

Firm-Level Litigation Risk: Measurement and Effects

Benjamin Bennett¹ Charles Ham² Todd Milbourn³ Zexi Wang⁴

September 2025

Abstract

Firms are exposed to litigation risk from various stakeholders, including investors, customers, and regulators, and prior research has shown that such risk shapes a range of corporate policies. In this study, we develop a simple and transparent measure of firm-level litigation risk by parsing SEC 10-K annual reports. This measure demonstrates strong predictability for lawsuits, even after controlling for firm characteristics, existing lawsuits, geographic factors, and various fixed effects, including firm fixed effects. This simple measure performs comparably to a more resource-intensive, AI-generated litigation risk measure. We further link our measure of litigation risk to a variety of corporate policies, including cash holdings, capital expenditures, R&D, and acquisitions, uncovering results consistent with the literature.

Keywords: Litigation risk, ChatGPT, Corporate policies, SEC 10-K filings

JEL: G30, K20

¹ Neeley School of Business, Texas Christian University, 2900 Lubbock Ave, Fort Worth, TX 76109, ben.bennett@tcu.edu

² Kelley School of Business, Indiana University, 1309 E. 10th St., Bloomington, IN 47405, USA, hamch@iu.edu

³ Cox School of Business, Southern Methodist University, 6214 Bishop Blvd., Dallas, TX 75275, USA, toddmilbourn@mail.smu.edu

⁴ Management School, Lancaster University, Lancaster, LA1 4YX, United Kingdom, zexi.wang@lancaster.ac.uk

We would like to thank the editor, the associate editor, two anonymous referees, Justin Birru, John Barrios, Kimball Chapman, Claudia Custodio, Rüdiger Fahlenbrach, Laurent Frésard, Todd Gormley, Justin Hopkins, Jared Jennings, Claudio Loderer, Swaminathan Kalpathy, Florian Peters, Anjan Thakor, Philip Valt, Alexander Wagner, Toni Whited, and the audience of the workshop at the University of Bern and SFI Research Days 2018 for helpful comments. We are grateful for excellent research assistance from Mengzhou Wang. Ham acknowledges financial support from the Blanche “Peg” Philpott Faculty Fellowship.

1. Introduction

We live in a highly litigious society, as evidenced by the fact that legal expenses in the United States have exceeded \$4 trillion over the past two decades. Remarkably, annual legal expenses average \$192 billion, accounting for approximately 1.4% of GDP (see Figure 1). Corporate America is certainly front and center in this battle. Litigation risk, the uncertainty of being involved in lawsuits and suffering from future legal costs and penalties, can have a significant impact on corporate policies (e.g., Arena and Julio, 2015; Ferris, Jandik, Lawless, and Makhija, 2007; Gande and Lewis, 2009; Huang, Ozkan, and Xu 2023; Kim and Skinner, 2012; and Lin, Liu, and Manso, 2021). In this paper, we propose a simple and intuitive measure for firm-level litigation risk, validating the measure by linking it to realized litigation. We also re-examine its relation with several corporate policies previously studied in the literature that are expected to vary with litigation risk, as well as future realized fines and settlement payments.

In both the finance and accounting literatures, two approaches are commonly employed to measure firm-level litigation risk. The first approach involves event studies based on exogenous shocks to firms' litigation risk, such as the adoption of universal demand (UD) laws and rulings by the Ninth Circuit Court of Appeals (Bourveau, Lou and Wang 2018; Foroughi, Marcus, Nguyen, and Tehranian 2022, Choi and Pritchard 2012; Houston, Lin, Liu, and Wei, 2019; Crane and Koch 2018; Black, Ham, Kimbrough, and Yee, 2022). One advantage of this method is to provide clean identification, leveraging regulation shocks to address relevant endogeneity concerns. The second approach uses more direct measures of firms' litigation risk, such as industry membership, predicted probabilities of litigation derived from regressions on industry dummies and firm characteristics (e.g., Francis, Philbrick, and Schipper, 1994; Johnson, Kasznik, and Nelson, 2001; Field, Lowry, and Shu, 2005; Rogers and Stocken 2005; Kim and Skinner 2012), and federal judge ideology (e.g., Huang, Hui, and Li, 2021). This approach offers a direct and systematic assessment of litigation risk across firms, enabling researchers to explore a broad range of economic questions. Our paper aims to complement the second approach with a firm-level litigation risk measure constructed by extracting firm-specific litigation-related information from firms' annual reports

(SEC form 10-K). Our measure thus has the potential to offer additional insights into the impacts of firm-level litigation risk.

It is also worth noting that much of the focus in the literature has been on specific types of litigation risk, such as securities and class action lawsuits.⁵ While these types of lawsuits are undoubtedly significant and relevant studies have deepened our understanding on the effects of litigation risk (e.g., Ferris, Jandik, Lawless, and Makhija 2007; Cheng, Huang, Li, and Lobo 2010; McTier and Wald 2011), in practice, firms face a broad spectrum of litigation threats, including lawsuits related to patents, fraud, and antitrust violations, among others. Figure 2 illustrates the annual frequencies with which firms in our sample are named in a lawsuit. As seen there, just over 15% of firms are named in at least one type of lawsuit in a given year, and many firms face more than one suit. However, we find that firms are actually more likely to be named in a non-securities lawsuit than a securities lawsuit: firms are named in at least one securities lawsuit with a frequency of 7.3% annually, whereas 9.7% of firms are named annually in at least one non-securities lawsuit. Of these latter types of lawsuits, patents (intellectual property), contract disputes, and antitrust cases occur regularly. This highlights that firms are exposed to a diverse range of legal issues, all of which can exert significant pressure on management and influence corporate decision-making. Therefore, if only focusing on one type of litigation, the existence of other types of litigation could lead to potential concerns about omitted variables or measurement error. Our approach offers a single, comprehensive firm-level litigation risk measure that provides a holistic view of the firm's litigation risk, potentially enabling additional insights into its economic impact.

We extract firm-level litigation-related information by parsing SEC 10-K filings for a large sample of U.S. public firms. We identify legal language via the Loughran-McDonald Master Dictionary (Loughran and McDonald, 2011), which includes a list of litigious words. We also separately examine legal words that

⁵ Given that such lawsuits can name executive officers or board directors as defendants, it is intuitive that the personal litigation risk influences corporate policies made by the relevant individuals. In practice, corporate directors and executive officers have liability insurance (D&O insurance), but their personal litigation risk may not be fully shielded (Houston, Lin, and Xie, 2018). For example, such insurance policies usually do not cover cases of dishonesty and intentional wrongdoing, and insurers may seek to deny coverage. Furthermore, D&O insurance helps little with reputational losses.

include a negative connotation, versus either a neutral or positive connotation. We expect that litigious words with a *negative* connotation are more closely related to litigation risk than those with a neutral or positive connotation. To account for this, we classify the litigious words into negative litigious words and non-negative litigious words, following the word connotation classifications provided in the Loughran-McDonald Master Dictionary. This results in three categories: all legal words, negative legal words, and non-negative legal words, where the latter two are subsets of the first. Our analysis confirms that only negative legal words show significant predictability for future lawsuits.

Our litigation risk measure is the count of negative litigious words in firms' entire 10-K filings, scaled by the total file size of the 10-K.⁶ We interpret higher values of this measure to indicate greater pressure on firms stemming from potential or ongoing legal issues. Figure 3 demonstrates why it is important to use the firm's entire 10-K filing, as negative legal words are distributed widely across the filing. In fact, only 15% of the negative legal words in our measure come from the Legal Proceedings (Section 3) portion of the filing. In contrast, 46% and 25% of the negative legal words, respectively, come from the Business & Risk Factors (Section 1) and Financials (Sections 8 & 15) portions of the annual report. Conceptually, our legal risk measure captures firms' concerns about legal risks, including the uncertainty surrounding potential litigation in the future and the unknown outcome and penalties of ongoing lawsuits. We believe that one powerful advantage of this measure lies in its simplicity, transparency, and replicability: it requires minimal data constraints, is straightforward to construct, can be easily applied across large samples of firms, and is intuitive. These features make it a useful and accessible tool for researchers, investors, and practitioners seeking to assess litigation risk in corporate settings.

We take several steps to validate this firm-level measure for litigation risk. First, we investigate the predictive ability of this measure for future lawsuits filed against the firm using both univariate and multivariate analyses. In the univariate analysis, we parse our sample into quintiles of litigation risk (using our word count-based measure) and document the percentage of firms within each quintile that are targeted

⁶ Our list of words thereby contains all negative and litigious keywords from the full Loughran and McDonald dictionary. We thank Loughran and McDonald for sharing the word lists at <https://sraf.nd.edu/>

by a lawsuit in the following year. The incidence of lawsuit filings increases monotonically from the bottom quintile (7.8%) to the top quintile (22.9%), representing a nearly three-fold rate of incidence increase. Next, we parse our sample using measures of litigation risk proposed by prior studies. Specifically, we split the sample based on industry membership (Francis, Philbrick, and Schipper, 1994), firm characteristics (Kim and Skinner, 2012, hereafter as KS), and federal judge ideology (Huang, Hui, and Li, 2021), respectively. In each case, after splitting the sample based on high vs. low litigation risk, the incidence of litigation continues to increase when moving from the lowest to the highest quintile based on our text-based measure.

In the multivariate analysis, we show that our measure has incremental explanatory power above and beyond the securities litigation risk measure proposed by KS, which is widely used in the literature. Specifically, we regress a lawsuit dummy in year $t+1$ (capturing lawsuits filed against the firm in the following year) on our simple text-based litigation risk measure in year t , controlling for a current period lawsuit dummy (capturing lawsuits filed against the firm in the current year) and additional firm characteristics proposed in KS. Our firm-level litigation risk measure also allows us to include a variety of fixed effects, such as firm fixed effects, industry-by-year fixed effects, and circuit court-by-year fixed effects (for federal judge ideology, following Huang, Hui and Li 2019). The results confirm that our simple text-based litigation risk measure has significant predictive ability for future lawsuits (including both securities and non-securities lawsuits). These findings collectively show that our measure has strong potential to serve as a measure of litigation risk. Highlighting that these risks have real effects, we also show that our measure positively predicts future cash fines and settlements imposed on firms over the next three years.

Next, we contrast the predictive ability relative to a litigation risk measure derived from generative artificial intelligence. Doing so enables us to contrast the predictive content of a relatively simple and intuitive text-based measure to one that is less transparent and is computationally and monetarily expensive to construct. We create a litigation risk measure using generative AI by feeding it a subset of the sections comprising companies' 10-K filings. We then compare this AI-generated litigation risk measure with our simple word-count based litigation risk measure. Our analysis shows that the simple word-count based

litigation risk measure performs comparably to the AI-generated litigation measure in predicting future lawsuits. While this suggests that the simple word-count based measure can serve as a cost-effective and efficient candidate for gauging firm-level litigation risk, we acknowledge that the AI-constructed measure was constructed using only a portion of the 10-K filing and leave to future research to further examine the viability of using AI to construct measures of litigation risk.

In our final set of analyses, we revisit existing studies examining how litigation risk is related to a variety of corporate policies (e.g., Arena and Julio 2015; Gormley and Matsa 2011; Huang, Ozkan, and Xu, 2023; Krishnan, Masulis, Thomas, and Thompson, 2012; Nguyen, Phan, and Sun 2018). In particular, we examine the relation between our litigation risk measure and corporate cash holdings as well as various investment policies. For example, Arena and Julio (2015) show that litigation risk can lead firms to reduce both short-term (R&D expenditures) and long-term (capital and acquisition-related expenditures) investments and increase cash holdings. Using our litigation risk measure, we find consistent results: controlling for firm and circuit court-by-year fixed effects, a one-standard-deviation increase in our litigation risk measure is associated with lower future firm investment (capital expenditure, R&D, and acquisitions) of approximately 0.4% (a reduction of \$2.1 million for a firm with average book assets of \$677 million), and the reduction is comparable with the increase in cash holdings. These findings remain robust when controlling for the firm characteristics in the KS litigation risk model. We also provide evidence that the significant associations between current litigation risk and future firm policies are primarily driven by the component of the litigation risk *unrelated* to the presence of current lawsuits.

Our paper offers the following contributions to the literature. First, we propose a simple and intuitive measure of litigation risk, which is parsimonious and widely applicable as it is based solely on textual analysis of firms' 10-K filings that are publicly accessible on the SEC's website. By focusing on negative litigious keywords, this measure has an intuitive link to litigation risk, and it is constructed at the firm-year level, allowing us to exploit within circuit court-by-year variation in litigation risk (Huang, Hui, and Li, 2019). Further, while many studies tend to focus on a specific type of litigation, such as securities litigation, our measure encompasses various types of lawsuits, is available for all US publicly traded firms, and can

thus be used in a variety of empirical applications. In relevant literature, De La Bruslerie and Le Maux (2018) use media coverage on lawsuits to measure firm-level litigation risk. Arguably, lawsuits covered by the financial media would predominantly include large and well-known firms attracting public attention. Chung, Wynn, and Yi (2013) measure litigation risk by the level of liability insurance of corporate directors and officers, but the relevant data are only publicly available for Canadian firms.

Second, we contribute to the literature on the relation between litigation risk and corporate policies, and our findings are consistent with prior work. Existing studies in this space cover topics that include M&A and growth (Chu and Zhao 2020; Chung, Kim, Rabarison, To, and Wu, 2020; Krishnan, Masulis, Thomas, and Thompson, 2012; and Huang, Ozkan, and Xu, 2023), innovation (Cohen, Gurun, and Kominers, 2016, 2018; and Lin, Liu, and Manso, 2021), employment (Appel, Farre-Mensa, and Simintzi, 2019), payout policy (Arena and Julio, 2023), capital structure (Nguyen, Phan, and Lee, 2020), cost of capital (Houston, Lin, and Xie, 2018 and Ni and Yin, 2018), IPO underpricing (e.g. Tinic, 1988; Alexander, 1991, Hughes and Thakor, 1992; Lowry and Shu, 2002; and Hanley and Hoberg, 2012), misreporting (Laux and Stocken, 2012), disclosure and reporting choices (Houston, Lin, and Xie, 2018 and Black, Ham, Kimbrough, and Yee, 2022), and corporate governance (Ferris, Jandik, Lawless, and Makhija, 2007 and Appel 2016).⁷ Our paper is closely related to the portion of this literature studying the relation between litigation risk and investment. Arena and Julio (2015) find that firms with greater exposure to securities litigation cut investment and hold more cash. Nguyen, Phan, and Sun (2018) find lower shareholder litigation risk leads firms to invest in riskier, but value-enhancing projects. Li, Monroe, and Coulton (2023) show that lower managerial litigation risk leads to lower investment efficiency. Chung, Wynn, and Yi (2013) find that litigation risk plays a role in the relation between accounting quality and investment efficiency. Our paper complements this literature by considering firm-level litigation risk beyond that from a specific type of lawsuits (e.g., securities class actions).

⁷ There is a branch of this literature that relies on the staggered adoption of universal demand (UD) laws as shocks to litigation risk related to derivative lawsuits, but the validity of this setting has been challenged (Donelson, Kettell, McInnis, and Toynbee, 2022).

2. Data and sample

The data for our empirical analysis are obtained from various sources. Corporate accounting data are from Compustat and stock-related data are from CRSP. Lawsuit-related data are from Audit Analytics. Firm-level enforcement (government fines) data comes from Violation Tracker, maintained by the Corporate Research Project of Good Jobs First. The data to construct our litigation risk measure is from firms' SEC 10-K filings drawn from the EDGAR website (sec.gov/edgar). Our main sample period spans from 2000 to 2022, and variable definitions are in Appendix A.⁸

The summary statistics for the main variables are reported in Table 1. At least one lawsuit is filed against the firm in 15.1% of the firm-years in our sample.⁹ The average annual sales growth is 17.2%, and the average Tobin's q (Cash/Assets, Debt/Assets) is 2.080 (19.5%, 24.5%). The construction and characteristics of our litigation risk measure are described in the following section.

3. Litigation risk measure

Litigation risk is an important topic in finance and accounting due to its significant impact on corporate decisions. In the literature, two approaches are widely used to measure corporate litigation risk. The first approach relies on event studies based on litigation-related shocks, such as the staggered adoption of universal demand (UD) laws or rulings by the Ninth Circuit Court, which can provide clean identification strategies (e.g., Bourveau, Lou and Wang 2018; Foroughi, Marcus, Nguyen, and Tehranian 2022, Choi and Pritchard 2012; Houston, Lin, Liu, and Wei, 2019; Crane and Koch 2018; Black, Ham, Kimbrough, and Yee, 2022). The second approach uses more direct measures, such as industry membership, lawsuit indicators, or predicted litigation likelihood based on firm and industry characteristics (e.g., Kim and Skinner 2012), offering potentially more direct measurement of litigation risk. While many studies focus

⁸ The government fine data are available from 2000 to 2019.

⁹ The percent of firm-years with a lawsuit filed against the firm is larger than prior studies (both Kim and Skinner 2012 and Huang, Hui, and Li, 2019) because we consider all types of litigation, not only securities class action lawsuits. For comparability, we separately examine securities lawsuits and non-securities lawsuits in Table 5.

on a specific type of legal issue (e.g., securities class action lawsuits), in practice firms face a broad range of legal threats, including patent, fraud, product liability, and antitrust lawsuits, among others. Based on lawsuit data from Audit Analytics, Figure 2 shows that securities lawsuits occur in 7.3% of firm-years, highlighting their importance. However, non-securities suits occur in 9.7% of firm-years, with patent, contract and antitrust lawsuits occurring regularly (4.1%, 1.6%, 0.9% of firm-years, respectively). These various lawsuits can impose substantial pressure on management and consequently influence corporate decisions. To better understand the total impact of litigation on corporate policies, we argue that it is meaningful to capture overall litigation risk with a single and comprehensive firm-level measure.

We thus propose a simple text-based measure of firm-level litigation risk by conducting textual analysis on the content of firms' SEC 10-K filings, which includes firm-level information related to litigation, both current and anticipated. Publicly-traded firms are required to disclose information regarding material pending litigation to shareholders, and companies are also required to disclose material risk factors, which oftentimes includes a discussion of legal and regulatory-related risks. Importantly, these disclosures are not restricted to any particular type of legal issue. Thus, these SEC filings may include valuable information on litigation risk stemming from all types of legal issues. The general counsel and executives of the firm are arguably the most informed regarding the firm's litigation risk, and as such, their mandatory disclosures are likely to be a good source of litigation-related information to construct a measure of litigation risk. In this paper, we thus use a simple word-count-based approach to construct a firm-level litigation risk measure.

To construct the measure, we extract litigation-related information by parsing 10-K filings for a large sample of US public firms. We identify an initial list of litigious words using the Loughran-McDonald Master Dictionary (Loughran and McDonald, 2011). The dictionary also denotes the words' connotation, which we incorporate into our measure because we expect litigious words with a negative connotation to be more strongly associated with litigation risk than those with a positive or neutral connotation. We thus split the litigious words into negative litigious words and non-negative litigious words, where the classification of word connotation also follows the Loughran-McDonald Master Dictionary. This leaves us

with three sets of litigious words: all legal words, negative legal words, and non-negative legal words, among which the last two sets are subsets of the first. For illustrative purposes, Figure 4 presents word clouds of the legal words in 10-K filings for firms with high levels of negative and non-negative legal words. Panel A (B) illustrates the word cloud figure for all legal words in 10-K filings of the five firms with the most negative (non-negative) legal words in our sample.¹⁰ The legal words with a negative connotation appear to be more closely related to concerns about legal issues expressed in firms' 10-K filings, though we explore this more directly in the empirical analyses.

To create our measure, we count the number of litigious words (in each of the three groups) and scale the word counts by the 10-K file size (in kilobytes).¹¹ We then have three candidates for measuring litigation risk and denote them by *All Legal*, *Neg Legal*, and *Non-Neg Legal*, respectively. To investigate the association between the candidate measures and future lawsuits, we conduct panel regressions using the following specification:

$$Lawsuit_{i,t+1} = \beta_0 + \beta_1 \cdot Candidate_{i,t} + X_{i,t} \cdot \Gamma + \mu_i + \gamma_k \times v_t + \varepsilon_{i,t+1}, \quad (1)$$

where i is the firm index, t is the year index, and k is the circuit court index. $Lawsuit_{i,t+1}$ is an indicator variable equal to one if a new lawsuit is filed against firm i in year $t+1$ (i.e., not lawsuits filed in the previous or current years), $Candidate$ is one of the candidate measures for firm-level litigation risk as defined above, X is a vector of control variables following KS, Γ is the vector of the corresponding coefficients, μ_i is firm fixed effects, $\gamma_k \times v_t$ is circuit court-by-year fixed effects (to control for federal judge ideology effect, following Huang, Hui, and Li, 2019), and $\varepsilon_{i,t+1}$ is the error term.

Table 2 reports the results. Columns 1, 2, and 3 present the results using the measure *All Legal*, *Neg Legal*, and *Non-Neg Legal*, respectively. Column 1 shows that the coefficient on *All Legal* is positive but not statistically significant at conventional levels, suggesting there is no significant predictive power for the candidate measure based on all litigious words. Column 2 shows that the coefficient on *Neg Legal*

¹⁰ Internet Appendix Table IA1 reports examples of common litigious words in each of the three categories.

¹¹ Loughran and McDonald (2014) propose the 10-K file size as a measure of financial document readability, which is the ability of investors and analysts to assimilate information from a financial disclosure. Our results are robust when scaled by either total words in the 10-K or total litigious words in the 10-K.

is positive and statistically significant at the 1% level, suggesting *Neg Legal* may be a valid candidate to measure firm-level litigation risk. Column 3 shows that the coefficient on *Non-Neg Legal* is negative but not statistically significant at conventional levels. In Column 4, we include both *Neg Legal* and *Non-Neg Legal* in the same regression. The results show that the coefficient on *Neg Legal* remains positive and statistically significant at the 1% level, and the coefficient on *Non-Neg Legal* remains negative but becomes statistically significant at the 1% level. This suggests that negative litigious words can still predict future lawsuits in a significant way, but *Non-Neg Legal* is actually negatively associated with the likelihood of future lawsuits once controlling for negative litigious words.

Based on the investigation above, we propose *Neg Legal* as our firm-level litigation risk measure and denote this measure as *Legal Risk* in our following analysis. In Appendix B, we provide several examples of negative litigious words appearing in 10-K filings, including the surrounding text to offer some additional context. It is important to note that our measure captures both current and future legal issues. Statistics of this litigation risk measure are reported in Table 1. Its mean and standard deviation across industries (Fama-French 12) are illustrated in Figure 5, demonstrating considerable variation in the measure even within industry. Figure 3 highlights the dispersion of negative legal words across the *entire* 10-K filing, with 46% of negative legal words appearing among the “Business & Risk Factors” section (Section 1), 25% in the “Financial Statements & Exhibits” section (Sections 8 & 15), and 15% in the “Legal Proceedings” section (Section 3).

4. Validating our litigation risk measure

In this section, we seek to validate our firm-level litigation risk measure. We conduct both univariate tests and multivariate regression analyses to test the relation between our measure of litigation risk and the incidence of being targeted in a lawsuit in future periods. Prior research predominantly focuses on one specific type of litigation risk, such as securities class action litigation. While these studies have significantly increased our understanding of how specific types of litigation can affect corporate decisions,

we seek to complement this line of research by focusing on overall firm-level litigation risk, which can be related to various types of firm litigation.

4.1. Univariate analyses

We begin by conducting univariate tests of the relation between our firm-level measure of litigation risk and the likelihood of being targeted in a future lawsuit. We partition the sample into quintiles based on the value of *Legal Risk* within a year and calculate the likelihood of lawsuits within each quintile in the following year, where the likelihood is calculated by the number of firm-years targeted in a lawsuit in the following year divided by the total number of firm-years within each quintile. Table 3 reports the results of the univariate tests.

Panel A displays the frequency of lawsuits across *Legal Risk* quintiles for all firm-years in our sample. In the bottom (top) quintile of *Legal Risk*, 7.8% (22.9%) of firms are targeted in a lawsuit in the next year, which indicates that the likelihood of new lawsuits next year increases by 193.6% from firms in the bottom quintile to those in the top quintile, and this difference is statistically significant at the 1% level.¹² These results suggest that the firms with greater values of *Legal Risk* are in fact more likely to be involved in future lawsuits.

Prior studies have proposed other litigation risk measures based on industry membership, firm characteristics, or judge ideology. We thus investigate whether our measure contains information content above and beyond prior measures of litigation risk. To do so, we first partition our sample into high-litigation risk and low-litigation risk groups using these prior measures. We then study the association between our litigation risk measure and lawsuit likelihood within each sub-group in a similar manner to the approach used in Panel A. Specifically, within the high-litigation risk and low-litigation risk subsamples, we further partition the subsamples into quintiles based on our measure and we present the likelihood of being sued within each quintile.

¹² $193.6\% = (0.229 - 0.078)/0.078$.

Panel B reports the results for the partition based on FPS industries, which represent more litigious industries. The left (right) partition of Panel B shows the lawsuit likelihood across *Legal Risk* quintiles for firms in FPS (non-FPS) industries. As expected, the lawsuit likelihood is greater for FPS industries than non-FPS industries across each of the corresponding quintiles, which is consistent with firms in FPS industries tending to have higher litigation risk. When focusing on the variation of lawsuit likelihood within FPS industries in the left partition, we observe that the lawsuit likelihood shows an increasing trend across *Legal Risk* quintiles (from bottom to top). The lawsuit likelihood is 14.6% (25.3%) in the bottom (top) quintile, which shows that the likelihood of new lawsuits next year increases by 73.3% from firms in the bottom quintile to those in the top quintile and this difference is statistically significant at the 1% level. The right partition shows consistent results within the non-FPS firms. That is, the likelihood of new lawsuits next year increases by 210.1% from firms in the bottom quintile to those in the top quintile. These results indicate that our measure of litigation risk still has significant explanatory power within FPS and non-FPS industries.

Panel C reports the results for the partition based on both industry membership and firm characteristics following the KS litigation risk prediction model. For each firm-year in our sample, we estimate the likelihood of being sued based on the firm characteristics used by KS. We then partition the sample using the median value of the likelihood of being sued in each year in the sample. The left (right) partition of Panel C reports the results for firm-years with a predicted litigation risk above (below) the annual median. As expected, firm-years in the left partition are more likely to be targeted by a lawsuit next year in each quintile. We focus on the variation across the quintiles that captures variation in our litigation risk measure while holding constant variation in litigation risk captured by the firm characteristics used in KS. The left partition of Panel C shows that lawsuit likelihood monotonically increases across *Legal Risk* quintiles (from bottom to top). The likelihood of a lawsuit is 13.8% (32.4%) in the bottom (top) quintile, suggesting the likelihood of new lawsuits increases by 134.8% from firms in the bottom quintile to those in the top quintile, and the difference is statistically significant at the 1% level. The right partition of Panel

C reports consistent results for the group with low predicted litigation risk, with the likelihood of new lawsuits next year increases by 268.4% from firms in the bottom quintile to those in the top quintile.

Panel D reports the results for the partition based on judge ideology following Huang, Hui, and Li (2019).¹³ The left (right) partition of Panel D reports the results for firm-years with the liberal court measure above (below) the annual median. Similar to the other panels, whether in the more liberal court partition or the less liberal court partition, we find monotonic increases in the likelihood of being sued across the quintiles of *Legal Risk*. In the left partition of Panel D the lawsuit likelihood is 7.1% (21.9%) in the bottom (top) quintile, which shows that the likelihood of new lawsuits next year increases by 208.5% from firms in the bottom quintile to those in the top quintile, and the difference is statistically significant at the 1% level. The right partition of Panel D reports consistent results for the group with less liberal courts as the firms in the top quintile are 181.1% more likely to be targeted in a lawsuit than those in bottom quintile.¹⁴

4.2 Multivariate regression analyses

To further validate our litigation risk measure, we conduct a series of multivariate regression analyses. Specifically, we estimate the following specification:

$$Y_{i,t+1} = \beta_0 + \beta_1 \cdot Legal\ Risk_{i,t} + X_{i,t} \cdot \Gamma + FEs + \varepsilon_{i,t+1}, \quad (2)$$

where i is the firm index and t is the year index. $Y_{i,t+1}$ is an indicator variable equal to one if a new lawsuit is filed against firm i in year $t+1$ (i.e., not lawsuits filed in previous or current years), *Legal Risk* is our firm-level litigation risk measure as defined earlier, X is a vector of control variables following KS, Γ is the

¹³ We thank Allen Huang for sharing the data on judge ideology.

¹⁴ We note the lawsuit likelihood is slightly higher for the less liberal courts (the right portion of Panel D) than for the more liberal courts (the left portion of Panel D). The reason appears to be that we focus on all types of lawsuits, whereas Huang, Hui, and Li, (2019) focus only on securities lawsuits. Internet Appendix Table IA2 presents Panel D of Table 3 separately for securities lawsuits and non-securities lawsuits, and firms are more likely to be targeted by securities lawsuits in the more liberal circuit courts, but less likely to be targeted by non-securities lawsuits in the more liberal circuit courts.

vector of corresponding coefficients, FEs represents the fixed effects, and ε is the error term. Our baseline analysis uses a linear probability model.¹⁵ Robust standard errors are clustered at the firm level.

4.2.1 All lawsuit types

In the first set of analyses, we consider all lawsuits against a firm, and then subsequently separate securities lawsuits and non-securities lawsuits. If the firm-level litigation risk measure carries unique information related to future lawsuits, we expect the coefficient on *Legal Risk* to be significantly positive.

Table 4 reports the results. The dummy variable *Lawsuit* includes all lawsuits filed against the firm in year $t+1$, including both securities-related and non-securities-related lawsuits. A set of control variables from KS are included: firm size ($\text{Log}(\text{Assets})$), firm growth (*Sales Growth*), market adjusted stock returns (*Excess Return*), stock return skewness (*Skewness*), stock return volatility (*Volatility*), and stock turnover (*Turnover*).¹⁶ We also control for lawsuits filed against the firm in year t . Our focus is the firm-level litigation risk measure, *Legal Risk*. Column 1 includes firm and year fixed effects. The coefficient on *Legal Risk* is positive and statistically significant at the 1% level. This suggests that after controlling for the relevant firm and stock characteristics in KS and current lawsuits, our firm-level litigation risk measure still has significant explanatory power for future lawsuits filed against the firm. The economic impact is also significant. The coefficient 0.143 indicates that when *Legal Risk* increases by one standard deviation (0.086), the probability of a new lawsuit next year increases by 1.2%, which is 8% of the average lawsuit probability in our sample.

Column 2 includes firm fixed effects and industry-by-year fixed effects, which further address potential concerns about within-industry, time-varying omitted variables. Column 3 includes firm fixed effects and circuit court-by-year fixed effects, which address within-circuit court time-varying omitted

¹⁵ Our results are robust when using a logit model (see Internet Appendix Table IA3). In our main analysis we use the linear probability model because the inclusion of fixed effects in nonlinear models may lead to biased coefficients and standard errors due to the incidental parameters problem.

¹⁶ KS also include a dummy variable for litigious industries, which is absorbed by the firm fixed effects in our model.

variables (e.g., judge ideology). The results are consistent with those in Column 1 and the coefficient on *Legal Risk* remains significantly positive.^{17,18}

4.2.2 Securities and non-securities lawsuits

Next, we separate the lawsuits into two groups. The first group includes securities and class action lawsuits that are more commonly covered by existing studies, and the second group includes the remaining types of lawsuits. We define an indicator variable *Securities (Non-Securities)* for the presence of a new lawsuit in the first (second) group filed against a firm in a year (i.e., again not lawsuits filed in current or previous years). We then study the relation between the firm-level litigation risk measure and the relevant types of future lawsuits using similar specifications as in the previous section.

Table 5 reports the results. Both specifications include firm fixed effects and circuit court-by-year fixed effects. The coefficient on *Legal Risk* is positive in both specifications, and statistically significant at the 10% and 1% level, respectively. It is worth noting that the coefficient on our litigation risk measure is larger when predicting non-securities lawsuits than when predicting securities lawsuits (the relevant R-squared values are much larger too, e.g., 0.305 in Column 2 vs. 0.217 in Column 1), which suggests our firm-level litigation risk measure may have incremental advantages in predicting non-securities lawsuits.¹⁹

4.3 Generative-AI-based litigation risk measure

In this section, we use a generative artificial intelligence (AI) tool, ChatGPT, to generate a firm-level litigation risk measure and compare the performance in predicting future lawsuits between the

¹⁷ In robustness tests, additional lags of *Lawsuit* are included as controls. The results are reported in the Internet Appendix Table IA4.

¹⁸ Alternatively, a litigation probability can be predicted via the Kim and Skinner (2012) litigation risk model and used as a separate litigation risk measure. We do so in Internet Appendix Table IA5 and our measure is incrementally predictive of future lawsuits even when controlling for the probability of being sued (Prob. Pred.) based on the Kim and Skinner (2012) model.

¹⁹ In robustness tests, we further examine finer subcategories of non-securities lawsuits by regressing the indicators of five subcategories of non-securities lawsuits (i.e., patent, contract, antitrust, injury, and ERISA) on *Legal Risk* and the results are reported in Internet Appendix Table IA6. The results show that the coefficients on *Legal Risk* are positive in all five specifications and statistically significant in four of them (except ERISA lawsuits).

generative AI measure and our simple word count measure. ChatGPT, developed by OpenAI, is a generative large language model (LLM) based on the Generative Pre-trained Transformer (GPT) architecture. It has undergone several iterations since its inception, evolving to become more sophisticated in understanding and generating human-like text. For example, one milestone is the GPT-3 model released in June 2020, which has 175 billion parameters and was trained on 45TB of data.

To extract useful litigious information from 10-K filings, generative AI offers a potential advantage by analyzing not only specific litigious words, but also incorporating context. This enables it to potentially capture nuanced discussions about legal issues. However, the output is less transparent as it is not clear how the model will estimate the firm's litigation risk. It is also very resource intensive whereas the word-count based method has the advantage of being cost-effective, highly replicable, transparent and intuitive, which may make it a practical tool for gauging litigation risk across a large sample of firms.

To construct the alternative litigation risk measure at the firm-year level, we ask ChatGPT to read firms' 10-K filings and provide a score between 0 and 100, where a higher score indicates greater litigation risk. Specifically, we use GPT-3.5 Turbo API that was released in November 2022.²⁰ It has a total limit of 4,096 tokens (about 3,000 words) per request, but firms' 10-K filings are usually much longer. For example, in our sample, the average (median) length of a 10-K filing is about 66,520 (57,028) words. Following relevant studies using ChatGPT to deal with corporate disclosure files (e.g., Armstrong 2023; Choi and Kim, 2023; Dasgupta, Li, and Wu, 2024), instead of using the whole 10-K filing, we focus on two sections closely related to firms' concerns about legal issues, which are Item 1A "Business/Risk factors" and Item 3 "Legal Proceedings".²¹ Specifically, we ask ChatGPT to read each of these two sections separately using the following prompt:

"You are a legal expert on corporate finance. Here is part of an SEC 10K filing about a company, can you please evaluate the legal risk of the company based on this piece of text?"

²⁰ When generating the litigation measure by ChatGPT, GPT-4 was available but would be prohibitively expensive for us to use for all firm-years in our sample.

²¹ As illustrated in Figure 3, Sections 1 and 3 of the 10-K capture 61% of the negative legal words we capture in our proposed measure.

Specifically, please create a measure of litigation risk, where 100 means very high legal risk and 0 means very low legal risk.”

We collect the scores generated by ChatGPT for each section and take the maximum value between the two scores as the final score for each firm-year. We then scale the final score by 100 and use the scaled value as the generative AI-based litigation risk measure, denoted by GAI. To compare our simple word-count based litigation measure with the generative AI measure, we re-run the tests predicting future lawsuits using Equation (2).

Table 6 reports the results. Column 1 (2) of Panel A reports the result only using our word-count based (generative AI) litigation risk measure. The coefficients on both measures are positive and statistically significant at the 1% level, suggesting both measures provide significant predictability for future lawsuits.²² The correlation between these two measures is 0.20, which means they are positively correlated, but the correlation is not particularly high. Seemingly, they tend to capture their own specific information on firms’ litigation risk. In Column 3, we include both measures in the same regression. The result shows that the coefficients on both measures remain positive and statistically significant at the 1% level, again suggesting that both litigation measures contain unique predictability for future lawsuits. To better compare the predictability, we standardize both measures in the results reported in Panel B. As shown in Column 3, a one standard deviation increase in the word-count based (AI-based) measure increases the likelihood of new lawsuits next year by 1.1% (0.4%), which is 7% (3%) of the average of *Lawsuit* in our sample. Our simple word-count litigation measure thus displays comparable predictability for future lawsuits as the generative AI measure.

We do not intend to suggest that the simple word-count approach is superior to generative AI in capturing litigation risk. Rather, our findings reflect that within the scope of this paper’s implementation, the word-count-based measure performs comparably to a resource-intensive AI-generated measure. We

²² The number of observations in Column 1 of Table 6 is slightly different from that of Column 3 of Table 4 because in Table 6 we restrict the sample to firm-year observations with the AI-based measure (GAI) available for comparison purposes.

emphasize that the AI-based litigation risk measure is subject to important limitations. Due to token constraints and computational cost, we feed only two sections of each 10-K (Items 1A and 3) to the model and use a general prompt with no supervised calibration. A more complete implementation, e.g., based on the full 10-K, purpose-trained models, or domain-adapted LLMs, may well yield stronger predictive performance. We leave such exploration to future research. In sum, rather than downplaying the promise of generative AI, our point is to show that a simple and transparent word-count approach continues to perform well, even when compared with a more advanced tool. Given the potential concern about LLM hallucination and the relative lack of transparency in generative outputs, our approach may serve as a useful benchmark, one that is low-cost, replicable, reliable, and interpretable.

5. Firm-level litigation risk and corporate policies

Prior literature has shown that litigation risk can affect various corporate policies, such as cash holdings and investment decisions. In this section, with our new measure in hand, we revisit the relation between litigation risk and several corporate policies, including cash holdings, capital expenditures, R&D expenditures, and acquisitions using our firm-level litigation risk measure.

5.1 Cash holdings, settlement costs, and government fines

Lawsuits can be financially costly to firms, so litigation risk can induce firms to hoard more cash for precautionary motives. In this section, using our litigation risk measure, we first investigate the relation between litigation risk and firms' future financial costs related to legal issues, such as settlements and government fines. We then study how firm-level litigation risk is associated with cash holdings. The following specification is used in our regression analysis.

$$Y_{i,t+1} = \beta_0 + \beta_1 \cdot Legal\ Risk_{it} + X_{it} \cdot \Gamma + \mu + \gamma_k \times v_t + \varepsilon_{i,t+1}, \quad (3)$$

where i is the firm index, t is the year index, k is the circuit court index, Y is settlement costs, government fines, or cash holdings (all scaled by total assets), $Legal\ Risk$ is our litigation risk measure as defined earlier,

X is a vector of control variables, Γ is the vector of the corresponding coefficients, μ is for firm (or industry) fixed effects, $\gamma_k \times v_t$ is circuit court-by-year fixed effects, and ε is the error term. Robust standard errors are clustered at the firm level.

Table 7 reports the results. The odd (even) columns include industry (firm) fixed effects. All columns include circuit court-by-year fixed effects. Columns 1 and 2 report the results for settlements. The coefficients on *Legal Risk* are both positive and statistically significant at the 1% level, which suggests that litigation risk is positively associated with firms' settlement costs in the subsequent year. Columns 3 and 4 report the results for government fines, which are consistent with those in the first two columns. The results show that litigation risk is positively associated with future financial burdens related to legal issues, and our firm-level litigation measure is useful to capture this intuitive relation.

Columns 5 and 6 report the results for cash holdings. The coefficients on *Legal Risk* are both positive and statistically significant at the 1% level, which is consistent with the findings in literature that firms facing higher litigation risk tend to hold more cash in the future (e.g., Arena and Julio 2015). The economic impact is also significant. For example, Column 6 (with firm and circuit court-by-year fixed effects) shows that a one standard deviation increase in *Legal Risk* is associated with a 0.4 percentage-point increase in future cash holdings. For a firm with average book assets of 677 million dollars, it corresponds to an increase in cash holdings of 2.6 million dollars.

5.2 Capital expenditures, R&D expenses, and acquisitions

As high litigation risk stimulates firms to hoard liquidity, their investment activities can also be restricted (e.g., Arena and Julio 2015). In this section, we investigate the relation between litigation risk and various components of corporate investments, including capital expenditures, R&D expenditures, and acquisitions. The specification in Equation (3) is used for the relevant regression analysis. Table 8 reports the results. The odd (even) columns include industry (firm) fixed effects. All columns include circuit court-by-year fixed effects. Columns 1 and 2 report the results for capital expenditures. The coefficient on *Legal*

Risk is negative and statistically significant at the 1% level in both columns, which suggests that firms with high litigation risk tend to reduce their future capital expenditures.

Columns 3 and 4 (5 and 6) report the results for R&D expenses (acquisitions). The findings are consistent with those for capital expenditures. The coefficients on *Legal Risk* are all negative and statistically significant at the 5% or 1% level. Regarding the economic impact, the result in Column 2 (4, 6) shows that a one standard deviation increase in *Legal Risk* is associated with a 9 (16.3, 12) basis-point decrease in future capital expenditures (R&D expenses, acquisitions). For a firm with average book assets of 677 million dollars, the total reduction in the three types of investments corresponds to 2.1 million dollars, which is comparable with the increase in cash holdings documented above. These findings suggest that a firm's litigation risk is negatively associated with various forms of investment activities, thereby shrinking investment and potentially curbing firm growth.

Next, we conduct a set of robustness tests on the relation between firm-level litigation risk and corporate policies. In the first set of robustness tests, we also control for the factors in the Kim and Skinner (2012) litigation risk model and re-estimate the models in Tables 7 and 8 (with firm fixed effects). Panel A of Table 9 reports the results. Specifically, besides the typical control variables used in the corporate finance literature, we *also* control for the variables included in the KS litigation risk model (see Column 3 in our Table 4). Firm fixed effects and circuit court-by-year fixed effects are also included. The results show that the coefficients on *Legal Risk* in all 6 columns remain statistically significant at the 1% level, with consistent signs to the results reported above. Furthermore, the magnitude of the coefficients on *Legal Risk* is similar to those in the corresponding tests, which suggests that our firm-level litigation risk measure tends to capture distinct information relative to the firm characteristics included in the KS litigation risk model.²³

In 10-K filings, firms can discuss concerns about litigation risk in their businesses as well as risks stemming from ongoing lawsuits. A natural question regarding our measure is whether it contains

²³ The predictive power of our litigation risk measure for corporate policies also remains significant when controlling for a lawsuit dummy and the probability of being sued per the Kim and Skinner (2012) model, as reported in Internet Appendix Table IA7.

information about ongoing lawsuits (e.g., uncertainty about the trial or settlement outcomes) or uncertainty related to the potential for future legal issues. To address this question, we seek to strip the information related to current lawsuits from our measure. Specifically, for each firm in our sample, we regress our litigation risk measure on a dummy variable capturing lawsuits filed against the firm in the current year. The fitted values serve as our proxy for litigation risk related to current lawsuits, which is denoted by *Legal Risk (Current)*. The residuals serve as our proxy for litigation risk related to future legal issues, which is denoted by *Legal Risk (Future)*. We then use both measures in the analysis of corporate policies and investigate how each component is associated with these firm decisions.

Panel B of Table 9 reports the results. Columns 1 and 2 show that for financial costs of legal issues (measured by settlements and government fines), the coefficients on both legal risk measures are statistically significant at the 1% level, which is intuitive as both ongoing lawsuits and future legal issues are likely to lead to significant financial costs. In contrast, when considering cash, capital expenditures, R&D expenses, and acquisitions, only the coefficients on *Legal Risk (Future)* are significant, suggesting the component related to future legal issues is the driving factor for future firm policies.

6. Conclusion

This paper proposes a simple firm-level litigation risk measure, which adopts a word-count-based approach using litigious words in firms' 10-K filings. This measure has the following advantages. First, it is intuitive and transparent, as it simply counts the occurrence of litigious words with a negative connotation in firms' 10-K annual reports. This straightforward methodology makes it easy to understand and interpret. Second, the measure is easy to replicate, as it relies on publicly available 10-K filings and a well-defined set of litigious words, ensuring consistency and reproducibility across studies. Third, it provides a forward-looking, firm-specific measure of litigation risk, capturing management's own assessment of potential legal exposure as reflected in the firm's disclosures. This allows for the detection of risk that may not yet have materialized into lawsuits but are embedded in the firm's narrative. Unlike historical lawsuit data, which are backward-looking, or industry-based measures, which generalize across firms, this word-count-based

approach reflects unique, firm-specific risk. Fourth, it is scalable and cost-effective, enabling the systematic analysis of a large number of firms' filings over time using automated textual analysis. In sum, by leveraging textual disclosures, the measure offers a transparent, replicable, and nuanced tool to evaluate litigation risk effectively and efficiently.

We validate this simple litigation risk measure through both univariate and multivariate analyses. We provide evidence that this measure shows significant predictability for future lawsuits after controlling for existing litigation risk measures (e.g., industry membership, current lawsuits, and firm characteristics related to litigation risk specified in the literature). We also construct an AI-generated litigation risk measure for our large sample of US firms and show that the simple word-count based litigation measure performs comparably to the AI-generated litigation measure in predicting future lawsuits. We further re-examine the relation between litigation risk and corporate policies, such as cash holdings and investment policies (capital expenditure, R&D, and acquisitions), using the simple firm-level litigation risk measure, and show findings consistent with the existing literature.

References

- Alexander J (1991) Do the merits matter? A study of settlements in securities class actions. *Stanford Law Review* 43: 497–588.
- Appel I (2019) Governance by litigation. *University of Virginia Working Paper*.
- Appel I, Farre-Mensa J, Simintzi E (2019) Patent trolls and startup employment. *Journal of Financial Economics*, 133(3): 708-725.
- Arena M, Julio B (2015) The effects of securities class action litigation on corporate liquidity and investment policy. *Journal of Financial and Quantitative Analysis*, 50(1-2): 251-275.
- Arena M, Julio B (2023) Litigation risk management through corporate payout policy. *Journal of Financial and Quantitative Analysis*, 58(1): 148-174.
- Armstrong D (2023) Measuring tax enforcement with generative AI. *University of Chicago Working Paper*.
- Black J, Ham C, Kimbrough M, Yee H (2022) Legal expertise and the role of litigation risk in firms' conservatism choices. *The Accounting Review*, 97(4): 105-129.
- Bourveau T, Lou Y, Wang R (2018) Shareholder litigation and corporate disclosure: Evidence from derivative lawsuits. *Journal of Accounting Research*, 56(3): 797-842.
- Cheng C, Huang H, Li Y, Lobo G (2010) Institutional monitoring through shareholder litigation. *Journal of Financial Economics*, 95(3). 356-383.
- Choi G, Kim A (2023) Economic footprints of tax audits: A generative ai-driven approach. *Chicago Booth Research Paper*, 23-23.
- Choi S, Pritchard A (2012) The Supreme Court's impact on securities class actions: An empirical assessment of Tellabs. *The Journal of Law, Economics, & Organization*, 28(4): 850-881.
- Chu Y, Zhao Y (2021) The dark side of shareholder litigation: Evidence from corporate takeovers. *Financial Management*, 50(3): 845-873.
- Chung C, Kim I, Rabarison M, To T, & Wu E (2020) Shareholder litigation rights and corporate acquisitions. *Journal of Corporate Finance*, 62: 101599.
- Chung H, Wynn J, Yi H (2013) Litigation risk, accounting quality, and investment efficiency. *Advances in Accounting*, 29(2): 180-185.
- Cohen L, Gurun U, Kominers S (2016) The growing problem of patent trolling. *Science*, 352(6285): 521-522.
- Cohen L, Gurun U, Kominers S (2019) Patent trolls: Evidence from targeted firms. *Management Science*, 65(12): 5461-5486.
- Crane A, Koch A (2018) Shareholder litigation and ownership structure: Evidence from a natural experiment. *Management Science*, 64(1): 5-23.
- Dasgupta S, Li E, Wu S (2023) Inferring Financial Flexibility: Do Actions Speak Louder than Words?. *SSRN Scholarly Paper*. Rochester, NY.
- de La Bruslerie H, & Le Maux J (2018) Litigation risk: Measurement and impact on M&A transaction terms. *Journal of Business Finance & Accounting*, 45(7-8): 952-996.
- Donelson D, Kettell L, McInnis J, Toynbee S (2022) The need to validate exogenous shocks: Shareholder derivative litigation, universal demand laws and firm behavior. *Journal of Accounting and Economics*, 73(1): 101427.
- Ferris S, Jandik T, Lawless R, Makhija A (2007) Derivative lawsuits as a corporate governance mechanism: Empirical evidence on board changes surrounding filings. *Journal of Financial and Quantitative Analysis*, 42(1): 143-165.
- Field L, Lowry M, Shu S (2005) Does disclosure deter or trigger litigation?. *Journal of Accounting and Economics*, 39(3): 487-507.
- Foroughi P, Marcus A, Nguyen V, Tehranian H (2022) Peer effects in corporate governance practices: Evidence from universal demand laws. *The Review of Financial Studies*, 35(1): 132-167.
- Francis J, Philbrick D, Schipper K (1994) Shareholder litigation and corporate disclosures. *Journal of Accounting Research*, 32(2): 137-164.

- Frankel R, Jennings J, Lee J (2022) Disclosure sentiment: Machine learning vs. dictionary methods. *Management Science*, 68(7): 5514-5532.
- Gande A, Lewis C (2009) Shareholder-initiated class action lawsuits: Shareholder wealth effects and industry spillovers. *Journal of Financial and Quantitative Analysis*, 44(4): 823-850.
- Gormley T, Matsa D (2011) Growing out of trouble? Legal liability and corporate responses to adversity. *Review of Financial Studies*, 24(201): 1.
- Hadlock C, Pierce J (2010) New evidence on measuring financial constraints: Moving beyond the KZ index. *Review of Financial Studies*, 23(5): 1909-1940.
- Hanley K, Hoberg G (2012) Litigation risk, strategic disclosure and the underpricing of initial public offerings. *Journal of Financial Economics*, 103(2): 235-254.
- Hassan T, Hollander S, Van Lent L, Tahoun A (2019) Firm-level political risk: Measurement and effects. *Quarterly Journal of Economics*, 134(4): 2135-2202.
- Houston J, Lin C, Xie W (2018) Shareholder protection and the cost of capital. *The Journal of Law and Economics*, 61(4): 677-710.
- Houston J, Lin C, Liu S, Wei L (2019) Litigation risk and voluntary disclosure: Evidence from legal changes. *The Accounting Review*, 94(5): 247-272.
- Huang A, Hui K, Li R (2019) Federal judge ideology: A new measure of ex ante litigation risk. *Journal of Accounting Research*, 57(2): 431-489.
- Huang C, Ozkan N, Xu F (2023) Shareholder litigation risk and firms' choice of external growth. *Journal of Financial and Quantitative Analysis*, 58(2): 574-614.
- Hughes P, Thakor A (1992) Litigation risk, intermediation, and the underpricing of initial public offerings. *Review of Financial Studies*, 5(4): 709-742.
- Johnson M, Kasznik R, Nelson K (2001) The impact of securities litigation reform on the disclosure of forward-looking information by high technology firms. *Journal of Accounting Research*, 39(2): 297-327.
- Kim I, Skinner D (2012) Measuring securities litigation risk. *Journal of Accounting and Economics*, 53(1-2): 290-310.
- Krishnan C, Masulis R, Thomas R, Thompson R (2012) Shareholder litigation in mergers and acquisitions. *Journal of Corporate Finance*, 18(5): 1248-1268.
- Laux V, Stocken P (2012) Managerial reporting, overoptimism, and litigation risk. *Journal of Accounting and Economics*, 53(3): 577-591.
- Lee G, Naughton J, Zheng X, Zhou D (2020) Predicting litigation risk via machine learning. *University of British Columbia Working Paper*.
- Li L, Monroe G, Coulton J (2023) Managerial litigation risk and corporate investment efficiency: Evidence from universal demand laws. *Journal of Empirical Legal Studies*, 20(1): 196-232.
- Lin C, Liu S, Manso G (2021) Shareholder litigation and corporate innovation. *Management Science*, 67(6): 3346-3367.
- 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.
- Loughran T, McDonald B (2014) Measuring readability in financial disclosures. *Journal of Finance*, 69(4): 1643-1671.
- Lowry M, Shu S (2002) Litigation risk and IPO underpricing. *Journal of Financial Economics*, 65(3): 309-335.
- McTier B, Wald J (2011) The causes and consequences of securities class action litigation. *Journal of Corporate Finance*, 17(3): 649-665.
- Nguyen H, Phan H, Sun L (2018) Shareholder litigation rights and corporate cash holdings: Evidence from universal demand laws. *Journal of Corporate Finance*, 52: 192-213.
- Nguyen N, Phan H, Lee E (2020) Shareholder litigation rights and capital structure decisions. *Journal of Corporate Finance*, 62: 101601.
- Ni X, Yin S (2018) Shareholder litigation rights and the cost of debt: Evidence from derivative lawsuits. *Journal of Corporate Finance*, 48: 169-186.

- Rogers J, Stocken P (2005) Credibility of management forecasts. *The Accounting Review*, 80(4): 1233-1260.
- Tinic S (1988) Anatomy of initial public offerings of common stock. *Journal of Finance*, 43(4): 789-822.

Appendix A: Variable Definitions

Acq/Assets	acquisition expenditures scaled by total assets
Cash/Assets	cash and short-term investments (CHE) scaled by total assets
Capex/Assets	capital expenditures scaled by total assets
CFO/Assets	cash flows from operating activities (OANCF in Compustat), scaled by total assets
Debt/Assets	long-term debt plus current debt, scaled by total assets
Excess Return	market adjusted stock returns within a firm-year
Fines/Assets	government fines scaled by total assets, available between 2000 and 2019
GAI	litigation risk measure generated by generative AI (ChatGPT)
Lawsuit	dummy variable equal to one if a firm experiences a new lawsuit in a year, and zero otherwise
Legal Risk	the number of litigious words with negative connotation in a firm's 10-K filing scaled by the 10-K file size (in kilobytes)
Log(Assets)	the natural log of total (book) assets
NonSecurities	a dummy variable equal to one if a firm has a new lawsuit that is neither securities lawsuit nor class action lawsuit in a year and zero otherwise
RD/Assets	research and development expenses scaled by total assets
Sales Growth	current year revenues less previous year revenues all over the previous year's revenues
Securities	a dummy variable equal to one if a firm has a new securities or class action lawsuit in a year and zero otherwise
Settl/Assets	settlement costs scaled by total assets. Settlement data from Compustat.
Skewness	skewness of a firm's stock returns within a year
Tobin's q	the market value of equity (shares outstanding multiplied by share price) plus the book value of debt, scaled by the book value of assets
Turnover	a firm's share trading volume within a year scaled by shares outstanding, then divided by 1000
Volatility	standard deviation of a firm's stock returns within a year

Appendix B. Examples of Negative Litigious Words in 10-Ks

This table illustrates examples of negative litigious words appearing in SEC 10-K annual reports. The first column shows the company name, fiscal year, Central Index Key (CIK), and business fields. The second column reports sentences including negative litigious words (in bold) in the relevant 10-K filings.

Company	Examples
<i>Eli Lilly and Company</i> 2007, 0000059478, Healthcare	We believe these claims are without legal merit and expect to prevail in this litigation ; however, it is not possible to determine the outcome. An unfavorable final outcome could have a material adverse impact on our consolidated results of operations, liquidity, and financial position.
<i>HCA INC.</i> 2002, 0000860730, Hospitals	While HCA is currently not aware of any material new investigations of the Company, it is possible that governmental entities could initiate investigations or litigation in the future at facilities operated by HCA and that such matters could result in significant penalties as well as adverse publicity. It is also possible that HCA's executives and managers could be included in governmental investigations or litigation or named as defendants in private litigation .
<i>NL INDUSTRIES, INC.</i> 2015, 0000072162, Chemicals	We and others have been named as defendants in various legal proceedings seeking damages for personal injury, property damage and governmental expenditures allegedly caused by the use of lead-based paints.... The plaintiffs in these actions generally seek to impose on the defendants responsibility for lead paint abatement and health concerns As with all legal proceedings, the outcome is uncertain. Any liability we might incur in the future could be material.
<i>PHILIP MORRIS INTERNATIONAL INC.</i> 2014, 0001413329, Tobacco	Investigations include allegations of contraband shipments of cigarettes, allegations of unlawful pricing activities within certain markets, allegations of underpayment of customs duties and/or excise taxes, allegations of false and misleading usage of descriptors and allegations of unlawful advertising.
<i>VERISK ANALYTICS, INC.</i> 2016, 0001442145, Data analytics	We are subject to antitrust , consumer protection and other litigation , and may in the future become further subject to such litigation ; an adverse outcome in such litigation could have a material adverse effect on our financial condition, revenues and profitability.

Figure 1: Time Series of Legal Services Expenses and GDP in the United States

This figure plots the time series of US GDP (left y-axis; dashed line) and legal services expenditures (right y-axis; solid line) from 2000 to 2022. Unit is in billions of US dollars. Legal service expenses include expenditures such as lawyer expenses and paralegals, but do not include costs related to legal settlements. Data is from the Bureau of Economic Analysis (BEA).

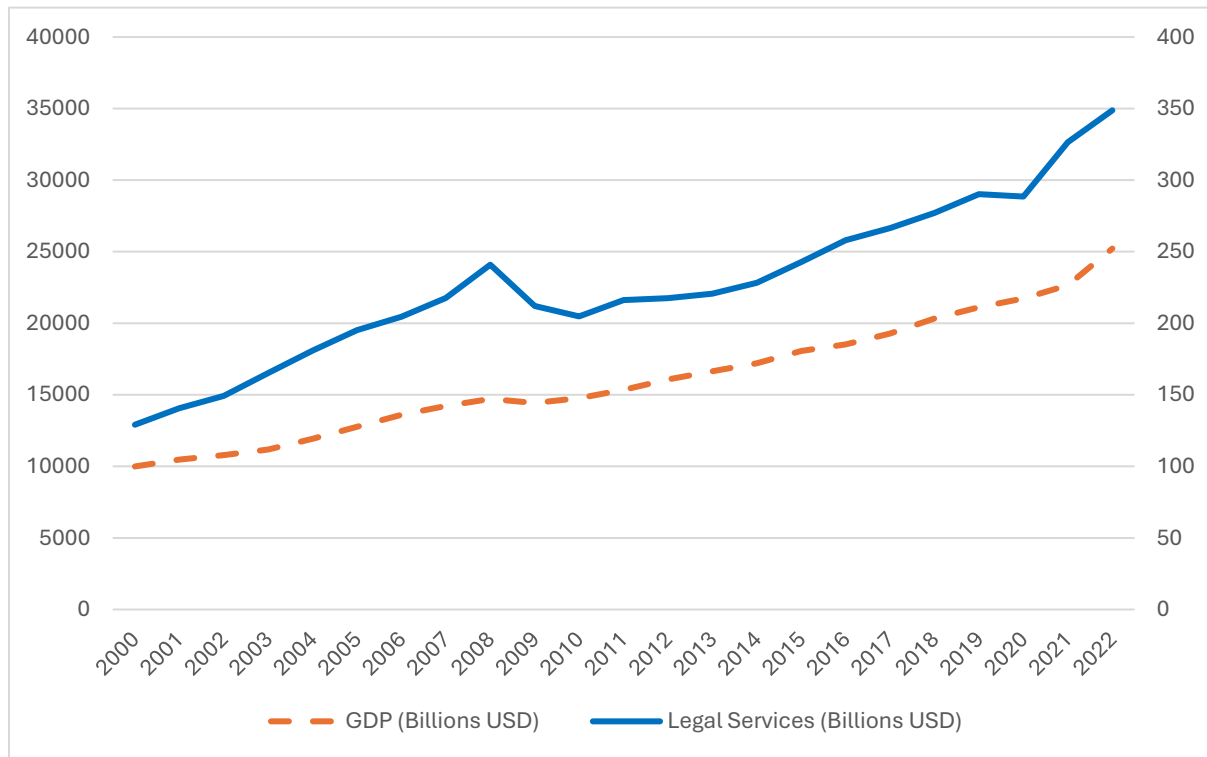


Figure 2. Distribution of Lawsuit Types

This figure illustrates the proportion of firms with at least one lawsuit of a given type in a year. It also shows the proportion of firms with at least one securities lawsuit, and respectively, non-securities lawsuit (and its five specific types) in a year. The data is from the Audit Analytics litigation database. The sample period is 2000-2022.

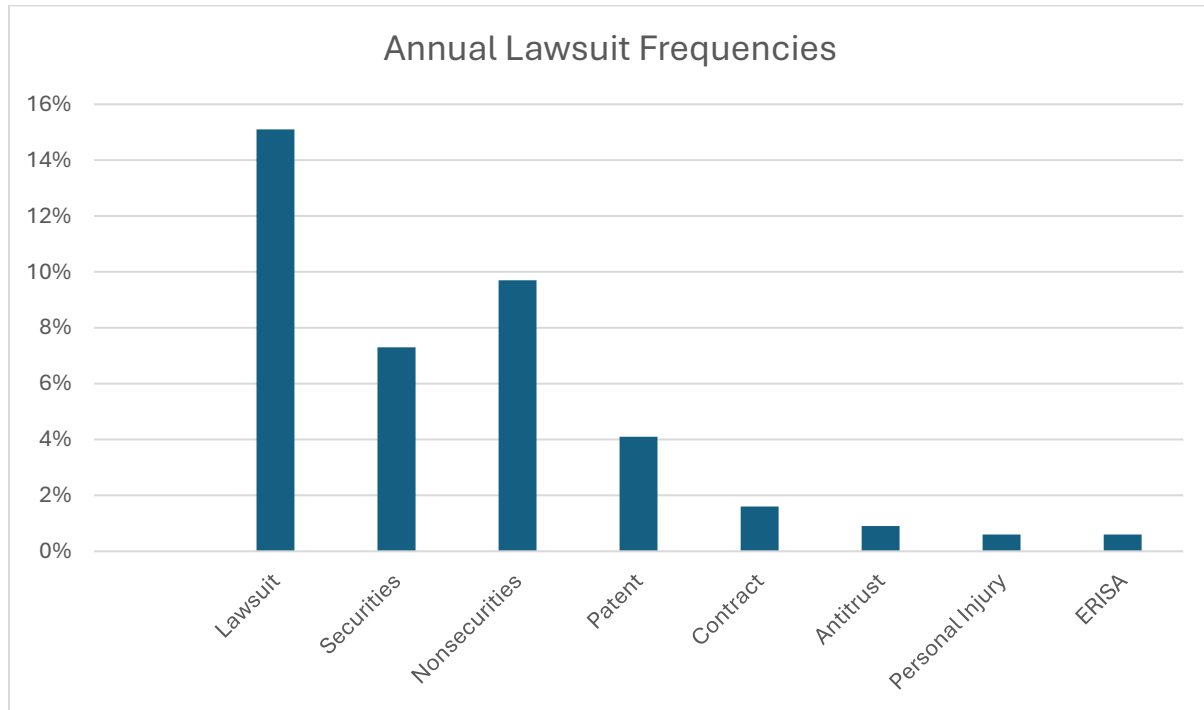


Figure 3. Distribution of Negative Legal Words Across Sections of 10-K Filings

This figure illustrates the distribution of negative litigious words across major sections of firms' 10-K filings in our sample.

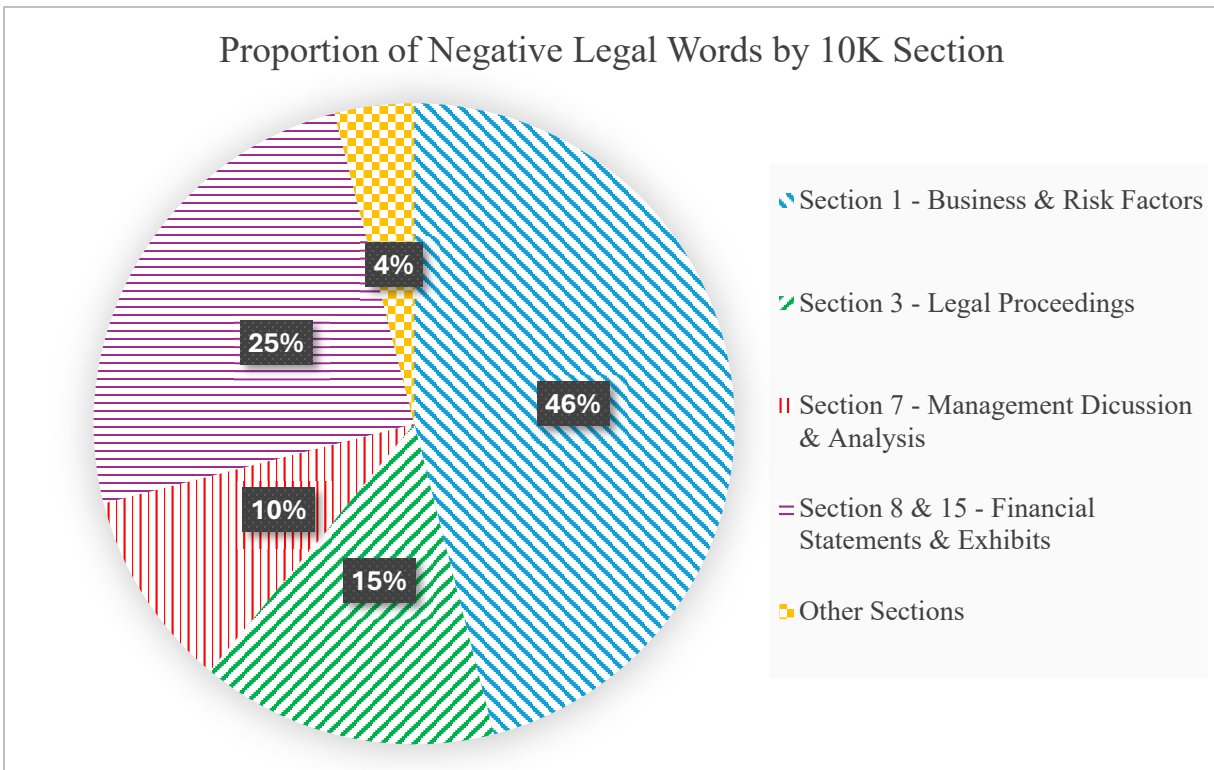


Figure 4. Example Legal Words: Firms with Many Negative Legal Words vs. Firms with Many Non-Negative Legal Words

Panel A. Legal words for firms with most negative legal words

Panel B. Legal words for firms with most non-negative legal words

Figure 5. Means and Standard Deviations of Our Litigation Risk Measure Across Industries

This figure reports means and standard deviations of our firm-level litigation risk measure (Legal Risk) across Fama-French 12 industries. The blue solid (orange striped) bars are for the industry-level means (standard deviations). Sample period is 2000-2022.

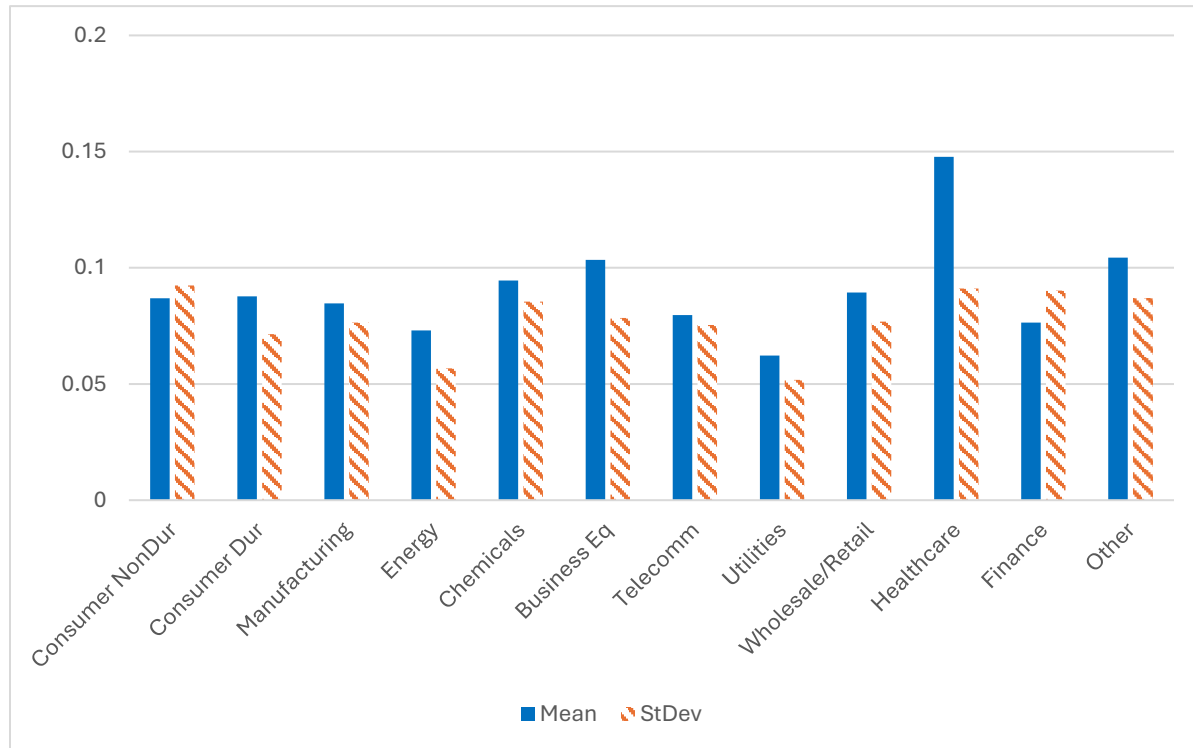


Table 1: Summary Statistics

This table presents summary statistics of the main variables used in our analysis. The sample consists of firms in both Compustat and CRSP with SEC filings available during the main sample period 2000-2022. Variables (except dummies) are winsorized at the 1st and 99th percentile values. Variable definitions are in Appendix A.

Variable	mean	sd	p25	p50	p75	N
Lawsuit	0.151	0.358	0	0	0	77,543
Legal Risk	0.097	0.086	0.035	0.072	0.132	77,543
GAI	0.802	0.093	0.750	0.800	0.850	77,402
Log(Assets)	6.509	2.068	5.025	6.520	7.912	77,543
Sales Growth	0.172	0.692	-0.036	0.068	0.204	77,543
Tobin's q	2.080	2.874	1.056	1.406	2.175	77,462
Cash/Assets	0.195	0.226	0.032	0.100	0.276	77,540
Debt/Assets	0.245	0.258	0.037	0.188	0.376	77,542
CFO/Assets	0.024	0.202	0.008	0.062	0.117	76,136
Capex/Assets	0.040	0.056	0.007	0.022	0.049	77,542
RD/Assets	0.052	0.120	0	0	0.047	77,542
Acq/Assets	0.020	0.053	0	0	0.006	77,542
Settl/Assets	0.001	0.003	0	0	0	77,542
Fines/Assets	0.011	0.066	0	0	0	67,827
Securities	0.073	0.261	0	0	0	77,543
NonSecurities	0.097	0.296	0	0	0	77,543
Excess Return	0.004	0.044	-0.017	0.002	0.023	77,543
Skewness	0.235	0.737	-0.250	0.194	0.677	77,543
Volatility	0.136	0.092	0.073	0.110	0.169	77,543
Turnover	0.185	0.241	0.059	0.124	0.219	77,543

Table 2. Future Lawsuits: Total, Negative, and Non-Negative Legal Words

This table reports the predictive power of candidate litigation risk measures for future lawsuits. The specification is $\text{Lawsuit}_{i,t+1} = \beta_0 + \beta_1 \cdot \text{Candidate}_{i,t} + X_{i,t} \cdot \Gamma + \mu_i + \gamma_k \times v_t + \varepsilon_{i,t+1}$, where $\text{Lawsuit}_{i,t+1}$ is an indicator variable equal to one if a new lawsuit is filed against firm i in year $t+1$, *Candidate* stands for a candidate litigation risk measure. *All Legal* (*Neg Legal*, *Non-Neg Legal*) is the number of all litigious words (negative litigious words, non-negative litigious words) scaled by 10-K file size. A linear probability model is applied. Standard errors are clustered at the firm level. The sample period is 2000-2022. Variable definitions are in Appendix A. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

VARIABLES	(1) Lawsuit(t+1)	(2) Lawsuit(t+1)	(3) Lawsuit(t+1)	(4) Lawsuit(t+1)
All Legal	0.006 [0.90]			
Neg Legal (Legal Risk)		0.138*** [4.19]		0.218*** [5.43]
Non-Neg Legal			-0.005 [-0.58]	-0.043*** [-4.23]
Log(Assets)	0.047*** [14.03]	0.047*** [14.07]	0.047*** [14.04]	0.047*** [14.12]
Sales Growth	0.003 [1.54]	0.004* [1.66]	0.003 [1.51]	0.004 [1.64]
Excess Return	-0.142*** [-4.09]	-0.140*** [-4.03]	-0.142*** [-4.10]	-0.139*** [-4.00]
Skewness	0.000 [0.12]	0.000 [0.15]	0.000 [0.12]	0.000 [0.17]
Volatility	0.116*** [4.59]	0.114*** [4.51]	0.116*** [4.61]	0.114*** [4.52]
Turnover	0.044*** [4.99]	0.043*** [4.90]	0.044*** [4.99]	0.042*** [4.82]
Lawsuit(t)	-0.008 [-1.40]	-0.010* [-1.82]	-0.007 [-1.36]	-0.011** [-2.01]
Observations	74,242	74,242	74,242	74,242
R-squared	0.300	0.300	0.300	0.301
Firm FE	Y	Y	Y	Y
Circuit-Year FE	Y	Y	Y	Y

Table 3: Lawsuit Likelihood and the Measure of Litigation Risk

This table presents the distributions of lawsuit likelihood in year $t+1$ across quintiles of *Legal Risk* in year t , as well as the tests for the significance of differences between the top and bottom quintiles. Panel A reports lawsuit likelihood for all lawsuits in the entire sample. Panel B reports lawsuit likelihood for FPS and non-FPS industries separately. FPS industries include biotech (SIC 2833–36, 8731–34), computer (SIC 3570–77, 7370–74), electronics (SIC 3670–74), and retails (SIC 5200–5961), and Non-FPS industries include all other industries. In Panel C, we first estimate the predicted probability of a lawsuit using the Kim-Skinner (2012) model (Table 7, specification 3 in KS), and then splits the sample based on the median predicted lawsuit probability. Panel D splits the sample based on the median of the liberal court measure from Huang, Hui, and Li (2019). Variable definitions are in Appendix A.

Panel A: Lawsuit likelihood across litigation risk quintiles

Legal Risk Quintile	Lawsuit likelihood
1 (bottom)	0.078
2	0.102
3	0.138
4	0.184
5 (top)	0.229
top - bottom	0.151
p value (top - bottom)	0.000

Panel B: Lawsuit likelihood across litigation risk quintiles: FPS vs. non-FPS industries

FPS Industries		Non-FPS Industries	
Legal Risk Quintile	Lawsuit likelihood	Legal Risk Quintile	Lawsuit likelihood
1 (bottom)	0.146	1 (bottom)	0.069
2	0.135	2	0.093
3	0.156	3	0.129
4	0.201	4	0.173
5 (top)	0.253	5 (top)	0.214
top - bottom	0.107	top - bottom	0.145
p value (top - bottom)	0.000	p value (top - bottom)	0.000

Panel C: Lawsuit likelihood across litigation risk quintiles: High vs. Low Litigation Risk per Kim and Skinner (2012)

High K&S Litigation Risk		Low K&S Litigation Risk	
Legal Risk Quintile	Lawsuit likelihood	Legal Risk Quintile	Lawsuit likelihood
1 (bottom)	0.138	1 (bottom)	0.038
2	0.164	2	0.052
3	0.202	3	0.075
4	0.253	4	0.110
5 (top)	0.324	5 (top)	0.140
top - bottom	0.186	top - bottom	0.102
<i>p</i> value (top - bottom)	0.000	<i>p</i> value (top - bottom)	0.000

Panel D: Lawsuit likelihood across litigation risk quintiles: More vs. Less Liberal Courts per Huang et al. (2019)

More Liberal Courts		Less Liberal Courts	
Legal Risk Quintile	Lawsuit likelihood	Legal Risk Quintile	Lawsuit likelihood
1 (bottom)	0.071	1 (bottom)	0.090
2	0.100	2	0.109
3	0.132	3	0.149
4	0.175	4	0.200
5 (top)	0.219	5 (top)	0.253
top - bottom	0.148	top - bottom	0.163
<i>p</i> value (top - bottom)	0.000	<i>p</i> value (top - bottom)	0.000

Table 4: Litigation Risk and Future Lawsuits

This table reports the predictability of our litigation risk measure for future new lawsuits using a linear probability model. The specification is $\text{Lawsuit}_{i,t+1} = \beta_0 + \beta_1 \cdot \text{Legal Risk}_{i,t} + X_{i,t} \cdot \Gamma + \text{FES} + \varepsilon_{i,t+1}$. *Lawsuit* is an indicator variable for the presence of a new lawsuit filed against a firm in year t+1. *Legal Risk* is the number of litigious words with a negative connotation in a firm's 10-K filing scaled by the 10-K file size. *X* is the vector of control variables and Γ is the vector of the corresponding coefficients. *FES* stands for various fixed effects as indicated in the bottom rows. Standard errors are clustered at the firm level. The sample period is 2000-2022. Variable definitions are in Appendix A. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

VARIABLES	(1) Lawsuit(t+1)	(2) Lawsuit(t+1)	(3) Lawsuit(t+1)
Legal Risk	0.143*** [4.44]	0.156*** [4.85]	0.138*** [4.19]
Log(Assets)	0.048*** [14.56]	0.046*** [13.71]	0.047*** [14.07]
Sales Growth	0.003 [1.43]	0.004* [1.85]	0.004* [1.66]
Excess Return	-0.146*** [-4.30]	-0.147*** [-4.14]	-0.140*** [-4.03]
Skewness	0.000 [0.24]	-0.000 [-0.25]	0.000 [0.15]
Volatility	0.109*** [4.41]	0.098*** [3.82]	0.114*** [4.51]
Turnover	0.044*** [5.20]	0.048*** [5.62]	0.043*** [4.90]
Lawsuit(t)	-0.009* [-1.72]	-0.017*** [-3.13]	-0.010* [-1.82]
Observations	76,053	75,970	74,242
R-squared	0.298	0.323	0.300
Firm FE	Y	Y	Y
Year FE	Y	N	N
Industry-Year FE	N	Y	N
Circuit-Year FE	N	N	Y

Table 5. Litigation Risk and Future Lawsuits by Type: Securities vs. Non-Securities

This table reports the predictability of our litigation risk measure for future new securities and non-securities lawsuits using a linear probability model. The specification is $Y_{i,t+1} = \beta_0 + \beta_1 \cdot \text{Legal Risk}_{i,t} + X_{i,t} \cdot \Gamma + \text{FES} + \varepsilon_{i,t+1}$. The dependent variable *Securities* is an indicator variable for the presence of a new securities or class action lawsuit filed against a firm in year $t+1$. *NonSecurities* is an indicator variable for the presence of a new lawsuit not covered by the indicator *Securities*. *Legal Risk* is the number of litigious words with a negative connotation in a firm's 10-K filing scaled by the 10-K file size. The sample period is 2000-2022. Standard errors are clustered at the firm level. Fixed effects are indicated in the bottom rows. Variable definitions are in Appendix A. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

VARIABLES	(1) Securities(t+1)	(2) NonSecurities(t+1)
Legal Risk	0.042* [1.68]	0.156*** [5.57]
Log(Assets)	0.037*** [15.13]	0.021*** [7.52]
Sales Growth	0.006*** [3.08]	-0.001 [-0.59]
Excess Return	-0.119*** [-4.32]	-0.070** [-2.56]
Skewness	0.000 [0.03]	0.001 [0.88]
Volatility	0.094*** [4.98]	0.050** [2.47]
Turnover	0.034*** [4.94]	0.018** [2.56]
Lawsuit(t)	-0.015*** [-3.84]	0.015*** [3.09]
Observations	74,242	74,242
R-squared	0.217	0.305
Firm FE	Y	Y
Circuit-Year FE	Y	Y

Table 6. Comparison to a Litigation Risk Measure Constructed Using Generative AI

This table reports the predictability for future new lawsuits of a generative AI-based litigation risk measure and the word-count based measure. The specification is $Lawsuit_{i,t+1} = \beta_0 + \beta_1 \cdot LR\ Measure_{i,t} + X_{i,t} \cdot \Gamma + \mu_i + \gamma_k \times v_t + \varepsilon_{i,t+1}$, where *Lawsuit* is an indicator variable for the presence of a new lawsuit filed against a firm in year $t+1$, *LR Measure* is either the word-count based measure *Legal Risk* or the generative AI-based measure *GAI*, X is the vector of control variables, Γ is the vector of the corresponding coefficients, μ_i is for firm fixed effects, $\gamma_k \times v_t$ is circuit court-by-year fixed effects. A linear probability model is used. Panel A shows the tests using the original litigation risk measures and Panel B shows the results using standardized litigation risk measures. Standard errors are clustered at the firm level. Variable definitions are in Appendix A. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Panel A. Comparison with a Generative AI-based Litigation Risk Measure

VARIABLES	(1) Lawsuit(t+1)	(2) Lawsuit(t+1)	(3) Lawsuit(t+1)
Legal Risk	0.138*** [4.20]		0.132*** [4.03]
GAI		0.052*** [3.76]	0.046*** [3.30]
Log(Assets)	0.047*** [14.02]	0.047*** [13.96]	0.047*** [13.99]
Sales Growth	0.004* [1.66]	0.003 [1.56]	0.004* [1.68]
Excess Return	-0.141*** [-4.04]	-0.139*** [-4.01]	-0.138*** [-3.95]
Skewness	0.000 [0.14]	0.000 [0.13]	0.000 [0.15]
Volatility	0.114*** [4.51]	0.114*** [4.53]	0.112*** [4.45]
Turnover	0.043*** [4.88]	0.043*** [4.96]	0.042*** [4.87]
Lawsuit(t)	-0.010* [-1.83]	-0.008 [-1.45]	-0.010* [-1.87]
Observations	74,101	74,101	74,101
R-squared	0.300	0.300	0.300
Firm FE	Y	Y	Y
Circuit-Year FE	Y	Y	Y

Panel B. Standardized litigation risk measures

VARIABLES	(1) Lawsuit(t+1)	(2) Lawsuit(t+1)	(3) Lawsuit(t+1)
Legal Risk (Standardized)	0.012*** [4.20]		0.011*** [4.03]
GAI (Standardized)		0.005*** [3.76]	0.004*** [3.30]
Observations	74,101	74,101	74,101
R-squared	0.300	0.300	0.300
p-value (diff coefs)			0.03
Controls	Y	Y	Y
Firm FE	Y	Y	Y
Circuit-Year FE	Y	Y	Y

Table 7. Settlements, Government Fines, Cash Holdings, and Litigation Risk

This table reports the predictability of our litigation risk measure for future legal settlements, government fines and cash holdings. The specification is $Y_{i,t+1} = \beta_0 + \beta_1 \cdot \text{Legal Risk}_{i,t} + X_{i,t} \cdot \Gamma + \text{FES} + \varepsilon_{i,t+1}$. *Legal Risk* is the number of litigious words with a negative connotation in a firm's 10-K filing scaled by the 10-K file size. X is the vector of control variables and Γ is the vector of the corresponding coefficients. *FES* stands for various fixed effects as indicated in the bottom rows. Standard errors are clustered at the firm level. The sample period is 2000-2022, except for columns 3 and 4 which end in 2019 due to fines data availability. Variable definitions are in Appendix A. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

VARIABLES	(1) Settl/Assets	(2) Settl/Assets	(3) Fines/Assets	(4) Fines/Assets	(5) Cash/Assets	(6) Cash/Assets
Legal Risk	0.008*** [20.01]	0.007*** [14.43]	0.037*** [5.24]	0.042*** [5.38]	0.247*** [11.54]	0.045*** [3.20]
Log(Assets)	-0.000 [-1.40]	0.000 [0.48]	0.006*** [13.47]	-0.001 [-0.75]	-0.014*** [-13.73]	-0.030*** [-13.38]
Sales Growth	-0.000*** [-3.15]	-0.000** [-2.15]	-0.001*** [-2.91]	-0.000 [-1.20]	0.009*** [6.03]	-0.003*** [-2.92]
Tobin's q	0.000 [0.46]	-0.000*** [-2.72]	0.001*** [5.55]	-0.000 [-0.36]	0.021*** [17.34]	0.006*** [6.69]
Debt/Assets	0.000 [1.57]	0.000 [0.91]	-0.007*** [-3.30]	-0.001 [-0.96]	-0.200*** [-22.40]	-0.077*** [-11.40]
CFO/Assets	0.000 [1.20]	0.000 [0.52]	-0.005*** [-3.11]	0.003 [1.50]	-0.203*** [-19.79]	0.014 [1.64]
Volatility	0.001*** [3.44]	0.000 [0.67]	-0.008** [-2.42]	0.003 [0.82]	0.160*** [11.00]	0.057*** [5.82]
Cash/Assets	-0.000** [-2.25]	-0.000 [-0.84]	-0.006*** [-3.59]	-0.003 [-1.24]		
Observations	63,977	62,930	64,981	63,845	68,859	67,504
R-squared	0.044	0.217	0.049	0.228	0.474	0.831
Industry FE	Y	N	Y	N	Y	N
Firm FE	N	Y	N	Y	N	Y
Circuit-Year FE	Y	Y	Y	Y	Y	Y

Table 8. Capital Expenditure, R&D, Acquisitions, and Litigation Risk

This table reports the predictability of our litigation risk measure for future firm-level investments (capital expenditures, research and development expenses, and acquisition-related expenditures). The specification is $Y_{i,t+1} = \beta_0 + \beta_1 \cdot \text{Legal Risk}_{i,t} + X_{i,t} \cdot \Gamma + \text{FES} + \varepsilon_{i,t+1}$. *Legal Risk* is the number of litigious words with a negative connotation in a firm's 10-K filing scaled by the 10-K file size. X is the vector of control variables and Γ is the vector of the corresponding coefficients. *FES* stands for various fixed effects as indicated in the bottom rows. Standard errors are clustered at the firm level. The sample period is 2000-2022. Variable definitions are in Appendix A. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

VARIABLES	(1) Capex/Assets	(2) Capex/Assets	(3) RD/Assets	(4) RD/Assets	(5) Acq/Assets	(6) Acq/Assets
Legal Risk	-0.011*** [-2.92]	-0.010*** [-3.44]	-0.018** [-2.24]	-0.019*** [-3.32]	-0.007** [-2.00]	-0.014*** [-3.22]
Log(Assets)	-0.001*** [-2.90]	-0.001 [-1.20]	-0.000 [-0.08]	-0.008*** [-8.14]	0.000** [2.18]	-0.005*** [-8.01]
Sales Growth	0.003*** [8.67]	0.002*** [7.09]	0.001 [1.53]	-0.001 [-0.96]	0.001*** [5.04]	0.000 [0.11]
Tobin's q	0.002*** [8.44]	0.002*** [8.20]	0.004*** [5.19]	-0.000 [-0.12]	0.001*** [4.36]	0.001*** [6.18]
Debt/Assets	-0.003** [-1.98]	-0.017*** [-10.27]	-0.010*** [-2.81]	-0.010*** [-2.67]	-0.001 [-1.06]	-0.017*** [-9.77]
CFO/Assets	0.027*** [14.79]	0.014*** [8.39]	-0.226*** [-32.84]	-0.091*** [-13.51]	0.025*** [17.06]	0.020*** [10.06]
Volatility	0.006* [1.88]	-0.013*** [-5.57]	0.021*** [2.89]	-0.023*** [-4.04]	-0.032*** [-11.29]	-0.033*** [-10.79]
Cash/Assets	-0.014*** [-9.03]	0.004** [2.45]	0.154*** [29.30]	0.023*** [4.45]	0.002 [1.31]	0.052*** [18.20]
Observations	70,141	68,812	70,141	68,812	70,141	68,812
R-squared	0.398	0.718	0.570	0.835	0.061	0.254
Industry FE	Y	N	Y	N	Y	N
Firm FE	N	Y	N	Y	N	Y
Circuit-Year FE	Y	Y	Y	Y	Y	Y

Table 9. Robustness Tests: Litigation Risk and Corporate Policies

This table presents robustness tests for the firm-level outcome tests in Tables 7 and 8. The specification is $Y_{i,t+1} = \beta_0 + \beta_1 \cdot \text{Legal Risk}_{i,t} + X_{i,t} \cdot \Gamma + \mu_i + \gamma_k \times v_t + \varepsilon_{i,t+1}$. *Legal Risk* is the number of litigious words with a negative connotation in a firm's 10-K filing scaled by the 10-K file size. X is the vector of control variables and Γ is the vector of the corresponding coefficients. μ_i stands for firm fixed effects and $\gamma_k \times v_t$ stands for circuit court-by-year fixed effects. Panel A re-runs the even specifications from Tables 7 and 8 and adds the KS control variables from Table 2 (denoted as KS Controls). Panel B decomposes our legal risk measure into current and future components where current is the predicted value when regressing legal risk on a contemporaneous lawsuit dummy variable. We then use the predicted value (*Legal Risk (Current)*) as the legal risk related to current lawsuits and the residual value (*Legal Risk (Future)*) as the legal risk associated with non-current lawsuits. Standard errors are clustered at the firm level. The sample period is 2000-2022, except for specification 2 which ends in 2019 due to data availability. Variable definitions are in Appendix A. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Panel A. Including KS model factors

VARIABLES	(1) Settl/Assets	(2) Fines/Assets	(3) Cash/Assets	(4) Capex/Assets	(5) RD/Assets	(6) Acq/Assets
Legal Risk	0.007*** [14.41]	0.042*** [5.36]	0.042*** [3.04]	-0.009*** [-3.35]	-0.019*** [-3.21]	-0.013*** [-3.09]
Observations	62,930	63,845	67,504	68,812	68,812	68,812
R-squared	0.217	0.228	0.831	0.719	0.836	0.256
KS Controls	Y	Y	Y	Y	Y	Y
Other Controls	Y	Y	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y	Y	Y
Circuit-Year FE	Y	Y	Y	Y	Y	Y

Panel B. Legal Risk decomposition: Current vs. future legal risk

VARIABLES	(1) Settl/Assets	(2) Fines/Assets	(3) Cash/Assets	(4) Capex/Assets	(5) RD/Assets	(6) Acq/Assets
Legal Risk (Current)	0.011*** [6.40]	0.060*** [2.87]	0.058 [1.64]	-0.006 [-0.89]	-0.000 [-0.00]	-0.002 [-0.15]
Legal Risk (Future)	0.007*** [13.71]	0.040*** [5.04]	0.043*** [3.08]	-0.010*** [-3.48]	-0.022*** [-3.63]	-0.015*** [-3.43]
Observations	62,930	63,845	67,504	68,812	68,812	68,812
R-squared	0.217	0.228	0.831	0.718	0.835	0.254
Controls	Y	Y	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y	Y	Y
Circuit-Year FE	Y	Y	Y	Y	Y	Y