Changes in Risk Factor Disclosures and the Variance Risk Premium

Matthew Lyle
Northwestern University

Edward J. Riedl *
Boston University

Federico Siano
Boston University

March 2021

Abstract

This paper examines how changes in firms' risk disclosures affect a key market measure of risk. Our proxy for changes in risk disclosures is the addition and deletion of individual risk factors to firms' 10-K annual filings, identified via textual analysis of the risk factors section. Our market measure proxy for risk is the variance risk premium (VRP), which captures the market's uncertainty about the risks the firm faces. Following the theoretical predictions of recent literature, we expect that newly disclosed signals of risk-factor exposure decrease the uncertainty surrounding firm risk, as proxied via the VRP. Empirical results strongly support these predictions: greater changes in individual risk factors are related to a lower VRP. Importantly, our new proxy based on individual risk factors offers incremental insights as compared to aggregate textual measures (including of risk) based on word counts. Collectively, our findings suggest that textually evaluating individual risk factors reveals information about the uncertainty regarding firm risk.

Key Terms: risk factors, disclosure, textual analysis, variance risk premium

Data Sources: SEC filings, and the Compustat, CRSP, and OptionMetrics databases

* Corresponding Author: eriedl@bu.edu

Changes in Risk Factor Disclosures and the Variance Risk Premium

Abstract

This paper examines how changes in firms' risk disclosures affect a key market measure of risk. Our proxy for changes in risk disclosures is the addition and deletion of individual risk factors to firms' 10-K annual filings, identified via textual analysis of the risk factors section. Our market measure proxy for risk is the variance risk premium (VRP), which captures the market's uncertainty about the risks the firm faces. Following the theoretical predictions of recent literature, we expect that newly disclosed signals of risk-factor exposure decrease the uncertainty surrounding firm risk, as proxied via the VRP. Empirical results strongly support these predictions: greater changes in individual risk factors are related to a lower VRP. Importantly, our new proxy based on individual risk factors offers incremental insights as compared to aggregate textual measures (including of risk) based on word counts. Collectively, our findings suggest that textually evaluating individual risk factors reveals information about the uncertainty regarding firm risk.

Key Terms: risk factors, disclosure, textual analysis, variance risk premium

Data Sources: SEC filings, and the Compustat, CRSP, and OptionMetrics databases

I. INTRODUCTION

This paper tests the predictions of recent research (Heinle and Smith 2017; Heinle, Smith, and Verrecchia 2018) that firm-provided signals about risk factors lead to lower factor-exposure *uncertainty* and, all else equal, lower uncertainty about firm risk. In particular, we examine whether changes in the composition of individual risk factors, identified via application of textual analysis algorithms to the risk factor section of the 10-K filings, trigger updated beliefs about firms' risk exposure, captured via the so called variance risk premium.[1]

To measure changes in firms' risk disclosures, we use the risk factors section of the 10-K filing, which provides a strong setting for several reasons. First, the US Securities and Exchange Commission (SEC) mandated effective 2005 that publicly-listed firms provide narrative discussion of all material risks the firm faces within the 10-K filing. This mandate mitigates issues of self-selection observed in other voluntary disclosure settings. Second, many sources of information about a firm's risk exposure exist, including those originating within the firm (such as the disclosure channels of 8-Ks, 10-Qs, MD&A sections, earnings announcements, and conference calls), as well as those provided by external users (such as analyst reports, industry reports, and competitor disclosures). However, the risk factors section of the 10-K remains the only *comprehensive* source of all material risks the firm faces. Third, as a firm-supplied disclosure, the risk factors section benefits from managers' private information that can help identify and characterize the firm's risks. Access to this private information should enable management to understand and present the risks faced by the firm in a way that external users are unlikely able to replicate on their own. Finally, the SEC mandate requires that each disclosed risk factor be labeled

1 As explained throughout the paper, the variance risk premium is the difference between the cost of purchasing a contract to hedge future (realized) returns volatility and the subsequent realized volatility (Carr and Wu 2009). That is, it represents the compensation traders require to bear volatility risk.

with a descriptive heading. Our analyses exploit this disclosure requirement through a textual analysis algorithm; this enables us to both identify the individual risk factors disclosed by the firm, as well as characterize the changes in these individual risk factors.

Prior academic research (Heinle and Smith 2017; Heinle, Smith, and Verrecchia 2018) predicts that signals provided by firms regarding their exposure to impending risks should reduce the *uncertainty* surrounding those risks. In particular, revealed signals about a risk factor provide inputs into market participants' assessment of the uncertainty about the variability of a firm's future cash flows. That is, the revelation of information regarding these risks should facilitate investors' ability to more precisely estimate their effects upon the firm, its performance, and the related stock price effects. In sum, risk-factor disclosures should reduce uncertainty about higher moments of the cash flows distribution, and thus lead to more precise estimates of the distribution of future realized outcomes (i.e., realized cash flows). We therefore hypothesize that firms providing greater changes in risk factors will exhibit lower uncertainty about their risk.

To test these predictions requires two key proxies: one for a risk-factor disclosure; and one for uncertainty about risk. We use as our proxy for changes in risk-factor disclosures the year-over-year *additions* and *deletions* of individual risk factors within 10-K filings. We use textual analysis techniques to identify the risk caption headings provided in each firm's annual report; we then use string similarity algorithms to compare the text of the individual risk factor captions. This allows us to isolate individual risk factors that have been added, and those that have been deleted, relative to the previous year. Our underlying assumption is that additions and deletions of risk factors reflect management insights into changes in the exposure to the risks a firm faces. We then partition each year of observations into those characterized as having high changes and low changes in individual risk factor disclosures. Our primary analyses denote as high (low) change firms those in the top (bottom) quintile of total added and deleted individual risk factors.

To proxy for investors' beliefs of uncertainty about risk, we use the variance risk premium (VRP), which represents an estimable value directly related to the uncertainty about risk.[2] Prior empirical research (Carr and Wu 2009) documents that the cost of purchasing contracts to hedge future realized stock return volatility generally exceeds subsequent realized volatility. That is, traders appear willing to pay for a higher level of anticipated volatility than the actual volatility that occurs. This difference—denoted the spread, or more formerly the VRP—represents the compensation traders require to bear volatility risk. In theory, investors are only willing to pay a premium to trade volatility if there is uncertainty about volatility itself. Prior empirical research provides compelling evidence consistent with this theoretical argument. Accordingly, we use the VRP as our measure of market participants' uncertainty regarding the risks faced by the firms.

Using a broad cross-sectional sample spanning 2006-2019, empirical tests are consistent with our expectation. We document a significantly negative association between changes in individual risk factors and the VRP. This finding is consistent with investors using changes in the disclosure of individual risk factors in firm's 10-Ks to update their beliefs about the uncertainty of the risks that firms face. We note two key research design choices. First, our tests employ recent matching procedures to minimize potential biases resulting from unobserved heterogeneity. That is, after identifying firms having high versus low changes in individual risk factor disclosures, we match the high and low firms on a large set of control variables. This matching mitigates issues

[2] Throughout the study, we use the term "uncertainty about risk" to align with the analytical settings of Heinle and Smith (2017) and Heinle, Smith, and Verrecchia (2018). In the option and asset pricing literatures, uncertainty about risk would sometimes be called the "volatility of volatility," as it represents the uncertainty around future volatility. Theory suggests that for the VRP to exist, volatility itself must have volatility; accordingly, VRP represents a direct measure of the uncertainty about risk, which we wish to examine.

of self-selection by minimizing heterogeneity when comparing firms across risk factor disclosure groups. Second, our analyses directly control for word-based textual disclosure of both the 10-K and the risk factors section, which have been used in prior research (e.g., Kravet and Muslu 2013; Campbell, Chen, Dhaliwal, Lu, and Steele 2014). Interestingly, these aggregate textual measures exhibit statistically insignificant associations with the VRP: this suggests our findings are not an artifact of the individual risk factors only reflecting these coarser textual measures. Overall, our results suggest that changes in individual risk factors are correlated with market perceptions of risk uncertainty, and that additions and deletions of these individual risk factors provide market participants with insights into the risk exposures faced by the firm beyond aggregate text disclosure measures (such as word counts).

We confirm that our findings are robust to a number of sensitivity analyses. First, we use alternative thresholds to characterize high versus low changes in individual risk factor disclosures: terciles and medians (as opposed to quintiles in the primary analyses). Inferences are unchanged; and suggest that the effects we document occur throughout the full range of available observations, not only in the tails of the distribution. Second, we use alternative methods to derive the VRP: we employ expected (versus realized) measures, and non-Garch (versus Garch) estimations. Results are again consistent. Third, we use alternative matching procedures for the samples designated as high and those designated as low changes in individual risk factor disclosures. In particular, we use both "Propensity Score" matching (versus "Mahalanobis Distance" matching in primary analyses), as well as matching with replacement (versus without replacement in primary analyses). Inferences remain unchanged. Finally, we employ a placebo test, in which we exclude our key experimental variable (changes in individual risk factors) and include aggregate change in text in the risk factor section. We do not find significance for the latter variable, again suggesting that the effects we document appear unique to individually disclosed risk factors.

Our findings provide three central contributions. First, we provide the first direct empirical tests of the predictions of recent related analytical research (Heinle and Smith 2017; Heinle et al. 2018). This research argues that the disclosure of risk signals should reduce market participants' *ex ante* perceptions of the uncertainty about the firm's risk. Our findings confirm these theoretical expectations, using a market-based proxy (i.e., the VRP) that reflects uncertainty about such risks. Second and related, we provide novel evidence about the decision-relevance of disclosed risk signals. In particular, we advance prior empirical papers, which document that aggregated risk disclosures are associated with higher perceived and realized riskiness (Kravet and Muslu 2013; Campbell et al. 2014), by demonstrating incremental information to identifying individual risk factors. Of note, our findings show that company-sourced changes in disclosed risk factors reduce the *ex ante* uncertainty surrounding future business outcomes. Third, we present a unique textual approach to identify and measure *individual* risk factors. Our findings further suggest that such individual risk factors allow for a more complete characterization of the information underlying risk disclosures, as the effects we document appear incremental (and seemingly unique) relative to more commonly-applied aggregate textual measures (such as word counts) and non-text measures (such as volatility of income).

Section II provides the background and hypothesis development. Section III presents the research design. Section IV discusses the sample selection and textual analysis methodology. Section V reports and describes the primary empirical results. Section VI presents sensitivity analyses, and Section VII presents the conclusion and directions for future research.

II. BACKGROUND AND HYPOTHESIS DEVELOPMENT

The Securities and Exchange Commission (SEC) dictates the reporting and disclosure requirements that publicly-traded firms on US exchanges must follow. In 2005, the SEC mandated that all firms provide a structured discussion of the risk factors they face within a dedicated section of their annual 10-K filings, referred to as section *Item 1A, Risk Factors* (Regulation S-K, Item 503(c); SEC 2005). Under the primary requirements, this section must list and comment on any material risk factors the firm faces that can impact future performance. Such factors reflect a wide-range of risk categories including those related to customers, competitors, suppliers, and the legal and regulatory environments. Note that the SEC guidance specifies that the provided disclosures should include risks specific to the registrant and exclude "generic" risks that could apply to any registrant (Kuntz, Ehrenberg, and Travis 2017).

Several other key features of these risk factor disclosures warrant highlighting. First, as the SEC requires these disclosures, the risk factors section represents a mandated disclosure. While firms have discretion in both the identification and characterization of "material" risk factors, the mandate mitigates endogenous reporting choices that can occur in other disclosure settings (e.g., other corporate narratives, or CSR reports). In particular, the requirement to disclose and discuss *any* material risks incentivizes full disclosure, as failure to do so with subsequent *ex-post* revelation can lead to significant costs (e.g., litigation, compensation and employment adjustments to senior management). Second, the SEC guidance requires that each individual risk factor be organized under a relevant sub-caption. This caption typically is a lead sentence, summarizing the nature of the risk; in most filings, these captions are offset and highlighted in some fashion (such as through bold-face, italics, or listed numerically). This latter feature is particularly useful: it facilitates our ability to identify and exploit the disclosure of individual risk factors (as described below).[3]

Prior research documents that the textual content of the risk factors section has increased substantially over time (Dyer, Lang, and Stice-Lawrence 2017). This research generally characterizes this information on an aggregate level, usually through word or sentence counts within the risk factors section. Using such approaches, several studies investigate the informativeness of aggregated risk disclosures. Campbell et al. (2014) provides initial evidence, finding that aggregate risk disclosures reduce information asymmetries (proxied using bid-ask spreads), and correlate with future systematic and idiosyncratic components of risk (proxied using CAPM Beta and returns volatility). Kravet and Muslu (2013) further documents that changes in the aggregate amount of textual risk disclosures are positively associated with proxies of increased risk perception (including returns volatility, trading volumes, and the volatility of analyst forecast dispersions). Chiu, Guan and Kim (2018) extends these findings to credit markets, finding that the SEC risk disclosure mandate decreases uncertainty about companies' credit risk, as measured through credit default swaps spread and volatility. Finally, Brown et al. (2018) documents that SEC risk monitoring—evidenced through the regulators' comment letters—induces the monitored firm's peers to change their risk disclosures, consistent with reducing the likelihood of subsequent regulatory scrutiny.[4]

We build on this literature by using textual analysis to identify ***individual*** risk factors within the risk factors section. In particular, using the mandated captions to discern changes in individual

[3] Note, the SEC recommends that the individual risk factors be ordered in terms of economic significance, with the 10-K preparation guidelines stating: "companies generally list the risk factors in order of their importance." While not mandated, best practice suggests many firms do so voluntarily—in part to minimize risks relating to litigation (Gelfond, Wechsler, and Cohen 2017). Further, the SEC approach to risk factor disclosures is consistent with similar guidance in other contexts, such as Form 20-F preparation for non-US firms: "companies are encouraged, but not required, to list the risk factors in the order of their priority to the company."

[4] Other papers have examined the role of accounting disclosures and option-implied risk measures, including Rogers, Skinner, and Van Buskirk (2009), Barth and So (2014), Billings and Jennings (2015), Neururer, Papadakis, and Riedl (2016), and Neururer, Papadakis and Riedl (2020).

risk factors over time, we identify those factors that are static (i.e., unchanged from one year to the next), added, or deleted. Doing so characterizes granular changes in the disclosed risks of the firm, which we can then relate to market measures. Critically, this is a necessary condition to better understand and identify whether it is the over-time increase in aggregate risk disclosure, or specific signals about risk-factor exposure, that influence investor uncertainty surrounding firm risk. That is, managers choose to disaggregate their firms' risk narratives into individual risk factors presumably to facilitate users' ability to identify and understand differential risks affecting the firm. Thus, we predict that the identification of individual risk factors conveys incremental information beyond aggregate text measures. As such, our research directly relates to prior work examining the textual attributes of corporate narrative disclosures and their effects on accounting, financial, and economic outcomes (Li 2008; Loughran and McDonald 2016).[5]

Prior theoretical research argues that signals about a firm's exposure to risks enable users of this information to adjust expectations of the variance of cash flows. In particular, Heinle and Smith (2017) and Heinle, Smith, and Verrecchia (2018) argue that disclosures about firm's risk factor exposures allow investors to better estimate the distribution of future cash flows and hence the distribution of future stock prices. These papers suggest that higher moments of stock price outcomes best reflect the nature of the information conveyed by risk factor disclosures, predicting that greater risk factor disclosure leads to lowered perceptions of the uncertainty about a firm's risk.

Prior literature (Carr and Wu 2009) has used the variance risk premium (VRP) to proxy for

5 Recent research characterizes the relevance of risk disclosures focusing on individual risk factors as the unit of analysis. In particular, Gaulin (2017) argues that added risk factors provide predictive power for estimating firm's future adverse operating outcomes (such as earnings declines or losses). Further, Chin, Liu, and Moffitt (2018) examines the ranking of risk factors in the context of credit risk, documenting that highly ranked factors are positively correlated with higher credit risk (as measured using S&P ratings and credit spread).

the market's perceived uncertainty about the risks a firm faces.[6] In particular, the VRP has been so named based on the findings of prior research that option implied volatility generally is higher than subsequent realized volatility: this difference leads to a "spread" between option implied and realized volatility, denoted the VRP. In asset pricing theory, VRP reflects the premium investors are willing to pay to hedge against shocks to risk in future stock prices. That is, the more uncertainty around this risk, the higher the VRP.

Thus, if the addition or removal of individual risk factors provides signals, which market participants can use to better characterize their expectations of the uncertainty of outcomes a firm faces, then we expect this will lead to lowered uncertainty, or a lower VRP. This leads to our following expectation:

H1 Enhanced signals of individual risk factor exposure (proxied via additions and deletions of individual risk factors to the risk factor section) are negatively associated with investor uncertainty of firm risk (proxied via the variance risk premium).

We note considerable tension in this expectation. In particular, while prior research provides some evidence of aggregate risk disclosures correlating with market outcomes (Kravet and Muslu 2013; Campbell et al. 2014), other research questions whether firms are simply satisfying the risk disclosure mandate through excessively long and boilerplate discussions (IRRC 2016; SEC 2016). Hope, Hu, and Lu (2016) argues that the informativeness of risk disclosures is directly related to their level of specificity, and documents larger market reactions to more specific disclosed risk factors. Beatty, Cheng, and Zhang (2019) challenges the informativeness of these disclosures, documenting a decrease in the relevance of risk discussions following the 2008

---

[6] We focus on the second moment of the return distribution, and not higher moments such as skewness or kurtosis, for several reasons. First, the prior research motivating our study (Smith and Heinle 2017; Heinle et al. 2018) models risk disclosures as information about the second moment. Second, the predicted associations are less clear with moments greater than two, and the empirical estimation of such moments is also more challenging (as it requires additional data and distributional assumptions). Related, we do not test empirical associations with respect to the first moment (i.e., the cost of capital), as Smith and Heinle (2017) and Heinle et al. (2018) argue *against* doing so due to the ambiguous relation between the risk disclosure and such measures.

financial crisis, as well as lowered associations between risk narratives and future measures of realized operating risk (Altman's Z-Score). The latter paper suggests that this decline in informativeness may reflect companies' heightened focus on litigation concerns. Consistent with this latter claim, Cazier, McMullin, and Treu (2019) confirms the effectiveness of long and boilerplate risk discussions as a shielding mechanism within judicial contexts. This collective research suggests that changes to risk disclosures will not be associated with our market outcomes measures.

III. RESEARCH DESIGN

Primary Model

To test how changes in risk-factor disclosures affect market perceptions of the firm's factor-exposure *uncertainty*, we use the following research design:

$$
\begin{aligned}
\textit{Log rVRP}_{jt} = \;& \alpha_1 + \beta_1 \textit{High RF Discl}_{jt} \\
& + \beta_2 \textit{MA Ret Volatility}_{jt} + \beta_3 \textit{Beta}_{jt} + \beta_4 \textit{CFO Volatility}_{jt} + \beta_5 \textit{Size}_{jt} \\
& + \beta_6 \textit{ROA}_{jt} + \beta_7 \textit{Loss}_{jt} + \beta_8 \textit{MB}_{jt} + \beta_9 \textit{Leverage}_{jt} \\
& + \beta_{10} \textit{Analyst Follow}_{jt} + \beta_{11} \textit{Inst Investors}_{jt} + \beta_{12} \textit{Text Length 10K wo Risk}_{jt} \\
& + \text{Year-Month F.E.} + \text{Industry F.E.} + \varepsilon_{jt} \qquad (1)
\end{aligned}
$$

The dependent variable is the variance risk premium (*Log rVRP*). It is measured as the natural log of a Garch(1,1)-based one-month ahead realized variance, less the natural log of the one-month ahead option implied variance.[7] Prior empirical research (Carr and Wu 2009) documents negative variance risk premiums, or VRPs; this implies that the cost of purchasing contracts to hedge future realized stock return volatility are generally more expensive than subsequent realized volatility. This "spread" between the cost of the "variance contract" and future realized variance is the VRP,

[7] Results are robust to alternative measures of the dependent variable (discussed in the sensitivity analyses).

which represents the compensation traders require to bear volatility risk. For a fixed level of expected future volatility, higher option implied volatility (and thus more negative VRP) indicates a higher perceived uncertainty about future volatility. To help foster interpretability, we multiply the dependent variable (i.e., *Log rVRP*) by –1; accordingly, lower (higher) values of *Log rVRP* indicate lower (higher) levels of expected uncertainty about future firm risk.

The experimental variable is the change in risk factors, *High RF Discl*. It is measured as an indicator variable equal to 1 if the sum of the added plus removed risk factors from firm *j*'s 10-K Item 1A for year *t* is within the top (i.e., 5th) quintile of the filing year distribution, and 0 if this sum is within the bottom (i.e., 1st) quintile. That is, the indicator turns on for those firms exhibiting the largest changes in the number of added plus removed risk factors, which effectively captures our treatment group; while those firms having the lowest changes in added plus removed risk factors capture our benchmark group. Thus, this variable proxies for the extent to which the risk factors section provides new signals of the firm's riskiness to financial statement users, evidenced through the addition and deletion of individual risk factors. As we expect that firms exhibiting the highest addition and deletion of individual risk factors provide investors with the most signals to parameterize expectations of uncertainty about a firm's risk, the predicted coefficient on *High RF Discl* is negative.

We include a variety of control variables to capture potential other drivers of observed risk. First, we include four measures of risk. *MA Ret Volatility* is the standard deviation of firm *j*'s monthly market-adjusted stock returns; it is measured over the preceding 12 months, with the value-weighted index applied as the market adjustment. We also include *Beta*, a proxy for firm *j*'s systematic risk; it is measured using the firm's beta derived from a one-factor market model estimated over the year preceding the 10-K filing. Third, we include *CFO Volatility*, as a proxy for the firm's operating risk; it is measured as the standard deviation of firm *j*'s annual operating cash flows (each scaled by beginning total assets) over the preceding five years. Finally, we include *Size*, the natural logarithm of firm *j*'s market capitalization at the beginning of the filing year *t*. As a firm's historical idiosyncratic risk, systematic risk, and cash flow-based operating risk should be positively related to realized volatility, the predicted coefficients for *MA Ret Volatility*, *Beta*, and *CFO Volatility* are all negative. In addition, as the scale of firms' economic transactions is negatively related to risk uncertainty, we also expect a negative coefficient on *Size*.

Second, we include variables to capture various attributes of the firm that may relate to our outcome measure. We include *ROA*, firm *j*'s net income for filing year *t* divided by total assets at the beginning of the same year. We also include *Loss*, an indicator variable equal to 1 if firm *j* reports negative net income for filing year *t*, and 0 otherwise. As we expect stronger performance to be negatively related to perceived riskiness, the predicted sign on both coefficients is negative. We include *MB*, firm *j*'s market-to-book ratio measured at the beginning of the filing year *t*. As higher growth opportunities generally reflect higher expected variance of outcomes—and thus higher expected risk, the predicted coefficient is positive. We include *Leverage*, firm *j*'s short-term plus long-term liabilities divided by total assets, all measured at the beginning of filing year *t*. Higher indebtedness may indicate more financial constraints; alternatively, higher debt may reflect more stable operations that allow the firm to support greater leverage, as well as the effective use of a tax shield. Accordingly, we do not predict the coefficient sign. Next, we include two measures of the firm's information environment: *Analyst Follow* (the natural logarithm of the maximum earnings estimates for firm *j* in filing year *t*), and *Inst Investors* (the number of shares held by institutional investors in firm *j* for year *t*, scaled by total shares outstanding). As enhanced information environments should lower perceived riskiness, the predicted coefficients for both variables are negative.

Third, we include *Text Length 10K wo Risk*, the natural log of the total words in firm *j*'s 10-K for year *t* less the total words in the same filing in the Item 1 A Risk Factors sections. This variable captures the propensity towards disclosure in the 10-K filing. We do not predict the coefficient sign for the following reason. The length of regulatory filings may reflect higher business and/or linguistic complexity, suggesting higher risk and thus a positive coefficient. Alternatively, this length may capture a higher number of business-related information signals, potentially providing enhanced disclosure to lower risk, and thus a negative coefficient. Accordingly, we do not predict the coefficient sign for *Text Length 10K wo Risk*. As a final control variable to augment our primary model, we also add in additional specifications *Change RF Text Length*. It is measured as the change in the natural log of the total number of words within "Item 1A" (i.e., "Risk Factors" section within 10-K reports). The addition of this variable is intended to confirm that added and removed risk factors capture a dimension of disclosures *distinct from changes in aggregate risk text*. To the extent aggregate changes in risk-related words also measure risk-factor exposure, we predict a negative coefficient.[8]

Finally, we include fixed effects for date, and industry. We cluster standard errors by firm.[9]

Matching Design

To rule out confounding differences between firms displaying a "high" and a "low" number of added and removed individual risk factors, we employ a matching procedure as

[8] Note that we expect *ex ante* that aggregate text will be a noisier measure of risk relative to the identification of individual risk factors used as our primary variable. Robustness tests (discussed later) confirm a negative but insignificant relationship between changes in aggregate risk text and the variance risk premium.

[9] In untabulated tests, we also examine two-way clustering by firm and year; results are unchanged.

follows. We use thirteen dimensions to match the "high" and "low" disclosure firms. We perform an exact match on industry membership, date, and the loss binary variable; this ensures running our analyses across comparable groups of firms belonging to the same industry and same date, and same profit versus loss characterization. We then match firms on the remaining ten control variables included in the primary regression, capturing (i) drivers of observed risk (*MA Ret Volatility*; *Beta*; *CFO Volatility;* and *Size*), (ii) fundamentals (*ROA*; *MB*; *Leverage*; *Analyst Follow*; and *Inst Investors*), and (iii) characteristics of the textual filing (*Text Length 10K wo Risk*). Our primary tests use a one-to-one "Nearest-Neighbor" matching, based on "Mahalanobis Distance," with no replacement.[10] We entropy-balance (Hainmueller 2012; Zhao and Percival 2016) the first and second moments (i.e., mean and variance) of continuous variables to level remaining differences between the two groups of firms.

Using the above procedure, we perform three alternative matches of companies having a "high" and a "low" change in risk-factor disclosures. The primary sample includes firms with added and removed risk factors in the top quintile of the yearly distribution matched with firms belonging to the bottom quintile. To provide robustness, as well as to assess the extent to which our predicted effects occur within the distribution of firms, we additionally match companies based on first/third terciles, and above/below the median of the yearly distribution of added and removed disclosed risk factors.

[10] We also test the sensitivity of our findings to alternative matching procedures. In particular, we use propensity score matching, as well as apply a protocol *with* replacement. Our findings are unchanged and discussed in the sensitivity analyses.

IV. DATA AND DESCRIPTIVE STATISTICS

This study employs a novel dataset of *individual* risk-factor disclosures and proposes a new textual measure that captures changes to those individual disclosures.[11] This section describes (i) the main data samples, (ii) the textual analysis methodology used to identify risk-factor disclosures and build the experimental (textual) measure, and (iii) the characteristics of and rationale behind our research design choices.

Data and Textual Analysis Methodology

We use three primary sources of data. Textual data is extracted from 10-K regulatory filings, which are downloaded from the *SEC EDGAR* database. Firms' financial fundamentals are obtained from *WRDS Compustat* (annual files) and *CRSP* (monthly files). Finally, variance risk premia (VRP) are estimated using *OptionMetrics*'s volatility surface files and CRSP daily files. OptionMetrics data is used to estimate option-implied volatility and the CRSP files are used to estimate expected and realized volatility.

We start by identifying and downloading 10-K reports filed between 2006 and 2019 with valid matches in financial databases (i.e., valid CIK, GVKEY, and PERMNO numbers).[12] We scrape each document to extract a company's identifiers, the filing date, the reporting period, "Item 1A" (i.e., the "Risk Factors" section), and individual risk-factor disclosures (i.e., distinct risk factors) within "Item 1A." To avoid introducing measurement error, we discard 10-Ks for

---

[11] As previously discussed, prior literature generally focuses on *aggregated* risk disclosures (either investigating the aggregated text within "Risk Factors" sections or extracting risk-related discussions within regulatory filings through statistical techniques) rather than on *individual* risk factors found in "Item 1A" of 10-K reports.

[12] As previously highlighted, the SEC required the disclosure of risk factors within annual reports in 2005; accordingly, we begin our sample period in 2006.

which "Item 1A" and/or individual risk factors are difficult to extract.[13] We also exclude a filing if the lagged (i.e., 1-year) risk-factor disclosures have not been identified, or the relevant financial items are missing.[14] Our sample of textual reports with available risk-factor disclosures and financial data comprises 33,797 annual filings. Each firm-year observation is then matched with monthly VRP data. This leads to our final sample of 180,449 monthly observations with available risk-factor disclosures, financial fundamentals, and VRP (see Table 1).

For our final sample, we match firms displaying a *high* number to those displaying a *low* number of added and removed risk-factor disclosures. As previously discussed, we use three alternative procedures to partition high and low-disclosing companies, classifying firms with a number of added and removed risk factors in the: (i) fifth quintile of the yearly distribution as "high-disclosing" (first quintile as "low-disclosing") (N = 50,330 matched-pairs observations); (ii) third (first) terciles (N = 81,142); and above (below) median (N = 131,634).

To quantify the total number of added and removed risk factors in a firm's 10-K report we develop textual analysis procedures that: (i) identify "Item 1A" within 10-K filings; (ii) extract individual risk factors within "Item 1A"; and (iii) classify each individual risk factor as *added* (i.e., if missing in the prior filing period) or *removed* (i.e., if missing in the current filing year). We first download the "HTML" version of all available 10-K reports from *SEC EDGAR*; we exclude 10-K Amendments, 10-KSB, and 10-KSB Amendments. We then scrape each document to identify a set of starting signals (e.g., "1A", "RISK FACTORS") and ending signals (e.g., "1B",

13 Difficult-to-extract Item 1A sections result from missing or non-standard starting/ending markers used to parse the textual reports. Difficult-to-extract disclosed risk-factors result from non-standard formatting and layout within Item 1A of 10-K reports.

14 To clarify: the main textual experimental variable in this study (the number of added and removed risk factors in each financial period and for a given 10-K report) requires the existence and identification of two consecutive (i) annual filings, (ii) "Item 1A" sections, and (iii) risk-factor disclosures (i.e., individual risk factors).

"UNRESOLVED"); this enables us to delimit and extract the "Risk Factors" section.[15]

To identify and extract individual risk factors within "Item 1A", we exploit the textual layout and formatting that is preserved in the "HTML" version of the file. Risk factors are generally introduced through a boldface and/or italicized header, which is typically followed by a plain-text paragraph(s) that contains an in-depth discussion of the individual risk factor. Accordingly, we classify individual risk factors as portions of the text within "Item 1A" that are both (i) introduced by a boldface and/or italicized header, and (ii) followed by a plain-text paragraph.[16]

A key novelty of our study is the classification of *added* and *removed* individual risk factors. To identify these factors, we compare each individual risk factor in the 10-K of a given firm with all the risk factors found in the prior year's 10-K of the same company (see Appendix B for an example). We exploit flexible string-similarity algorithms to assess whether a risk factor is repeated within two consecutive 10-K reports.[17] We then classify a risk factor as "added" whenever that risk factor was not disclosed in the prior financial period, but it is disclosed in the present year; conversely, a "removed" risk factor was disclosed in the prior year, but it is no longer present in the current financial period. Figure 1 provides descriptive evidence of how the number of individual risk factors has increased dramatically across all quintiles between 2006

---

[15] We also develop tailored conditional statements to handle cases of multiple starting and/or ending signals. Whenever the "Item 1A" cannot be identified with sufficient reliability, we exclude the document from the sample; this leads to the exclusion of 17% of the total downloaded 10-Ks. In addition, we confirm the reliability of the parsing strategy by manually checking the extracted "Item 1A" for 50 distinct companies in different years.

[16] In some cases, boldface and/or italicized headers introduce the "type" or "class" of risk factors (e.g., "Risks Related to Our Business and Industry") rather than an individual risk factor. Our algorithm excludes such cases, as we require a plain-text paragraph to follow the header for it to be classified as an individual risk factor.

[17] Specifically, we combine several algorithms to assess string similarity: (i) regardless of the position of the tokens found within two textual sequences; (ii) despite differences in two strings that are solely driven by repeated tokens; and (iii) even if minor other differences are present between the two sequences we compare. Each algorithm outputs a similarity score between 0% and 100%; we take the average of all similarity scores and identify a "match" for scores above 85%. We manually inspect 100 "matched" risk factors and 100 "unmatched" risk factors for different firm-years and confirm that our strategy is highly accurate (i.e., 99%).

and 2019. Furthermore, Figure 2 highlights that firm-years in the top quintile added or removed an average of 15.8 risk factors over the full sample period, while those in the bottom quintile added or removed an average of less than 1 risk factor over the same period. The 15 added risk factors represent a change of about 50% relative to the mean of 32 risk factors disclosed across the sample period.

Descriptive Statistics

Table 2 presents the descriptive statistics for the primary sample of matched-pairs observations using disclosure quintiles. Statistics are presented for the full sample (Column 1), as well as the subsamples of "high-disclosing" firms for Quintile 5 (Column 2) and "low-disclosing" firms for Quintile 1 (Column 3). Consistent with data requirements for options to derive our dependent variable, the sample is tilted towards larger firms, where the mean market capitalization of our sample firms is approximately $6.1 billion. The mean *ROA* is 1%, with 24% of firms making losses. The mean market-to-book ratio is 3.22, and sample firms have moderate leverage.

Of primary note, Column (4) presents differences in means and standard deviations across the "high-disclosing" and "low-disclosing" subsamples. Consistent with the matching protocol, we observe no differences for any control variables in either the mean or standard deviation; this should mitigate potential confounds arising due to other characteristics that may be correlated with our experimental variable. We also observe, consistent with expectations, that our dependent variable (*Log rVRP*) has a significantly lower mean for the "high disclosing" subsample (0.41) relative to the "low-disclosing" subsample (0.46).[18]

[18] As we report in Section III and in Appendix A, we multiply the VRP by –1 to foster its interpretability.

Table 3 provides further univariate evidence, as the experimental variable *High RF Discl* has a significantly negative correlation with *Log rVRP* (as well as our two alternative dependent variables of *Log eVRP* and *Log rVRPng*). Both univariate associations are consistent with lower perceived uncertainty about firm risk for those firms exhibiting the highest number of added or removed risk factors.

V. EMPIRICAL RESULTS

Table 4 presents the primary empirical results. Column (1) presents unconditional results, including only the experimental variable *High RF Discl*, and fixed effects. Consistent with expectations, we find a significantly negative coefficient (–5.474; *t*-stat = –3.7). Column (2) repeats the results, now including the four control variables to capture the firm's risk environment (*MA Ret Volatility*, *Beta*, *CFO Volatility*, and *Size*). The coefficient on *High RF Discl* remains significantly negative (–5.245; *t*-stat = –3.5); further, the coefficients on all four risk variables attain the predicted negative signs. Column (3) then presents results including the additional firm-level control variables (*ROA*, *Loss*, *MB*, *Leverage*, *Analyst Follow*, *Inst Investor*, and *Text Length 10K wo RF*). As previously, the coefficient on *High RF Discl* is significantly negative (–4.193; *t*-stat = –3.2); the coefficients on the other risk variables also remain significantly negative. We further find (as predicted) significantly negative coefficients on the control variables of *Loss*, *Analyst Follow*, and *Inst Investors*; coefficients on the remaining control variables are insignificant.

Finally, Column (4) presents results including one additional control variable: *Change RF Text Length*. As discussed previously, this latter variable measures the change in *aggregate* risk text within "Item 1A" of 10-K reports. We include this variable to discern whether changes in perceived risk are due to the change in individual risk factors (as captured by *High RF Discl*), and/or due to the change in aggregate text disclosure in the risk factors section (as captured by *Change RF Text Length*). The coefficient on *High RF Discl* remains significantly negative (–4.011; *t*-stat = –3.0); in addition, the coefficient significance levels on the other control variables are unchanged from the Column (3) specification. Of note, the coefficient on *Change RF Text Length* is negative but insignificant (–0.014; *t*-stat = –0.6). Combined with the insignificant coefficient on *Text Length 10K wo RF*, this suggests neither the aggregate text disclosures in the filed 10-K, nor the aggregate text disclosures in the risk factors sections, are the primary drivers of perceived uncertainty about risk.[19] Rather, it appears that market participants reassess the firm's riskiness conditional on observing the disclosure of newly added and removed individual risk factors. Restated, our results are broadly consistent with the predictions of Heinle and Smith (2017) and Heinle, Smith, and Verrecchia (2018) that the disclosure of additional risk signals (proxied via added and removed risk factors) serves to reduce perceptions of forward-looking riskiness (proxied via the VRP). Thus, the results are in support of H1.

VI. SENSITIVITY ANALYSES

Alternative Definitions of Risk-Factor Disclosures

Our first sensitivity analyses assess the extent to which our documented associations occur throughout the distribution of available observations. In particular, the primary analyses define the change in individual risk factor disclosure using the top and bottom quintiles of the sample distribution: this places our primary sample in the tails of the distribution of individual risk factor

[19] Note also that Table 2 reveals positive, but not large, correlations between *High RF Discl* and both *Text Length 10K wo RF* (0.15), as well as *Change RF Text Length* (0.16). This suggests that the changes in individual risk factors captures information incremental to that with the aggregate text. In the sensitivity analyses, we test and discuss the relationship between changes in *aggregate* risk text (i.e., *Change RF Text Length*) and the VRP (i.e., *Log rVRP*) without including added or removed individual risk factors (i.e., *High RF Discl*).

changes. We now assess two alternative derivations, both of which expand the sample of included observations. In both derivations, we continue to match observations as previously; again, this is done to minimize concerns of systematic differences beyond changes in individual risk factors driving the observed results. First, we define the alternative experimental variable *High RF Discl Tercile*, an indicator variable equal to 1 if the sum of the added plus removed risk factors from firm *j*'s 10-K Item 1A for year *t* is within the top (i.e., 3rd) tercile of the filing year distribution, and 0 if this sum is within the bottom (i.e., 1st) tercile. This alternative definition expands the sample to now 81,142 matched-pairs observations.

Table 5 presents the results. Of note, the inferences remain consistent with our primary findings of Table 4. In particular, we find a significantly negative coefficient on *High RF Discl Tercile*, both unconditionally in Column (1) (–3.027; *t*-stat = –2.6), and conditionally in Column (2) including all control variables (–2.655; *t*-stat = –2.6). Further, coefficient signs and significance levels on the control variables are generally unchanged from the primary specification. Of note, the coefficients on the two variables to control for total 10-K text length (*Text Length 10K wo RF*) and the risk factors section of the 10-K (*Change RF Text Length*) remain insignificant.

Second, we further expand the sample using the alternative experimental variable *High RF Discl Median*, an indicator variable equal to 1 if the sum of the added plus removed risk factors from firm *j*'s 10-K Item 1A for year *t* is above the median of the filing year distribution, and 0 otherwise (i.e., is below the median). This alternative definition includes the full available sample, now consisting of 131,634 matched-pairs observations. Results are presented in Column (3) (unconditional), and Column (4) (conditional). As above, the coefficients on *High RF Discl Median* remain significantly negative (in Column 3, –2.007, with *t*-stat = –2.6; and in Column 4,

–1.793, with *t*-stat = –2.3). The control variables also attain the same predicted signs as in the primary analysis.

Overall, these results suggest our findings—of lower perceived uncertainty of risk for firms with higher added and removed individual risk factors—are robust to alternative definitions to define changes in individual risk factors. We note that the coefficient magnitude and significance on the experimental variable does appear to decrease as the sample is broadened. That is, in moving from the samples using *High RF Discl*, *High RF Discl Tercile*, and *High RF Discl Median*, the coefficient declines in value and significance despite the increased power due to the inclusion of more observations. This suggests that, while the effects appear to occur through the full distribution of available observations, they are strongest in the tails of the distribution of total added and removed individual risk factor observations.

Alternative Definitions of the Dependent Variable

Our primary analyses use the *realized* VRP, as the dependent variable. One potential concern with this derivation is that our results are based on our choice of how we measure stock return variance. To alleviate this potential concern, we next examine two alternative dependent variable estimations of the variance risk premium. First, we use expected (i.e., forecasted) VRP (*Log eVRP*); this is based on the one-month ahead forecast of variance minus the natural logarithm of Option Implied Variance. From a theoretical standpoint, VRP is the difference between option-implied and expected future variance. In our primary analyses, we follow prior research and used realized variance as our proxy for expected variance (e.g., Carr and Wu 2009). However, using expected VRP provides an alternative specification that is based entirely on *ex ante* information, and thus avoids any potential confounds of using realized (*ex post*) information in the tests.

Table 6 presents the results, with Column (1) showing the unconditional specification, and Column (2) the conditional specification. Inferences are unchanged from our primary specifications. In particular, we continue to find a significantly negative coefficient on *High RF Discl* (–3.747; *t*-stat = –2.5); further, the coefficient signs and significance levels on the control variables remain unchanged from our Table 4 specification.

Second, we examine a realized VRP based on a non-Garch estimation, using the variable *Log rVRPng*.[20] In particular, this variable is the natural logarithm of one-month ahead (non-Garch) realized variance based on the sum of daily squared log-returns minus the natural logarithm of Option Implied Variance. While GARCH-based model volatility estimation is common, and provides calibrated models that allow for forecasting, it is also common to use "model free" volatility estimation of realized volatility. Accordingly, to ensure our results are not driven by our choice of volatility model, "model free" results are presented in Column (3) (the unconditional estimation), and Column (4) (the conditional estimation). Once again, the results remain unchanged. In particular, we continue to find a significantly negative coefficient on *High RF Discl* (–3.654; *t*-stat = –2.4).

Overall, these sensitivity analyses confirm that our inferences are robust to alternative derivations of the dependent variable of the variance risk premium.

Alternative Matching Strategies

Next, we examine whether our results are robust to alternative matching strategies. Our primary research design uses one-to-one "Nearest Neighbor" protocol, estimated without

[20] We note that the correlations across the three proposed dependent variables (i.e., *Log rVRP*; *Log eVRP*; *Log rVRPng*) are high, but range from 0.43 to 0.77. This suggests that the measures—while similar—also capture unique measurement of uncertainty regarding risk.

replacement. Again, exact matching is done with respect to year, month, industry membership, and the loss indicator; and matched control variables are adjusted through entropy-balancing (Hainmueller 2012; Zhao and Percival 2016) the first and second moments of each variable's distribution (see Table 2).

We now consider two alternative matching techniques. First, we conduct propensity score matching (see, for example, Hirano, Imbens, and Ridder 2003) with no replacement. While propensity score matching imposes a structure on the data (i.e., a first-stage dimensionality reduction between 0 and 1), it continues to represent a popular matching strategy; accordingly, we corroborate our findings employing alternative distance measures (i.e., "Mahalanobis Distance", "Propensity Score"). As in the primary analysis, we choose the closest match (i.e., one-to-one matching, based on propensity scores) and perform exact matching on year, month, industry, and whether the firm exhibits a loss. Weights on the other control variables are adjusted through entropy-balancing (Hainmueller 2012; Zhao and Percival 2016) the first and second moments of each variable's distribution. Table 7 presents the results in Columns (1) and (2), with inferences remaining robust. Focusing on the conditional results in Column (2), we continue to find a significantly negative coefficient on *High RF Discl* (–4.365; *t*-stat = –3.4); the coefficient signs and significance on the other variables are generally unchanged from our primary analyses.

Second, we again estimate one-to-one nearest neighbor based on "Mahalanobis Distance"; in contrast to the primary specification done without replacement, this alternative estimation is done with replacement. Allowing replacement provides for potentially stronger matches; however, by allowing the same observations to be used multiple times, it can lead to potentially overstated significance due to increased correlations across the observations. Nonetheless, results remain unchanged in Columns (3) and (4). Focusing on Column (4), the coefficient on *High RF Discl* remains significantly negative (–4.280; *t*-stat = –2.7).

Overall, we find that our inferences are robust to alternative matching strategies for firms disclosing a high versus those disclosing a low number of added and removed risk factors within Item 1A of 10-K filings.

Placebo Using Changes in Risk Text Length

The primary experimental variable throughout the analyses focuses on measuring *individual* risk-factor signals, rather than aggregate risk text. Interestingly, the results of Tables 4 through 7 consistently reveal that changes in risk text (i.e., *Change RF Text Length*) are unrelated to our outcome variable of VRP. We further note that the correlation between *High RF Discl* and *Change RF Text Length* is 0.16 (see Table 2); this moderately low correlation suggests the two variables appear to capture different characteristics. Nonetheless, as additional evidence, we next estimate unconditional and conditional regression tests using *Change RF Text Length* as the main experimental variable, while excluding changes in individual risk factors (i.e., *High RF Discl*). This analysis will confirm whether aggregate risk factor text length may spuriously drive the observed relationships.

Table 8 presents the results. Of note, we find that the coefficient on *Change RF Text Length* remains insignificant in both unconditional analysis of Column 1 (–0.011, *t*-stat = –0.5) and all three conditional settings in Columns 2-4.[21] This provides additional support that the identification

[21] Here we provide evidence based on the same matching criteria and variable definitions used for our primary tests of Table 4. In untabulated analyses, we also use an experimental variable equal to 1 for observations in the top quintile (i.e., 5th) of the *Change RF Length* distribution, and 0 for observations in the bottom quintile (i.e., 1st) of the same distribution. The tenor of our results does not change.

of individual risk factors within 10-K "Item 1A" does not appear subsumed by (or even significantly correlated in regression analysis with) aggregated changes in risk text.

VII. CONCLUSION

This paper provides the first empirical test of predictions of recent analytical work examining risk disclosures (i.e., Heinle and Smith 2017 and Heinle et al. 2018) by exploring whether additional signals of risk facilitate investors' ability to formulate expectations of the uncertainty surrounding the risks the firm faces. In particular, we develop a novel technique to measure and characterize changes in the disclosed risk signals within corporate regulatory filings. This technique evaluates the text within Item 1A of firm's annual 10-K reports; this process enables identification of firm-level additions and deletions of *individual* risk factors. Our analyses assume that these additions and deletions serve as additional signals of uncertainty about the firm's risk. Using the variance risk premium (VRP) a theoretically-motivated *ex ante* measure of investor perceptions of the uncertainty of firm risks, we then test and find evidence consistent with risk-factor disclosures lowering uncertainty about risk. Our results document that firms' exhibiting the highest changes to individual risk-factor disclosures have lower VRP relative to those having the lowest changes to individual risk-factor disclosures. Critically, these findings are robust to the inclusion of word-based aggregate text measures (used in recent research) of both the 10-K filing and risk factor section. Thus, consistent with the theory, we infer that market participants appear to use additions and deletions of individual risk factors to form expectations of uncertainty about future risk outcomes. These results are robust to a variety of sensitivity analyses using alternative samples, alternative derivations of the dependent variable (i.e., VRP), and alternative matching strategies to mitigate the effects of correlated omitted variables.

Our work offers new insights regarding the decision-relevance of risk-factor disclosures: company-sourced risk signals not only map into future realized variance, as prior literature shows, but also reduce *ex ante* uncertainty surrounding future risks. Critically, our findings confirm that the granular measurement of individual risk factors we propose is not subsumed by previous aggregate textual measures of risk introduced by prior studies. This suggests that our measure is capturing a distinct dimension of risk relative to these other proxies. Future work may explore additional characteristics of individually disclosed risk signals. This can include understanding attributes of the individual risk factors themselves (such as their relative positioning within the risk factor section, or the specific contents of their disclosures); time-series trends in how these individual risk factors are reported; benchmarking of individual risk factors across relevant peer-firms; and the consequences of these disclosures on other outputs by market participants such as institutional investors, creditors, and intermediaries like analysts.

REFERENCES

Barth, M., and E. So. 2014. Non-diversifiable volatility risk and risk premiums at earnings announcements. *The Accounting Review* 89 (5): 1579-1607.

Beatty, A., L. Cheng, and H. Zhang. 2019. Are risk factor disclosures still relevant? Evidence from market reactions to risk factor disclosures before and after the financial crisis. *Contemporary Accounting Research* 36 (2): 805–838.

Billings, M., and R. Jennings. 2011. The option market's anticipation of information content in earnings announcements. *Review of Accounting Studies* 16 (3): 587-619.

Brown, S., X. Tian, and J. 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, J., H. Chen, D. Dhaliwal, H. Lu, and L. Steele. 2014. The information content of mandatory risk factor disclosures in corporate filings. *Review of Accounting Studies* 19 (1): 396–455.

Carr, P., and L. Wu. 2009. Variance risk premiums. *Review of Financial Studies* 22 (3): 1311-1341.

Cazier, R., J. McMullin, and J. Treu. 2019. Are lengthy and boilerplate risk factor disclosures inadequate? An examination of judicial and regulatory assessments of risk factor language. *The Accounting Review* (forthcoming).

Chin, M., Y. Liu, and K. Moffitt. 2018. Voluntary disclosure through the ranking of risk factors in the 10-K. *Working paper*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3142990.

Chiu, T., Y. Guan, and J. Kim. 2018. The effect of risk factor disclosures on the pricing of credit default swaps. *Contemporary Accounting Research* 35 (4): 2191–2224.

Code of Federal Regulations, Title 17, Section 229, Item 503(c). 17 CFR 229.503.

Dyer, T., M. Lang, and L. Stice-Lawrence. 2017. The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting and Economics* 64 (2-3): 221–245.

Gaulin, M. 2017. Risk fact or fiction: The information content of risk factor disclosures. *Doctoral dissertation*. Rice University.

Gelfond, S., J. Wechsler, and H. Cohen. 2017. Top 10 practice tips by experts: risk factors. LexisNexis Practice Advisor.

Hainmueller, J. 2012. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis* 20 (1): 25–46.

Heinle, M., and K. Smith. 2017. A theory of risk disclosure. *Review of Accounting Studies* 22 (4): 1459-1491.

Heinle, M., K. Smith, and R. Verrecchia. 2018. Risk-factor disclosure and asset prices. *The Accounting Review* 93 (2): 191-208.

Hirano, K., G. Imbens, and G. Ridder. 2003. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* 71 (4): 1161–1189.

Hope, O., D. Hu, and H. Lu. 2016. The benefits of specific risk-factor disclosures. *Review of Accounting Studies* 21 (4): 1005-1045.

Investor Responsibility Research Center Institute (IRRC). 2016. The corporate risk factor disclosure landscape.

Kravet, T., and V. Muslu. 2013. Textual risk disclosures and investors' risk perceptions. *Review of Accounting Studies* 18 (4): 1088–1122.

Kuntz, L., P. Ehrenberg, and S. Travis. 2017. Market trends: risk factors. Lexis Practice Advisor.

Li, F. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45 (2-3): 221–247.

Loughran, T., and B. McDonald. 2016. Textual analysis in accounting and finance: A survey. *Journal of Accounting Research* 54 (4): 1187–1230.

Neururer, T., G. Papadakis, and E. Riedl. 2016. Tests of investor learning models using earnings innovations and implied volatilities. *Review of Accounting Studies* 21: 400-437.

__________. 2020. The effect of reporting streaks on ex ante uncertainty. *Management Science* 66: 3771-3787.

Rogers, J., D. Skinner, and A. Van Buskirk. 2009. Earnings guidance and market uncertainty. *Journal of Accounting and Economics* 48 (1): 90-109.

Securities and Exchange Commission (SEC). 2005. Release no. 33-8591 (FR-75).

Securities and Exchange Commission (SEC). 2016. Business and financial disclosure required by regulation S-K, release no. 33-10064; 34-77599; File No. S7-06-16: 1-341.

Zhao, Q., and D. Percival. 2016. Entropy balancing is doubly robust. *Journal of Causal Inference* 5(1): 1–23.

APPENDIX A – Variable Descriptions

	Description	Source
Dependent Variables		
Log rVRP	Natural logarithm of Garch-based one-month ahead realized variance minus natural logarithm of Option Implied Variance (estimated using Gatheral's SVI) [the VRP is multiplied by –1 in the empirical analyses]	OptionMetrics
Log eVRP	Natural logarithm of Garch-based one-month ahead forecast of variance minus natural logarithm of Option Implied Variance (estimated using Gatheral's SVI) [the VRP is multiplied by –1 in the empirical analyses]	OptionMetrics
Log rVRPng	Natural logarithm of one-month ahead (non-Garch) realized variance based on sum of daily squared log-returns minus natural logarithm of Option Implied Variance (estimated using Gatheral's SVI) [the VRP is multiplied by –1 in the empirical analyses]	OptionMetrics
Experimental Variables		
High RF Discl	Indicator variable equal to 1 if the sum of a 10-K *Item 1A* (i.e., “Risk Factors”) added and removed risk factors (from filing year *t-1* to *t*) is in the highest (i.e., 5^{th}) quintile of the filing year distribution, and 0 if it belongs to the lowest (i.e., 1^{st}) quintile	EDGAR
High RF Discl Tercile	Indicator variable equal to 1 if the sum of a 10-K *Item 1A* (i.e., “Risk Factors”) added and removed risk factors (from filing year *t-1* to *t*) belongs to the highest (i.e., 3^{rd}) tercile of the filing year distribution, and 0 if it belongs to the lowest (i.e., 1^{st}) tercile	EDGAR
High RF Discl Median	Indicator variable equal to 1 if the sum of a 10-K *Item 1A* (i.e., “Risk Factors”) added and removed risk factors (from filing year *t-1* to *t*) exceeds the median of the filing year distribution, and 0 otherwise	EDGAR
Control Variables		
MA Ret Volatility	12-month standard deviation of a firm’s monthly market-adjusted stock returns, with market return proxied via the value-weighted index including dividend distributions	CRSP
Beta	CAPM beta computed using monthly returns over the past 12 months	CRSP
CFO Volatility	5-year standard deviation of annual operating cash-flows divided by total assets	Compustat
Size	Natural logarithm of a firm’s market capitalization	CRSP
ROA	Net income/total assets	Compustat

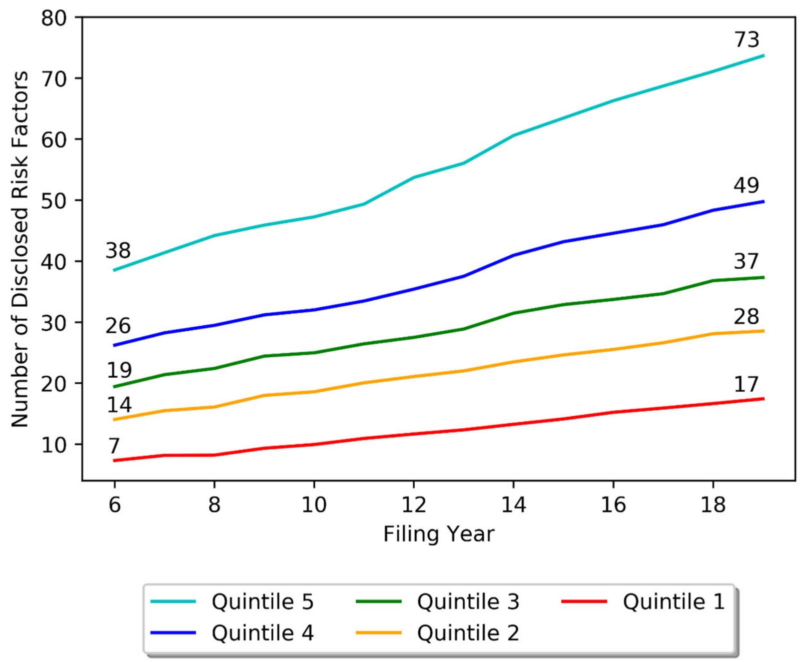
Loss	Indicator variable equal to 1 if net income is negative, and 0 otherwise	Compustat
MB	Fiscal closing price / book value per share	Compustat
Leverage	(Short-term liabilities + long term liabilities) / total assets	Compustat
Analyst Follow	Natural logarithm of the maximum number of estimates in a filing year divided by the natural logarithm of total assets; missing observations are replaced with a zero value	I/B/E/S
Inst Investors	Numbers of shares held by institutional investors each year divided by the total number of shares outstanding	Reuters 13-F
Length 10K wo RF	Natural logarithm of the total number of words in a 10-K minus the number of words within *Item 1A* (i.e., "Risk Factors")	EDGAR
Change RF Length	Change in the natural logarithm of the number of words within *Item 1A* (i.e., "Risk Factors") from filing year *t-1* to *t*	EDGAR

APPENDIX B – Example of Textually-Evaluated Changes in Risk Factor Disclosures

Chevron Corporation (CIK: 93410) – Individual Risk Factors Within 10-K Item 1A

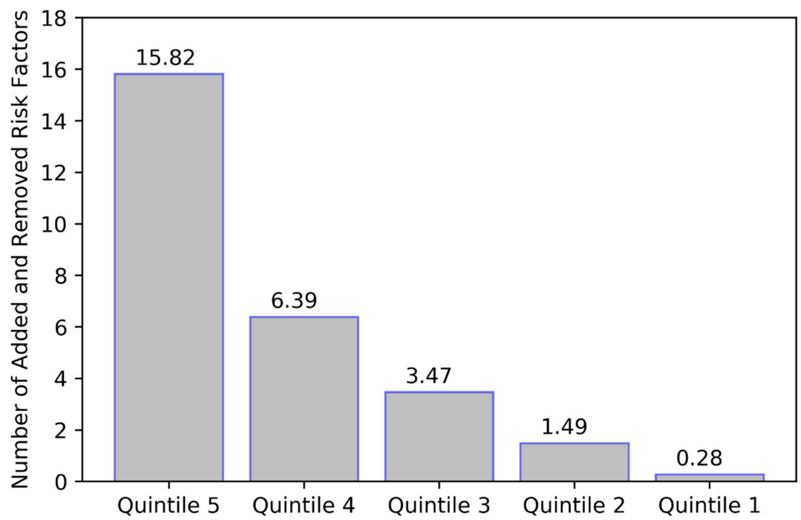
	2009	2010	2011
(1)	*Chevron is exposed to the effects of changing commodity prices.*	*Chevron is exposed to the effects of changing commodity prices.*	*Chevron is exposed to the effects of changing commodity prices.*
(2)	*The scope of Chevron's business will decline if the company does not successfully develop resources.*	*The scope of Chevron's business will decline if the company does not successfully develop resources.*	*The scope of Chevron's business will decline if the company does not successfully develop resources.*
(3)	*The company's operations could be disrupted by natural or human factors.*	*The company's operations could be disrupted by natural or human factors.*	*The company's operations could be disrupted by natural or human factors.*
(4)	*Chevron's business subjects the company to liability risks.*	*Chevron's business subjects the company to liability risks from litigation or government action.*	**The company's operations have inherent risks and hazards that require significant and continuous oversight.**
(5)	*Political instability could harm Chevron's business.*	*Political instability could harm Chevron's business.*	*Chevron's business subjects the company to liability risks from litigation or government action.*
(6)	*Regulation of greenhouse gas emissions could increase Chevron's operational costs and reduce demand for Chevron's products.*	*Regulation of greenhouse gas emissions could increase Chevron's operational costs and reduce demand for Chevron's products.*	*Political instability could harm Chevron's business.*
(7)		**Changes in management's estimates and assumptions may have a material impact on the company's consolidated financial statements and financial or operations performance in any given period.**	*Regulation of greenhouse gas emissions could increase Chevron's operational costs and reduce demand for Chevron's products.*
(8)			*Changes in management's estimates and assumptions may have a material impact on the company's consolidated financial statements and financial or operations performance in any given period.*
Additions and Deletions (bolded items above):			
	0	**1**	**1**

FIGURE 1
Number of Disclosed Risk Factors Within 10K Item 1A (2006-2019)



This figure shows the over-time trend of the average number of disclosed individual risk factors found within Item 1A (i.e., "Risk Factors") of SEC EDGAR 10-K reports, filed 2006-2019. The total number of firm-year observations is 54,773. Five curves are displayed, one for each quintile of the cross-sectional (i.e., by filing year) distribution of disclosed risk factors. The number of disclosed risk factors is winsorized at the 1st and 99th percentiles of its cross-sectional distribution.

FIGURE 2
Number of Added and Removed Risk Factors Within 10K Item 1A



This figure displays the average number of added and removed risk factors found within Item 1A (i.e., "Risk Factors") of SEC EDGAR 10-K reports, filed 2007-2019. The total number of firm-year observations is 49,740 (representing 54,773 available firm-year disclosures less 5,033 missing lagged disclosures). Five bars are displayed, one for each quintile of the cross-sectional (i.e., by filing year) distribution of added and removed risk factors, assessed over the full sample period. The number of added and removed risk factors is winsorized at the 1st and 99th percentiles of its cross-sectional distribution.

TABLE 1
Sample Selection

	Number
EDGAR 10-K filings (2006-2019) by companies with valid CIK, GVKEY, and PERMNO	72,322
Less: 10-K filings with:	
- difficult-to-extract Item 1A (i.e., "Risk Factors")	(12,551)
- difficult-to-extract disclosed risk factors within Item 1A	(4,998)
Total EDGAR 10-K filings with available risk-factor disclosures	54,773
Less: observations with missing:	
- Compustat items	(8,220)
- CRSP items	(7,723)
- lagged disclosures to compute textual variables	(5,033)
Total firm-years with available risk-factor disclosures and financial data	33,797
Total firm-year-months with available risk-factor disclosures, and financial data	180,449
Matched-pairs observations (by quintiles) – "Primary Sample"	**50,330**
"High-disclosing" (i.e., top quintile) firm-year-months	25,165
"Low-disclosing" (i.e., bottom quintile) firm-year-months	25,165
Matched-pairs observations (by terciles)	**81,142**
"High-disclosing" (i.e., top tercile) firm-year-months	40,571
"Low-disclosing" (i.e., bottom tercile) firm-year-months	40,571
Matched-pairs observations (above or below median)	**131,634**
"High-disclosing" (i.e., above median) firm-year-months	65,817
"Low-disclosing" (i.e., below median) firm-year-months	65,817

This table presents the sample selection for the final sample of matched-pairs observations (i.e., firm-year-months) with available data, including (i) available risk disclosures from SEC EDGAR 10-K Item 1A, (ii) financial fundamentals, and (iii) and variance risk premia ("VRP," obtained from OptionMetrics). "Difficult-to-extract Item 1A sections" result from missing or non-standard starting/ending markers used to parse the textual reports. "Difficult-to-extract disclosed risk-factors" result from non-standard formatting and layout within Item 1A of 10-K reports. Missing "lagged disclosures to compute textual variables" result from the absence of firm-specific prior period disclosed risk factors, which are necessary to compute the total number of added and removed risk-factor disclosures between consecutive filing years. Yearly SEC EDGAR textual disclosures are joined with monthly OptionMetrics VRP.

The primary matching strategy is a one-to-one "Nearest-Neighbor" protocol (based on "Mahalanobis Distance") and estimated without replacement (see Table 2). The primary matched-pairs sample comprises 50,330 observations, and includes firms disclosing a *high* number (5^{th} quintile of the yearly distribution) and *low* number (1^{st} quintile of the yearly distribution) of added and removed risk factors within Item 1A ("Risk Factors") of 10-K filings. Alternative matched-pairs samples based on tercile and median distributions are investigated in Table 5.

TABLE 2
Descriptive Statistics: Primary Sample of Matched-pairs Observations by Disclosure Quintiles

	Full Matched Sample (N = 50,330)		"High-disclosing" Quintile 5 Sample (N = 25,165)		"Low-disclosing" Quintile 1 Sample (N = 25,165)		Difference "High" – "Low" (Post Entropy-Balancing)	
	(1)		(2)		(3)		(4)	
	Mean	**St. Dev.**	**Mean**	**St. Dev.**	**Mean**	**St. Dev.**	**Mean**	**St. Dev.**
Dependent Variables								
Log rVRP	0.43	0.65	0.41	0.66	0.46	0.65	–0.05***	0.01***
Log eVRP	0.26	0.65	0.24	0.67	0.28	0.64	–0.04***	0.03***
Log rVRPng	0.79	0.86	0.78	0.87	0.81	0.86	–0.03***	0.01***
Experimental Variable								
High RF Discl	0.50	0.50	1.00	0.00	0.00	0.00	1.00***	-
Matched Control Variables								
MA Ret Volatility	0.11	0.07	0.11	0.01	0.11	0.01	0.00	0.00
Beta	1.06	0.89	1.06	0.89	1.06	0.89	0.00	0.00
CFO Volatility	0.05	0.06	0.05	0.06	0.05	0.06	0.00	0.00
Size	7.33	1.53	7.33	1.53	7.33	1.53	0.00	0.00
ROA	0.01	0.18	0.01	0.18	0.01	0.18	0.00	0.00
Loss	0.24	0.18	0.24	0.18	0.24	0.18	0.00	0.00
MB	3.22	7.01	3.22	7.01	3.22	7.01	0.00	0.00
Leverage	0.23	0.21	0.23	0.21	0.23	0.21	0.00	0.00
Analyst Follow	0.29	0.11	0.29	0.11	0.29	0.11	0.00	0.00
Inst Investors	0.83	0.25	0.83	0.25	0.83	0.25	0.00	0.00
Text Length 10K wo RF	10.82	0.47	10.82	0.47	10.82	0.47	0.00	0.00

This table presents descriptive statistics for the primary sample of 50,330 matched-pairs observations; the matching reflects firm-year-month observations chosen within the highest and lowest quintiles of added and removed risk factors within Item 1A (i.e., "Risk Factors") of 10-Ks filed between 2007-2019. All variables are defined in Appendix A. Continuous variables are winsorized at the 1st and 99th percentiles. "Dependent" ("Experimental") Variables reflect the primary dependent and independent variables. *Log rVRP*, *Log eVRP*, and *Log rVRPng* are multiplied by –1 to facilitate inferences. "Matched Control Variables" include the covariates used to match "high" risk-factor to "low" risk-factor observations. The matching protocol is one-to-one "Nearest-Neighbor" (based on "Mahalanobis Distance") and estimated without replacement. Exact matching is performed on year, month, industry membership, and the loss indicator; the mean and variance of the "Matched Control Variables" are entropy-balanced. Columns 1 (2) [3] present mean and standard deviations for the full matched sample (sample disclosing a *high* number of added and removed risk factors) [sample disclosing a *low* number of added and removed risk factors]; Column (4) presents tests of mean and standard deviations across the two subsamples. ***, **, * indicate significance at <0.01, <0.05, <0.10, respectively (two-tailed tests).

TABLE 3
Correlations: Primary Sample of Matched-pairs Observations by Disclosure Quintiles

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
(1)	*High RF Discl*	1														
(2)	*Log rVRP*	**–0.04**	1													
(3)	*Log eVRP*	**–0.04**	**0.77**	1												
(4)	*Log rVRPng*	**–0.03**	**0.75**	**0.43**	1											
(5)	*MA Ret Volatility*	**0.12**	**–0.06**	**–0.07**	**–0.02**	1										
(6)	*Beta*	**0.02**	**–0.06**	**–0.07**	**–0.06**	**0.23**	1									
(7)	*CFO Volatility*	**0.04**	**–0.02**	**–0.04**	**0.02**	**0.29**	**0.09**	1								
(8)	*Size*	**–0.04**	**–0.16**	**–0.15**	**–0.16**	**–0.45**	**–0.06**	**–0.30**	1							
(9)	*ROA*	**–0.05**	**–0.03**	**–0.03**	**–0.05**	**–0.31**	**–0.06**	**–0.49**	**0.30**	1						
(10)	*Loss*	**0.05**	0.00	0.00	**0.02**	**0.40**	**0.09**	**0.26**	**–0.37**	**–0.53**	1					
(11)	*MB*	0.01	**–0.05**	**–0.05**	**–0.03**	0.00	**–0.02**	**0.04**	**0.08**	**0.02**	**0.02**	**1**				
(12)	*Leverage*	**0.05**	**0.01**	**0.02**	**–0.01**	**0.04**	**0.05**	**–0.03**	**0.11**	**–0.06**	**0.04**	**–0.08**	1			
(13)	*Analyst Follow*	**–0.04**	**–0.13**	**–0.13**	**–0.08**	**–0.03**	**–0.02**	**0.01**	**0.15**	**–0.01**	**0.01**	**0.09**	**–0.09**	1		
(14)	*Inst Investors*	0.00	**–0.18**	**–0.16**	**–0.13**	**–0.07**	**0.03**	**–0.11**	**0.16**	**0.17**	**–0.13**	–0.01	**0.06**	**0.31**	1	
(15)	*Text Length 10K wo RF*	**0.15**	0.00	0.00	**–0.02**	**–0.07**	**0.02**	**–0.10**	**0.30**	**–0.02**	**0.01**	**–0.05**	**0.21**	**–0.06**	0.00	1
(16)	*Change RF Text Length*	**0.16**	**–0.02**	–0.01	**–0.02**	–0.01	**–0.02**	**–0.04**	**0.03**	**0.01**	**–0.01**	**–0.01**	**0.01**	**–0.02**	**0.02**	**0.02**

This table reports Pearson correlations for the primary sample of 50,330 matched-pairs observations; the matching reflects firm-year-month observations chosen within the highest and lowest quintiles of added and removed risk factors within Item 1A (i.e., "Risk Factors") of 10-Ks filed between 2007-2019. *Log rVRP*, *Log eVRP*, and *Log rVRPng* are all multiplied by –1 to facilitate inferences. All variables are defined in Appendix A. Continuous variables are winsorized at the 1st and 99th percentiles. Bold coefficients indicate significance at <0.01 (two-tailed tests).

TABLE 4
Risk-Factor Disclosures and the Variance Risk Premium

		Dependent Variable: *Log rVRP*			
	Pred. Sign	**(1)**	**(2)**	**(3)**	**(4)**
High RF Discl	(–)	**–5.474 *****	**–5.245 *****	**–4.193 *****	**–4.011 *****
		(–3.7)	**(–3.5)**	**(–3.2)**	**(–3.0)**
MA Ret Volatility	(–)		–1.160 ***	–1.272 ***	–1.273 ***
			(–8.0)	(–8.7)	(–8.7)
Beta	(–)		–0.019 ***	–0.019 ***	–0.019 ***
			(–2.4)	(–2.6)	(–2.6)
CFO Volatility	(–)		–0.402 ***	–0.470 ***	–0.472 ***
			(–3.2)	(–4.0)	(–4.0)
Size	(–)		–0.124 ***	–0.109 ***	–0.109 ***
			(–22.2)	(–18.4)	(–18.4)
ROA	(–)			–0.070	–0.071
				(–1.4)	(–1.4)
Loss	(–)			–0.033 **	–0.033 **
				(–1.7)	(–1.7)
MB	(+)			–0.000	–0.000
				(–0.5)	(–0.5)
Leverage	(+/–)			0.004	0.004
				(0.1)	(0.1)
Analyst Follow	(–)			–0.543 ***	–0.544 ***
				(–5.9)	(–5.9)
Inst Investors	(–)			–0.159 ***	–0.159 ***
				(–4.4)	(–4.4)
Text Length 10K wo RF	(+/–)			–0.013	–0.013
				(–0.8)	(–0.8)
Change RF Text Length	(–)				–0.014
					(–0.6)
Fixed Effects		Date Industry	Date Industry	Date Industry	Date Industry
N		50,330	50,330	50,330	50,330
Adjusted R^2		0.164	0.218	0.238	0.238

This table presents OLS results for the primary sample of matched-pair observations; the matching reflects firm-year-month observations chosen within the highest and lowest quintiles of added and removed risk factors within Item 1A (i.e., "Risk Factors") of 10-Ks filed 2007-2019 ("filing years"). The dependent variable is the logged realized variance risk premium (*Log rVRP*); it has been multiplied by –1 to facilitate inferences. The experimental variable (*High RF Discl*) is an indicator equal to 1 if the sum of a 10-K Item 1A (i.e., "Risk Factors") added and removed risk factors is within the highest quintile of the filing year distribution, and 0 if it is within the lowest quintile. The coefficients of the experimental variable are multiplied by 100. All other variables are defined in Appendix A.

The matching is performed using a one-to-one "Nearest Neighbor" protocol (based on "Mahalanobis Distance"), estimated without replacement. Exact matching is carried out with respect to year, month, industry membership, and the loss indicator; matched control variables (see Table 2) are adjusted through entropy-balancing the first and second moments of each variable's distribution. Continuous variables are winsorized at the 1st and 99th percentiles by filing year. All regressions are estimated with an intercept (not reported). "Date" fixed effects are based on year and month. "Industry" fixed effects are based on the Fama-French 48 classification. Cluster-robust-to-heteroskedasticity *t*-statistics are reported in parentheses; standard-errors are clustered by firm. ***, **, * indicate significance at <0.01, <0.05, <0.10 (two-tailed tests).

TABLE 5
Sensitivity Analyses: Alternative Definitions of Risk-Factor Disclosures

		Dependent Variable: *Log rVRP*			
	Pred. Sign	**(1)**	**(2)**	**(3)**	**(4)**
High RF Discl Tercile	**(–)**	**–3.027 *****	**–2.655 *****		
		(–2.6)	**(–2.6)**		
High RF Discl Median	**(–)**			**–2.007 ****	**–1.793 ****
				(–2.3)	**(–2.3)**
MA Ret Volatility	(–)		–1.229 ***		–1.214 ***
			(–10.1)		(–11.8)
Beta	(–)		–0.015 **		–0.012 **
			(–2.3)		(–2.3)
CFO Volatility	(–)		–0.434 ***		–0.429 ***
			(–4.2)		(–4.9)
Size	(–)		–0.109 ***		–0.103 ***
			(–21.1)		(–23.9)
ROA	(–)		–0.071		–0.071 **
			(–1.5)		(–2.0)
Loss	(–)		–0.032 **		–0.023 **
			(–2.0)		(–1.8)
MB	(+)		–0.000		–0.001
			(–0.0)		(–1.0)
Leverage	(+/–)		–0.006		–0.007
			(–0.2)		(–0.3)
Analyst Follow	(–)		–0.528 ***		–0.623 ***
			(–6.8)		(–9.5)
Inst Investors	(–)		–0.144 ***		–0.135 ***
			(–4.9)		(–5.4)
Text Length 10K wo RF	(+/–)		–0.005		–0.010
			(–0.3)		(–0.8)
Change RF Text Length	(–)		–0.019		–0.026
			(–1.0)		(–1.6)
Fixed Effects		Date Industry	Date Industry	Date Industry	Date Industry
N		81,142	81,142	131,634	131,634
Adjusted R^2		0.167	0.236	0.164	0.233

This table presents sensitivity OLS results for alternatively defined matched-pairs samples of firms disclosing a *high* number and firms disclosing a *low* number of added and removed risk factors within Item 1A of 10-Ks filed 2007-2019 ("filing years"). The dependent variable is the logged realized variance risk premium (*Log rVRP*); it is multiplied by –1 to facilitate inferences. In Columns (1) and (2), the experimental variable is *High RF Discl Tercile*, an indicator variable equal to 1 if the sum of a 10-K Item 1A added and removed risk factors is within the highest (i.e., 3rd) tercile of the filing year distribution, and 0 if it is within the lowest (i.e., 1st) tercile. In Columns (3) and (4), the experimental variable is *High RF Discl Median*, an indicator variable equal to 1 if the sum of a 10-K Item 1A added and removed risk factors is above the median of the filing year distribution, and 0 otherwise. The coefficients of the experimental variables are multiplied by 100. All other variables are defined in Appendix A.

The matching is performed using a one-to-one "Nearest Neighbor" protocol (based on "Mahalanobis Distance"), estimated without replacement. Exact matching is carried out with respect to year, month, industry membership, and the loss indicator; matched control variables (see Table 2) are adjusted through entropy-balancing the first and second moments of each variable's distribution. Continuous variables are winsorized at the 1st and 99th percentiles by filing year. All regressions are estimated with an intercept (not reported). "Date" fixed effects are based on year and month. "Industry" fixed effects are based on the Fama-French 48 classification. Cluster-robust-to-heteroskedasticity *t*-statistics are reported in parentheses; standard-errors are clustered by firm. ***, **, * indicate significance at <0.01, <0.05, <0.10 (two-tailed tests).

TABLE 6
Sensitivity Analyses: Alternative Definitions of the Dependent Variable

		Dependent Variables:			
		Log eVRP		***Log rVRPng***	
	Pred. Sign	**(1)**	**(2)**	**(3)**	**(4)**
High RF Discl	(–)	**–4.145 *****	**–3.747 *****	**–4.041 *****	**–3.654 *****
		(–2.6)	**(–2.5)**	**(–2.4)**	**(–2.4)**
MA Ret Volatility	(–)		–1.197 ***		–1.102 ***
			(–7.7)		(–6.6)
Beta	(–)		–0.028 ***		–0.030 ***
			(–3.4)		(–2.9)
CFO Volatility	(–)		–0.563 ***		–0.384 ***
			(–4.0)		(–2.5)
Size	(–)		–0.106 ***		–0.123 ***
			(–15.1)		(–17.7)
ROA	(–)		–0.144 ***		–0.015
			(–2.4)		(–0.3)
Loss	(–)		0.011		–0.043 **
			(0.5)		(–1.7)
MB	(+)		0.000		0.001
			(0.3)		(0.7)
Leverage	(+/–)		–0.027		–0.022
			(–0.5)		(–0.5)
Analyst Follow	(–)		–0.455 ***		–0.431 ***
			(–4.4)		(–4.1)
Inst Investors	(–)		–0.109 ***		–0.208 ***
			(–2.6)		(–4.8)
Text Length 10K wo RF	(+/–)		–0.023		–0.007
			(–1.2)		(–0.4)
Change RF Text Length	(–)		0.013		–0.009
			(0.6)		(–0.4)
Fixed Effects		Date Industry	Date Industry	Date Industry	Date Industry
N		50,266	50,266	50,330	50,330
Adjusted R^2		0.127	0.199	0.151	0.204

This table presents sensitivity OLS results for alternatively defined matched-pairs samples of firms disclosing a high number and firms disclosing a low number of added and removed risk factors within Item 1A of 10-Ks filed 2007–2019 ("filing years"). In Columns (1) and (2), the dependent variable is *Log eVRP*, the expected or forecasted (rather than realized) Garch-based variance risk premium. In Columns (3) and (4), the dependent variable is *Log rVRPng*, the non-Garch (rather than Garch-based) realized variance risk premium. Both *Log eVRP* and *Log rVRPng* have been multiplied by –1 to facilitate inferences. Across all columns, the experimental variable is *High RF Discl*, an indicator variable equal to 1 if the sum of a 10-K Item 1A added and removed risk factors is within the highest (i.e., 5th) quintile of the filing year distribution, and 0 if it is within the lowest (i.e., 1st) quintile. The coefficients of the experimental variable are multiplied by 100. All other variables are defined in Appendix A.

The matching is performed using a one-to-one "Nearest Neighbor" protocol (based on "Mahalanobis Distance"), estimated without replacement. Exact matching is carried out with respect to year, month, industry membership, and the loss indicator; matched control variables (see Table 2) are adjusted through entropy-balancing the first and second moments of each variable's distribution. Continuous variables are winsorized at the 1st and 99th percentiles by filing year. All regressions are estimated with an intercept (not reported). "Date" fixed effects are based on year and month. "Industry" fixed effects are based on the Fama-French 48 classification. Cluster-robust-to-heteroskedasticity *t*-statistics are reported in parentheses; standard-errors are clustered by firm. ***, **, * indicate significance at <0.01, <0.05, <0.10 (two-tailed tests).

TABLE 7
Sensitivity Analyses: Alternative Matching Strategies

		Dependent Variable: *Log rVRP*			
		PSM Without Replacement		1:1 NN With Replacement	
	Pred. Sign	**(1)**	**(2)**	**(3)**	**(4)**
High RF Discl	(–)	**–4.757 *****	**–4.365 *****	**–4.728 *****	**–4.280 *****
		(–3.2)	**(–3.4)**	**(–2.7)**	**(–2.7)**
MA Ret Volatility	(–)		–1.153 ***		–1.189 ***
			(–8.5)		(–7.4)
Beta	(–)		–0.023 ***		–0.022 **
			(–3.1)		(–2.0)
CFO Volatility	(–)		–0.327 ***		–0.178
			(–3.4)		(–1.1)
Size	(–)		–0.108 ***		–0.105 ***
			(–18.4)		(–14.9)
ROA	(–)		–0.060		–0.052
			(–1.3)		(–1.0)
Loss	(–)		–0.029		–0.038
			(–1.5)		(–1.4)
MB	(+)		0.000		–0.000
			(0.4)		(–0.1)
Leverage	(+/–)		–0.010		0.027
			(–0.3)		(0.5)
Analyst Follow	(–)		–0.552 ***		–0.432 ***
			(–6.2)		(–3.5)
Inst Investors	(–)		–0.132 ***		–0.137 ***
			(–3.8)		(–3.4)
Text Length 10K wo RF	(+/–)		0.000		–0.015
			(0.0)		(–0.7)
Change RF Text Length	(–)		–0.011		–0.009
			(–0.5)		(–0.4)
Fixed Effects		Date Industry	Date Industry	Date Industry	Date Industry
N		50,330	50,330	56,814	56,814
Adjusted R^2		0.159	0.232	0.156	0.223

This table reports sensitivity OLS results for alternatively matched samples of firms disclosing a high number (i.e., 5th quintile of the yearly distribution) and firms disclosing a low number (i.e., 1st quintile of the yearly distribution) of added and removed risk factors within Item 1A (i.e., "Risk Factors") of 10-Ks filed 2007-2019 ("filing years"). The dependent variable (*Log rVRP*) is the logged realized variance risk premium; it has been multiplied by –1 to facilitate inferences. The experimental variable (*High RF Discl*) is an indicator equal to 1 if the sum of a 10-K Item 1A (i.e., "Risk Factors") added and removed risk factors is within the highest quintile of the filing year distribution, and 0 if it is within the lowest quintile. The coefficients of the experimental variable are multiplied by 100. All other variables are defined in Appendix A.

Columns (1) and (2) reflect "Propensity Score Matching" (rather than one-to-one "Nearest Neighbor" based on "Mahalanobis Distance") with no replacement. Columns (3) and (4) reflect one-to-one "Nearest Neighbor Matching" with replacement (rather than without). In both cases, exact matching is carried out with respect to year, month, industry membership, and the loss indicator; weights for matched control variables (see Table 2) are adjusted through entropy-balancing the first and second moments of each variable's distribution. Continuous variables are winsorized at the 1st and 99th percentiles by filing year. All regressions are estimated with an intercept (not reported). "Date" fixed effects are based on year and month. "Industry" fixed effects are based on the Fama-French 48 classification. Cluster-robust-to-heteroskedasticity *t*-statistics are reported in parentheses; standard-errors are clustered by firm. ***, **, * indicate significance at <0.01, <0.05, <0.10 (two-tailed tests).

TABLE 8
Sensitivity Analyses: Placebo Using Changes in Risk Text Length

		Dependent Variable: *Log rVRP*			
	Pred. Sign	**(1)**	**(2)**	**(3)**	**(4)**
Change RF Text Length	**(–)**	**–0.011**	**–0.024**	**–0.024**	**–0.024**
		(–0.5)	**(–1.1)**	**(–1.1)**	**(–1.1)**
MA Ret Volatility	(–)		–1.169 ***	–1.194 ***	–1.194 ***
			(–8.1)	(–8.4)	(–8.4)
Beta	(–)		–0.020 **	–0.019 **	–0.019 **
			(–2.4)	(–2.5)	(–2.5)
CFO Volatility	(–)		–0.408 ***	–0.461 ***	–0.461 ***
			(–3.2)	(–3.5)	(–3.5)
Size	(–)		–0.124 ***	–0.107 ***	–0.107 ***
			(–22.1)	(–17.5)	(–17.5)
ROA	(–)			–0.080	–0.080
				(–1.6)	(–1.6)
Loss	(–)			–0.030	–0.030
				(–1.5)	(–1.5)
MB	(+)			0.000	0.000
				(0.2)	(0.2)
Leverage	(+/–)			–0.000	–0.000
				(–0.0)	(–0.0)
Analyst Follow	(–)			–0.518 ***	–0.518 ***
				(–5.4)	(–5.4)
Inst Investors	(–)			–0.157 ***	–0.157 ***
				(–4.2)	(–4.2)
Text Length 10K wo RF	(+/–)			–0.014	–0.014
				(–0.8)	(–0.8)
Fixed Effects		Date Industry	Date Industry	Date Industry	Date Industry
N		50,330	50,330	50,330	50,330
Adjusted R^2		0.151	0.208	0.231	0.231

This table presents sensitivity analyses of OLS regressions for the primary sample of matched-pair observations; the matching reflects firm-year-month observations chosen within the highest and lowest quintiles of added and removed risk factors within Item 1A (i.e., "Risk Factors") of 10-Ks filed 2007-2019 ("filing years"). The dependent variable is the logged realized variance risk premium (*Log rVRP*)); it has been multiplied by –1 to facilitate inferences. The experimental variable is *Change RF Text Length*, the change in the natural log of the total number of words within 10-K Item 1A (i.e., "Risk Factors") between two consecutive filing years. All other variables are defined in Appendix A.

The matching is performed using a one-to-one "Nearest Neighbor" protocol (based on "Mahalanobis Distance"), estimated without replacement. Exact matching is carried out with respect to year, month, industry membership, and the loss indicator; matched control variables (see Table 2) are adjusted through entropy-balancing the first and second moments of each variable's distribution. Continuous variables are winsorized at the 1st and 99th percentiles by filing year. All regressions are estimated with an intercept (not reported). "Date" fixed effects are based on year and month. "Industry" fixed effects are based on the Fama-French 48 classification. Cluster-robust-to-heteroskedasticity *t*-statistics are reported in parentheses; standard-errors are clustered by firm. ***, **, * indicate significance at <0.01, <0.05, <0.10 (two-tailed tests).