# A FIRM'S OPERATIONAL RISK: DATA SET AND EMPIRICAL EVIDENCE

Forthcoming in *Manufacturing & Service Operations Management*

Vivek Astvansh
Associate Professor of Quantitative Marketing and Analytics,
Desautels Faculty of Management, McGill University

Academic Director, Bensadoun School of Retail Management, McGill University

Adjunct Associate Professor of Data Science, Department of Informatics,
Luddy School of Informatics, Computing, and Engineering, Indiana University Bloomington

Affiliate, Environmental Resilience Institute, Indiana University
vivek.astvansh@mcgill.ca

Joseph Simpson
Collegiate Associate Professor,
Department of Management, Pamplin College of Business
Virginia Polytechnic Institute and State University
jjsimpson@vt.edu

# A FIRM'S OPERATIONAL RISK: DATA SET AND EMPIRICAL EVIDENCE


## ABSTRACT

**Problem Definition**. A firm's stakeholders would benefit from a quantitative measure of the managerial disclosure of its operational risk and the implications of this disclosure. However, a quantitative and scalable measure of a firm's disclosed operational risk is unavailable. Thus, the firm's stakeholders remain unaware of its disclosed operational risk and the risk's implications.

**Methodology/Results**. U.S. law requires a public firm to textually disclose its operational and nonoperational risks in Item 1A of its annual report (i.e., Form 10-K). We train 64 transformer models on U.S. public firms' Item 1A text to score 131,920 firm-years (16,959 firms, 2005 to 2024) on eight risk factors: (1) accounting, (2) finance, (3) international, (4) legal, (5) management, (6) marketing, (7) operations, and (8) technology. We measure each transformer's performance on eight metrics. Next, our Python code retains each risk factor's best-performing transformer (among the eight). Subsequently, our regression estimates report that a firm's disclosed operational risk is positively associated with its operational cost and that its disclosed nonoperational risk strengthens this positive association, thus supporting our two hypotheses. Our OSF repository (https://osf.io/gz93b/files/osfstorage?view_only=9086f51f8704462d8933a8131b22da25) includes an Excel file that contains a data dictionary (dataset_dictionary.xlsx) and count and probability scores of the eight risk factors for 131,920 firm-years (16,959 firms, 2005 to 2024) (dataset.xlsx). The repository also includes our Python code file (01_prepare_data.py) and trained models' files (risk_classifier-1.0.0.zip).

**Managerial Implications**. First, our empirical evidence informs managers and corporate stakeholders that a firm's disclosed operational and nonoperational risks are associated with its operational cost, thus showcasing disclosure's relevance. Second, our Excel data file provides stakeholders with eight risk factors for 131,920 firm-years (16,959 firms, 2005 to 2024). Third, one can also use our Python code files and trained transformers to measure risk reflected in other sources of firm-generated text (e.g., managers' answers in earnings calls, CEO interviews, and press releases).


*Keywords*: dataset article, risk factors, Item 1A, disclosure, Form 10-K

## INTRODUCTION

Managers often disclose their perceptions of their employer firm's risks. These managerial disclosures of firm risks have long been a major source of intrigue for academics (Bao and Datta 2014; Schnatterly, Calvano, Berns, and Deng 2021; Schmidt and Raman 2022), practitioners (PricewaterhouseCoopers 2024), stakeholders (Bloomberg Law 2023), and regulators (SEC 2020). Managers' risk disclosures can lead investors to withdraw support, reducing both long-term and short-term market performance (Schnatterly, Calvano, Berns, and Deng, 2021) or increasing volatility (Karvet and Muslu, 2013). For example, a firm facing supply-chain disruptions but lacking adequate safeguards may cause security analysts to lower their earnings-per-share forecasts for the firm, in part because they anticipate higher costs and other adverse outcomes. Indeed, Houthis' numerous attacks on shipping vessels in 2024 disrupted several firms' supply chains (Feingold 2024). Indeed, many OM academics have advocated that the stakeholder significance of a firm's operational and nonoperational risks cannot be overstated (Simchi-Levi et al. 2015).

Research on a firm's operational and nonoperational risk disclosure is theoretically interesting (Kalatha, Dontis-Charitos, Kosmidou, and Andriosopoulos 2023). On the one hand, stakeholders may interpret a firm's narrative/textual disclosure of operational risk as managers' inability to prevent it sufficiently. On the other hand, the disclosure may suggest that the managers are aware of and have the know-how to reduce the risk's consequences (Schmidt and Raman 2022). In addition to these two-sided theoretical arguments, risk disclosures provide a practical utility. Many commentators have argued that a firm's annual reports are largely "boilerplate" (Cazier, McMullin, and Treu 2021) and meant to fulfill the legal mandate and lower the firm's litigation risks rather than boost transparency. This cynical perspective suggests that a firm's risk disclosure may be biased at best or fraudulent at worst (Gaulin 2017). Conversely, others have noted that firms accurately and honestly report their material risks (Cazier, McMullin, and Treu 2021). Regardless of which perspective one takes, the legal mandate offers reason to believe that a firm's risk disclosure may carry relevant information that can determine the firm's operational cost (Lopez-Lira 2021). Therefore, our empirical manuscript answers the question: *How is a firm's disclosed operational risk*[1] *associated with its operational cost?*

The foremost impediment to answering this question is measuring operational risk for a representative set of firms for a representative number of periods. The United States Securities Exchange Act of 1934 requires a U.S. public firm to report its annual financial performance by filing Form 10-K prescribed by the U.S. Securities and Exchange Commission (SEC) (https://www.sec.gov/files/form10-k.pdf). The form is organized into multiple "items." In 2005, the SEC announced Regulation S-K, which introduced a new item/section in the Form 10-K: Item 1A, titled "Risk Factors." Via this announcement,

[1] We focus on managers' disclosure of their firm's risks based on their perceptions. In other words, our research is unrelated to a firm's realized risks.

the SEC mandated a U.S. public firm to disclose "the most significant factors that make the company speculative or risky. The risk factor item is intended to provide investors with a clear and concise summary of the material risks to an investment in the issuer's securities" (SEC 2005, p. 257).[2] The mandate is legally mentioned in Item 503(c) of Regulation S-K (also known as Section 229.503(c) of Title 17 of the U.S. Code for Federal Regulations) (see online appendix for a narrative summary of research that has discovered a firm's risk factors from Form 10-K).

We extract Item 1A from the near-universe of 131,920 firm-years' Form 10-Ks, covering 16,959 firms, observed from 2005 to 2024. Next, we follow Schnatterly, Calvano, Berns, and Deng's (2021) eight risk factors of (1) accounting, (2) finance, (3) international, (4) legal, (5) management, (6) marketing, (7) operations, and (8) technology. We apply Schnatterly, Calvano, Berns, and Deng's (2021) eight risk factor dictionaries to a firm-year-specific Item 1A text, receiving the number of words in the dictionary *and* the focal item. This number is the quintessential count measure that many academics use (Schnatterly, Calvano, Berns, and Deng, 2021; Campbell, Chen, Dhaliwal, Lu, and Steele, 2014). It is face-valid because a user can check whether the count matches the input text.

Next, we employ the following five-step procedure to build machine learning-based transformer models (hereafter, transformers, for brevity) to quantify firm-year-specific eight risk factors (Astvansh 2025a, 2025b; Damavandi, Mai, and Astvansh 2025). First, we input Schnatterly, Calvano, Berns, and Deng's (2021) dictionary to ChatGPT (version o1 pro) and prompt it to generate 761 sentences that include the dictionary terms. Second, we wrote Python code to label the simulated sentence on the presence (vs. absence) of each of the eight risk factors. Third, we use the labeled sentences to train eight transformers for each risk factor (thus, 64 transformers). We obtained eight performance metrics (e.g., accuracy, precision, recall, F-1 score, and loss) for each transformer. Fourth, our Python code selects the best-performing transformer (among the eight we trained) for each risk factor. Fifth, we execute the best-performing trained transformer on 131,920 firm-year-specific Item 1A, spanning 16,959 firms from 2005

[2] See research on Item 1A by Bao and Datta (2014); Campbell, Cecchini, Cianci, Ehinger, and Wener (2019); Kravet and Muslu (2013); Matsumura, Prakash, and Vera-Muñoz (2024); Schnatterly, Calvano, Berns, and Deng (2021)

to 2024. This execution outputs the predicted probability of an Item 1A mentioning a risk factor (eight probabilities for each firm-year).

We hypothesize that a firm's disclosed operational risk in a year is positively associated with its operational cost in the following year. Further, its nonoperational risk strengthens this positive association. Estimates from fixed-effects regressions provide robust support for the two hypotheses, thus showcasing the relevance of our measures of firm-year-specific operational and nonoperational risks.

We offer two contributions to the empirical literature on operational risk (e.g., Simchi-Levi et al. 2015). First, the bulk of this literature has considered (1) singular, rare, and exogenous shocks that disrupt a firm's supply chain (Hendricks, Jacobs, and Singhal 2020) or (2) firm announcements of supply-chain glitches (Hendricks and Singhal 2003, 2014). We provide novel, count and machine-learning-based probability measures of a firm's *disclosed operational* risk (Bai, Gao, and Sarkis 2020). Following prior dataset articles (e.g., Astvansh, Ball, and Josefy 2022; Damavandi and Astvansh 2024), our data file provides measures of operational and seven types of nonoperational risks for 131,920 firm-years, spanning 16,959 firms from 2005 to 2024. Our measure is broader in scope because it encompasses operational risks, rather than supply-chain glitches/disruptions. Further, our measure generalizes to multiple firms over multiple years. Lastly, our data capture managers' mandated firm risk disclosures based on their perceptions, rather than voluntary press releases, and is thus representative of the population.

Second, prior research has shown that a firm's supply-chain disruptions impact its stock returns (e.g., Hendricks, Jacobs, and Singhal 2020), stock return risk (Hendricks and Singhal 2014), and accounting outcomes (Hendricks and Singhal 2005a). We complement this evidence by demonstrating that a firm's disclosed operational risk is positively associated with the firm's operational cost, a key OM outcome. The substantive insight is that a firm's risk disclosure conveys OM-relevant information. Stated differently, the firm's disclosure is not a boilerplate (Campbell, Chen, Dhaliwal, Lu, and Steele 2014; Hope, Hu, and Lu 2016; Kravet and Muslu 2013) but is associated with its actual OM outcome.

# THEORY AND HYPOTHESES

## Operational and Nonoperational Risk

*Operational risk* refers to the possibility of losses or performance shortfalls arising from disruptions in a firm's core processes, production systems, and supply chain activities (Hora and Klassen 2013; Sodhi, Son, and Tang 2012). This risk arises from internal inefficiencies, supplier delays, quality-control failures, human errors, or external shocks, such as geopolitical instability and natural disasters. The empirical OM literature has documented operational risk's consequences. For example, a firm's supply-chain disruptions lower its stock returns (Hendricks and Singhal 2003), raise its stock-return risk (Hendricks and Singhal 2005b, 2014), and impede its accounting metrics (Hendricks and Singhal 2005a).

Anecdotes support empirical evidence. For example, when the 2011 earthquake and tsunami in Japan disrupted Toyota's supply chain, the firm faced extensive production halts, resulting in global inventory shortages and a substantial increase in costs (Webb 2016). Similarly, unexpected bottlenecks in the semiconductor supply chain led to production shortfalls, delaying manufacturing operations worldwide (Wishart-Smith 2024). Such examples underscore that operational risks directly affect a firm's operational cost, often requiring substantial financial resources to mitigate the consequences of production slowdowns, inefficiencies, or supplier failures.

A firm's operational risks directly impact its ability to produce goods or deliver services efficiently, making them highly salient for operations managers. In contrast, *nonoperational risk* originates outside a firm's immediate operational processes and typically includes financial, strategic, legal, and technological factors (Kalatha, Dontis-Charitos, Kosmidou, and Andriosopoulos 2023). More concretely, it includes (1) regulatory changes that impact a firm's accounting, (2) changes in customer tastes and preferences, (3) exchange rate volatility, (4) geopolitical instability, (5) litigation, and (6) industry-wide regulatory changes (Peker, Murat, Erol, and Searcy 2023; Schmidt and Raman 2022). Whereas operational risks affect production and supply chain processes, nonoperational risks impose financial or strategic constraints that influence managerial decision-making.

For instance, currency fluctuations in emerging markets pose significant nonoperational risks for multinational firms that rely on global supply chains. Similarly, policy changes such as trade restrictions, tariffs, or environmental regulations impose financial burdens on companies, which must adjust their strategies accordingly (Astvansh, Eshghi, Shahriari, and Shi 2025; Schmidt and Raman, 2022). During the U.S.–China trade war, firms in the manufacturing sector faced higher import tariffs on raw materials, increasing operational costs without any direct change in production efficiency. Likewise, industries such as pharmaceuticals are subject to legal risks due to patent disputes and changing regulatory requirements, which can force firms to redirect financial resources away from core operations.

Operational and nonoperational risks can increase firms' costs, but they do so through different mechanisms. Operational risk affects cost efficiency by disrupting production and supply chains, whereas nonoperational risks impose external constraints that force firms to divert resources toward compliance, legal proceedings, or financial restructuring (Schmidt and Raman, 2022). Given these distinctions, understanding how firms disclose these risks provides insights into their expected future costs.

## Signaling Theory

Signaling theory provides a framework for understanding how firms communicate private information to external stakeholders in contexts characterized by information asymmetry. Spence (1973) argued that decision-makers use signals to infer unobservable qualities in times of uncertainty. A key feature of effective signals is that they must be costly or difficult for lower-quality firms to imitate. For example, a person's education credentials signal productivity since obtaining a degree requires substantial effort and investment (Spence 1973).

Academics have applied signaling theory to explain how firms communicate with investors, regulators, and other stakeholders (Astvansh, Eshghi, Shahriari, and Shi 2025; Connelly, Certo, Ireland, and Reutzel 2011). A firm's managers possess superior knowledge about its risks and prospects. External stakeholders, such as investors and suppliers, lack direct access to this information. Consequently, the firm's managers use disclosures as signals to manage stakeholder expectations and reduce uncertainty (Connelly, Certo, Reutzel, DesJardine, and Zhou 2025). According to signaling theory, a firm with stronger risk management practices is more likely to voluntarily disclose risks because such disclosure can credibly demonstrate managers' ability to mitigate risks. In contrast, weaker firms may withhold information to avoid scrutiny (Heil and Robertson 1991).

A signal is effective if it is observable and credible to external stakeholders. In the case of risk disclosures, investors and analysts interpret a firm's reported risks as indicators of potential operational and nonoperational instability. However, not all disclosures carry the same weight. A firm that systematically underreports risks may be perceived as untrustworthy, whereas a firm that overstates risks may raise unnecessary alarms among stakeholders.

## Linking Signaling Theory to Disclosed Operational and Nonoperational Risks

Risk disclosures serve as strategic signals that shape stakeholder perceptions of a firm's performance and capabilities. Operational and nonoperational risk disclosures convey different types of information to investors, suppliers, and regulators. The decision to disclose risks is nonrandom. A firm chooses how much information to reveal based on the expected costs and benefits of signaling (Connelly, Certo, Ireland, and Reutzel 2011).

Disclosing operational risk signals that the firm faces substantial internal or supply chain uncertainties that could result in higher costs. Given that operational risk directly impacts operational efficiency, a firm that discloses significant operational risk is likely preparing stakeholders for future disruptions. Prior research has shown that operational failures, such as supply-chain bottlenecks, increase operational costs due to recovery efforts, lost sales, and reputational damage (Hendricks and Singhal 2003). Investors interpret such disclosures as indicators that firms may face challenges in maintaining cost efficiency, reinforcing the expectation that higher disclosed operational risk is associated with increased future operational costs.

In contrast, disclosed nonoperational risks provide broader contextual information about external pressures the firm faces. A firm that discloses high operational and nonoperational risks signals stakeholders that it operates in an environment with compounding challenges. Consider firm A, which faces supply-chain disruptions (operational risk) and geopolitical uncertainty (nonoperational risk).

Contrast this firm with its peer firm B, which faces only supply-chain disruptions. Investors may anticipate that firm A's risks will yield a stronger impact on its operational costs than firm B's. Furthermore, a firm with high nonoperational risks may have fewer financial or strategic resources to absorb operational disruptions, amplifying the cost implications of disclosed operational risks. These insights lead to two testable hypotheses on how a firm's risk disclosures relate to subsequent operational costs.

*Hypothesis 1: A firm's disclosed operational risk is positively associated with its operational cost.*

*Hypothesis 2. The higher a firm's disclosed nonoperational risk, the stronger the positive association between its operational risk and operational cost.*

## DICTIONARY AND MACHINE LEARNING-BASED MEASURES OF FIRM RISK

### Obtaining Firm-Specific Risk Disclosure Text

Each public firm in the United States files annually a Form 10-K with the U.S. Securities and Exchange Commission (SEC) (https://www.sec.gov/files/form10-k.pdf). This form constitutes the firm's annual report to its shareholders. The SEC uses the Central Index Key (CIK) to identify each firm uniquely.

We followed a three-step procedure to download the near-universe of Form 10-Ks with an Item 1A. First, in January 2025, we downloaded 29,912 CIKs from Standard & Poor's Compustat – Capital IQ North America Fundamentals Annual database. Second, we used https://pypi.org/project/python-edgar/ to download the Form 10-Ks for 21,821 (of the 29,912) CIKs. Third, the SEC mandated firms to disclose risk factors in Item 1A (rather than anywhere) in Form 10-Ks filed effective December 1, 2005 (Nelson and Pritchard 2007). Therefore, we excluded Form 10-Ks with a filing date earlier than December 31, 2004. This step led us to 133,394 Form 10-Ks filed between 2005 and 2024. Further, 9,539 Form 10-Ks (approximately 7%) filed between 2005 and 2024 did not mention "Item 1a" (case insensitive).[3] We

[3] Of the 9,539 Form 10-Ks that did not include Item 1A, 5,480 (57%) were filed in 2005. We speculate two reasons for this missingness. First, the SEC made Item 1A mandatory for Form 10-Ks filed effective December 1, 2005. So, Form 10-Ks that were filed between January 1, 2005 and November 31, 2005 may not have included Item 1A. Second, "small reporting firms" are not required to fill Item 1A (https://www.sec.gov/resources-small-businesses/going-public/smaller-reporting-companies). We manually inspected approximately 10% of the filings that did not contain Item 1A to ensure that we captured these sections correctly.

removed duplicate Form 10-Ks for CIK-Year.[4] The procedure left us with 131,920 firm-year Form 10-Ks, spanning 16,959 CIKs/firms, observed from 2005 to 2024.

Next, we collected text on firm-year-specific *risk factors* in three steps. First, we "cleaned" the Form 10-K HTML files by removing HTML tags to make the files easier to parse using the Python libraries *beautifulsoup4* and *re* (Lesmy, Muchnik, and Mugerman 2019). While *beautifulsoup4* helps extract data from markup files, the *re* library ("re" is short for regular expression) helps search for regular expressions and extract results. Second, we converted the cleaned HTML files to their text (TXT) equivalents. Third, following Schnatterly, Calvano, Berns, and Deng's (2021) procedure[5], we wrote Python code to extract *Item 1A. Risk Factors* from each TXT file.

**Dictionary-based measures of eight dimensions of risk**

We used Schnatterly, Calvano, Berns, and Deng's (2021) dictionaries of eight risk factors (see Table A1 in the authors' web appendix). These factors are (1) accounting, (2) finance, (3) international, (4) legal, (5) management, (6) marketing, (7) operations, and (8) technology. More precisely, we counted the terms matching between a risk factor's dictionary and Item 1A (Short, Broberg, Cogliser, and Brigham 2010). We also recorded the (1) *CIK*, (2) *Year*, (3) *Item 1A word count*, and (4) *Item 1A optional* values for each Item 1A. *Item 1A word count* is the number of words in an Item 1A. Table 1 presents the variables' descriptive statistics. *Item 1A optional* equals 1 when a registrant is legally permitted to omit the risk-factor section (e.g., smaller reporting companies under Regulation S-K §229.10(f)(1)), and 0 otherwise.

[Insert Table 1 about here.]

**Generating Synthetic Data**

We generated synthetic data (specifically, 761 risk-mentioning sentences) using ChatGPT. Web Appendix B in the online appendix describes our procedure. Next, we wrote Python code to label each

---

[4] A U.S. public firm files a Form 10-K each year. However, the firm may later amend Form 10-K voluntarily or upon being asked by the SEC. These revised Form 10-Ks may lead to more than one Form 10-K for a CIK-year.

[5] Schnatterly, Calvano, Berns, and Deng's (2021) procedure involves searching for "Item 1A" as the marker for the beginning of the item and either "Item 1B" or "Item 2" as the marker for the end of the item.

sentence on eight risk factors. Table 2 reports six example sentences and their eight labels' 0 versus 1 value. It illustrates that some sentences can include multiple risk factors. For example, the sentence "Technological integration challenges during mergers or acquisitions could disrupt our operations" is labeled 1 for the technology risk, management risk, and operations risk factors but 0 for all remaining six risk factors.

[Insert Table 2 about here.]

**Eight Transformers for Each Risk Factor and Eight Performance Metrics**

Following Miric, Jia, and Huang (2023)[6], we built eight transformers to measure each of the eight firm-year-specific risk factors; thus, we trained 64 transformers. Next, we used eight performance metrics to evaluate the transformers (Miric, Jia, and Huang 2023; Rathje, Katila, and Reineke 2024; Fyffe, Lee, and Kaplan 2024). Specifically, we used (1) accuracy, (2) precision (micro and macro), (3) recall (micro and macro), (4) F-1 score (micro and macro), and (5) loss. Web Appendix B provides the details.

Table 3A reports the eight performance metrics for the eight transformers for the operational risk factor (as an example).[7] Six transformers performed well, displaying near-perfect performance across multiple metrics, with our models outperforming those in prior management research (e.g., Miric, Jia, and Huang 2023), likely due to our synthetically created training data set. The variation in loss across different models underscores the importance of model selection, as some architectures significantly outperform others in their ability to classify firm-year-specific risks. More specifically, nlpaueb/legal-bert-base-uncased and roberta-large score a zero on seven of the eight metrics and a loss of about .5. These models perform similarly poorly for the other seven risk factors. We conjecture that these two transformers are ill-suited for classifying firm risk factors using our code.

[Insert Table 3A about here.]

---

[6] We removed PatentSBERTa and BioBERT because they are less relevant to our context.
[7] In the interest of space, we do not provide similar tables for the other seven risk factors.

We wrote a Python code that computed the average of the eight performance metrics for each of the 64 transformers. Next, the code selected the transformer with the highest average for each of the eight risk factors. Table 3B reports the selected transformer for each risk factor.

[Insert Table 3B about here.]

For each firm-year-specific Item 1A and each risk factor, the best-performing transformer produced a probability ranging from 0.00 to 1.00, indicating the model's confidence that the firm-year-specific Item 1A is associated with the focal risk factor. In addition, the transformer assigned a binary value of 0 if the probability was below 0.5 or 1 if it was at or above .5. This threshold reflects the convention in prior research that a probability of at least 0.5 means the positive class is assigned. The predicted probabilities provide our data set users with continuous variables and thus allow them to explore determinants and consequences of the variation in the eight risk factors (1) within an annual report, (2) across annual reports (over the years) within a firm, and (3) across annual reports (within a year) across an industry or a geographical region of headquarters. Tables 4a and 4b below display the descriptive statistics for the variables included in our data set.

[Insert Tables 4A and 4B about here.]

Figure 1 summarizes the nine steps in our procedure and Table B1 (in the online appendix) describes each step.

[Insert Figure 1 about here.]

## DATA AND METHOD

### Sample

We merged the firm-year-specific eight risk probabilities predicted by our best-performing transformers with data from Standard & Poor's Compustat-Capital IQ North America Fundamentals Annual. The merger resulted in 131,920 firm-years, encompassing 16,959 firms (identified by CIK) observed from 2005 to 2024.

**Dependent Variable (DV): Operational cost$_{i,y}$**

We measure firm-year-specific operational costs by the *Cost of Goods Sold* (COGS) (Hu, Skowronski, Dong, and Shou 2022; Li, Lam, Ho, and Yeung 2022; Rumyantsev and Netessine 2007). We do not divide COGS by a firm fundamental variable (e.g., firm size or gross margin) because methodologists (Certo, Busenbark, Kalm, and LePine 2020; Certo, Jeon, Raney, and Lee 2024) have expressed concerns with ratio DVs. In brief, an increase in a ratio variable's values could be caused by an increase in the numerator, a decrease in the denominator, or both, thus preventing unambiguous interpretation.

We natural logarithm-transformed (ln) the DV to reduce skewness and kurtosis (Rumyantsev and Netessine 2007). Additionally, we winsorized the values at the 1$^{st}$ and 99$^{th}$ percentiles to limit the influence of outliers. Lastly, we measured our DV in fiscal year $y + 1$, where $y$ is the year we measure all regressors. The lagged regression allows the DV to follow the regressors temporally, thus reducing concerns about reverse causality.

**Explanatory Variables (EV): Operational risk$_{i,y}$**

*Operational risk*$_{i,y}$ is the probability that Item 1A in the firm *i*'s Form 10-K in fiscal year *y* mentions words related to operational risk.

**Moderator: Nonoperational risk$_{i,y}$**

*Nonoperational risk*$_{i,y}$ is the sum of the probabilities that Item 1A in the firm *i*'s Form 10-K in fiscal year *y* mentions words related to five nonoperational risks: (1) accounting, (2) finance, (3) international, (4) legal, (5) management, (6) marketing, and (7) technology.

**Control Variables**

Our regression includes four firm-year-specific control variables that account for confounding factors and mitigate potential bias caused by correlated omitted variable(s). First, larger firms are likely more efficient and capable of reducing operational costs (Lévesque, Joglekar, and Davies 2012). Therefore, our regression controls for *Firm size*$_{i,y}$, the natural log of the firm *i*'s revenue in year *y*. Second, firms with higher *R&D* likely incur higher costs (Lévesque, Joglekar, and Davies 2012). Therefore, we control for *R&D*$_{i,y}$. Third, our regression includes *Item 1A optional*$_{i,y}$, which equals 1 if the filling Item 1A is optional for the firm-year because of small firm size (PricewaterhouseCoopers 2023; U.S. Securities and Exchange Commission n.d.). Fifth, we control for *Item 1A word count*$_{i,y}$, the number of words in firm *i*'s Item 1A in year *y*.

Table 5 lists the variables, their measures, reasons for inclusion (if regressor), and reference research. Table 6 reports the variables' descriptive statistics.

[Insert Tables 5 and 6 about here.]

**Regression Specification**

First, our regressions include firm-specific-(*CIK*)-fixed effects to control for firms' unobserved, time-invariant characteristics that may influence a firm's operational risk and/or operational cost. Second, including fiscal year-(*fyear*) fixed effects accounts for macroeconomic or systemic factors that vary across time but affect all firms in a year, ensuring that our results are not driven by temporal shocks or trends. Third, we address potential serial correlation and heteroscedasticity in the data by estimating standard errors (SEs) clustered at the firm level (i.e., CIK). Clustering accounts for the possibility that observations within the same firm are not independent, ensuring that our statistical inferences are robust and appropriately conservative. This approach aligns with common practice in panel data analysis (e.g., Huang, Li, Shen, and Wang 2022; Peterson 2009), particularly when working with firm-level data sets where cross-sectional unit- and time-unit fixed effects and clustered standard errors are crucial for obtaining unbiased and reliable estimates. We estimate the regression using an OLS estimator as follows:

$$\ln(\text{COGS})_{i,y+1} = \beta_0 + \beta_1 \times \ln(\text{Firm Size})_{I,y} + \beta_2 \times \text{Item 1A Optional}_{i,y} + \beta_3 \times \ln(\text{Item 1A Word Count})_{i,y} + \beta_4 \times (\text{R\&D Intensity})_{i,y} + \beta_5 \times \ln(\text{Operational Risk})_{i,y} + \alpha \times \text{CIK}_i + \gamma \times \text{Firm Year}_y + \epsilon_{i,y+1}$$

## RESULTS

**Model-Free Evidence**

We grouped our firm-years into four groups, arranged by operational risk quartile. Next, we created a bar graph in which a bar's height is the average operational cost for firm-years in the specific quartile. Figure 2 (the bar graph) reports that as one moves from firm-years in a lower operational risk quartile to a higher quartile, the bars become taller, suggesting a positive association between operational risk and operational cost.

[Insert Figure 2 about here.]

**Regression Estimates**

We estimate four regressions that test the robustness of the coefficients of our regressors (see Table 7). Regression 1 includes all covariates plus firm- and year-FEs. Regression 2 builds on Regression 1 by including *Operational risk*. Regression 3 further includes *Nonoperational risk*. Lastly, Regression 4 includes *Operational risk* × *Nonoperational risk*. The associations of control variables are consistent with intuition. The larger a firm, the greater its operational cost. Firms for which filling out Item 1A is optional incur lower operational costs. Word count and R&D are insignificantly associated with operational cost.

We use estimates from Regression 4 to report our hypothesis tests. A firm's *Operational risk* in a year is positively associated with its *Operational cost* in the following year ($b = .247$, $p < .01$), thus supporting our $H_1$. This relationship supports the argument that firms acknowledging higher operational risk in their disclosures tend to experience increased operational costs in the following year. Thus,  risk disclosures are not merely compliance-driven boilerplate statements but provide meaningful signals about the firm's underlying vulnerabilities. Further, its *Nonoperational risk* amplifies this positive association ($b = .298$, $p < .01$). This result illustrates a compounding effect; when firms disclose high operational and nonoperational risks, the cost implications are amplified. This finding highlights the interdependent nature of risk management, where challenges in nonoperational domains (e.g., financial, legal, or technological risks) may exacerbate operational inefficiencies, potentially due to resource constraints or managerial attention being diverted across multiple risk areas.

[Insert Table 7 about here.]

## ROBUSTNESS CHECKS

**Robustness to Count (In Place of Probability) Measures and Alternative Fixed Effects**

We report two robustness checks. First, we estimate our four regressions using dictionary-based count measures of *Operational risk* and *Nonoperational risk*. *Operational risk* continues to be positively associated with *Operational cost* ($b$ = .117, $p$ < .01). However, *Nonoperational risk*'s moderation does not reach statistical significance ($b$ = .000, $p$ > .10). Second, our main regression includes *Firm*-FEs and *Year*-FEs. We test robustness to the inclusion of (1) *Industry* (classified by four-digit Standard Industrial Classification [SIC] codes)-FEs and *Year*-FEs and (2) *Industry* × *Year* FEs and *Year* FEs. The coefficients of *Operational Risk*'s and *Operational Risk* × *Nonoperational risk*'s coefficients' sign and significance are consistent with those in Table 7. Third, we control for *Operational risk*'s and *Nonoperational risk*'s endogeneity by estimating a two-stage least-squares regression that uses heteroskedasticity-based instruments. The coefficients maintain signs and significance.

[Insert Table 8A and 8B about here.]

**Robustness to an Instrumental Variable Two-Stage Least-Squares Regression**

A firm's operational and nonoperational risks' endogeneity can bias our estimates and threaten causality claims (Ho, Lim, Reza, and Xia 2017; Ketokivi and McIntosh 2017; Lu, Ding, Peng, and Chuang 2018; Yılmaz, Son, Shang, and Arslan 2024). However, we believe endogeneity poses minimal risk to our results' validity for three reasons (e.g., Busenbark, Yoon, Gamache, and Withers 2022; Frank, Maroulis, Duong, and Kelcey 2013; Gamache, Devers, Klein, and Hannigan 2023). First, we employed high-dimensional fixed effects regressions that absorb firm-level (*CIK*) and year-level (*fyear*) fixed effects. Together, these fixed effects mitigate the risk of omitted variable bias, a central component of endogeneity.

Second, we designed our analysis to ensure temporal precedence by measuring the DV at $y$ +1. A lagged (vs. contemporaneous) regression ensures that the regressors temporally precede the DV, mitigating reverse causality concerns. Also, this temporal sequencing reduces simultaneity bias, a key source of endogeneity in empirical research.

Third, we conducted a formal test of the Impact Threshold of a Confounding Variable (ITCV) (e.g., Busenbark, Langer, and Certo 2017; Frank, Maroulis, Duong, and Kelcey 2013), which provides robust evidence that omitted variable bias is unlikely to threaten the validity of our findings. The ITCV assesses the robustness of the observed relation by quantifying the percentage of statistically significant treated observations that would need to be replaced by null effects under an omitted variable to bias the overall relationship. Our results show that 51.54% (33,785) of significant cases would need to be altered for the findings to become spurious, representing a high invalidation threshold. This finding provides strong evidence that the observed relations are robust to omitted variables, even under scenarios of potential confounding. We conducted the ITCV analysis at a significance level of $\alpha = 0.10$, validating our confidence in the results. Collectively, these methodological choices, high-dimensional fixed effects, temporal sequencing, and ITCV analysis work together to address endogeneity concerns. These measures suggest that our results are robust, minimizing the likelihood that endogeneity significantly biases our findings.

Notwithstanding the above reasons, we employ an instrumental variable (IV) approach to address potential endogeneity in a firm's disclosed operational risk. Specifically, we instrument each firm's operational risk measure with the corresponding industry-year average, assuming that industry-level trends in risk language correlate with the firm's disclosure (thus satisfying relevance) yet are exogenous to the firm's idiosyncratic cost shocks (satisfying exclusion). We follow prior OM research (e.g., Adamopoulos and Todri 2024; Bonnett and Heim 2024; Sharma and Goradia 2023) to also include heteroskedasticity-based instruments. The first-stage regression yields an F-statistic of 1224.78 ($p < 0.01$), indicating that the instruments meet the relevance criterion. The coefficient on instrumented operational risk is 0.222 ($p < 0.01$), suggesting that the estimate in Table 7 is robust to controlling for endogeneity.

[Insert Table 9 about here.]

## ILLUSTRATIVE OM-RELEVANT RESEARCH QUESTIONS THAT OUR DATA SET CAN HELP ANSWER

### Does Disclosed Operational Risk Predict Firm Performance Outcomes?

We show that a firm's disclosed operational and nonoperational risks are associated with its operational costs. Future research can extend our empirical evidence to other firm performance outcome variables (e.g., Campbell, Chen, Dhaliwal, Lu, and Steele 2014; Certo, Jeon, Raney, and Lee 2024; Kravet and Muslu 2013). This research would benefit from measuring the content, frequency, timing, and strategic intent behind operational risk disclosures, potentially combining our data file with real-time data from earnings calls, social media, or press releases. By integrating stakeholder theory and signaling theory, scholars could investigate whether certain patterns of operational risk disclosure build or erode stakeholder trust, alter investment behavior, or influence the extent of regulatory scrutiny. Such insights would help elucidate the broader consequences of risk disclosure strategies and offer a more comprehensive understanding of how firms can navigate complex stakeholder environments. For example, does a healthcare firm's disclosed operational risk predict a decline in patients' demand for the firm's healthcare offerings in the subsequent year? Similarly, does a retailer's operational risk explain its inventory holdings? Last, how does a firm's disclosed operational risk impact its suppliers' decisions toward the focal firm? Do managerial perceptions of the firm's risk raise the firm's resilience and adjust the firm's governance of and responses to supply-chain disruptions?

**Does Risk Disclosure Vary by COO Tenure?**

A Chief Operating/Operations Officer's (COO's) tendency to disclose negative information might vary by tenure. For example, a COO—particularly one who has been hired from outside the firm as opposed to internally promoted—may be more likely to disclose operational risk in the early stage of their career so they could implicitly pass the risk to the earlier COO (Darby, Ketchen, Ball, and Mukherjee 2023; Kothari, Shu, and Wysocki 2009). However, as the COO's tenure increases, they may become less likely to disclose, lest the disclosure hurt their compensation and future career and raise the odds of the board firing the COO. Future research could relate our data file to the COO's tenure and their other characteristics.

**Are the SEC's Changes to Item 1-A Facilitating Disclosure or Impeding It?**

Starting in 2005, the SEC mandated firms to disclose their risk in Item 1-A. The underlying aim was to induce firms to provide shareholders with more information about their risk (SEC 2013, 2016). However, some evidence suggests that, driven by litigation concerns, firms are reported to have increased the quantity of text while not improving its quality (Beatty, Cheng, and Zhang 2019; Dyer, Lang, and Stice-Lawrence 2017). In addition, firms are accused of using "too broad and generic" words and thus not insightful for shareholders (Johnson 2010). The resulting "disclosure overload" is impeding rather than facilitating transparency between firms and shareholders. In contrast, other evidence has suggested an unintended benefit of the disclosure mandate (Huang, Shen, and Zang 2022). Interestingly, the SEC amended Item 1A in 2020, requiring firms to provide "summary risk factor disclosure of no more than two pages if the risk factor section exceeds 15 pages" and organize risk factors "under relevant headings" (SEC 2020, p. 1). Future research could test whether Item 1A has become more informative starting in 2021.

**Do Pre-Filing Firm Announcements and COO Behaviors Predict a Firm's Disclosed Operational Risk?**

Campbell, Chen, Dhaliwal, Lu, and Steele (2014) showed that a firm's fundamentals before its Form 10-K filing are positively associated with its disclosed risk factors. For example, a firm's size is positively associated with the magnitude of legal risk it discloses in Item 1A. Future research may use our data set to extend Campbell, Chen, Dhaliwal, Lu, and Steele's (2014) findings by considering firm announcements and COO behaviors as precursors to firm risk. For example, consider a firm that anticipates a supply-chain disruption. Such anticipation will be reflected in an increase in the firm's disclosed operational risk in Item 1A, likely eliciting a penalty from its shareholders. The firm's managers anticipate such a reaction. They may strategically time their announcements of opening new stores or manufacturing units, thereby releasing positive news before disclosing operational risks in their Form 10-K. Future research could utilize our firm-year-specific risk factors to assess whether managers follow this impression management strategy to weaken the shareholder penalty for disclosing operational risk.

**Forward-Looking Statements in Item 1A**

Prior research has shown that the SEC's mandatory risk disclosure has elicited firms' voluntary disclosure of optimistic performance in the future (Huang, Shen, and Zang 2022). If a firm makes an optimistic statement about its future performance (e.g., revenue, profit, dividend, earnings per share), and the performance is inconsistent with the prior statement, its shareholders may sue it for securities fraud. The shareholders attribute their losses to managers' optimistic disclosures. The firm's litigation risk has curbed its forward-looking, optimistic statements, thus impeding transparency in the market. In 1995, the U.S. Congress enacted the Private Securities Litigation Act in response to this impediment. The law included a safe harbor provision for firms that made forward-looking statements. Future research could use our XLSX supplementary data file to measure a firm's Item 1A on verb tense, thus determining the proportion of forward-looking (or future-oriented) statements.

### Item 1A's Other Characteristics

A firm could create subheadings for each type of risk factor. Such subheadings could improve transparency in the capital market (Huang, Wang, and Yen 2023) and hopefully improve firm outcomes. Alternatively, future research could measure the text on semantics and style (SEC 2013, 2016). Semantics relates to words that convey meaning (e.g., nouns, adjectives, adverbs, main verbs), and style relates to how the writer orders the words. Such research would test whether firms have followed the SEC's handbook (SEC 1998). Research in cognitive linguistics has shown that the writer's psychological state can manifest in their choice of words and the order of these words. For example, what triggers a firm to use passive voice while describing risk factors in passive voice, and does such usage impact its outcomes?

## DISCUSSION

### Contributions

First, we contribute to the empirical literature on operational risk (e.g., Babich and Kouvelis 2018; Craighead et al. 2007; Simchi-Levi et al. 2015). Most of this literature has examined low-probability, high-impact events (e.g., an earthquake) that disrupt a firm's supply chain (e.g., Hendricks, Jacobs, and Singhal 2020) or the firm's announcements of supply-chain disruptions (Hendricks and Singhal 2003; Hendricks, Singhal, and Zhang 2009). We contribute by providing a novel, transformer model-based measure of firm-year-specific operational and nonoperational risks. Our measure captures *all* types of operational risk, not merely one triggered by supply-chain disruptions. Furthermore, it utilizes legally required disclosure, thereby covering almost all U.S. public firms from 2005 to 2024, rather than merely firms that voluntarily reported their risk. Further, prior research's focus has been on disruptions' impacts on the firm's accounting metrics (Hendricks and Singhal 2005a), stock returns (Hendricks and

Singhal 2003; Hendricks et al 2009), and stock return volatility (Hendricks and Singhal 2014). We contribute by hypothesizing and empirically demonstrating that a firm's operational risk is positively related to its future operational costs. We reason that a firm's operational costs are more relevant to operations managers than accounting and stock market metrics.

Second, we contribute methodologically to the corporate risk literature by providing a large-scale, transformer model-derived data set on firm-year-specific risk disclosures. Traditional approaches to measuring corporate risk (e.g., dictionary methods, topic modeling, word2vec) (Li, Mai, Shen, and Yan 2021; Wu 2024) are limited in capturing nuanced patterns in large-scale text data. By applying transformers to quantify eight risk dimensions across 131,920 firm-years, we provide a scalable and replicable dataset that enables future research on risk disclosures and their consequences. Our dataset enhances scholars' ability to analyze firm risk narratives more precisely, offering a foundation for further empirical and theoretical advancements in risk disclosure studies.

Third, our research contributes to a growing body of literature that shows a firm's legally required regulatory filings have meaningful, substantive impacts on its outcomes (e.g., Schmidt and Raman 2022). Prior research has debated whether risk disclosures in regulatory filings serve as legal safeguards or provide meaningful insights into firm behavior and performance (Cazier, McMullin, and Treu 2021). Our findings reinforce emerging evidence that these disclosures, particularly those within Form 10-K's Item 1A, extend beyond compliance requirements and influence a firm's operational cost. By demonstrating a systematic relationship between disclosed risks and subsequent cost, we add to the literature that challenges the notion of corporate filings as boilerplate documents (Cazier, McMullin, and Treu 2021). Instead, our study supports the view that these disclosures carry informational value, shaping stakeholders' expectations and managerial decision-making in ways that have tangible economic consequences. We also demonstrate that nonoperational risk amplifies the operational risk's positive association with operational cost. The interactive effects of operational and other risks provide a fertile avenue for future research into cascading failures and risk interdependencies.

Fourth, prior research has predominantly applied signaling theory to investor relations and market-based responses, often focusing on how firms signal financial strength, governance quality, or strategic direction (Connelly, Certo, Reutzel, DesJardine, and Zhou 2025; Spence 1973). We contribute to signaling theory by showing that disclosed risk factors function as signals that firms must navigate internally, not just externally. The distinction between operational and nonoperational risk signals is fundamental, as firms may respond to these signals differently depending on how they prioritize and allocate resources.

**Limitations**

Although our study offers valuable insights, it is not without limitations. First, our analysis centers on operational cost as the DV, which may not capture the full range of implications stemming from risk disclosures. Second, relying on textual disclosures in Form 10-Ks assumes that these documents accurately reflect firms' risk profiles. However, firms may engage in strategic obfuscation, omission, or selective disclosure (Carlos and Lewis 2018; Suslava 2021), potentially limiting the validity of these measures. Third, our machine learning-based measures, although sophisticated, are subject to inherent limitations, such as model bias and reliance on the quality of training data (Fyffe, Lee, and Kaplan 2024). Although we took steps to ensure robust model performance, the measures may still fail to capture nuanced or emergent risks. Fourth, the study's focus on U.S. publicly traded firms limits the generalizability of the findings to other contexts, such as privately held firms or international markets with different regulatory and disclosure standards.

## REFERENCES

Adamopoulos P, Todri V (2024) Consumer Social Connectedness and Persuasiveness of Collaborative-Filtering Recommender Systems: Evidence From an Online-to-Offline Recommendation App. *Production and Operations Management*, Forthcoming.

Astvansh, V. (2025a). Classifying an incoming customer message into spam versus ham. *Marketing Intelligence & Planning*. Forthcoming.

Astvansh, V. (2025b). Insights from customers' chats with bots and human agents. *Marketing Intelligence & Planning*. Forthcoming.

Astvansh, V., Ball, G. P., & Josefy, M. (2022). The recall decision exposed: Automobile recall timing and process data set. *Manufacturing & Service Operations Management*, *24*(3), 1457-1473.

Astvansh, V., Eshghi, K., Shahriari, H., & Shi, W. (2025). How Can a Firm Suppress Shareholders' Punitive Reaction to Its Disengagement from a Geopolitically Uncertain Market? *Journal of Marketing*, Forthcoming

Babich V, Kouvelis, P (2018) Introduction to the special issue on research at the interface of finance, operations, and risk management (iFORM): Recent contributions and future directions. *Manufacturing & Service Operations Management,* 20(1): 1-18.

Bansal T, Jha R, Munkhdalai T, McCallum A (2020) Self-supervised meta-learning for few-shot natural language classification tasks. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing* (EMNLP): 522-534.

Bao Y, Datta A (2014) Simultaneously discovering and quantifying risk types from textual risk disclosures. *Management Science*, 60(6): 1371-1391.

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

Bertrand M, Schoar A (2003) Managing with style: The effect of managers on firm policies. *Quarterly Journal of Economics*, 118(4): 1169-1208.

Bloomberg Law (2023) Capital markets, drafting guide – drafting effective SEC-complaint risk factors. https://www.bloomberglaw.com/external/document/X7FUB9DK000000/capital-markets-drafting-guide-drafting-effective-sec-compliant-

Bonnett AW, Heim GR (2024) Do mid-level providers enhance hospital cost per discharge or triple aim performance efficiency? An exploratory analysis. *Journal of Operations Management*, *70*(5): 793-830.

Busenbark JR, Lange D, Certo ST (2017) Foreshadowing as impression management: Illuminating the path for security analysts. *Strategic Management Journal*, 38(12): 2486-2507.

Busenbark, JR, Yoon H, Gamache DL, Withers MC (2022) Omitted variable bias: Examining management research with the impact threshold of a confounding variable (ITCV). *Journal of Management*, 48(1): 17-48.

Campbell JL, Chen H, Dhaliwal D, Lu HM, Steele LB (2014) The information content of mandatory risk factor disclosures in corporate filings. *Review of Accounting Studies*, *19*, 396-455.

Carlos WC, Lewis BW (2018) Strategic silence: Withholding certification status as a hypocrisy avoidance tactic. *Administrative Science Quarterly,* 63(1): 130-169.

Cazier RA, McMullin JL, Treu JS (2021) Are lengthy and boilerplate risk factor disclosures inadequate? An examination of judicial and regulatory assessments of risk factor language. *The Accounting Review*, 96(4): 131-155.

Certo ST, Busenbark JR, Kalm M, LePine JA (2020) Divided we fall: How ratios undermine research in strategic management. *Organizational Research Methods,* 23(2): 211-237.

Certo ST, Jeon C, Raney K, Lee W (2024) Measuring what matters: Assessing how executives reference firm performance in corporate filings. *Organizational Research Methods*, *27*(1): 140-166.

Connelly BL, Certo ST, Ireland RD, Reutzel CR (2011) Signaling theory: A review and assessment. *Journal of Management,* 37(1): 39-67.

Connelly BL, Certo ST, Reutzel CR, DesJardine MR, Zhou YS (2025) Signaling theory: state of the theory and its future. *Journal of Management*, 51(1): 24-61.

Craighead CW, Blackhurst J, Rungtusanatham MJ, Handfield RB (2007) The severity of supply chain disruptions: design characteristics and mitigation capabilities. *Decision Sciences*, 38(1): 131-156.

Damavandi, H., & Astvansh, V. (2025). Unveiling Regulatory Operations: A Data Set of the Determinants, Process, and Outcomes of Product Defect Investigations by the US Automotive Safety Regulator. *Manufacturing & Service Operations Management*, *27*(1), 181-199.

Damavandi, H., Mai, F., & Astvansh, V. (2025). A new technique for measuring a firm's marketing emphasis. *Marketing Letters*, 1-15.

Darby JL, Ketchen Jr DJ, Ball GP, Mukherjee U (2023) CEO stock ownership, recall timing, and stock market penalties. *Manufacturing & Service Operations Management*, 25(5): 1909-1930.

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

Feldman R, Govindaraj S, Livnat J, Segal B (2010) Management's tone change, post earnings announcement drift and accruals. *Review of Accounting Studies*, 15: 915-953.

Frank KA, Maroulis SJ, Duong MQ, Kelcey BM (2013) What would it take to change an inference? Using Rubin's causal model to interpret the robustness of causal inferences. *Educational Evaluation and Policy Analysis,* 35: 437-460.

Frankel R, Jennings J, Lee J (2022) Disclosure sentiment: Machine learning vs. dictionary methods. *Management Science*, 68(7): 5514-5532.

Fyffe S, Lee P, Kaplan S (2024) "Transforming" personality scale development: Illustrating the potential of state-of-the-art natural language processing. *Organizational Research Methods*, 27(2): 265-300.

Gamache DL, Devers CE, Klein FB, Hannigan T (2023) Shifting perspectives: How scrutiny shapes the relationship between CEO gender and acquisition activity. *Strategic Management Journal*, *44*(12): 3012-3041.

Gaulin MP (2017) *Risk fact or fiction: The information content of risk factor disclosures*. https://core.ac.uk/download/pdf/85161266.pdf

Government Information (2007) Code of Federal Regulations. https://www.govinfo.gov/content/pkg/CFR-2007-title17-vol2/html/CFR-2007-title17-vol2-sec229-503.htm

Gormley TA, Matsa DA (2014) Common errors: How to (and not to) control for unobserved heterogeneity. *Review of Financial Studies*, 27(2): 617-661.

Halder K, Akbik A, Krapac J, Vollgraf R (2020) Task-aware representation of sentences for generic text classification. *Proceedings of the 28th International Conference on Computational Linguistics*, 3202-3213.

Hassan TA, Hollander S, Van Lent L, Tahoun A (2019) Firm-level political risk: Measurement and effects. *Quarterly Journal of Economics*, 134(4): 2135-2202.

Heil O, Robertson TS (1991) Toward a theory of competitive market signaling: A research agenda. *Strategic Management Journal,* 12(6): 403-418.

Hendricks KB, Jacobs BW, Singhal VR (2020) Stock market reaction to supply chain disruptions from the 2011 Great East Japan Earthquake. *Manufacturing & Service Operations Management*, *22*(4): 683-699.

Hendricks KB, Singhal VR (2003) The effect of supply chain glitches on shareholder wealth. *Journal of Operations Management,* 21(5): 501-522.

Hendricks KB, Singhal VR (2005a) An empirical analysis of the effect of supply chain disruptions on long-run stock price performance and equity risk of the firm. *Production and Operations Management*, *14*(1): 35-52.

Hendricks KB, Singhal VR (2005b) Association between supply chain glitches and operating performance. *Management Science*, *51*(5): 695-711.

Hendricks KB, Singhal VR (2014) The effect of demand–supply mismatches on firm risk. *Production and Operations Management*, *23*(12): 2137-2151.

Hendricks KB, Singhal VR, Zhang R (2009) The effect of operational slack, diversification, and vertical relatedness on the stock market reaction to supply chain disruptions. *Journal of Operations Management*, *27*(3): 233-246.

Ho TH, Lim N, Reza S, Xia X (2017) OM forum—Causal inference models in operations management. *Manufacturing & Service Operations Management*, *19*(4): 509-525.

Hope OK, Hu D, Lu H (2016) The benefits of specific risk-factor disclosures. *Review of Accounting Studies*, 21(4): 1005–45.

Hora M and Klassen RD (2013) Learning from others' misfortune: Factors influencing knowledge acquisition to reduce operational risk. *Journal of Operations Management, 31*(1-2): 52-61.

Hu, W., Skowronski, K., Dong, Y., & Shou, Y. (2023). Mergers and acquisitions in supply bases. *Production and Operations Management*, *32*(4): 1059-1078.

Huang KW, Li Z (2011) A multilabel text classification algorithm for labeling risk factors in SEC form 10-K. *ACM Transactions on Management Information Systems (TMIS)*, 2(3): 1-19.

Huang KG, Li MC, Shen CHH, Wang Y (2024) Escaping the patent trolls: The impact of non-practicing entity litigation on firm innovation strategies. *Strategic Management Journal*, *45*(10): 1954-1987.

Huang AH, Shen J, Zang, AY (2022) The unintended benefit of the risk factor mandate of 2005. *Review of Accounting Studies*, 1-37.

Huang, F, Wang, T, Yen, JC (2023) Using relevant headings in risk factor disclosures: What is the impact on information processing? *Journal of Accounting and Public Policy*, 42(5): 107123.

Johnson, S. (2010). SEC purchases companies for more risk information. https://www.cfo.com/news/sec-pushes-companies-for-more-risk-information/669152/

Ketokivi M, McIntosh CN (2017) Addressing the endogeneity dilemma in operations management research: Theoretical, empirical, and pragmatic considerations. *Journal of Operations Management*, 52: 1-14.

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

Knight FH (1921) *Risk, uncertainty and profit* (Vol. 31). Houghton Mifflin.

Kothari SP, Shu S and Wysocki PD (2009) Do managers withhold bad news? *Journal of Accounting Research*, *47*(1): 241-276.

Lesmy D, Muchnik L, Mugerman Y (2019) Doyoureadme? temporal trends in the language complexity of financial reporting. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3469073

Li F (2006) Do stock market investors understand the risk sentiment of corporate annual reports? https://papers.ssrn.com/sol3/papers.cfm?abstract_id=898181

Li H, Lam HK, Ho W, Yeung AC (2022) The impact of chief risk officer appointments on firm risk and operational efficiency. *Journal of Operations Management*, *68*(3): 241-269.

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

Loughran T, McDonald B (2011) When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1): 35-65.

Lu G, Ding XD, Peng DX Chuang HHC (2018) Addressing endogeneity in operations management research: Recent developments, common problems, and directions for future research. *Journal of Operations Management*, 64: 53-64.

Matsumura EM, Prakash R, Vera-Muñoz SC (2024) Climate-risk materiality and firm risk. *Review of Accounting Studies*, *29*(1), 33-74.

McKenny AF, Aguinis H, Short JC, Anglin AH (2018) What doesn't get measured does exist: Improving the accuracy of computer-aided text analysis. *Journal of Management*, 44(7): 2909-2933.

McKinsey & Company (2022) The CEO's risk agenda: An insurance perspective. Retrieved from https://www.mckinsey.com/industries/financial-services/our-insights/the-ceos-risk-agenda-an-insurance-perspective

Miric M, Jia N, Huang KG (2023) Using supervised machine learning for large-scale classification in management research: The case for identifying artificial intelligence patents. *Strategic Management Journal*, 44(2): 491-519.

Oliver AG, Krause R, Busenbark JR, Kalm M (2018) BS in the boardroom: Benevolent sexism and board chair orientations. *Strategic Management Journal,* 39(1): 113-130.

Peker I, Ar IM, Erol I, Searcy C (2023) Leveraging blockchain in response to a pandemic through disaster risk management: an IF-MCDM framework. *Operations Management Research*, 16(2): 642-667.

Petersen MA (2009) Estimating standard errors in finance panel data sets: *Comparing approaches. Review of Financial Studies,* 22(1): 435–480.

PricewaterhouseCoopers (2023) *SEC 2160 - Registration and reporting by smaller reporting companies*. PwC Viewpoint. Retrieved from https://viewpoint.pwc.com/dt/us/en/pwc/pwc_sec_volume/pwc_sec_volume_US/2000_registration_un_US/sec_2160_registratio_US.html

Rathje J, Katila R, Reineke P (2024) Making the most of AI and machine learning in organizations and strategy research: Supervised machine learning, causal inference, and matching models. *Strategic Management Journal.* DOI:10.1002/smj.3604

Ridge JW, Hill AD, Ingram A, Kolomeitsev S, Worrell DL (2024) Avoidance and Aggression in Stakeholder Engagement: The Impact of CEO Paranoia and Paranoia-Relevant Cues. *Academy of Management Journal*, (ja): amj-2021.

Rumyantsev S, Netessine S (2007) What can be learned from classical inventory models? A cross-industry exploratory investigation. *Manufacturing & Service Operations Management*, *9*(4): 409-429.

Schick T, Schütze H (2020) It's not just size that matters: Small language models are also few-shot learners. *arXiv preprint arXiv:2009.07118*.

Schmidt W, Raman A (2022) Operational disruptions, firm risk, and control systems. *Manufacturing & Service Operations Management*, *24*(1): 411-429.

Schnatterly K, Calvano F, Berns JP, Deng C (2021) The effects of board expertise-risk misalignment and subsequent strategic board reconfiguration on firm performance. *Strategic Management Journal*, *42*(11): 2162-2191.

SEC (2005). Securities offering reform. https://www.sec.gov/files/rules/final/33-8591.pdf

SEC. (2013). The Path forward to disclosure. https://www.sec.gov/news/speech/spch101513mjw

SEC. 2016. Business and financial disclosure required by Regulation S-K. https://www.sec.gov/files/rules/concept/2016/33-10064.pdf

SEC (2020) Modernization of Regulation S-K items 101, 103, and 105: A small entity compliance guide. https://www.sec.gov/corpfin/modernization-regulation-s-k-compliance-guide

Sharma L, Goradia D (2023) Will your insurance type influence clinical quality outcomes? An investigation of contributing factors, underlying mechanism, and consequences. *Production and Operations Management*, *32*(7): 2207-2226.

Short JC, Broberg JC, Cogliser CC, Brigham KH (2010) Construct validation using computer-aided text analysis (CATA) an illustration using entrepreneurial orientation. *Organizational research methods*, *13*(2): 320-347.

Simchi-Levi D, Schmidt W, Wei Y, Zhang PY, Combs K, Ge Y, Gusikhin O, Sanders M, Zhang D (2015) Identifying risks and mitigating disruptions in the automotive supply chain. *Interfaces.* 45(5): 375-390.

Spence M (1973) Job market signaling. *The Quarterly Journal of Economics*, 87(3): 355-374.

Sodhi MS, Son BG, Tang CS (2012) Researchers' perspectives on supply chain risk management. *Production and Operations Management*. 21(1): 1-13.

Suslava K (2021) "Stiff Business Headwinds and Uncharted Economic Waters": The Use of Euphemisms in Earnings Conference Calls. *Management Science,* 67(11): 7184–7213. https://doi.org/10.1287/mnsc.2020.3826

Tang CS (2006) Perspectives in supply chain risk management. *International Journal of Production Economics,* 103(2): 451-488.

U.S. Securities and Exchange Commission (n.d.) *General instructions as to financial statements* (17 C.F.R. § 229.10). Legal Information Institute. Retrieved from https://www.law.cornell.edu/cfr/text/17/229.10

Verrecchia RE (2001) Essays on disclosure. *Journal of Accounting and Economics*, 32(1-3): 97-180.

Webb J (2016) Toyota's 'Quake-Proof' Supply Chain that Never Was. Forbes. https://www.forbes.com/sites/jwebb/2016/04/26/toyotas-quake-proof-supply-chain-that-never-was/

Wishart-Smith H (2024) The Semiconductor Crisis: Addressing Chip Shortages and Security. Forbes. https://www.forbes.com/sites/heatherwishartsmith/2024/07/19/the-semiconductor-crisis-addressing-chip-shortages-and-security/

Yılmaz Ö, Son Y, Shang G, Arslan HA (2024) Causal inference under selection on observables in operations management research: Matching methods and synthetic controls. *Journal of Operations Management*, *70*(5): 831-859.

**Figure 1: Our Procedure**

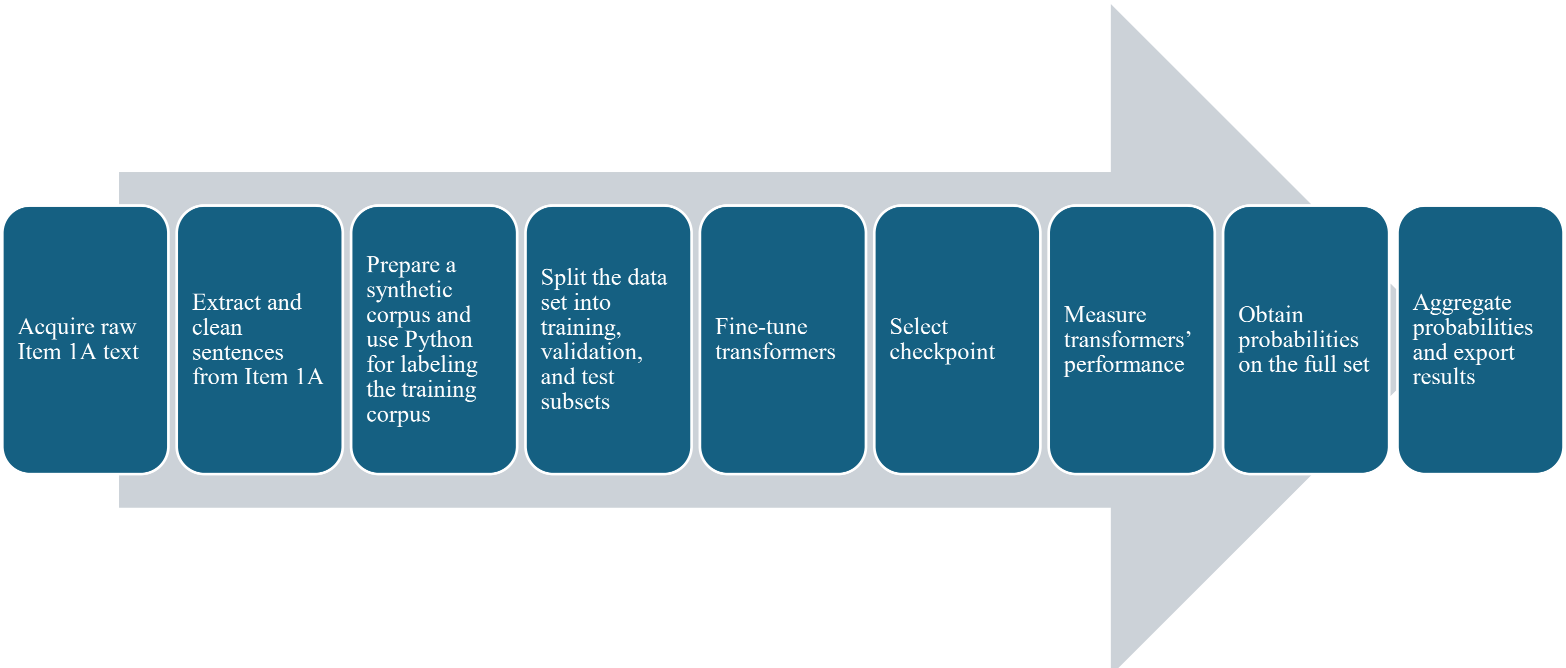
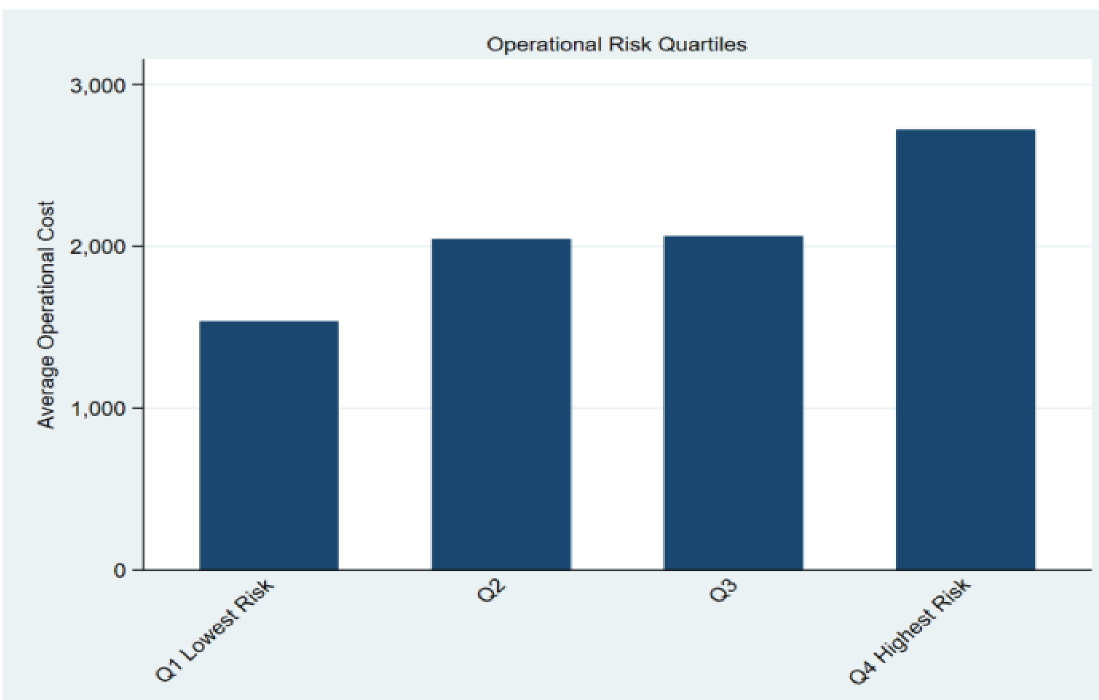



**Figure 2: Average Operational Cost by Operational Risk Quartiles**



## Table 1: Descriptive Statistics for Dimensions of Risk Factors Data Set (Item 1A and Word Counts)

| | **Item 1A word count** | **Item 1A optional** | **Accounting** | **Finance** | **International** | **Legal** | **Management** | **Marketing** | **Operations** | **Technology** |
|---|---|---|---|---|---|---|---|---|---|---|
| **count** | 131920 | 131920.00 | 131920.000 | 131920.00 | 131920.00 | 131920.00 | 131920.00 | 131920.00 | 131920.00 | 131920.00 |
| **mean** | 17651.21 | 0.098 | 93.112 | 421.003 | 53.316 | 168.474 | 288.008 | 161.593 | 175.998 | 249.849 |
| **std** | 21871.03 | 0.297 | 152.255 | 644.999 | 88.000 | 201.815 | 343.677 | 185.540 | 221.168 | 313.688 |
| **min** | 0 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **25%** | 2899 | 0.000 | 3.000 | 62.000 | 2.000 | 26.000 | 56.000 | 31.000 | 27.000 | 32.000 |
| **50%** | 9317 | 0.000 | 21.000 | 182.000 | 19.000 | 107.000 | 171.000 | 102.000 | 101.000 | 138.000 |
| **75%** | 25014.25 | 0.000 | 112.000 | 541.000 | 66.000 | 232.000 | 401.000 | 228.000 | 236.000 | 355.000 |
| **max** | 505360 | 1.000 | 2873.000 | 11977.000 | 2519.000 | 3608.000 | 7499.000 | 10014.000 | 5304.000 | 9523.000 |

## Table 2: Example of Labels for Training Data Set

*Note*: Acc = Accounting. Fin = Finance. Mgt = Management. Ops = Operations. Intl = International. Tech = Technology. Mktg = Marketing

| **Example Sentence** | **Acc** | **Fin** | **Mgt** | **Ops** | **Intl** | **Legal** | **Tech** | **Mktg** |
|---|---|---|---|---|---|---|---|---|
| Technological integration challenges during mergers or acquisitions could disrupt our operations. | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 |
| Evolving consumer technology preferences may outpace our development capabilities. | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| We are exposed to risks associated with off-balance sheet arrangements that could impact our financial condition. | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Changes in credit ratings could impact our ability to raise capital and increase our borrowing costs. | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Supply chain disruptions could significantly impact our ability to manufacture and deliver products on time. | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Political instability in the Middle East, particularly in countries like Syria and Iraq, could disrupt our oil and gas operations in the region. | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |

## Table 3A: Eight Performance Metrics for Our Eight Trained Transformers for Operational Risk Factor

| **Model Name** | **Accuracy** | **Precision Micro** | **Precision Macro** | **Recall Micro** | **Recall Macro** | **F-1 Micro** | **F-1 Macro** | **Loss** |
|---|---|---|---|---|---|---|---|---|
| bert-base-uncased | 0.961 | 0.992 | 0.990 | 0.983 | 0.985 | 0.987 | 0.987 | 0.067 |
| roberta-base | 0.961 | 0.992 | 0.990 | 0.983 | 0.985 | 0.987 | 0.987 | 0.057 |
| distilbert-base-uncased | 0.974 | 1.000 | 1.000 | 0.983 | 0.985 | 0.992 | 0.992 | 0.040 |
| microsoft/deberta-base | 0.987 | 1.000 | 1.000 | 0.992 | 0.993 | 0.996 | 0.996 | 0.018 |
| google/electra-base-discriminator | 0.908 | 1.000 | 1.000 | 0.924 | 0.915 | 0.961 | 0.955 | 0.094 |
| yiyanghkust/finbert-tone | 0.974 | 1.000 | 1.000 | 0.983 | 0.985 | 0.992 | 0.992 | 0.033 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| nlpaueb/legal-bert-base-uncased | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.497 |
| roberta-large | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.497 |

## Table 3B: Best-Performing Transformer for the Eight Risk Factors

| **Risk factor** | **Best-Performing Transformer** |
|---|---|
| Accounting | bert-base-uncased |
| Finance | bert-base-uncased |
| Management | microsoft/deberta-base |
| Operations | bert-base-uncased |
| International | bert-base-uncased |
| Legal | bert-base-uncased |
| Technology | roberta-base |
| Marketing | bert-base-uncased |

## Table 4A: Descriptive Statistics for Risk Factors Data Set (Transformer-Determined Probabilities on Training Subset)

| | **Risk factor probability** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Accounting | Finance | International | Legal | Management | Marketing | Operations | Technology |
| **Count** | 131920.00 | 131920.00 | 131920.00 | 131920.00 | 131920.00 | 131920.00 | 131920.00 | 131920.00 |
| **Mean** | 0.160 | 0.285 | 0.053 | 0.200 | 0.063 | 0.047 | 0.177 | 0.051 |
| **SD** | 0.149 | 0.193 | 0.043 | 0.125 | 0.046 | 0.044 | 0.129 | 0.051 |
| **Min** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **25%** | 0.068 | 0.143 | 0.024 | 0.115 | 0.034 | 0.019 | 0.089 | 0.018 |
| **50%** | 0.120 | 0.284 | 0.047 | 0.187 | 0.063 | 0.037 | 0.163 | 0.038 |
| **75%** | 0.215 | 0.418 | 0.074 | 0.279 | 0.087 | 0.064 | 0.258 | 0.068 |
| **Max** | 0.978 | 0.933 | 0.487 | 0.982 | 0.920 | 0.695 | 0.871 | 0.988 |

## Table 4B: Descriptive Statistics for Risk Factors Data Set (Transformer Predictions on Test Subset)

| | **Risk factor prediction** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Accounting | Finance | International | Legal | Management | Marketing | Operations | Technological |
| **Count** | 131920.00 | 131920.00 | 131920.00 | 131920.00 | 131920.00 | 131920.00 | 131920.00 | 131920.00 |
| **Mean** | 0.042 | 0.140 | 0.000 | 0.015 | 0.001 | 0.000 | 0.013 | 0.000 |
| **SD** | 0.201 | 0.347 | 0.000 | 0.122 | 0.030 | 0.007 | 0.113 | 0.010 |
| **Min** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **25%** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **50%** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **75%** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **Max** | 1.000 | 1.000 | 0.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

**Table 5: Variable Key**

| Variable name | Variable measure | Reason for inclusion, if regressor | Reference research |
|---|---|---|---|
| **Operational cost$_{i,y}$** | Natural log of firm *i*'s cost of goods sold in fiscal year *y* (Compustat: COGS) | N/A | Hu, Skowronski, Dong, and Shou (2022); Li, Lam, Ho, and Yeung (2022) |
| **Operational risk$_{i,y}$** | Machine-learning-based probability of the firm *i*'s operational risk disclosed in Item 1A of fiscal year *y*'s Form 10-K | Operational risks impact firms' everyday processes, supply chain coordination, and cost structures, leading to higher operational costs. | Modified from Schnatterly, Calvano, Berns, and Deng (2021) |
| **Nonoperational risk$_{i,y}$** | The sum of machine-learning-based-probabilities of the firm *i*'s nonoperational risk disclosed in Item 1A of fiscal year *y*'s Form 10-K | Nonoperational risks require managerial attention and resources, diverting focus from operational efficiency and potentially increasing costs. | Modified from Schnatterly, Calvano, Berns, and Deng (2021) |
| **Firm size$_{i,y}$** | Natural log of the dollar value of firm *i*'s total Assets in year y (Compustat: AT) | Firm size is a significant factor influencing the firm's operational costs and financial reporting practices | Hu, Skowronski, Dong, and Shou (2022); Li, Lam, Ho, and Yeung (2022) |
| **Item 1A optional$_{i,y}$** | = 1 if the firm *i* identified itself as a small firm for the SEC's reporting purposes in year *y* | Smaller firms report less risk | The authors' variable |
| **Item 1A word count$_{i,y}$** | Natural log of the number of words in the firm *i*'s Item 1A of fiscal year *y*'s Form 10-K | The longer the Item 1A, the more could be disclosed risks. | McKenny, Aguinis, Short, and Anglin (2018) |
| **R&D$_{i,y}$** | Firm *i*'s R&D expenses in fiscal year *y* divided by its total revenue in year *y* (Compustat: XRD ÷ AT) | R&D intensity can influence production processes and efficiency and thus impact operational costs. | Hu, Skowronski, Dong, and Shou (2022); Li, Lam, Ho, and Yeung (2022) |

**Table 6: Descriptive Statistics (Raw Variables)**

| Variable | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|
| Operational cost | 2091.954 | 11032.218 | -8016 | 452776 |
| Operational risk | .192 | .127 | 0 | .861 |
| Nonoperational risk | .875 | .32 | 0 | 2.185 |
| Firm size | 3186.73 | 15250.135 | -6389.9 | 608481 |
| Item 1A optional | .053 | .225 | 0 | 1 |
| Item 1A word count | 18998.225 | 22048.863 | 0 | 505360 |

| | | | | |
|---|---|---|---|---|
| R&D | .229 | 10.953 | -28.444 | 1981.5 |

**Table 7: Operational Risk → Operational Cost: OLS Estimates**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | DV = Operational $cost_{i,y+1}$ | | | |
| Firm size | .487*** | .486*** | .486*** | .486*** |
| | (.014) | (.014) | (.014) | (.014) |
| Item 1A optional | -.104*** | -.093*** | -.089*** | -.083*** |
| | (.025) | (.025) | (.025) | (.025) |
| Item 1A word count | -.006** | -.007*** | -.004 | .001 |
| | (.003) | (.003) | (.003) | (.004) |
| R&D | .004 | .004 | .004 | .004 |
| | (.014) | (.014) | (.014) | (.014) |
| Operational risk | | .222*** | .225*** | .247*** |
| | | (.038) | (.038) | (.041) |
| Nonoperational risk | | | -.034 | -.046* |
| | | | (.022) | (.024) |
| Operational risk × | | | | .298** |
| Nonoperational risk | | | | (.125) |
| Year-fixed effects | Yes | Yes | Yes | Yes |
| Firm-fixed effects | Yes | Yes | Yes | Yes |
| Constant | 2.429*** | 2.447*** | 2.42*** | 2.373*** |
| | (.081) | (.081) | (.083) | (.085) |
| Observations | 67727 | 67727 | 67727 | 67727 |
| R-squared | .963 | .963 | .963 | .963 |
| *Standard errors are in parentheses* | | | | |
| **** p<.01, ** p<.05, * p<.1* | | | | |
| *We mean-centered Operational risk and Nonoperational risk.* | | | | |

**Table 8A. Operational Risk → Operational Cost: Robustness to Using Risk Word Counts in Place of Probabilities**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Operational $cost_{i,y+1}$ | | | |
| Firm size | .487*** | .486*** | .486*** | .486*** |
| | (.014) | (.014) | (.014) | (.014) |
| Item 1A optional | -.104*** | -.093*** | -.089*** | -.083*** |
| | (.025) | (.025) | (.025) | (.025) |
| Item 1A word count | -.006** | -.007*** | -.004 | .001 |
| | (.003) | (.003) | (.003) | (.004) |
| R&D Intensity | .004 | .004 | .004 | .004 |
| | (.014) | (.014) | (.014) | (.014) |
| Operational risk | | .029*** | .029*** | .031*** |
| | | (.005) | (.005) | (.005) |
| Nonoperational risks | | | -.011 | -.017** |
| | | | (.007) | (.008) |

| | | | | |
|---|---|---|---|---|
| Operational risk × Nonoperational risk | | | | .013** |
| | | | | (.005) |
| Year-fixed effects | Yes | Yes | Yes | Yes |
| Firm-fixed effects | Yes | Yes | Yes | Yes |
| Constant | 2.429*** | 2.443*** | 2.417*** | 2.37*** |
| | (.081) | (.081) | (.083) | (.085) |
| Observations | 67727 | 67727 | 67727 | 67727 |
| R-squared | .963 | .963 | .963 | .963 |
| *Standard errors are in parentheses* | | | | |
| **** p<.01, ** p<.05, * p<.1* | | | | |
| *We standardized Operational risk and Nonoperational risk.* | | | | |

**Table 8B: Operational Risk → Operational Cost: Robustness to Alternative Fixed Effects**

| | (1) | (2) |
|---|---|---|
| | DV: Operational $cost_{i,y+1}$ | |
| Firm size | .486*** | .486*** |
| | (.079) | (.079) |
| Item 1A optional | -.083** | -.083** |
| | (.035) | (.035) |
| Item 1A word count | .001 | .001 |
| | (.008) | (.008) |
| R&D Intensity | .004 | .004 |
| | (.014) | (.014) |
| Operational risk | .247** | .247** |
| | (.088) | (.088) |
| Nonoperational risk | -.046 | -.046 |
| | (.043) | (.043) |
| Operational risk × Nonoperational risk | .298** | .298** |
| | (.122) | (.122) |
| Year-fixed effects | Yes | No |
| Industry (SIC4)-fixed effects | Yes | No |
| Firm-fixed effects | No | Yes |
| Industry × Year-fixed effects | No | Yes |
| Constant | 2.373*** | 2.373*** |
| | (.428) | (.428) |
| Observations | 67727 | 67727 |
| R-squared | .963 | .963 |

**Table 9: Operational Risk → Operational Cost: Robustness to Controlling for Endogeneity**

| | (1) | (2) |
|---|---|---|
| | Operational $cost_{i,y+1}$ | Operational $cost_{i,y+1}$ |
| Firm size | .884*** | .880*** |
| | (.052) | (.006) |
| Item 1A optional | -.129 | .331*** |

| | | |
|---|---|---|
| | (.085) | (.051) |
| Item 1A word count | -.017 | .282*** |
| | (.014) | (.051) |
| R&D Intensity | .181** | .130** |
| | (.067) | (.050) |
| Operational risk | 1.716*** | .622*** |
| | (.399) | (.135) |
| Non-operational risk | | -2.942*** |
| | | (.128) |
| Year-fixed effects | Yes | Yes |
| Firm-fixed effects | Yes | Yes |
| Constant | .265 | -2.25*** |
| | (.117) | (.124) |
| Observations | 68968 | 68968 |
| R-squared | .85 | .84 |
| Kleibergen-Paap LM Test Underidentification Test | 16.736*** | 1328.15*** |
| Kleibergen-Paap Wald F-Test (Weak IV Test) | 1221.17 | 593.10 |
| Stock-Yogo 10% Maximal IV Size | 16.38 | 7.03 |

**A FIRM'S OPERATIONAL RISK: DATA SET AND EMPIRICAL EVIDENCE**

**Vivek Astvansh and Joseph J. Simpson**

## ONLINE APPENDIX

## APPENDIX A: PRIOR RESEARCH ON DISCOVERING FIRM RISK FACTORS FROM FORM 10-K

Although the SEC mandates that a U.S. public firm describe its risk factors, it does not list them. Therefore, academics have used narrative Item 1A to perform two tasks: (1) discover risk factors *and* (2) measure the magnitude of each risk factor.[8] Our literature review revealed only four studies that used Item 1A text to *discover and measure* a firm's risk factors. These studies are closest to our research; therefore, we summarize these four studies in reverse chronological order.

First, Huang and Li (2011) manually identified 25 risk factors from the headings (and not descriptions) of Item 1A. They named each factor (see Table III in the article's appendix on pp. 15-17 for the names). Next, they hired student assistants who labeled Item 1A against each of the 25 factors. The authors used this labeled data to train a multilabel classifier.

Second, Bao and Datta (2014) followed up by using a latent Dirichlet allocation (LDA) and discovered 30 risk factors (or topics), which included Huang and Li's (2011) 25 types plus five "new" factors. The authors regressed a firm's stock return volatility on the magnitude (i.e., number of sentences) of each of the 30 risk factors. They found that 22 types of risk were not associated with the return volatility, suggesting that disclosure of not each risk type is informative. Three types—macroeconomic risk, funding risk, and credit risk—are positively associated with stock return volatility, whereas five types of risks—human resources, shareholders' interests, environment, information security, and infrastructure disruption—are negatively associated with return volatility.

---

[8] We acknowledge but do not summarize the vast literature that has considered (1) Form 10-K's nonrisk-related characteristics or (2) risk measured from Form 10-K's items other than Item 1A (e.g., Item 7A on management discussion and analysis). This research has considered the following Form 10-K characteristics: (1) the number of sentences in all items that include at least one risk-related word (Kravet and Muslu 2013), (2) the number of words on risk factors (Nelson and Pritchard 2007), (3) the proportion of positive words and negative words in the Item 7A (Feldman, Govindaraj, Livnat, and Segal 2010), (4) future-related sentences in the Item 7A (Muslu et al. 2009), (5) number of risk-related words (Li 2006), (6) complexity (specifically, length and readability; Guay, Samuels and Taylor 2005; Lesmy, Muchnik, and Mugerman 2019), (7) negative words in Item 7A and the entire Form 10-K (Loughran and McDonald 2011), and (8) boilerplate language, stickiness, redundancy, hard information (Dyer, Lang, and Stice-Lawrence 2017), (9) specificity (Dyer, Lang, and Stice-Lawrence 2017; Hope, Hu, and Lu 2016), and (10) headings (Huang, Wang, and Yen 2023).

Third, using Nelson and Pritchard's (2007) list of keywords and an LDA model, Campbell, Chen, Dhaliwal, Lu, and Steele (2014) identified five risk factors: (1) financial, (2) tax, (3) legal, (4) systematic, and (5) idiosyncratic. They showed that a firm's pre-filing fundamentals are positively associated with the number of words on a risk factor. They also reported that the firm's disclosed systematic risk (idiosyncratic risk) is positively associated with future stock-market risk (stock return volatility). The full measure of disclosed risk (as opposed to a measure of a specific type of risk) is negatively associated with (1) bid-ask spread, which proxies information asymmetry among shareholders, and (2) cumulative abnormal (buy-and-hold) stock return in the $[-1,1]$ window centered on the firm's date of release of Form 10-K.

Fourth, Schnatterly, Calvano, Berns, and Deng (2021) considered eight risk factors that align with managers' functional expertise, such as finance and operations. They sourced the dictionaries from those provided by Bao and Datta (2014), Campbell, Chen, Dhaliwal, Lu, and Steele (2014), and Gaulin (2017), combining the keywords and recategorizing Bao and Datta's (2014) and Gaulin's (2017) lists of 30 and 20 risk types, respectively. The authors focused on the misalignment between the firm's eight types of risks and its directors' skills that can help the firm manage these risks. They reported that the sum of misalignment on the eight risk types is associated with a drop in the firm's stock return in the following year.

These studies exhibit methodological divergence in their approach to discovering and measuring risk factors. Some researchers employed machine learning algorithms to inductively identify and classify risk typologies (Huang and Li, 2011; Bao and Datta, 2014). Others utilized a deductive framework, examining predetermined risk categories identified a priori (Campbell, Chen, Dhaliwal, Lu, and Steele, 2014; Schnatterly, Calvano, Berns, and Deng, 2021). We follow both methodologies and measure a firm's risk based on word count-based measures and probabilities predicted by transformer models.

## APPENDIX B: TECHNICAL DETAILS OF OUR METHOD

### Using ChatGPT to Generate Synthetic Data

We supplied ChatGPT (version o1 pro) with the dictionaries of Schnatterly, Calvano, Berns, and Deng's (2021) eight risk factors and prompted it to generate sentences for each factor. We instructed ChatGPT that each statement could mention multiple risk factors and asked it to attempt to balance the number of sentences for the eight factors. Although machine-learning models can perform well with only a few labels per category (Fyffe, Lee, and Kaplan 2024; Schick and Schütze 2020), we generated 761 sentences and stored them in a comma-separated values (CSV) file. This count far exceeds prior research's (e.g., Bansal, Munkhdalai, and McCallum 2020; Fyffe, Lee, and Kaplan 2024; Halder, Akbik, Krapac, and Vollgraf 2020). Next, we wrote Python code to determine whether a sentence mentions each of the eight risk factors. Thus, the CSV file contained 761 rows × eight columns (one for each risk factor). The file includes an average of 145 labeled sentences per risk factor. We use this CSV file to train our transformers.

Consider an example. Suppose we input Schnatterly, Calvano, Berns, and Deng's (2021) technology risk factor dictionary into ChatGPT. Assume ChatGPT generated 10 sentences. Consider sentence #1. It will include at least one word in the technology risk dictionary and thus takes 1 for the technology risk factor column. Assume this sentence contains a word from the accounting risk factor dictionary but no word from any of the other risk-factor dictionaries. Therefore, our Python code will set this sentence's accounting risk factor column as 1 and the remaining six risk factors as 0.

### Transformers and Performance Metrics

Transformers are deep-learning models specializing in processing and generating sequential data, such as text, where words are interdependent, and context determines a word's meaning. Unlike other models, such as recursive neural networks and long short-term memory, which read words sequentially, transformers read all words in a sentence at once, comparing a word with every other word in the sentence. This processing method, called self-attention, allows a word to "attend" to other words and thus assign words different weights and understand the context. We trained eight popular transformers: (1)

BERT-base-uncased, (2) RoBERTa-base, (3) distilBERT-base-uncased, (4) microsoft/deBERTa-base, (5) google/electra-base-discriminator, (6) yiyanghkust/finbert-tone, (7) nlpaueb/legal-BERT-base-uncased, and (8) roBERTa-large.[9]

Our procedure begins with a corpus of risk-factor sentences generated by ChatGPT (after we input ChatGPT Schnatterly, Calvano, Berns, and Deng's [2021] eight dictionaries) and labelled by our Python code for the presence or absence of each of the eight risk factors. After a single shuffle (random_state = 42), we divided the corpus into three mutually exclusive subsets: approximately 80% constituted the training subset, 10% the validation subset, and 10% the test subset. We updated the model parameters only on the training subset. Each model was applied to the validation subset at the end of every epoch. We retained the epoch with the highest average micro-F-1 score as the final checkpoint. The test subset remained unseen until the models were completely trained, and the best checkpoint was selected and used. Next, each model was run exactly once on this "untouched" test subset. Table 3 reports the resulting precision, recall, and F-1 scores. Because the probability threshold is held at 0.50 for every label and no parameters are tuned on the test data, these scores provide an unbiased estimate of out-of-sample performance, albeit on a deliberately balanced and lexically clear set of sentences that favors high overall F-values.

The F-scores for most transformers are high because the classification task is relatively straightforward. Each sentence is short and clear in our labeled data set, the key words that signal a given risk category are typically strong, and examples are balanced for all eight labels. Pre-trained transformers perform exceptionally well under those conditions, so near-ceiling values are expected and are unlikely to imply leakage or threshold tuning.

*Accuracy* measures the model's proportion of correct predictions. *Precision* indicates the model's proportion of "true positive" predictions, thus reflecting how well the model avoids false positives. *Recall*

---

[9] https://huggingface.co/google-bert/bert-base-uncased; https://huggingface.co/FacebookAI/roberta-base; https://huggingface.co/distilbert/distilbert-base-uncased; https://huggingface.co/google/electra-base-discriminator; https://huggingface.co/yiyanghkust/finbert-tone; https://huggingface.co/nlpaueb/legal-bert-base-uncased; https://huggingface.co/FacebookAI/roberta-large

(or sensitivity) measures the proportion of actual positives correctly predicted, focusing on minimizing false negatives. The *F-1 score* is the harmonic mean of precision and recall, providing a balanced metric when the class distribution is imbalanced.

In addition to these classification metrics, we report the *loss*, quantifying the model's prediction error. Lower loss values indicate that the model's predictions are closely aligned with actual outcomes, whereas higher loss values suggest greater deviations. Table 3 reports that the models that demonstrate low loss scores, such as *microsoft/deBERTa-base* (0.018) and *distilBERT-base-uncased* (0.040), tend to achieve high accuracy and F-1 scores, indicating strong predictive performance. Conversely, models with high loss values, such as *google/electra-base-discriminator* (0.367) and *roBERTa-large* (0.499), perform poorly across all metrics, suggesting they struggle to generate reliable classifications.

We wrote a Python program to select the best-performing model based on the average of the eight metrics (including macro/micro variants of recall, precision, and F-1 score; see Table 3A) to ensure that the chosen model scores consistently well on all metrics. We used this best-performing model to evaluate documents against eight risk factors. We set the chunk size to 510 tokens to avoid exceeding the input size constraints of many transformer-based models. This setting leaves space for special tokens so that the total sequence length stays within the typical limit of 512 tokens. We wrote code that took the text, split the token indices into chunks of up to 510 tokens each, and inserted special tokens to conform to the model's expected input structure. Each chunk was forwarded through the model, and the code stored the predicted probabilities for all target labels. After the code gathered predictions across all chunks, it averaged the probabilities for Item 1A for a firm-year. This approach ensured that when documents exceeded the token length limit, the code produced a cohesive result by combining chunk-level probabilities. In other words, we ensured that the entire firm-year specific Item 1A was scored rather than limiting the scoring to 512 tokens. Table B1 lists our nine-step procedure.

**Table B1: Our Nine-Step Procedure**

| Step | Step name | Step description |
|---|---|---|

| | | |
|---|---|---|
| **1** | Acquire raw Item 1A text | Download Form 10-Ks from EDGAR and store them as TXT files with "item" headings intact. |
| **2** | Extract and clean sentences from Item 1A | Isolate Item 1A, split its text into sentences, remove HTML tags, normalize Unicode, and discard boiler-plate headers. |
| **3** | Prepare a synthetic corpus and use Python for labeling the training corpus | Input Schnatterly, Calvano, Berns, and Deng's (2021) dictionaries of eight risk factors into ChatGPT and ask it to generate synthetic sentences mentioning one or more of them. Save the ChatGPT output as a CSV file. Write a Python program that classifies each sentence into eight binary variables, one for each risk factor. Set the variable to 1 if the focal sentence mentions terms of the focal risk factor, and 0 otherwise. Save the new eight columns in the CSV file. |
| **4** | Split the data set into training, validation, and test subsets | Shuffle the synthetic corpus once (random_state = 42) and divide it into 80% training, 10% validation, and 10% test. Use only the training subset for weight updates. |
| **5** | Fine-tune transformers | Fine-tune eight base checkpoints (BERT, RoBERTa, DistilBERT, DeBERTa, ELECTRA, FinBERT, Legal-BERT, RoBERTa-large) for four epochs on the training subset with binary cross-entropy loss. |
| **6** | Select checkpoint | After every epoch, compute performance metrics on the validation subset. Keep the epoch with the highest score as the final checkpoint for that base model. |
| **7** | Measure transformers' performance | Apply each retained checkpoint exactly once to the test subset. Record values of the eight performance metrics (see Table 3A for an example of operational risk factor). |
| **8** | Obtain probabilities on the full set | Run the best checkpoint on every Item 1A. Split text into 510-token chunks, obtain label probabilities for each chunk, and average them across Item 1A |
| **9** | Aggregate probabilities and export results | Threshold averaged probabilities at 0.50 to generate binary risk indicators and write the resulting firm-year risk-factor matrix to a CSV for subsequent analyses. |

## REFERENCES

Bao Y Datta A (2014) Simultaneously discovering and quantifying risk types from textual risk disclosures. *Management Science*, *60*(6), 1371-1391.

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

Campbell JL, Chen H, Dhaliwal DS, Lu HM, Steele LB (2014) The information content of mandatory risk factor disclosures in corporate filings. *Review of Accounting Studies*, *19*, 396-455.

Darby JL, Ketchen Jr DJ, Ball GP Mukherjee U (2023) CEO stock ownership, recall timing, and stock market penalties. *Manufacturing & Service Operations Management*, *25*(5), 1909-1930.

Dyer T, Lang M, Stice-Lawrence L (2016) Do managers really guide through the fog? On the challenges in assessing the causes of voluntary disclosure. *Journal of Accounting and Economics*, 62(2-3), 270-276.

Feldman R, Govindaraj S, Livnat J, Segal B (2010) Management's tone change, post earnings announcement drift and accruals. *Review of Accounting Studies*, 15, 915-953.

Gaulin MP (2017) *Risk fact or fiction: The information content of risk factor disclosures*. https://core.ac.uk/download/pdf/85161266.pdf

Guay W, Samuels D, Taylor D (2016) Guiding through the fog: Financial statement complexity and voluntary disclosure. *Journal of Accounting and Economics*, *62*(2-3), 234-269.

Hope O.-K, Hu D, Lu, H (2016) The benefits of specific risk-factor disclosures. *Review of Accounting Studies*, 21(4): 1005–45.

Huang KW, Li Z (2011) A multilabel text classification algorithm for labeling risk factors in SEC form 10-K. *ACM Transactions on Management Information Systems (TMIS)*, *2*(3), 1-19.

Huang AH, Shen J, Zang AY (2022) The unintended benefit of the risk factor mandate of 2005. *Review of Accounting Studies*, 1-37.

Huang F, Wang T, Yen JC (2023) Using relevant headings in risk factor disclosures: What is the impact on information processing? *Journal of Accounting and Public Policy*, *42*(5), 107123.

Kothari SP, Shu S, Wysocki PD (2009) Do managers withhold bad news? *Journal of Accounting Research*, 47, 241–276.

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

Lesmy D, Muchnik L, Mugerman Y (2019) Doyoureadme? temporal trends in the language complexity of financial reporting. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3469073

Li F (2006) Do stock market investors understand the risk sentiment of corporate annual reports? https://papers.ssrn.com/sol3/papers.cfm?abstract_id=898181

Loughran T, McDonald B (2011) When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance*, 66(1), 35-65.

Muslu V, Mutlu S, Radhakrishnan S, Tsang A (2019) Corporate social responsibility report narratives and analyst forecast accuracy. *Journal of Business Ethics*, *154*, 1119-1142.

Nelson KK Pritchard AC (2007) Litigation risk and voluntary disclosure: The use of meaningful cautionary language. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=998590

SEC (2013) The Path forward to disclosure. https://www.sec.gov/news/speech/spch101513mjw

SEC (2016) Business and financial disclosure required by Regulation S-K. https://www.sec.gov/files/rules/concept/2016/33-10064.pdf