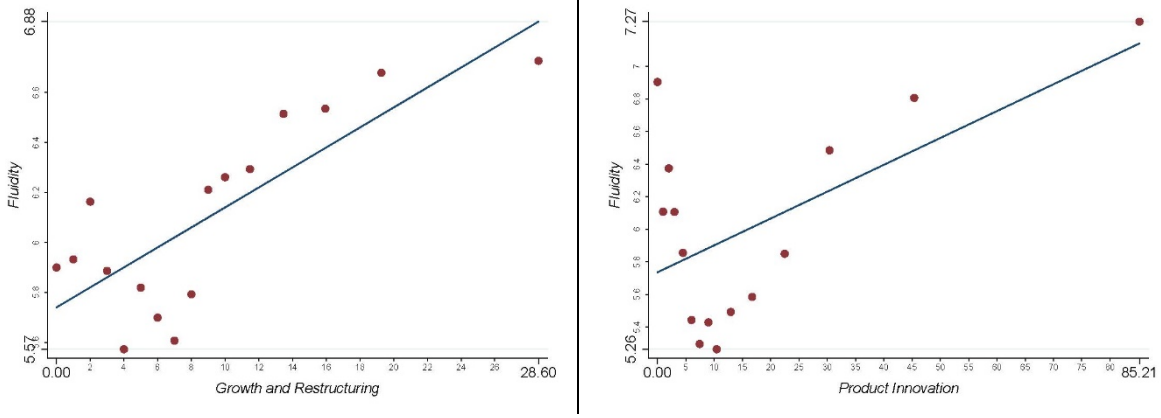
Modigliani, Franco and Merton H. Miller, 1958, The cost of capital, corporation finance and the theory of investment, *American Economic Review* 48(3), 261–297.

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

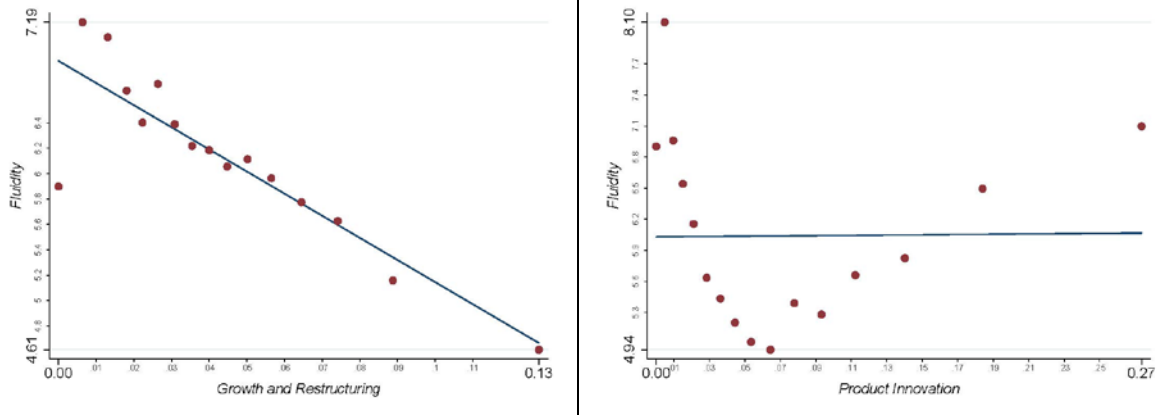
Figure 1: Average Fluidity per Risk-Factor Hexadecile

The Figure consists of 3 Panels each containing two plots. Each panel uses a different quantification of the risk-factor sentence-counts. The left plot shows the relation between *Growth and Restructuring* risk and Fluidity. The right plot shows the relation between *Product Innovation* risk and Fluidity. The X-axis shows the average of the risk-factor quantification of firms in each hexadecile of the quantification. The first hexadecile contains all firms with a zero value for the risk-factor. The Y-axis shows the average Fluidity of the firms in a hexadecile.

Panel A: Risk-factor sentence-count.



Panel B: Risk-factor sentence-count as a fraction of the total Item 1A sentence-count.



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

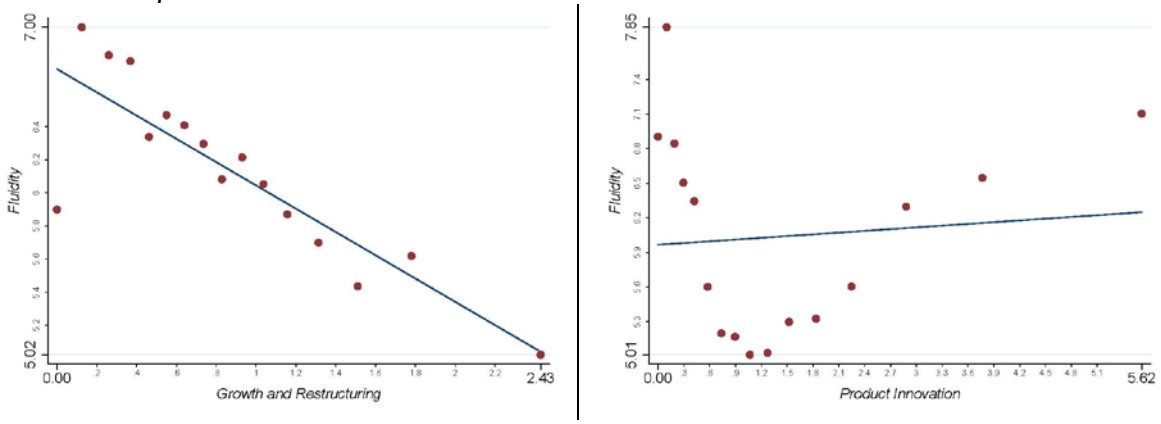


Table I: Descriptive Statistics

Sample summary statistics. Variable definitions are contained in Appendix Table A2. Total assets are in million USD. Book Lev. is the book value of long- plus short-term debt over book value of total assets. HHI is the Herfindahl-Hirschman Index. Cust. Dummy is a dummy variable equal to one if a single corporate customer is responsible for at least 10% of firm sales. Fluidity is the measure of the intensity of product market change due to Hoberg, Phillips and Prabhala (2014). (TA–Net PP&E)/TA is a measure of *Asset Intangibility* calculated as 1–Net property plant and equipment / total assets. R&D Expense/TA is a proxy for research and development intensity. Asset Beta is the Choi and Richardson (2016) 12 months average of monthly asset volatilities preceding fiscal year-end. 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). σ(Asset Return) is the 12-month average of 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(Stock Return) is the annualized stock return volatility estimated using monthly stock returns over the 60 months preceding the fiscal year end.

	N	Mean	p25	Median	p75	Std Dev	min	max
Total Assets	13,581	8,268.3	587.0	1,667.3	5,319.5	30,374	10.012	797,769
Book Lev.	13,581	0.203	0.037	0.191	0.319	0.171	0	0.852
HHI	13,581	0.085	0.039	0.053	0.1	0.091	0.017	1
RD Expense/TA	13,581	0.031	0.000	0.001	0.041	0.058	0.000	0.887
(TA–Net PP&E)/TA	13,575	0.743	0.634	0.822	0.916	0.226	0.017	1
Fluidity	13,388	6.089	3.825	5.478	7.638	3.146	0.416	24.493
Cust. Dummy	10,545	0.498	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
Credit Spread	4,583	2.959	1.274	2.206	4.054	2.238	0.033	9.995
σ(Asset Return)	5,171	0.322	0.210	0.291	0.409	0.158	0.060	1.670
σ(Cost)	13,153	0.061	0.022	0.041	0.078	0.070	0.000	3.083
σ(NCF)	11,589	0.037	0.011	0.019	0.037	0.075	0.000	4.187
σ(Revenue)	13,155	0.087	0.038	0.064	0.11	0.082	0.002	3.072
σ(Stock Return)	13,479	0.403	0.275	0.369	0.485	0.189	0.026	2.712

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	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
Volatile Costs	3.56	0	0	3	11.19	0	872
Volatile Net Cash Flows	11.75	5	10	16	8.36	0	76
Volatile Revenue	14.02	5	11	20	12.13	0	92
Volatile Stock Price	4.77	0	2	7	6.51	0	60

Table III: Means and Medians of Risk Proxies of Firms with *High* vs. *Low* Levels of LDA-identified Risks

The Table examines whether the mean and median values of risk proxies differ between firms with *High* vs. *Low* quantifications of the associated risk-factor. The risk factor quantification is labeled *High* (*Low*) if the quantification is above (below) its sample median value, or above (equal to) zero if the median quantified value is zero. The two right-most columns report the difference in the means and medians of the sets of firms with *High* vs. *Low* quantifications of the LDA risk factor. Significance levels are based on two-tailed, two-sample *t*-tests of the difference in means and Wilcoxon-Mann-Whitney tests of the difference in medians. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

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

LDA-identified Risk Factor	Risk Proxy	Risk Proxy High LDA Risk-Factor quantification (1)			Risky Proxy Low LDA Risk-Factor quantification (2)			Difference (1) − (2)	
		# Obs	Mean	Median	# Obs	Mean	Median	Mean	Median
Customer Concentration	Cust. Dummy	4,811	0.61	1.00	5,640	0.41	0.00	0.20***	1.00***
Economic Conditions	Asset Beta	2,245	1.17	1.03	2,908	1.01	0.86	0.16***	0.17***
Growth and Restructuring	Fluidity	6,197	6.39	5.86	7,084	5.82	5.17	0.57***	0.69***
	–HHI	6,280	–0.08	–0.05	7,190	–0.09	–0.06	0.02***	0.01***
Financing I	Book Lev.	5,287	0.22	0.21	8,183	0.19	0.18	0.03***	0.03***
	Credit Spread	2,086	3.18	2.40	2,427	2.80	2.11	0.38***	0.29***
Financing II	Book Lev.	6,654	0.26	0.26	6,816	0.15	0.11	0.12***	0.15***
	Credit Spread	2,397	3.67	3.13	2,116	2.19	1.60	1.48***	1.53***
Financing III	Book Lev.	6,610	0.26	0.26	6,860	0.15	0.12	0.12***	0.14***
	Credit Spread	2,612	3.54	2.91	1,901	2.21	1.61	1.33***	1.30***
Human Capital	(TA–Net PP&E)/TA	6,002	0.78	0.86	7,462	0.71	0.79	0.07***	0.07***
Intangible Assets	(TA–Net PP&E)/TA	6,164	0.77	0.86	7,300	0.72	0.78	0.06***	0.08***
Intellectual Property	R&D Exp/TA	6,465	0.06	0.04	7,005	0.01	0.00	0.05***	0.04***
Product Approval	R&D Exp/TA	4,289	0.04	0.02	9,181	0.03	0.00	0.02***	0.02***
Product Innovation	Fluidity	6,590	6.00	5.62	6,691	6.17	5.30	–0.17***	0.32***
	R&D Exp/TA	6,681	0.05	0.03	6,789	0.01	0.00	0.04***	0.03***
Volatile Costs	σ(Cost)	6,365	0.05	0.03	6,674	0.05	0.03	0.00***	0.00***
Volatile Net Cash Flow	σ(NCF)	5,629	0.04	0.02	5,874	0.03	0.02	0.01***	0.00***
Volatile Revenue	σ(Revenue)	6,416	0.08	0.06	6,625	0.06	0.04	0.02***	0.02***
Volatile Stock Price	σ(Stock Return)	6,216	0.46	0.42	7,154	0.36	0.32	0.11***	0.10***

Table III (cont'd)

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

LDA-identified Risk Factor	Risk Proxy	High LDA Risk-Factor (1)			Low LDA Risk-Factor (2)			Difference (1) – (2)	
		Obs	Mean	Median	Obs	Mean	Median	Mean	Median
Customer Concentration	Cust. Dummy	4,934	0.57	1.00	4,821	0.41	0.00	0.16***	1.00***
Economic Conditions	Asset Beta	2,343	1.12	1.01	2,371	1.02	0.85	0.10***	0.15***
Growth and Restructuring	–HHI	6,258	-0.08	-0.05	6,268	–0.09	–0.06	0.00***	0.01***
	Fluidity	6,166	5.73	5.26	6,185	6.50	5.81	–0.77***	–0.55***
Financing I	Book Lev.	4,962	0.22	0.21	7,564	0.19	0.18	0.03***	0.03***
	Credit Spread	1,959	3.15	2.37	2,245	2.79	2.10	0.36***	0.28***
Financing II	Book Lev.	6,257	0.27	0.26	6,269	0.14	0.11	0.12***	0.15***
	Credit Spread	2,285	3.61	3.04	1,919	2.17	1.61	1.44***	1.43***
Financing III	Book Lev.	6,261	0.27	0.27	6,265	0.14	0.11	0.13***	0.16***
	Credit Spread	2,612	3.42	2.76	1,592	2.18	1.62	1.25***	1.14***
Human Capital	(TA–Net PP&E)/TA	6,261	0.77	0.84	6,259	0.72	0.80	0.06***	0.04***
Intangible Assets	(TA–Net PP&E)/TA	6,261	0.77	0.86	6,259	0.72	0.79	0.05***	0.07***
Intellectual Property	R&D Exp/TA	6,263	0.06	0.03	6,263	0.01	0.00	0.05***	0.03***
Product Approval	R&D Exp/TA	4,010	0.04	0.02	8,516	0.03	0.00	0.02***	0.02***
Product Innovation	Fluidity	6,167	5.66	5.35	6,184	6.56	5.72	–0.90***	–0.37***
	R&D Exp/TA	6,262	0.05	0.02	6,264	0.02	0.00	0.03***	0.02***
Volatile Costs	σ(Cost)	5,937	0.05	0.03	6,168	0.05	0.03	0.00***	0.00***
Volatile Net Cash Flow	σ(NCF)	5,375	0.03	0.02	5,303	0.04	0.02	–0.00***	–0.00***
Volatile Revenue	σ(Revenue)	6,053	0.08	0.06	6,054	0.06	0.04	0.02***	0.02***
Volatile Stock Price	σ(Stock Return)	6,184	0.45	0.42	6,244	0.35	0.32	0.10***	0.10***

Table III (cont'd)

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

LDA-identified Risk Factor	Risk Proxy	High LDA Risk-Factor (1)			Low LDA Risk-Factor (2)			Difference (1) – (2)	
		Obs	Mean	Median	Obs	Mean	Median	Mean	Median
Customer Concentration	Cust. Dummy	5,010	0.58	1.00	4,745	0.41	0.00	0.17***	1.00***
Economic Conditions	Asset Beta	2,272	1.12	1.01	2,442	1.02	0.85	0.10***	0.16***
Growth and Restructuring	–HHI	6,263	–0.08	–0.05	6,263	–0.09	–0.06	0.01***	0.01**
	Fluidity	6,174	5.79	5.32	6,177	6.43	5.74	–0.64***	–0.41***
Financing I	Book Lev.	4,962	0.22	0.21	7,564	0.19	0.18	0.03***	0.03***
	Credit Spread	1,959	3.15	2.37	2,245	2.79	2.10	0.36***	0.28***
Financing II	Book Lev.	6,262	0.26	0.26	6,264	0.15	0.11	0.12***	0.15***
	Credit Spread	2,272	3.61	3.04	1,932	2.18	1.61	1.43***	1.43***
Financing III	Book Lev.	6,262	0.27	0.27	6,264	0.14	0.11	0.13***	0.15***
	Credit Spread	2,580	3.44	2.79	1,624	2.18	1.62	1.27***	1.17***
Human Capital	(TA–Net PP&E)/TA	6,259	0.77	0.85	6,261	0.71	0.80	0.06***	0.05***
Intangible Assets	(TA–Net PP&E)/TA	6,262	0.77	0.86	6,258	0.72	0.79	0.06***	0.07***
Intellectual Property	R&D Exp/TA	6,263	0.06	0.03	6,263	0.01	0.00	0.05***	0.03***
Product Approval	R&D Exp/TA	4,010	0.04	0.02	8,516	0.03	0.00	0.02***	0.02***
Product Innovation	Fluidity	6,172	5.72	5.41	6,179	6.50	5.65	–0.78***	–0.23***
	R&D Exp/TA	6,262	0.05	0.02	6,264	0.01	0.00	0.03***	0.02***
Volatile Costs	σ(Cost)	5,937	0.05	0.03	6,168	0.05	0.03	0.00***	0.00***
Volatile Net Cash Flow	σ(NCF)	5,426	0.04	0.02	5,252	0.04	0.02	–0.00*	–0.00**
Volatile Revenue	σ(Revenue)	6,056	0.08	0.06	6,051	0.06	0.04	0.02***	0.02***
Volatile Stock Price	σ(Stock Return)	6,195	0.45	0.42	6,233	0.35	0.32	0.10***	0.10***

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

This Table reports the annual regression of the risk proxies on the corresponding LDA-identified risk-factor sentence counts for the years 2005—2015. Except for the regressions involving the HHI risk proxy, all regressions include industry fixed effects. Robust *t*-statistics based on standard errors clustered at the SIC 2-digit level are shown in parentheses and adjusted R^2 statistics are shown in brackets. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. The weights used to calculate the annual weighted-aver coefficient and weighted-aver adjusted R^2 are the yearly total number of sentences across all firms that are devoted to the risk factor relative to the total the number of sentences in Item 1A across all firms that year.

LDA-identified Risk-Factor	Risk Proxy	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Customer Concentration	Cust. Dummy	0.00088 (0.57) [0.212]	0.00600*** (2.86) [0.167]	0.00650*** (4.53) [0.187]	0.00313 (1.52) [0.189]	0.00629*** (3.93) [0.195]	0.00535*** (3.16) [0.198]	0.00433** (2.32) [0.165]	0.00385* (1.87) [0.202]	0.00528*** (3.10) [0.211]	0.00306 (1.45) [0.195]	0.00691*** (3.01) [0.189]
Economic Conditions	Asset Beta	0.02820** (2.07) [0.182]	0.00476 (0.52) [0.161]	0.01270 (1.17) [0.037]	0.00837* (1.86) [0.163]	0.00349 (0.64) [0.163]	–0.00234 (-0.58) [0.239]	0.00603 (1.07) [0.243]	–0.01440** (–2.46) [0.080]			
Growth & Restructuring	–HHI	0.00101 (1.52) [0.005]	0.00110* (1.78) [0.007]	0.00076 (1.06) [0.003]	0.00100 (1.31) [0.005]	0.00099 (1.44) [0.006]	0.00138** (2.18) [0.011]	0.00137** (2.43) [0.010]	0.00108* (1.94) [0.007]	0.00093* (1.73) [0.006]	0.00097** (2.25) [0.007]	0.00063 (1.57) [0.003]
Growth & Restructuring	Fluidity	0.121*** (5.68) [0.355]	0.0551*** (7.57) [0.309]	0.0748*** (7.76) [0.325]	0.0756*** (7.67) [0.316]	0.0542*** (4.99) [0.324]	0.0514*** (6.89) [0.329]	0.0478*** (5.93) [0.313]	0.0552*** (4.13) [0.372]	0.0343*** (4.15) [0.363]	0.0274*** (4.22) [0.231]	0.0298** (2.41) [0.370]
Financing I	Book Lev.	0.00402*** (4.93) [0.231]	0.00288*** (3.91) [0.225]	–0.00097 (-0.79) [0.211]	0.00159 (1.59) [0.235]	0.00222** (2.09) [0.229]	0.00417*** (3.00) [0.228]	0.00401*** (3.24) [0.213]	0.00484*** (6.27) [0.217]	0.00507*** (6.28) [0.194]	0.00349*** (3.59) [0.173]	0.00533*** (5.39) [0.163]
Financing I	Credit Spread	–0.00344 (-0.12) [0.064]	–0.01740 (-0.69) [0.123]	0.03900 (1.27) [0.196]	0.00734 (0.22) [0.149]	0.03010 (1.16) [0.157]	0.02090 (0.69) [0.161]	0.02920 (1.09) [0.193]	0.02560 (0.86) [0.143]	–0.00652 (-0.43) [0.123]	0.00658 (0.27) [0.198]	0.05120** (2.21) [0.312]
Financing II	Book Lev.	0.00938*** (9.97) [0.316]	0.00941*** (8.64) [0.310]	0.00351* (1.83) [0.255]	0.00845*** (6.90) [0.331]	0.00808*** (9.93) [0.328]	0.00812*** (10.37) [0.328]	0.00809*** (9.93) [0.311]	0.00797*** (9.06) [0.317]	0.00539*** (3.13) [0.257]	0.00753*** (10.35) [0.271]	0.00680*** (8.13) [0.231]
Financing II	Credit Spread	0.06510*** (2.72) [0.156]	0.06410*** (3.68) [0.233]	0.03130 (1.52) [0.250]	0.14300*** (6.38) [0.300]	0.11700*** (4.71) [0.284]	0.10700*** (6.89) [0.306]	0.15200*** (7.16) [0.427]	0.11500*** (5.24) [0.318]	0.03780 (1.25) [0.187]	0.10000*** (4.47) [0.381]	0.09520*** (4.11) [0.432]
Financing III	Book Lev.	0.01030*** (9.86) [0.326]	0.00975*** (9.64) [0.313]	0.01020*** (6.87) [0.308]	0.00907*** (9.09) [0.339]	0.00837*** (9.93) [0.327]	0.00790*** (9.36) [0.311]	0.00780*** (9.52) [0.297]	0.00781*** (10.46) [0.300]	0.00757*** (9.33) [0.279]	0.00732*** (8.26) [0.264]	0.00697*** (6.92) [0.240]
Financing III	Credit Spread	0.06870*** (2.71) [0.133]	0.07060*** (4.33) [0.235]	0.11800*** (4.27) [0.336]	0.12100*** (5.72) [0.262]	0.10700*** (5.31) [0.281]	0.10000*** (6.86) [0.288]	0.13400*** (7.26) [0.379]	0.11400*** (7.28) [0.320]	0.08270*** (4.90) [0.289]	0.09050*** (3.99) [0.360]	0.09010*** (6.28) [0.423]

Table IV (cont'd)

Human Capital	(TA–Net PP&E)/TA	0.00201*** (3.05) [0.620]	0.00128* (1.86) [0.622]	0.00076 (0.91) [0.644]	0.00096 (1.17) [0.639]	0.00060 (0.79) [0.644]	0.00040 (0.43) [0.671]	0.00040 (0.49) [0.638]	0.00021 (0.31) [0.643]	0.00046 (0.57) [0.647]	0.00089 (1.10) [0.653]	0.00188** (2.16) [0.665]
Intangible Assets	(TA–Net PP&E)/TA	0.00097 (0.74) [0.616]	0.00188 (1.62) [0.623]	0.00014 (0.43) [0.644]	0.00208*** (2.75) [0.642]	0.00111 (1.34) [0.645]	0.00115** (2.00) [0.671]	0.00085 (1.06) [0.638]	0.00112* (2.00) [0.645]	0.00042 (0.46) [0.647]	0.00206** (2.04) [0.656]	0.00136* (1.98) [0.666]
Intellectual Property	R&D Exp/TA	0.00257*** (16.09) [0.443]	0.00253*** (7.27) [0.440]	0.00261*** (6.31) [0.455]	0.00276*** (6.45) [0.406]	0.00259*** (7.38) [0.406]	0.00229*** (7.24) [0.452]	0.00220*** (7.96) [0.399]	0.00215*** (7.03) [0.370]	0.00199*** (7.46) [0.342]	0.00212*** (5.75) [0.360]	0.00184*** (5.75) [0.358]
Product Approval	R&D Exp/TA	0.00033*** (20.29) [0.327]	0.00044*** (16.53) [0.351]	0.00048*** (11.98) [0.362]	0.00039*** (5.67) [0.305]	0.00051*** (7.13) [0.325]	0.00047*** (7.41) [0.385]	0.00047*** (7.78) [0.338]	0.00047*** (4.89) [0.313]	0.00049*** (4.89) [0.289]	0.00046*** (5.20) [0.285]	0.00043*** (4.23) [0.294]
Product Innovation	Fluidity	0.02260** (2.58) [0.344]	0.0322*** (4.64) [0.332]	0.0222*** (3.32) [0.321]	0.02680*** (6.07) [0.317]	0.0188*** (2.98) [0.323]	0.02110*** (2.91) [0.333]	0.02470*** (4.52) [0.320]	0.02060*** (5.45) [0.369]	0.01820*** (5.66) [0.368]	0.01350*** (4.32) [0.232]	0.01150*** (2.72) [0.366]
Product Innovation	R&D Exp/TA D	0.00035** (2.12) [0.308]	0.00081*** (4.67) [0.351]	0.00082*** (6.19) [0.358]	0.00100*** (4.70) [0.332]	0.00100*** (6.27) [0.359]	0.00087*** (5.20) [0.396]	0.00105*** (4.28) [0.373]	0.00079*** (3.05) [0.325]	0.00098*** (3.11) [0.332]	0.00109*** (3.08) [0.344]	0.00086*** (5.79) [0.308]
Volatile Costs	σ(Cost)	–0.00007 (-0.30) [0.173]	–0.00010 (-0.35) [0.218]	–0.00002 (-0.27) [0.219]	–0.00012 (-0.54) [0.236]	–0.00027 (-1.20) [0.220]	–0.00029 (-0.91) [0.189]	–0.00026 (-0.93) [0.183]	–0.00015 (-0.66) [0.140]	0.00010 (0.37) [0.125]	0.00018 (0.57) [0.142]	0.00003 (0.16) [0.255]
Volatile Net Cash Flow	σ(NCF)	0.00056** (2.07) [0.019]	0.00043** (2.34) [0.022]	0.00292 (1.06) [0.013]	0.00103 (0.85) [0.002]	–0.00004 (-0.11) [0.019]	–0.00003 (-0.13) [0.031]	0.00012 (0.52) [-0.004]	0.00020 (1.11) [-0.009]	0.00045** (2.11) [0.017]	0.00026 (1.34) [0.026]	0.00022 (0.96) [0.052]
Volatile Revenue	σ(Revenue)	0.00058** (2.45) [0.212]	0.00071*** (2.93) [0.223]	0.00056*** (2.83) [0.204]	0.00056** (2.59) [0.219]	0.00042** (2.48) [0.217]	0.00033* (1.78) [0.200]	0.00037 (1.51) [0.189]	0.00030 (1.40) [0.133]	0.00035* (1.90) [0.132]	0.00036* (1.73) [0.156]	0.00029 (1.05) [0.255]
Volatile Stock Price	σ(Stock Return)	0.01450*** (11.04) [0.281]	0.0117*** (8.17) [0.232]	0.00913*** (13.30) [0.237]	0.00974*** (7.20) [0.249]	0.00909*** (6.47) [0.192]	0.00667*** (5.82) [0.181]	0.00599*** (4.30) [0.194]	0.00615*** (5.03) [0.172]	0.00547*** (7.10) [0.183]	0.00538*** (7.35) [0.249]	0.00569*** (9.28) [0.290]
Simple-aver coefficient		0.01333	0.00960	0.01238	0.01507	0.01317	0.01168	0.01503	0.01173	0.00773	0.01030	0.01199
Simple-aver adjusted R^2		0.28032	0.28520	0.28555	0.28940	0.28905	0.30418	0.29346	0.26886	0.27073	0.28224	0.31767
Weighted-aver coefficient		0.01129	0.00912	0.01036	0.01253	0.01032	0.00936	0.01191	0.00939	0.00747	0.00872	0.00947
Weighted-aver adjusted R^2		0.28817	0.28739	0.28718	0.27920	0.28066	0.29484	0.27751	0.25533	0.25977	0.25966	0.29627

Table V: Verbosity and Informativeness of Risk Disclosure through Time

This Table presents the results of cross-sectional regressions of the annual values of the simple-average and weighted-average verbosity-related coefficient measure and informative measure given by the adjusted-R^2 values reported in Table IV on time measured as the excess of the calendar year over 2004. The averages are calculated over the 15 unique risk factors of Table IV. *t*-statistics based on Huber/White robust standard errors are shown in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Simple Average		Weighted Average	
	Coefficient	Adjusted R^2	Coefficient	Adjusted R^2
Constant	0.0132***	0.2828***	0.0113***	0.2898***
	(9.97)	(47.26)	(14.17)	(47.64)
Time	–0.0002	0.0009	–0.0002*	–0.0019
	(–1.07)	(0.56)	(–2.03)	(–1.20)
# Obs	11	11	11	11
Adjusted R^2	–0.005	0.065	0.157	0.090

Table VI: Contemporaneous Changes in Risk-Proxies and Risk-Factors

The Table reports the results of panel regressions of changes in risk-proxies on contemporaneous changes in associated risk-factor sentence-counts. The *Simple Fin Risk Index* is the mean of the *Financing I, II* and *III* risk sentence-counts. The *Weighted Fin Risk Index* weights each *Financing* risk sentence-count by the fraction of the total sentence-count across all three *Financing* risk-factors and all firm-years that is due to the *Financing* risk. A constant is included but not reported. *t*-statistics are based on robust standard errors. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Regression of change in risk-proxy on contemporaneous change in risk-factor sentence-count		
Risk-Factor	Risk-Proxy	Coefficient	*t*-stat
Customer Concentration	Cust. Dummy	0.00074**	2.31
Economic Conditions	Asset Beta	0.00022	0.13
Growth & Restructuring	–HHI	–0.00012	–0.54
	Fluidity	0.00738*	1.81
Financing I	Book Lev.	0.00061	1.54
	Credit Spread	11.26***	6.56
Financing II	Book Lev.	0.00201***	4.03
	Credit Spread	0.639	0.78
Financing III	Book Lev.	0.00473***	12.84
	Credit Spread	10.95***	7.34
Human Capital	(TA–Net PP&E)/TA	0.00011	0.79
Intangible Assets	(TA–Net PP&E)/TA	–0.00005	–1.22
Intellectual Property	R&D Exp/TA	0.00046	1.58
Product Approval	R&D Exp/TA	0.00004	1.04
Product Innovation	Fluidity	0.00979***	3.25
Product Innovation	R&D Exp/TA	0.00001	0.19
Volatile Costs	σ(Cost)	0.00010	0.87
Volatile Net Cash Flow	σ(NCF)	–0.00000	–0.00
Volatile Revenue	σ(Revenue)	–0.00014	–0.96
Volatile Stock Price	σ(Equity)	0.00524***	5.42
Simple Fin Risk Index	Leverage	0.00575***	7.44
	Credit Spread	7.645***	2.84
Weighted Fin Risk Index	Leverage	0.00563***	8.83
	Credit Spread	7.407***	3.25

Table VII: Changes in Risk-Proxies and Lagged Changes in Risk-Factors

The Table reports the results of a panel regression of changes in risk-proxies on lagged changes in associated risk-factor sentence-counts. The *Simple Fin Risk Index* and *Weighted Fin Risk Index* are as defined in Table VI. A constant is included but not reported. *t*-statistics are based on robust standard errors. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Regression of change in risk-proxy on lagged change in risk-factor sentence-count		
Risk-Factor	Risk-Proxy	Coefficient	*t*-stat
Customer Concentration	Cust. Dummy	–0.00039	–1.26
Economic Conditions	Asset Beta	0.00029	0.19
Growth & Restructuring	–HHI	–0.00047*	–1.93
	Fluidity	–0.00664	–1.45
Financing I	Book Lev.	–0.00020	–0.36
	Credit Spread	–6.932***	–4.53
Financing II	Book Lev.	0.00016	1.22
	Credit Spread	0.779	1.18
Financing III	Book Lev.	0.00018	0.54
	Credit Spread	–5.730***	–5.04
Human Capital	(TA–Net PP&E)/TA	–0.00001	–0.08
Intangible Assets	(TA–Net PP&E)/TA	–0.00007**	–2.10
Intellectual Property	R&D Exp/TA	–0.00029	–0.97
Product Approval	R&D Exp/TA	–0.00010*	–1.89
Product Innovation	Fluidity	–0.00247	–1.06
Product Innovation	R&D Exp/TA D	0.00001	0.21
Volatile Costs	σ(Cost)	0.00001	0.09
Volatile Net Cash Flow	σ(NCF)	0.00075***	4.16
Volatile Revenue	σ(Revenue)	0.00034**	2.57
Volatile Stock Price	σ(Equity)	0.00350***	3.31
Simple Fin Risk Index	Leverage	0.00028	0.81
	Credit Spread	–2.085	–1.03
Weighted Fin Risk Index	Leverage	0.00029	0.95
	Credit Spread	–2.062	–1.22

Table VIII: Asset Volatility and Operating and Financing Risk Indices

Panel regressions of Asset Volatility on LDA-identified *Operating Risk* and *Financing Risk* indices with industry fixed-effects (Panel A) and without industry fixed-effects (Panel B). The *Simple Financing Risk Index* and *Weighted Financing Risk Index* are as defined in Table VI. The *Simple* and *Weighted Operating Risk* indices are constructed using the 26 LDA-identified operating risk factors in an analogous manner to the construction of the Financing Risk indices. *Market Lev.* 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. Robust *t*-statistics based on clustered standard errors at the SIC 2-digit industry level are shown in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel A: With Industry Fixed Effects

	Simple Risk Indices			*Weighted Risk Indices*		
Constant	0.2651***	0.2666***	0.3385***	0.2500***	0.2523***	0.3259***
	(33.57)	(36.70)	(29.58)	(30.51)	(31.54)	(27.26)
Operating Risk Index	0.0076***	0.0091***	0.0065***	0.0072***	0.0078***	0.0059***
	(5.50)	(6.59)	(4.71)	(7.22)	(7.47)	(5.70)
Financing Risk Index		–0.0036***	0.0019**		–0.0022***	0.0024***
		(–3.61)	(2.31)		(–3.13)	(3.29)
Market Lev.			–0.3432***			–0.3468***
			(–11.72)			(–11.61)
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,105	5,105	5,105	5,105	5,105	5,105
Adjusted R^2	0.374	0.382	0.467	0.387	0.392	0.478

Panel B: Without Industry Fixed Effects

	Simple Risk Indices			*Weighted Risk Indices*		
Constant	0.2504***	0.2566***	0.3548***	0.2256***	0.2360***	0.3368***
	(13.95)	(13.94)	(24.90)	(10.21)	(11.59)	(21.36)
Operating Risk Index	0.0104***	0.0135***	0.0085***	0.0103***	0.0115***	0.0076***
	(5.33)	(7.19)	(5.67)	(7.31)	(7.56)	(6.09)
Financing Risk Index		–0.0085***	0.0006		–0.0058***	0.0015
		(–3.57)	(0.26)		(–3.29)	(0.85)
Market Lev.			–0.4537***			–0.4485***
			(–12.60)			(–12.05)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,105	5,105	5,105	5,105	5,105	5,105
Adjusted R^2	0.097	0.153	0.343	0.143	0.184	0.363

Table IX: The Relation of Equity Volatility, Operating Risk, and Leverage

Panel regressions of $ln\sigma$(equity) on a constant, *ln*(*Market Lev.*) and *ln*(*Operating Risk Index*). σ(equity) is estimated as the annualised daily return volatility over the fiscal year. The *Market Lev.* measure 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. The *Simple Operating Risk Index* and *Weighted Operating Risk Index* are as defined in Table VIII. 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 10%, 5%, and 1% levels, respectively.

	Simple Operating Risk Index		*Weighted Operating Risk Index*	
Constant	–1.4084***	–1.4042***	–1.5504***	–1.5189***
	(–35.39)	(–45.77)	(–30.33)	(–40.09)
ln(Market Lev.)	0.1014	0.2389***	0.1217	0.2464***
	(0.84)	(6.82)	(1.03)	(7.05)
ln(Operating Risk Index)	0.1473***	0.1273***	0.1877***	0.1591***
	(8.58)	(7.52)	(10.05)	(9.15)
Industry FE	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	11,507	11,507	11,507	11,507
Adjusted R^2	0.367	0.502	0.375	0.506

.

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.

Catastrophe | *Corporate Governance* | *Country* | *Customer Concentration* | *Economic Conditions* | *Energy Sector*

Financing I | *Financing II* | *Financing III* | *Growth and Restructuring* | *Healthcare Spending* | *Human Capital*

Incomplete Contracts | *Information Systems* | *Intangible Assets* | *Intellectual Property* | *Product Approval* | *Product Defects*

Product Innovation | *Real Estate* | *Regulatory Change* | *Regulatory Compliance* | *Reporting Accuracy* | *Reporting Compliance*

Supply Chain | *Tax Uncertainty* | *Volatile Costs* | *Volatile Net Cash Flows* | *Volatile Revenue* | *Volatile Stock Price*

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*	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.

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.

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	*Volatile Costs*	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	*Volatile Net Cash Flows*	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	*Volatile Revenue*	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	*Volatile Stock Price*	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
(TA–Net PP&E)/TA	(Total assets – Net property plant and equipment) ÷ Total assets.	CRSP Compustat Merged (CCM)
σ(Asset Return)	Average of monthly stacked EGARCH estimates of asset volatility over 12 months ending at the fiscal year-end.	Website of Jaewon Choi
Asset Beta	Asset betas estimated by regressing firms' asset returns on CRSP VWRETD using a 12-month rolling window regression.	Website of Jaewon Choi
Book Lev.	Sum of book value of long- and short-term debt relative to the book value of assets total.	CCM
Market Lev.	Sum of book value of debt and market value of equity relative to the market value of equity.	CCM
Cash Flow	Sum of earnings before extraordinary items and depreciation divided by total assets.	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(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
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
HHI	Sum of the firms' squared market shares for firms in a 2-digit SIC industry. Market share is based on sales.	CCM
Fluidity	Intensity of product market change as defined in Hoberg, Phillips and Prabhala (2014).	Hoberg-Phillips Data Library
Cust. Dummy	Dummy variable equal to one if a corporate customer makes up at least 10% of the firm's sales.	Compustat Segments
σ(Stock Return)	Annualized stock return volatility estimated using monthly stock returns over the 60 months preceding the fiscal year end.	CRSP
σ(equity)	Annualized stock return volatility using daily stock returns estimated over the 250 trading days preceding the fiscal year end.	CRSP
RD Expense/TA	Research & development (R&D) expense divided by total assets. Set to zero if R&D expense is missing.	CCM
σ(Revenue)	Annualized standard deviation of quarterly revenue divided by total assets estimated over the 12 quarters preceding the fiscal year-end.	CCM

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.