Corporate Annual Filings and Default Risk: Does Readability Matter?

Abstract

We examine the association between the readability of corporate annual filings (10-K reports) and the default risk of publicly listed firms in the United States. While greater readability may reduce default risk by lowering information risk, it can also heighten default risk by exposing firms' financial vulnerabilities to credit investors. Using a large sample of 69,768 firm-year observations from 1994 to 2020, we find a statistically significant positive association between poor readability in 10-K reports and default risk. Less readable disclosures hinder investors' information processing, thereby increasing information risk and, consequently, default risk. Further analysis indicates that the degree of association between readability and default risk varies with firm-specific characteristics such as business strategy, financial performance, risk profile, monitoring intensity, and managerial ability. Channel analysis identifies information asymmetry, agency costs, risk-taking behaviour, cost of capital, and financial constraints as key mechanisms linking readability to default risk. Additionally, machine learning algorithms confirm the importance of readability in predicting default risk.

Keywords: default risk; disclosures; distance-to-default; information risk; readability

1. Introduction

Corporate annual reports serve as essential conduits for listed companies to communicate key information to investors, partners, and the public. In the United States, the Securities Exchange Act of 1934 mandates disclosure of comprehensive information through Form 10-K filings, which constitute crucial resources for capital market participants. The US Securities and Exchange Commission (SEC) emphasizes the value of 10-K reports for potential investors, stating: *'…if you want to follow or invest in a U.S. public company, you can find a wealth of information in the company's annual report on Form 10-K. Among other things, the 10-K offers a detailed picture of a company's business, the risks it faces, and the operating and financial results for the fiscal year'* (SEC Office of Investor Education and Advocacy, 2021). Despite the critical role of annual reports in conveying material information, the aspect of readability often remains under-addressed in accounting practices (Besuglov & Crasselt, 2021). Less readable information in these reports can create information asymmetry by making it challenging for investors to process the information, thus contributing to the information risk of a company. Recognizing the significance of 10-K report readability for the information climate and its perceived importance for the credit market, this study investigates how the readability of 10-K reports enhances corporate default risk.

Corporate default risk commonly refers to the probability of the debtor's failure to repay the required amount of debt to lenders (Duan et al., 2012). Default risk is detrimental to both debtors and creditors as it creates uncertainty in the financial market and causes the loss of valuable resources for the related parties. More precisely, the higher default risk of debtholders imposes stringency on future borrowings and requires a higher cost of debt, leading to limited access to capital and reduced credit score, and creditors may pursue legal action to recover debt, potentially leading to wage garnishment, asset seizure, or bankruptcy, putting them in a disadvantageous position. Therefore, default risk carries significant implications for market stability and resource allocation. Reflecting such importance of default risk, it has long been a topic of scholarly interest, especially during the post-great Financial Crisis period (Contessi et al., 2014; Dieckmann & Plank, 2012; Switzer et al., 2018). The COVID-19 pandemic further amplified these concerns, as evidenced by both academic and practitioner research (Choi et al., 2020; Yin et al., 2022).

Efforts to improve disclosure quality and readability have been a continuous endeavor. Notably, the 1995 Report of the Task Force on Disclosure Simplification suggested recommendations for improving disclosure quality and readability. Building upon these efforts, the 1998 Plain English Rule 421(d) mandated the use of plain English in financial disclosures, aiming to improve accessibility to average investors. However, despite these regulatory initiatives aimed at improving annual report readability, recent studies indicate that issues with the use of complex or vague language persist (Ataullah et al., 2018; Lehavy et al., 2011; Li, 2008; Lo et al., 2017; Nadeem, 2022), often attributed to management's self-interest (Nadeem, 2022). This ongoing concern highlights the importance of readability in financial reporting and its potential impact on investors and the broader credit market.

Annual report readability reflects the quality of communication between firm management and external stakeholders. Effective communication is essential for a firm's proper valuation, incorporating the assessment of its existing and potential risks. However, less readability acts as a barrier to effective communication (Loughran & McDonald, 2014), leading to adverse consequences in the capital market. For instance, less readability contributes to an asymmetric information environment and creates a barrier to the proper credit risk evaluation (Liao et al., 2009). Evidence indicates that management may intentionally make the annual report complicated or less understandable to obfuscate adverse information (Li, 2008). The complex nature of annual reports renders the information costly to analyze for investors and reduces the likelihood of it being completely revealed to the market in a timely manner (Bloomfield, 2002). Therefore, lower readability creates uncertainty for market participants in accurately assessing a firm's risk exposure and increases information processing costs to credit investors (Courtis, 2004; Fisher et al., 2019). To compensate for this uncertainty, credit investors demand higher risk premiums with stringent credit terms (Bonsall & Miller, 2017; Ertugrul et al., 2017; Rjiba et al., 2021). Consequently, this increases borrowing costs, reduces financial flexibility, and makes firms more vulnerable to financial distress and default risk (Baghdadi et al., 2020; Dalwai et al., 2021).

However, there is a contrasting view that higher readability may also contribute to higher default risk. Li (2008) found that firms may intentionally obfuscate unfavorable information by making their communication more complex. When financial reports are more readable, investors and analysts can more easily identify a firm's financial weaknesses, inconsistencies, or potential risks. Existing evidence suggests that increased transparency may expose vulnerabilities, such as risks, that were previously overlooked or difficult to detect in complex, less readable disclosures or the absence of disclosures (Kravet & Muslu, 2013). As a result, credit investors may reassess their risk exposure and adjust their lending terms accordingly. If a firm is exposed to higher risk due to factors, investors may demand higher risk premiums to compensate for the perceived risk. Consequently, the firm may face increased borrowing costs, shorter credit periods, and limited access to external financing. These financial constraints can reduce the firm's liquidity and financial flexibility, making it more vulnerable to financial distress and ultimately increasing its default risk.

We primarily measure the readability using the Bog Index (Bonsall et al., 2017), which builds upon the limitations identified in the previously widely used Fog Index for analyzing the readability of business text from financial documents (Biddle et al., 2009; Lawrence, 2013; Li, 2008). Recent literature suggests that the BOG index is the appropriate metric for evaluating the readability of corporate narrative disclosures (Hasan, 2020; Rjiba et al., 2021). The Bog index incorporates characteristics of plain English writing, such as the use of passive and hidden verbs, as well as complex, abstract, and legal terminology (Bonsall et al., 2017). This alignment with the SEC's plain English writing guidelines makes the Bog Index a well-suited tool for capturing readability in corporate disclosures. A higher Bog Index score indicates that annual filings are more complex and less readable. Similar to prior literature (Kabir et al., 2021; Nadarajah et al., 2021), this study primarily uses Merton's (1974) distance-to-default (DD) measure to calculate a firm's default risk. DD measures how far a firm is from being a defaulter in meeting debt obligations. Higher DD values indicate a lower risk of default, while lower values suggest a higher likelihood of default.

Drawing upon a large sample of 69,768 firm-year observations spanning 1994 to 2020, our results indicate a negative association between the readability of 10-K reports and the default risk of US firms. The association is also economically meaningful: a one standard deviation increase in the BOG Index corresponds to a 0.45 point drop in DD, translating to a 10.78% decrease in DD relative to the mean of DD. Thus, our results highlight the unfavorable effects of poor readability of narrative disclosures on default risk. Furthermore, the results indicate that less readable annual reports exacerbate default risk for firms characterized by less monitoring, lower organizational capital, and lower managerial ability. The cross-sectional analyses further reveal that the association between readability and default risk varies based on the business strategies (Prospectors versus Defenders), debt concentration (Bank debt versus Public debt), economic performance, overall risk, existence of SEC 1998 Plain English Rule, and board governance characteristics.

Additionally, our study identifies information asymmetry, risk-taking strategy, cost of capital, and financial constraints as the channel mechanisms for the association between readability and default risk. To address potential endogeneity concerns, we apply a multi-pronged approach. This includes two-stage least squares (2SLS) regression with an instrumental variable (IV) approach, propensity score matching (PSM), and entropy balancing techniques. Machine learning algorithms, specifically Random Forest (RF) and Extreme Gradient Boosting (XGBoost), offer additional evidence for the importance of readability in assessing default risk. Finally, we leverage alternative measures of both readability and default risk to ensure the robustness of our findings and mitigate potential biases arising from specific proxy selections.

This research contributes to the expanding scholarly discourse on the higher level of readability of narrative disclosures and their capital market implications. Prior research views lower readability act as a catalyst for corporate information risk, leading to price adjustments by both equity investors (Lee, 2012; Rjiba et al., 2021) and credit investors (Bonsall & Miller, 2017; Chen et al., 2024; Hu et al., 2018). Our investigation extends this research by demonstrating that credit markets incorporate readability assessments into pricing models, with default risk calculations reflecting readability deficiencies. For instance, Chen et al. (2024) documented that a lower level of readability in pension narratives correlates with higher bond yield spreads. While our research similarly emphasizes the readability of narrative disclosures, it distinctively evaluates comprehensive 10-K report readability and demonstrates how better readability can facilitate the mitigation of default risk. Although our findings align with Bonsall and Miller's (2017) work, which indicates that poor readability leads to poor bond ratings and wider bond spreads, an important distinction warrants emphasis: bond ratings and spreads reflect the market perception of risk associated with a specific bond, whereas default risk pertains to the actual risk of the issuer failing to meet its debt obligations. Despite bond metrics providing substantial credit risk insights, they may not fully capture underlying raw default probabilities (Hilscher & Wilson, 2017). Consequently, employing Merton's (1974) distance-to-default (DD) as a direct market-based measure of default risk in this study demonstrates the relevance of readability to default risk.

This research further adds to the literature on the determinants of default risk. Corporate default risk is detrimental to a borrower, as it further increases the cost of borrowing, accelerates financial distress, impairs stock market performance, and tarnishes the corporate reputation. Considering the importance and pervasive effect of default risk, prior literature identifies multifaceted factors of default risk, such as stock liquidity (Brogaard et al., 2017), governance quality (Ali et al., 2018), board composition (Baghdadi et al., 2020), ESG disclosure (Atif & Ali, 2021), and carbon performance (Kabir et al., 2021). This study not only identifies the readability of narrative disclosures as a key factor in determining default risk but also takes a step further by demonstrating its association with actual firm bankruptcies. In addition, we demonstrate that firms can mitigate the adverse effects of poor readability on default risk by implementing strong monitoring mechanisms, enhancing managerial ability, and increasing organizational capital.

Finally, it is implicit in the existing literature that better economic performance and a lower level of exposure to overall risk are useful in the credit market (Atif & Ali, 2021; Chava & Purnanandam, 2010; Francis et al., 2021; Valta, 2012). Our research builds on these studies by identifying that despite having better economic performance or overall lower risk, firms may still face higher default risk due to poor readability of narrative disclosures, thereby increasing investors' information risk. A lower level of readability enhances investor risk perceptions and offsets the favorable effects of better firm performance and a lower level of risk in the credit market. Even if firms exhibit strong economic performance and low overall risk, they might still incur costs in the credit market merely due to poor readability in their narrative disclosures, highlighting the importance of effective communication. Thus, these findings also offer practical implications for firm management, emphasizing the importance of effective communication through narrative disclosures in reducing default risk.

This paper is structured as follows. Section 2 reviews relevant prior research and develops the research hypothesis. Section 3 describes our variables and empirical models. Section 4 presents the results of empirical analyses. Section 5 explores the potential channels. Section 6 examines the results of our cross-sectional analysis, while Section 7 discusses the robustness checks. Section 8 provides the concluding summary of the paper.

2. Hypothesis development

A firm's annual 10-K reports serve as a key communication channel between firms and investors, providing critical financial and non-financial information. The readability of these reports plays a critical role in shaping investor perceptions of a firm's performance and assessing firm risk. Less readable reports make it challenging for investors to evaluate a firm's financial health and underlying risks. Empirical evidence demonstrates the multifaceted impacts of report readability on firm operations and performance. Dalwai et al. (2021) document that reduced readability in annual filings negatively affects firm value, while Kim et al. (2019) establish a relationship between poor readability of annual reports and anticipated stock price crash risk, attributing this to heightened information asymmetry.

The difficulty in reading annual reports often reflects underlying agency problems (Plumlee, 2003). The management obfuscation hypothesis provides a theoretical framework for understanding reporting complexity, positing that managers tend to be more transparent during periods of strong performance (Schrand & Walther, 2000), but may deliberately complicate disclosures to obscure unfavorable information during weaker periods to shape external perceptions (Li, 2008). They make the annual filing complex and less readable, allowing them to manipulate how investors, analysts, or the public perceive their overall performance. Taking this into account, in October 1998, the SEC recommended the implementation of 'plain English' disclosure rules that promote clear and straightforward language. The rationale behind this suggestion was that companies might employ ambiguous language to conceal unfavorable information, and ordinary investors might struggle to comprehend intricate financial reports, leading to inefficiencies in the capital market (Dempsey et al., 2012). However, despite these regulatory efforts, research indicates that management may rely on complex reporting for strategic purposes, as management aims to protect self-interests and suppress critical information (Goswami et al., 2023; Hassan et al., 2019). Supporting this perspective, Goswami et al. (2023) and Hassan et al. (2019) document positive associations between report complexity and agency costs while Luo et al. (2018) find an inverse relationship between readability and agency costs in Chinese A-share listed firms. Overall, less readable annual filings are characterized by higher agency costs with enhanced information asymmetry, leading to greater information risk.

Within credit markets, narrative disclosure readability carries significant implications. Bonsall and Miller (2017) demonstrate that complex narratives raise processing costs, leading to greater disagreement among rating agencies and affecting credit ratings. This complexity translates into tangible market outcomes, including elevated borrowing costs. Hoffmann and Kleimeier

(2021) argue that firms with less readable reports face higher information risk, which is priced into the debt market (Aldamen & Duncan, 2013). As a result, investors demand stricter loan terms (Ertugrul et al., 2017), charge higher risk premiums (Hu et al., 2018), and require higher returns on debt (Bonsall & Miller, 2017), all of which compound the firm's default risk.

Nevertheless, an alternative perspective suggests that enhanced readability might paradoxically increase default risk, and we draw the perspective from the literature on the informativeness of risk disclosures. Although existing research provides evidence of the effectiveness of risk disclosures in improving information transparency and the favorable effects of such disclosures on credit risk (Chiu et al., 2018), others research highlights how such disclosures can enhance the risk perception of credit investors. For instance, Kravet and Muslu (2013) find that informative risk disclosures unveil previously unexposed risks to investors, leading to enhanced risk perceptions. Similarly, Bao and Datta (2014) demonstrated that disclosing systematic and liquidity risks can increase investors' risk perceptions. Campbell et al. (2014) further show that investors use risk factor disclosures in their investment decisions, suggesting that higher readability can reveal financial vulnerabilities that may have remained unnoticed in less readable reports.

Building on this perspective, firms that prioritize transparency may produce highly readable reports regardless of financial performance or risk exposure. In such cases, higher readability allows credit investors to assess risks and uncertainties more accurately (Hope et al., 2016). Firms with stable growth and risk coupled with greater readability are likely to boost investor confidence, enabling access to credit under favorable terms (Hu et al., 2018). However, for riskier firms, high readability may have unintended consequences. If a firm exhibits weak performance or significant exposure to risk, higher readability may amplify investor concerns by making vulnerabilities more transparent. Especially as credit investors may be more cautious about financial risks (Chiu et al., 2018), they may impose stricter credit terms, such as higher interest rates and shorter loan periods, to mitigate or offset perceived risks. These tighter conditions can, in turn, restrict financial flexibility and increase default risk.

In addition, if a firm's credit financing is primarily reliant on bank loans, the association between readability and default risk may become more ambiguous. Banks, as sophisticated investors, have access to private information beyond publicly available data in annual reports and play a monitoring role (Ben-Nasr et al., 2021; Liao, 2015). Through direct lending relationships, due diligence processes, and ongoing monitoring, banks obtain non-public insights into a firm's financial health. Unlike other credit investors who rely heavily on narrative disclosures, banks can supplement this information with internal financial records, loan covenants, and management discussions, making them less dependent on disclosure readability when assessing credit risk. Additionally, banks' ability to process information gathered through both private and public channels is superior to that of other credit investors (Chen, 2016; Diamond, 1991). As a result, banks' exposure to information risk and their approach to pricing borrowing firms' credit risk differ from those of other credit investors (Chen, 2016). Therefore, if a firm relies more on bank borrowing than other sources of financing, the readability of narrative disclosures is likely to have a limited effect on default risk (Denis & Mihov, 2003).

In summary, the theoretical tension between the alternative perspectives informs our hypothesis development. Less readable disclosures can heighten information risk, compelling credit investors to demand higher risk premiums and impose stricter credit terms, ultimately increasing default risk. Alternatively, greater readability may reveal previously unknown risks and uncertainties, especially for financially vulnerable firms, potentially resulting in unfavorable credit adjustments and a higher likelihood of default. Finally, the extent to which readability affects default risk might depend on borrowers' reliance on bank loans versus non-bank loans. Therefore, based on these alternative perspectives, we formulate the following non-directional hypothesis:

Hypothesis: *Readability of narrative disclosures in 10-K reports is associated with corporate default risk.*

3. Data, variables, and summary statistics

3.1 Data

Our sample begins with the Bog index data from Bonsall et al. (2017) study, as reported on Miller's website from 1994 to 2020. We began our sample in 1994, as the Bog index score is available from this year, the initial sample yields 189,665 firm-year observations. We subsequently collect default risk data from the Credit Research Initiative (CRI) database of the Risk Management Institute (RMI), National University of Singapore (NUS), merging it to obtain matched 93,405 firm-year observations. Further data cleaning for missing control variables and excluding financial and utility firms led to the exclusion of 23,627 observations, leaving a final sample of 69,768 firm-year observations.[1] The sample selection process is reported in Panel A of Table 1. To avoid the undesirable impact of outliers, all variables are winsorized at the extreme 1% of the distributions.

[Insert Table 1 Here]

[1] The number of observations varies in cross-sectional and additional analyses based on the data availability across the databases.

3.2 Dependent variable

This study operationalizes default risk through Merton's distance-to-default (DD) metric as the primary dependent variable, consistent with existing empirical literature. The DD framework incorporates market-based information derived from Robert C. Merton's (1974) theoretical model and Black-Scholes theory. The formula is:

$$DD_t = \frac{ln\left(\frac{V_{A,t}}{X_t}\right) + \left(\mu - \frac{1}{2}\sigma_A^2\right)(T-t)}{\sigma_A \sqrt{T-t}} \tag{1}$$

here in Equation 1, *DD* measures a firm's distance from its default threshold, with a lower value of *DD* indicating a higher probability of default and *vice versa*. $V_{A,t}$ indicates an asset's market value with the assumption on geometric Brownian motion; X_t indicates a firm's default point value; σ_A measures the volatility, $\sqrt{T} - t$ is set to 1 year; μ indicates the average return on the asset.

The original *DD* measure has challenges, such as difficulties in accurately measuring μ and determining the market value of assets. Duan et al. (2012) propose a refined model:

$$\mathrm{DD} = \frac{ln\left(\frac{V_{A,t}}{X_t}\right)}{\sigma_A \sqrt{T-t}} \tag{2}$$

This improved model provides the default point as X_t = current liability+0.5×Long-term Liability + δ × other liability, where δ is calibrated according to sector-specific parameters. Empirical validation demonstrates that Equation (2) provides higher measurement accuracy compared to the original formulation (Duan et al., 2012). For empirical implementation, the CRI database provides us with the required DD data, which employs Duan et al. (2012) method. The CRI database's reliability is well-regarded in recent empirical literature (Kabir et al., 2021; Nadarajah et al., 2021).

3.3 Independent variable

This study adopts the Bog index (*Bog Index)* for assessing the readability of annual reports, an index validated by Bonsall et al. (2017). This index measures readability by focusing on elements recommended by the SEC for transparent communication, including passive voice usage, the presence of hidden verbs, and the complexity of language. A higher Bog index indicates more complex language within annual reports, signifying potentially lower readability. This approach to assessing readability aligns with prior research (Alam et al., 2024; Blanco et al., 2021; Habib & Hasan, 2020; Hasan, 2020), which also employs the Bog index as a proxy of the readability of annual reports. The *Bog Index* data is retrieved from the website of Brian Miller's, where higher values on this index correspond to greater difficulty in reading the annual report.[2]

3.4 Control variable

In examining the association between readability and *DD*, this study uses ordinary least squares (OLS) regressions, controlling for other determinants of *DD* as identified in previous research (Chen et al., 2023; Kabir et al., 2021). These include firm *Size*, measured by the logarithm of total assets. We also control for Capital Expenditure (*CAPEX*) (capital expenditure to total assets), as firms with higher capital expenditure are anticipated to exhibit lower default risk. *Liquidity* (liquid assets to total assets) and *Leverage* (long-term debt to total assets) are included, with higher leverage anticipated to be positively related to default risk and increases in firms' external liability. Additionally, we control the market-to-book ratio *(MTB)*, measured as the market-to-book value of a share that captures the growth of the firm. Profitability is controlled by taking return on assets *(ROA),* calculated as net income divided by total assets, expecting a negative association with default risk. Finally, we control for *Tangibility*, measured as net property,

[2] [2] Access here for details: https://sites.google.com/iu.edu/professorbrianpmiller/bog-data

plant, and equipment divided by total assets, to account for firms' investment in physical assets, which is expected to be negatively associated with default risk. Control variables are sourced from the Compustat database. Detailed definitions of the variables are in Appendix A1.

3.5 Summary statistics

Panel B of Table 1 presents descriptive statistics for variables employed in our analysis; see Appendix A1 for variable definitions. The mean (median) value of *DD* is 4.17 (3.68), indicating that firms included in the sample are fairly stable. *Bog Index* has a mean (median) of 84.75 (85), with an interquartile range from 79.00 to 90.00. This average *Bog Index* score indicates that, on average, the annual reports of the firms in our sample are relatively difficult to read. Furthermore, the substantial standard deviation of 8.59 for the *Bog Index* highlights considerable variation in readability across the firms in our sample. Overall, these descriptive statistics for the *Bog Index* are consistent with those reported in Rjiba et al. (2021).

Regarding the control variables, the mean (median) of *size* is 5.77 (5.69). Capital expenditure (*CAPEX*) has a mean value of 5.3% with a standard deviation of 0.061. On average, the sample firms have *Liquidity* of 21.6%, and the mean value of *Leverage* is 65.4%, indicating the average firms in the sample have a high level of borrowing. The market-to-book ratio (*MTB*) exhibits a mean value exceeding the book value among sampled firms. *ROA* has a mean (median) value of –10.9% (0.026) with a standard deviation of 0.613. Finally, on average, 25% of assets are tangible assets for sample firms.

Panel C, Table 1 presents the Pearson correlation matrix for all variables included in our regressions. As we predict, the *Bog Index* shows a significant negative correlation with *DD*, providing preliminary support for our hypothesized relationship between annual report readability and default risk. Most of the control variables exhibit a significant correlation with the dependent variable *DD*. None of the correlations among independent variables are high enough to warrant concerns about multicollinearity.

4. Empirical analysis

4.1 Baseline results

This section investigates the relation between annual report readability and default risk by estimating the panel regression model:

$$DD_{i,t} = \alpha_0 + \beta_1 Bog\ Index_{i,t} + \sum \gamma_k\ Control_{i,t} + \delta_i + \omega_t + \varepsilon_{i,t} \quad (3)$$

where *DD* is the Distance-to-Default – an inverse measure of default risk derived using Merton (1974). The *Bog Index* is the readability index. Control denotes a vector of firm characteristics encompassing *Size*, *Capex*, *Liquidity*, *Leverage*, *MTB*, *ROA*, *Tangibility*, *industry-fixed effects* (δ_i), and *year-fixed effects* (ω_t).

Table 2 presents the results of regression analyses for the association between readability and default risk, while controlling for other potential determinants that may influence default risk. In all regression estimation techniques, we additionally control for year-fixed and industry-fixed effects unless otherwise stated. To address potential heteroscedasticity and serial correlation in the data, we employ four distinct estimation techniques following the approach of Rjiba et al. (2021). In Column (1), we use a pooled cross-sectional time-series regression with robust standard errors clustered by firm. In Column (2), we use the Fama-McBeth two-step procedure to estimate Equation (3). This estimation technique is especially suitable for panel data where the dependent variable, in this case, *DD*, varies across both individuals and time. It also helps us address the potential issues of cross-sectional dependence and time-varying factor loadings. We also use the Prais-Winsten methodology in Column (3). This estimation technique is particularly useful for correcting first-order autocorrelations, allowing us to account for the dependence of current error terms on past errors. Finally, we use the firm fixed effect that accounts for unobserved heterogeneity across firms, ensuring that the estimation focuses on the within-firm variation over time in Column (4). Across all the different estimation techniques, the *BOG Index* is significantly negatively correlated with *DD* at a 1% significant level, which supports our hypothesis that default risk increases when the readability of the annual report decreases. Following Kim et al. (2019), we use the results of Column (1) as our reference baseline regression results for the purpose of comparison in the following sections.

Our results are also economically significant. Specifically, a one standard deviation increase in the *BOG Index* (8.59), indicating lower readability, is associated with a 0.45 point decrease in *DD* (calculated as 8.59 × –0.052 from Column 1), which is 10.78% (calculated as 0.45/4.173) decrease in *DD* relative the mean of *DD*. Additionally, an interquartile shift in the *BOG Index* from the lower to the upper quartile (25th to the 75th percentile) of the *Bog Index* distribution results in an average decrease in *DD* of 0.57 points. For comparative context, Column (5) of Table 2 presents the economic impact of an interquartile shift for each variable included in our regression model.

Focusing on the control variables, the results indicate a significant positive association between *Size* and *DD*. Consistent with intuition, firms with more capital expenditure, firms with more profitability, greater liquidity, and higher growth prospects have higher stability in terms of default risk. Conversely, firms with higher leverage and higher tangibility face higher default risk.

Overall, we conclude that the model presented in Table 2 provides a good basis for identifying the association between readability and default risk.

Overall, our findings align with prior research on how capital markets price the attributes of narrative disclosures. For example, Bonsall and Miller (2017) show that less readable financial disclosures are associated with less favorable bond ratings, greater rating agency disagreement, and higher borrowing costs, highlighting the financial risks of poor disclosure quality. Similarly, Chen et al. (2024) show how a lower level of readability in pension narratives is associated with higher bond yield spreads. Rjiba et al. (2021) further document that greater textual complexity is associated with a higher cost of equity capital, consistent with the notion that poor disclosure quality increases investors' risk perceptions. Thus, our study extends the extant literature by demonstrating a direct association between lower readability of narrative disclosures in 10-K reports and higher default risk.

[Insert Table 2 Here]

4.2 Addressing endogeneity: Instrumental variable approach

There are concerns about potential endogeneity issues in the baseline results, such as, reverse causality might be at play: firms facing default risk could make their annual reports less readable to obscure their financial difficulties. To address this, we use a two-stage least squares (2SLS) regression analysis with an instrumental variable (IV) approach. An IV excludes exogenous variation in the independent variable, mitigating the bias induced by endogeneity. The industry average Bog Index (*Ind_Avg_Bog*) is adopted as the instrument. Previous research supports using the industry average of the key independent variable as an instrument in similar settings (Carvajal & Nadeem, 2023; Hossain et al., 2020; Kim et al., 2023; Kong et al., 2021), given that identifying a perfect instrument for reporting-related variables is often challenging, if not impossible. Given the idiosyncratic nature of firm-level readability, the industry average is likely to influence a firm's readability without directly affecting the firm's default risk, fulfilling the requirements of a valid instrument.

Table 3 presents the results of the first-stage regression analysis, which examines the association between the instrument, *Ind_Avg_Bog*, and the key variable of interest, *Bog Index*. The results reported in Column (1) indicate a significant association between the two variables. To assess the validity of the instrument, we employ standard diagnostic tests. The Anderson-Canon under-identification test statistic (59.37) rejects the null hypothesis of under-identification. Additionally, the Cragg-Donald Wald F-statistic (44.04) exceeds the Stock-Yogo critical value (maximum 16.38 at the 10% level), providing strong evidence of instrument relevance. These findings collectively suggest that the selected instrument is appropriately specified and robust.

Column (2) displays the second-stage regression results, controlling for the endogeneity between firm-level readability and default risk. The predicted value of *Bog Index,* the independent variable, from the first-stage regression has a significant negative association with *DD*. These 2SLS results support our baseline estimates, reinforcing the finding that less readable 10-K reports are associated with higher levels of default risk.

[Insert Table 3 Here]

4.3 Addressing endogeneity: Propensity score matching

We employ propensity score matching (PSM) to triangulate our previous findings further, which helps address potential limitations of our OLS analysis. PSM enables us to test the sensitivity of our findings to correlated omitted variable bias, thereby relaxing the need to assume linear or other specific functional relationships between variables (Armstrong et al., 2010). As a result, we construct a new, matched sample of firms that share similar characteristics relevant to default risk but are distinguished by the readability of their annual reports. By creating this comparable group, we alleviate misspecification that occurs when the primary OLS design potentially assumes an incorrect functional form of relationship between our variables of interest (*Bog Index* and control variables) and dependent variable (*DD*) (Armstrong et al., 2010). Accordingly, differences in default risk can be attributed to the level of readability.

To perform the PSM analysis, our first step was to estimate a logistic model that regresses a dichotomous variable (*BOGDUMMY*), coded one (zero) if the *BOG Index* score is above (below) the median value, on potential determinants of the default risk, which are the control variables in Equation (3). The results of this logistic regression are presented in Column (1) of Panel A of Table 4. If the readability score in a given year is above the median value, we treat the firm as a high-readable; otherwise, it is a low-readable in that year. Using the coefficients from the logistic regression, we then calculated a propensity score for each firm-year observation. We subsequently matched each observation with a *BOGDUMMY* = 1, with replacement, to a unique firm-year observation where *BOGDUMMY*= 0, based on the closest propensity score within a caliper of 0.001, allowing for replacement. This matching process yielded a sample of 17,806 matched pairs of low-readable and high-readable firms, totaling 35,612 firm-year observations. A successful matching should have two key features: first, the variables used to predict readability are no longer statistically significant after matching in distinguishing the groups, evidencing balanced covariate distribution across treatment conditions; and second, the pseudo-R^2 from the initial logistic regression significantly dropped, indicating reduced systematic variation between high-readability and low-readability firms (Caliendo & Kopeinig, 2008).

The results presented in Column (2) of Panel A in Table 4 confirm that our matching process meets both these criteria successfully. A further examination of the covariate balance between high-readable and low-readable firms confirms that matched samples are comparable across all other dimensions except for the *BOG Index*. Panel B of Table 4 further supports that our matching algorithm has successfully achieved covariate balance, as evidenced by no significant difference in the mean values of control variables between high-readable and low-readable firms. In addition to validating the matching process, we also conducted a post-matching univariate test to examine whether default risk differs systematically between high- and low-readability firms (Panel C, Table 4). The results show that firms with lower readability (higher Bog Index) exhibit significantly higher default risk (lower *DD*), with a mean difference of –0.287 ($t = 12.039$, $p < 0.01$).

We re-estimate Equation (3) employing the propensity score-matched sample and the results shown in Column (1) of Panel C of Table 4. In addition to the neighboring matching method, we also used the radius and Kernel matching methods to re-estimate Equation (3) in Columns (2) and (3) of Panel D of Table 4, respectively, further supporting our hypothesis. In all three models, the coefficient for the *DD* is negative and statistically significant at the 1% level for the *BOG Index*.

[Insert Table 4 Here]

4.4 Addressing endogeneity: Entropy balancing

A firm's decision to make the annual reports more (less) readable may not be random—the decision likely reflects management's intention and trade-off between the benefits and costs of readability. Therefore, systematic differences may exist between high-readable and low-readable firms, exposing our results to potential self-selection bias. Although propensity score matching also addresses self-selection concerns, at least to some extent, the method has some limitations: the unmatched samples are dropped, resulting in a loss of observations; imbalance may exist in the covariate characteristics (mean, variance, and skewness) between treated and non-treated groups; PSM results are sensitive to design choices (Hainmueller, 2012). To address the limitations of PSM and provide further robustness to our findings, we follow existing research (Alam et al., 2024; Bonsall & Miller, 2017; McMullin & Schonberger, 2020) to employ an entropy balancing approach. This approach aims to achieve a superior covariate balance between the treated and non-treated firms (Hainmueller, 2012). The covariate balance is achieved by assigning weights to observations such that the post-weighing distributions of the treatment group become identical to those of the control group. This process adjusts for both random and systematic inequalities across the variables of interest (Hainmueller, 2012). Particularly, none of the observations are lost in the balancing process.

To further validate our findings, we employ entropy balancing as an additional robustness check by splitting our sample into two groups based on the median *Bog Index* value, similar to our PSM approach. Firms with a *Bog Index* above the median are classified as the treatment group, while others constitute the control group. Panels A and B of Table 5 show that covariate balance is achieved based on the distribution metrics before and after entropy balancing. Panel C of Table 5 presents the regression results. The coefficient on the predicted *Bog Index* remains negative and significant, reinforcing our initial findings, suggesting the association between readability and default risk is not spurious and is robust to the application of entropy balancing.

[Insert Table 5 Here]

5. Channel mechanism

In this section, we describe the mechanisms through which the readability of annual reports influences default risk. As discussed in Section 2, we hypothesize that firms with less readable annual reports may face higher risks, potentially increasing their default risk. We identify four pathways by which annual report readability can impact default risk: information asymmetry, risk-taking strategy, cost of capital, and financial constraints. In our analysis, a direct association refers to the direct effect of readability on default risk, whereas an indirect association describes a pathway from readability to default risk that involves at least one intermediary variable (e.g., information asymmetry, risk-taking strategy, cost of capital, and financial constraints). Following the established methodology, we adopt a two-step process to demonstrate the mediation effect. First, we examine whether the key independent variable (*Bog Index*) has a significant relationship with the mediator variables, such as proxies for information asymmetry, agency cost, risk-taking strategy, cost of capital, and financial constraint. Second, we regress the dependent variable (*DD*) on both the independent variable and the mediators, while controlling other relevant variables. For a variable to be considered a mediator in the association between readability and default risk, it should demonstrate a statistically significant association with *DD*. Moreover, the inclusion of the mediator in the model should attenuate the strength of the association between the key independent variable (readability) and the dependent variable (*DD*), relative to the baseline regression.

5.1 Information asymmetry and agency cost channel

We argue that a less readable annual report exacerbates information asymmetry, thereby elevating the default risk of firms. The obfuscation hypothesis posits that managers may intentionally conceal poor performance or negative information from investors by using sophisticated terminology and lengthy words (Li, 2008), leading to a poor information environment, as reflected in stock liquidity. In a liquid stock market where information asymmetry is minimum, a firm can readily sell its stock to manage debts with minimal impact on stock price. Conversely, in less liquid markets, issuance equity becomes more costly because increased stock supply exerts a more pronounced effect on price (Nadarajah et al., 2021). Therefore, enhanced liquidity may reduce default risk by improving the information environment in the process of increasing price efficiency. Empirical studies by Brogaard et al. (2017) in the US and Nadarajah et al. (2021) in international settings show that a lack of stock liquidity contributes to higher default risk. Aligned with these premises, we hypothesize that the Bog Index will exhibit a negative correlation with information asymmetry, which, in turn, is anticipated to inversely relate to default risk.

Beyond its impact on information asymmetry, readability is likely to affect agency costs, further mediating the readability–default risk relationship. A lower level of readability weakens monitoring, enabling managerial opportunism, inefficient investments, and risk-shifting, which heighten financial instability (Healy & Palepu, 2001; Jensen & Meckling, 1976). Elevated agency costs erode governance, increase capital constraints, and amplify default risk (Anderson et al., 2004). Thus, we hypothesize that a higher BOG index increases agency costs, which in turn elevates default risk, underscoring the importance of financial transparency in corporate stability.

Consistent with previous studies (Brogaard et al., 2017; Nadarajah et al., 2021), we consider Bid-Ask spread (*Spread*) and Amihud's stock illiquidity (*Amihud*) as proxies for information asymmetry. Additionally, we calculate agency costs (*Agency Cost*) using the methodology proposed by Obeng et al. (2021). The details of *Spread*, *Amihud, and Agency Cost* are presented in Appendix A2.

The regression results from the mediation analysis are presented in Table 6. We present the baseline regression result in Column 1 for the sake of comparison. The results presented in Column 2 indicate a positive and significant correlation between the *Bog Index* and *Spread*, indicating that less readability of annual reports is positively associated with heightened information asymmetry. When we include the *Bog Index* and *Spread*, along with other control variables, in Column (3), we find a significant negative association of *Bog Index* and *Spread* with *DD*. This provides evidence that the *Bog index* affects default risk by increasing the bid-ask spread. Importantly, *Bog Index* has a negative significance at the 1% level. In addition, the size of the *Bog Index* coefficient in Column (3) of Table 6 is smaller than that in the baseline regression shown in Column (1) of Table 6. Column (3) of Table 6 presents the results of the Sobel test, confirming the presence of a significant mediation effect at the 1% level.

In Column (1), the total effect of the *Bog Index* on *DD* is 0.052, and in Column (3), the direct effect is 0.030. This yields an indirect effect of 0.022 (0.052 – 0.030), accounting for 42% (0.022/0.052) of the total effect and implying *Spread* as a significant mediator in the readability and default risk association. The other information asymmetry proxy, *Amihud*, has qualitatively similar and statistically significant results. In Column (5), the direct effect is 0.032, and the indirect effect calculates to 0.02 (0.052 – 0.032), representing 38% (0.02/0.052) of the total effect. Columns (3) and (5) of Table 6 present the results of the Sobel test, confirming the presence of significant mediation effects at the 1% level. While the specific magnitudes of *Spread* and *Amihud* marginally differ, the overall pattern of results remains consistent, with information asymmetry acting as a key channel in the association between readability and default risk.

To further examine agency costs as a mediating factor, we first assess the relationship between *Bog Index* and *Agency Cost* in Column (6). The results indicate a significant positive association at a 1% significance level, suggesting that firms with less readable annual reports experience higher agency costs, consistent with the notion that complex disclosures provide greater opportunities for managerial opportunism and inefficiencies. Next, in Column (7), we regress *DD* on both the *Bog Index* and *Agency Cost* to test whether agency costs mediate the association between readability and default risk. Both the *Bog Index* and *Agency Cost* exhibit significant negative associations with *DD*, indicating that higher agency costs contribute to increased default risk. Notably, the Bog Index coefficient in Column (7) is smaller in magnitude than in Column (1), suggesting a partial mediation effect through agency costs. The total effect of the *Bog Index* on *DD* is 0.052 (Column (1)), while the direct effect, after accounting for *Agency Cost*, is 0.048 (Column (7)). The indirect effect, capturing the mediation via agency costs, is 0.004 (0.052 – 0.048), constituting 7.7% (0.004 / 0.052) of the total effect. While agency costs partially mediate the relationship between readability and default risk, the economic magnitude of this effect is smaller compared to information asymmetry proxies. Thus, the mediation analysis results in Table 6 demonstrate that information asymmetry and agency costs act as key channels in the association between readability and default risk.

[Insert Table 6 Here]

5.2 Risk taking strategy channel

Low readability in annual reports leads to increased risk-taking by obscuring key financial details by the management. When reports are complex or difficult to understand, stakeholders may find it challenging to assess a firm's financial health and strategic direction accurately, which could lead to misaligned risk-taking decisions by management. As discussed in the earlier section that low readability results in heightened information asymmetry, giving managers more discretion to engage in risky ventures that may not align with the firm's long-term stability (Ertugrul et al., 2017). When management is not clear about its risk exposure and financial performance, it can pursue speculative investments or risky strategies without effective oversight from investors or creditors, increasing the firm's overall default risk.

To examine the risk-taking channel, we utilize two proxies of risk following Coles et al. (2006) and Baghdadi et al. (2020): R&D expenditure (*R&D*) and *ROA volatility*. *R&D* expenditure is quantified as the ratio of research and development spending to total assets. *ROA volatility* is calculated by computing the rolling standard deviation of *ROA* over the preceding three years.

The mediation analysis results are detailed in Table 7. Following the identical procedure as in Section 5.1, we repeat the baseline regression result in Column 1 for comparison purposes. The results displayed in Column 2 reveal a positive and significant association between the *Bog Index* and *R&D*, indicating that a poorer readability of annual reports is linked to a greater level of *R&D* expenditure. When we include the *Bog Index* and *R&D*, along with other control variables, in Column 3, we find that *Bog Index* and *R&D* are negatively associated with *DD*, indicating that *Bog Index* affects default risk by increasing the *R&D* expenditures. In Table 7, the size of the coefficient on *Bog Index* in Column (3) is smaller than the baseline coefficient in Column (1) and the association is significant at the 1% level. The total effect of *Bog Index* on *DD* is 0.052 (Column (1)), while the direct effect presented in Column (3) is 0.028. The indirect effect is 0.024, which is calculated as 0.052 – 0.028, representing 46% (0.024/0.052) of the total effect, indicating that *R&D* mediates the association between readability and default risk, and the mediation effect is economically large.

Our further analysis involving ROA Volatility as a mediator reaffirms risk taking strategy as a potential channel, where the total effect of *Bog Index* on *DD* is 0.052 (Column (1)), with a direct effect of 0.032 (Column (5)). The indirect mediation effect of 0.02 (0.052 – 0.032) accounts for 38% (0.02/0.052) of the total effect, implying that *ROA Volatility* also serves as a mediator between readability and default risk association. The Sobel test results presented in Columns (3) and (5) confirm that both mediation effects are significant at the 1% level. Overall, the results in Table 7 indicate that risk-taking plays an important role in how readability affects *DD*.

[Insert Table 7 Here]

5.4 Cost of capital channel

Financial statement readability significantly affects a firm's cost of capital (Hoffmann & Kleimeier, 2021; Rjiba et al., 2021), encompassing both the cost of equity and the cost of debt, which is likely to mediate the relationship between readability and default risk. As argued earlier, firms with less readable annual reports create greater information asymmetry, increasing uncertainty and risk perception among investors and creditors (Li, 2008; Rennekamp, 2012). This opacity raises financing costs, as both equity and debt providers demand higher risk premiums to compensate for the difficulty in assessing firm fundamentals and default risk (Easley & O'Hara, 2004; Lambert et al., 2007). From an equity financing perspective, lower readability reduces investor confidence, increases opinion divergence, and impairs the reliability of financial signals, all of which result in a higher cost of equity (Bai et al., 2019; Bloomfield & Fischer, 2010). Empirical evidence confirms that greater textual complexity in financial reports leads to a higher cost of equity capital, as investors perceive such disclosures as increasing information risk and valuation uncertainty (Rjiba et al., 2021).

Similarly, from a debt financing perspective, poor readability increases credit risk perceptions, making it more challenging for rating agencies and creditors to assess default probabilities and long-term solvency (Livingston & Zhou, 2010). Firms with lower financial disclosure readability tend to experience higher credit spreads and borrowing costs, as uncertainty surrounding firm fundamentals raises default expectations (Hoffmann & Kleimeier, 2021). Recent empirical studies confirm that reduced financial transparency leads to increased borrowing costs, particularly for firms operating in environments with higher information asymmetry and limited prior lending relationships (Bonsall & Miller, 2017). As higher financing costs constrain liquidity and investment capacity, firms become more vulnerable to financial distress and default risk (Francis et al., 2005). Thus, the cost of capital serves as a key transmission channel, linking low financial statement readability to heightened default risk, emphasizing the critical role of transparent disclosures in corporate financial stability.

To estimate the cost of capital, we first calculate the cost of equity using the implied cost of equity approach. Following Rjiba et al. (2021), we derive four separate estimates using the models by Claus and Thomas (2001), Gebhardt et al. (2001), Easton (2004) and Ohlson and Juettner-Nauroth (2005). Since each model makes different assumptions regarding forecast horizons and growth expectations, we compute the arithmetic average of these estimates, denoted as r_E, to obtain a comprehensive measure of the cost of equity. Next, we estimate the cost of debt as r_D, measured as the ratio of interest expense to total debt, which reflects the firm's effective borrowing rate. To compute the firm's weighted average cost of capital (WACC), we assign weights to debt and equity based on their respective book values. We further adjust the cost of debt by incorporating the effective tax rate (τ_C), obtained from the database, to reflect the tax shield on interest payments. Finally, we compute WACC using the standard formula:

$$\text{WACC} = \frac{E}{V} r_E + \frac{D}{V} r_D (1 - \tau c)$$

where $WAAC$ represents the firm's weighted average cost of capital, E and D represent the book values of equity and debt, V is the total firm value (E + D), r_E is the cost of equity, r_D is the cost of debt, and τ_C is the effective tax rate. This approach provides a robust measure of the firm's overall cost of capital, which is crucial for evaluating investment and financing decisions. The regression results are presented in Table 8.

Our mediation analysis reveals that WACC, cost of equity, and cost of debt significantly mediate the relationship between readability and default risk (*DD*). First, in Column (2) of Table 8, we find that *Bog Index* is negatively associated with $WACC$, indicating that firms with less readable financial reports face higher capital costs. When WACC is included in the default risk regression in Column (3), the direct effect of *Bog Index* on *DD* reduces from 0.052 to 0.035, implying an indirect effect of 0.017, which accounts for 32.7% (0.017/0.052) of the total effect.

Similarly, in Column (4), we find that *Bog Index* is positively associated with cost of equity (r_E), suggesting that investors demand higher equity risk premiums from firms with lower readability. When both *Bog Index* and cost of equity are included in Column (5), the direct effect of *Bog Index* on *DD* decreases from 0.052 to 0.042, indicating an indirect effect of 0.010, which accounts for 19.2% (0.010/0.052) of the total effect.

Finally, in Column (6), *Bog Index* is positively associated with cost of debt (r_D), implying that creditors charge higher interest rates to firms with less transparent disclosures. In Column (7), after including cost of debt, the direct effect of *Bog Index* on *DD* drops from 0.052 to 0.039, reflecting an indirect effect of 0.013, which represents 25% (0.013/0.052) of the total effect. Across all three mediation channels, the Sobel tests ($p < 0.01$) confirm the statistical significance of these indirect effects, reinforcing the conclusion that higher financing costs—arising from poor readability—heighten firms' default risk. Overall, our results indicate that the Bog index increases both the cost of equity and the cost of debt, which increases the default risk of the firm.

[Insert Table 8 Here]

5.5 Financial constraint channel

In this section, we explore whether financial constraints could be a channel for the association between readability and default risk. The lack of readability could lead to increased financial constraints due to diminished trust and transparency with investors and lenders. For instance, Ertugrul et al. (2017) show that firms with less readable annual reports have higher costs of debt; Rjiba et al. (2021b) also show that lower annual report readability is associated with a higher cost of equity, which is likely to intensify financial constraints. Such financial constraints may exacerbate a firm's default risk by reducing its ability to secure funding and maintain financial stability (Zhang et al., 2020). Based on the above arguments, we expect *Bog Index* to be positively associated with proxies of financial constraint, whereas both *Bog Index* and financial constraint proxies to be negatively associated with *DD*. To test this conjecture, we measure financial constraint by using *KZ Index* as in Kaplan and Zingales (1997) and *WW Index* as in Whited and Wu (2006). *We* present the regression results in Table 9.

Following the process similar to earlier channels For the *KZ Index*, the direct effect (Column (3)) and indirect effect are 0.035 and 0.017 (0.052 – 0.035), respectively. The indirect effect represents 33% (0.017/0.052) of the total effect. Similarly, for the *WW Index*, the direct effect (Column (5)) and indirect effect are 0.034, and 0.018 (0.052 – 0.034), respectively. The indirect effect accounts for 35% of the total effect. These results provide evidence that financial constraint serves as a mediating channel through which readability affects default risk.

[Insert Table 9 Here]

6. Cross-sectional analysis

6.1 Role of monitoring

This section examines how the moderating role of monitoring, measured by institutional investors and analyst coverage, influences the association between the readability of annual reports and firms' default risk. Generally, monitoring is considered important to ensure a greater level of readability of narrative disclosures (Tunyi et al., 2024). Prior literature suggests that institutional owners and financial analysts, characterized by significant stakes and investment expertise, play a pivotal role as a governance mechanism (Rjiba et al., 2021; Wang & Zhang, 2009). These parties act as diligent monitors and influencers of firm behavior, particularly in financial reporting practices. They demand high-quality, transparent and reliable financial reporting from management, fostering practices that align with the interests of broader shareholder groups and potentially enhancing the quality of readability in annual reports (Lehavy et al., 2011). While the readability of annual reports primarily impacts the informational challenges faced by general investors, it may exert a relatively minor direct impact on the risk assessment capabilities of more sophisticated stakeholders, such as institutional investors and analysts. These stakeholders, acting as informed intermediaries, are expected to alleviate the information asymmetry exacerbated by poorly readable annual reports and can reasonably assess firms' risk, thus lowering the default risk. On the other hand, in firms with less external scrutiny, managers might use complex financial disclosures to obscure poor performance or risky financial practices, thus increasing the default risk. This risk is likely understated due to the lack of comprehensive analysis by fewer analysts and the passive nature of less involved institutional investors.

Research by Lo et al. (2017) observe that the increase in disclosure complexity for potentially problematic firms is three times greater for companies lacking analyst coverage compared to those with analyst following. Rjiba et al. (2021) report that the effect of the BOG index on the cost of equity is less significant for firms with greater institutional ownership and more extensive analyst coverage. Similarly, Luo et al. (2018) find that the impact of readability on agency cost is lower for firms with higher analyst coverage. Drawing upon these arguments, we propose that the impact of annual report readability on default risk would be lesser for firms with higher institutional ownership and greater analyst coverage. To test this hypothesis, we first created dummy variables based on the median values of institutional ownership and analyst coverage. These dummy variables are assigned a value of 1 if institutional ownership or the number of analysts is above the median and a value of 0 if they are at or below the median.

Table 10 presents the regression results examining the impact of the *BOG Index* on default risk *(DD),* with the analysis split by institutional ownership (*Institutional Ownership*) and analyst coverage (*Analyst*). Columns (1) and (2) focus on the effect of the *BOG Index* on default risk for firms with below-median and above-median institutional ownership, respectively. The results indicate that the negative impact of the *BOG Index* on *DD* is significant in both cases. However, the magnitude of the impact is smaller for firms with above-median institutional ownership compared to those with below-median institutional ownership. This suggests that higher institutional ownership mitigates the adverse effect of the *BOG Index* on default risk. Columns (3) and (4) show the impact of the *BOG Index* on *DD* for firms with below-median and above-median analyst coverage, respectively. Similar to the institutional ownership analysis, the negative impact of the *BOG Index* on *DD* is present in both groups. Nonetheless, the impact is less pronounced for firms with above-median analyst coverage compared to those with below-median analyst coverage. This indicates that greater analyst coverage also reduces the negative effect of the *BOG Index* on *DD*. These findings confirm our hypothesis that the negative impact of the *BOG Index* on *DD* is mitigated for firms with higher levels of institutional ownership and greater analyst coverage.

[Insert Table 10 Here]

6.2 Role of organizational capital and managerial ability

In this section, we explore the role of organizational capital and managerial ability in moderating the relationship between annual report readability and default risk. Organizational competencies enhance a firm's capacity to manage and utilize its resources effectively, leading to improved operational efficiencies and risk management practices (Eisfeldt & Papanikolaou, 2013). Similarly, managerial competencies refer to a set of skills, behaviors, and attributes that are crucial for effective management and leadership in organizations (Demerjian et al., 2012). These competencies encompass a broad spectrum of capabilities that enable managers to handle their roles efficiently and contribute to organizational success. We argue that firms with higher organizational capital typically exhibit better governance structures and are more adept at navigating financial complexities, thereby potentially reducing default risk. In addition, robust internal capabilities and competent management can compensate for the informational disadvantages that might arise from less readable reports (Panta & Panta, 2023), thereby increasing the firm's stability. This aligns with findings that superior management practices, as part of organizational capital, decrease a firm's investment-cash flow sensitivity by reducing financing frictions (Attig & Cleary, 2014).

In a similar vein, managerial ability may significantly moderate the relationship between the readability of annual reports and the perceived default risk of firms. This moderation is predicated on the premise that highly skilled managers leverage their profound understanding of the industry and the macroeconomic landscape to enhance the quality of financial reporting and estimates (Demerjian et al., 2013), thereby mitigating the potential adverse effects associated with less readable financial reports. Moreover, the strategic acumen of skilled managers enables them to balance the requisite technical detail with the clarity needed to ensure comprehensive stakeholder understanding. This balance is crucial for reducing the default risk that might be exacerbated by opaque financial reporting (Bonsall et al., 2017). To test this prediction, we utilize the organizational capital as in Peters and Taylor (2017) and retrieve the managerial ability data from Demerjian et al. (2012). We divide the sample based on the median value of organizational capital and managerial ability. A firm belongs to the above-median group if the organizational capital and managerial ability scores are higher than the median; otherwise, it belongs to the below-median group. We then rerun the baseline regression separately on both high and low samples. Our results are reported in Table 11. Columns (1) and (2) of Table 11 show that the impact of the *BOG Index* on *DD* is significantly negative for both groups of organizational capital. However, the impact is less pronounced for firms with above-median organizational capital, suggesting that higher organizational capital mitigates the adverse effects of the *BOG Index* on *DD*. Similarly, results reported in Columns (3) and (4) show that the *BOG Index* negatively impacts *DD* for firms with both below-median and above-median groups of managerial ability. However, the effect is less pronounced for firms with above-median managerial ability, indicating that strong managerial ability mitigates some adverse effects of the *BOG index* on *DD.*

[Insert Table 11 Here]

6.3 Role of business strategy

Extant literature suggests that firms' strategic orientation shapes the readability of their annual reports (Habib & Hasan, 2020; Lim et al., 2018), as business strategies drive variations in information complexity, investor perceptions, and financial stability. For instance, while prospector firms produce less readable disclosures, defender firms generate clearer narratives (Habib & Hasan, 2020; Lim et al., 2018). Building on this research, we examine how business strategy moderates the association between readability and default risk.

Prospector firms, driven by innovation and high growth, operate in volatile markets with frequent technological shifts, leading to heightened information asymmetry and financial distress (Bentley et al., 2013). These firms may produce less readable reports to obscure poor performance (Li, 2008) or to protect proprietary information (Bagnoli & Watts, 2010), which exacerbates default risk as investors demand transparency, raising financing costs and distress likelihood. In contrast, defender firms prioritize efficiency and stability, operating in predictable environments with lower investor uncertainty and reduced reliance on external financing. While defenders may strategically limit disclosures to protect cost structures, their financial stability weakens the link between low readability and default risk, as investors rely more on historical performance than narrative disclosures.

To test this proposition, we classify firms' business strategy following the methodology of Bentley et al. (2013), which builds on the work of Miles et al. (1978) typology. Specifically, we construct a composite strategy score based on six firm characteristics: (i) R&D intensity (R&D expenditures scaled by sales) to capture innovation-seeking behavior, (ii) employee-to-sales ratio to measure operational efficiency, (iii) employee fluctuations (standard deviation of total employees) as a proxy for workforce stability, (iv) historical sales growth to reflect firm expansion,

(v) marketing intensity (SG&A expenses scaled by sales) to assess emphasis on market positioning, and (vi) capital intensity (net PPE scaled by total assets) to indicate focus on production efficiency. Each of these variables is computed as a five-year rolling average and ranked into quintiles within each two-digit SIC industry-year, with firms in the highest quintile receiving a score of 5 and those in the lowest receiving a score of 1 (except for capital intensity, which is reverse-scored). The strategy score, ranging from 6 to 30, is then used to categorize firms: Prospectors (≥24), Defenders (≤12), and Analyzers (remaining observations). This classification enables a systematic, data-driven approach to capturing firms' strategic orientation, allowing us to examine its moderating role in the relationship between readability (*Bog Index*) and default risk (*DD*).

Consistent with our expectations, we find that the *Bog Index* has a stronger negative impact on *DD* for Prospector firms than for Defender firms, indicating that low readability exacerbates default risk in firms with high strategic uncertainty and external financing reliance.

[Insert Table 12 Here]

6.4 Role of debt concentration

Building on extant literature highlighting the adverse effects of poor readability (Bonsall & Miller, 2017; Ertugrul et al., 2017), we demonstrate that poor readability adversely affects default risk. However, this effect is likely to vary depending on the firm's debt ownership structure. Public bondholders, lacking access to private firm data, rely heavily on public disclosures, making firms with low readability more susceptible to risk premiums and higher default probabilities. In contrast, firms with high bank debt concentration benefit from superior monitoring, as banks access private information and mitigate the adverse effects of poor readability through direct oversight (Ben-Nasr et al., 2021; Boot et al., 1993).

Empirical evidence supports this differential effect, as relationship lending allows banks to assess risks through private channels rather than relying solely on public disclosures (Bharath et al., 2011). This mitigates managerial opportunism and financial distress, weakening the link between readability and default risk for firms with intensive bank debt (Denis & Mihov, 2003). Conversely, firms dependent on public debt face greater investor uncertainty, where low readability amplifies information asymmetry, eroding investor confidence and thus increasing the default risk. This theoretical underpinning motivates us to examine how the relationship between readability and default risk differs across firms with high versus low debt concentration.

To empirically examine this hypothesis, we collect firm-level capital structure data from Capital IQ, with a particular focus on the amount of bank debt and the percentage of total debt sourced from banks. Using this data, we categorize firms into high and low debt concentration groups based on the median value of bank debt percentage within our sample. Firms with above-median bank debt concentration are classified as high debt concentration firms, while those below the median are categorized as low debt concentration firms. To assess whether debt concentration moderates the relationship between readability and default risk, we re-estimate our baseline regression separately for each group, using a debt concentration dummy as the classification criterion. The regression results, presented in the subsequent section, provide empirical evidence on the differential impact of readability on default risk based on the firm's debt ownership structure.

The regression results presented in Panel B of Table 12 indicate that the extent of the effects of *Bog Index* on default risk (*DD*) varies significantly based on debt concentration, supporting the hypothesis that debt ownership structure moderates the readability-default risk relationship. In firms with high debt, the coefficient on *Bog Index* is –0.049 and statistically significant, indicating that lower readability is associated with increased default risk. However, this effect is attenuated compared to firms with low debt concentration, where the coefficient on the *Bog Index* is –0.058 and is also highly significant. The stronger effect in low-debt-concentration firms suggests that public debt holders rely more on readability when assessing firm risk, whereas bank lenders compensate for poor disclosure quality through direct monitoring and access to private information.

6.5 Role of SEC 1998 Plain English Rule

In 1998, the U.S. Securities and Exchange Commission (SEC) introduced the Plain English Disclosure Rule, requiring firms to enhance the clarity and transparency of their financial disclosures by eliminating complex language and ambiguous formatting (Dempsey et al., 2012; Li, 2008). This regulatory change aimed to improve the readability of corporate reports, enabling investors to better assess firm fundamentals and mitigate information asymmetry. Following Hasan (2020), we examine whether this regulatory intervention influenced the relationship between readability and default risk by splitting our sample into two periods: pre-SEC rule (1994–1997) and post-SEC rule (1999–2002) and re-estimating our baseline regression separately for each period. We consider 1999–2002 as the post-SEC rule period to ensure balance with the pre-SEC rule period.

The results, presented in Columns (1) and (2) of Table 13, indicate that the negative association between readability (*Bog Index*) and default risk (*DD*) remains statistically significant in both periods, reinforcing the notion that poor readability contributes to default risk. However, we observe that the *Bog Index* coefficient is smaller in the post-SEC period (–0.033, $p < 0.01$)

compared to the pre-SEC period (–0.057, $p < 0.01$), suggesting that the adverse impact of low readability on default risk weakened following the implementation of the SEC rule. These findings imply that regulatory efforts to enhance disclosure readability may have partially mitigated the financial consequences of opaque reporting, reducing firms' exposure to default risk.

[Insert Table 13 Here]

6.6 Role of firm performance and overall risk

In this section, we are interested to see whether poor readability still adversely affects default risk if the firm's performance is better or the overall risk level is lower than their counterparts. There is a possibility that better financial performance or a lower low-risk profile may offset the adverse effects of poor readability on default risk because firms' financial performance and risks are considered important factors for firms' cost of debt and overall solvency (Atif & Ali, 2021; Brogaard et al., 2017; Francis et al., 2021).

We measure firm performance and overall firm risk based on Tobin's Q (*Tobin's Q*) and stock volatility (*OverallRisk*), respectively (Atif & Ali, 2021; Guenther et al., 2017). Our results show that the adverse effect of *Bog Index* on *DD* is significantly more pronounced for firms with higher *Tobin's Q*. Thus, the results suggest that poor readability is an important factor in predicting default risk despite a firm having better financial performance, further indicating investors of such firms are likely to rely on transparent financial disclosures to assess future prospects and reduce investment risk. Turning to the overall risk, the results reveal that the negative association between *Bog Index* and *DD* is more pronounced for the firms with lower firm risk. Therefore, the results suggest that even if the firms have a low-risk profile, poor readability can still exacerbate the default risk, further reinforcing our baseline findings that poor readability increases default risk.

Overall, our results suggest that even if a firm has strong financial performance or a lower level of overall risk, a lower readability score in its narrative disclosures increases its default risk. Thus, the results reinforce our baseline findings that poor readability increases default risk.

[Insert Table 14 Here]

6.7 Role of board governance attributes

We examine the moderating effect of key governance characteristics—board size, board independence, and gender diversity—on the relationship between *Bog Index* and default risk (*DD*), considering the importance of governance characteristics for corporate disclosures and default risk (Liu et al., 2024; Schultz et al., 2017; Switzer et al., 2018). We split the sample based on the median values of these governance metrics and re-estimate our baseline regression. Our results show that the negative impact of *Bog Index* on *DD* is more pronounced in firms with smaller boards, lower board independence, and lower gender diversity, suggesting that weaker governance exacerbates the consequences of poor readability. The results are presented in Appendix A3.

7. Additional tests

7.1 Sensitivity analysis: Actual bankruptcy

To strengthen the empirical foundation of our analysis and ensure that our findings reflect real-world corporate failure, we incorporate actual bankruptcy filings as a complementary measure. While market-based indicators can offer useful early signals, the inclusion of formal Chapter 7 and Chapter 11 bankruptcy events allow us to directly capture instances of financial collapse (Bris et al., 2006), providing a more grounded validation of our results.[3] We obtain firm-

[3] See Bris et al. (2006) for details about Chapter 7 and Chapter 11.

level bankruptcy data from Audit Analytics, identifying 1,069 unique firms that filed for bankruptcy during our sample period.

These filings translate into a total of 8,032 firm-year observations, allowing us to analyze how persistent issues in readability influence actual bankruptcy events rather than market-based perceptions of default risk. This approach enables us to assess whether low readability, as measured by the Bog Index, contributes to the likelihood of firms entering bankruptcy proceedings.

To examine whether readability affects the likelihood of bankruptcy, we create a dummy variable (Bankruptcy = 1) if a firm has filed for bankruptcy in a given year and 0 otherwise. Since financial distress is often influenced by firm-specific factors beyond readability, we construct a control group of non-bankrupt firms that are similar in size, leverage, liquidity, and other key financial characteristics using Propensity Score Matching (PSM).

Using a nearest-neighbor matching approach, we match each bankrupt firm to a non-bankrupt firm with a similar propensity score based on size. After confirming balance in control variables, we first conduct a mean comparison test (t-test) to examine whether Bog Index differs systematically between bankrupt and non-bankrupt firms (results omitted). We then estimate a logistic regression model where bankruptcy status is regressed on *Bog Index* and other controls from the baseline regression model. Additionally, recognizing that financial distress is often the result of sustained low readability rather than a one-time reporting issue, we extend our analysis by constructing three-year and five-year rolling averages of *Bog Index* to capture the cumulative effect of poor readability over time. This approach allows us to evaluate whether persistent disclosure challenges have a more pronounced effect on default risk compared to short-term variations in readability.

Results are presented in Table 15. The results provide empirical evidence that firms with lower readability in financial disclosures face a significantly higher likelihood of bankruptcy. In Model (1), the negative and statistically significant coefficient on *Bog Index* indicates that firms with higher *Bog Index* (i.e., lower readability) are more prone to bankruptcy filings. This finding suggests that poor disclosure quality exacerbates financial distress. The effect of readability on bankruptcy risk becomes stronger over longer horizons, as evidenced by the three-year and five-year rolling averages in Columns (2) and (3), respectively. The coefficient on three-year rolling average (*Bog Index 3yr)* remains negative and significant, while the five-year rolling average (*Bog Index 5yr*) suggests an even larger and more persistent effect. Our results confirm that poor readability increased the probability of bankruptcy for firms.

[Insert Table 15 Here]

7.2 Sensitivity analysis: Alternative proxy of default risk

To check the robustness of our baseline regression, we employ alternative default risk proxies. Following Kabir et al. (2021), we implement two supplementary default risk indicators: probability of default (*PD*) and Altman Z score (*Altman Z*). The Probability of Default (*PD*) denotes the likelihood that a borrower is failing to meet their debt obligations. NUS-RMI utilizes a comprehensive system to calculate PDs for a broad spectrum of listed firms globally, offering daily updates based on available data. A higher score of the probability of default indicates a higher default risk. On the other hand, the *Altman Z*-score uses multiple discriminant analyses in assessing a borrower's likelihood of going bankrupt over the following two years. The Z-score is calculated by adding the weighted ratios of the five financial ratios that are used in this model. Each weighted ratio has a distinct value. A lower Z-score reflects a greater bankruptcy risk and *vice versa*. We expect the *Bog Index* to be passively associated with the probability of default (*PD*) and negatively associated with the *Altman Z* score.

Panel B of Table 16 displays regression results with alternative proxies for default risk. The *BOG Index* indicates a significant positive relationship with *PD* and a significant negative relationship with the *Altman Z*-score, both at the 1% significance level. These findings are consistent with our baseline results, confirming the robustness of our initial findings regarding the impact of the *BOG Index* on default risk.

[Insert Table 16 Here]

7.3 Sensitivity analysis: Alternative proxy of readability

7.3.1 Complexity

To further check the robustness of our baseline regression results, we use alternative proxies of readability from extant literature. While our main analysis uses the Bog Index to measure readability, this section utilizes alternative measures to ensure the robustness of our findings. As our first alternative, we consider the '*Complexity*' measure developed by Loughran and McDonald (2023), who define firm complexity using a textual approach. They define the concept of firm complexity using a textual measure. This measure, derived from the methods of Loughran and McDonald (2011), is based on the analysis of the language used in Form 10-Ks (annual reports) filed with the SEC. It encompasses the multifaceted nature of complexity and is easily calculable with machine learning techniques. Specifically, they identify 374 words that most significantly contribute to firm complexity. These words, such as 'lease,' 'merger,' 'foreign,' 'patent,' and 'contract,' are tallied for their unique occurrences in a firm's annual report. On average, a typical firm mentions 81 of these 374 complex words at least once in their 10-K filings, which serves as a proxy for measuring their complexity. We expect *complexity* to be negatively correlated with *DD*.

7.3.2 Accounting Reporting Complexity (ARC)

The second alternative measure of firm annual report readability is Accounting Reporting Complexity (*ARC*) developed and measured by Hoitash and Hoitash (2018) by employing XBRL filings. XBRL, mandated by the SEC for translating 10-K financial statements, is ideal for this purpose. *ARC* is calculated by counting the number of distinct accounting concepts (like revenues, net inventory, etc.) disclosed in these filings. Each concept, governed by specific standards and regulations, adds to the complexity of preparing financial reports. This complexity arises as more diverse information needs to be gathered, organized, and analyzed, requiring extensive knowledge of accounting standards. We expect *ARC* to be negatively associated with *DD*.

7.2.3 Gross file size

We incorporate the 10-K file size proposed by Hasan (2020), as a third measure of financial report readability. He demonstrates the relevance and robustness of 10-K file size (in megabytes) as an indicator of readability. Notably, file size offers an advantage in terms of measurement efficiency, as it avoids the need for complex text parsing and reduces the potential for measurement errors. We employ the natural logarithm of the file size *(Gross File Size)*, with higher values indicating lower readability. Accordingly, we expect a negative relationship between *Gross File Size* and *DD*.

7.3.3 Net file size

Following Hasan (2020), we also use the log of net file size (*Net File Size*) of the 10-K report as the fourth alternative measure of financial report readability. Net file size excluded the graph and table and considered actual text in the report. Similar to gross file size, net file size is expected to have a negative relation with *DD*.

Regression results are reported in Panel A of Table 16. Consistent with our predictions, all four alternative proxies of readability—*Complexity, ARC, Gross File Size, and Net File Size*—are negatively correlated with DD at the 1% significance level. These results confirm the baseline findings, reinforcing the importance of readability in reducing default risk.

7.4 Evidence from machine learning algorithm

We employ Random Forest (RF) and Extreme Gradient Boosting (XGBoost) algorithms, which provided additional evidence supporting the evidence that readability is a significant factor in default risk, corroborating findings in the current accounting and finance literature (Jones et al., 2023; Rahman et al., 2024). Contrary to the extant literature that relies on parametric statistical models, such as Ordinary Least Squares (OLS), for inference-methods, which have recently come under scrutiny due to concerns over "*p*-hacking" (Kim et al., 2018; Ohlson, 2015). However, these machine learning approaches offer a novel perspective. Breiman (2001) introduced RF algorithm, which operates by constructing multiple decision trees and then combining their individual predictions. XGBoost, another ensemble method, leverages the concept of boosting, focusing on improving predictions for instances that previous models misclassified (Bogousslavsky et al., 2024). Importantly, XGBoost operates independently of *p-values*, relying instead on the collective predictive power of multiple models, a characteristic of ensemble learning. Figure 1 presents the variable importance ranking based on RF and XGBoost models. In terms of importance for predicting default risk, the Bog Index (*Bog_Index*) ranks seventh by using the RF algorithm in Panel A and ranks eighth by using the XGBoost model in Panel B. These results strongly support the findings from our regression analysis. Both models demonstrate that readability is a significant factor influencing default risk.

[Insert Figure 1 Here]

7.5 Change regression

Following Kabir et al. (2021), we conduct the change regression by taking the first difference of all variables. By taking the first differences, we eliminate any time-invariant characteristics specific to each firm are removed. This is particularly useful when these characteristics are not observable or measurable but are known to influence the dependent variable-default risk. This transformation simplifies the model by focusing only on variations within each unit over time, removing the need to model these fixed effects explicitly. Regression results presented in Panel A of Table 17 reinforce our baseline findings of the negative association between readability and default risk.

7.6 Lag of independent variables

We also conduct lead-lag analysis by taking the lag of all independent variables in the regression model. This method helps determine if changes in the readability of financial reports (as captured by the *BOG Index*) precede changes in default risk, thus clarifying the causal direction between these variables. In addition, lead-lag analysis can address endogeneity issues by separating the effects of simultaneous changes in the *BOG Index* and other control variables. Regressions results presented in Panel B of Table 17, provide further robustness to our baseline results.

7.7 Inclusion of additional control variables

To further enhance the robustness of our findings and address concerns regarding omitted variable bias, we incorporate additional control variables into our baseline estimation, and the results are presented in Table 18. Column (1) extends the baseline estimation by including corporate governance characteristics, including board size (*Board_Size*), board independence (*Board_Independence*), director qualifications (*Director_Qualification*), gender diversity (*Gender_Diversity*), and director networks (*Director_Network*), which are likely to influence managerial oversight and financial stability. Column (2) further accounts for firm age (*Age*), loss (*Loss*), segment diversification (*BusSegment* and *GeoSegnent*), opacity (*Opaqueness*), the existence of special items (*Special Items*), and being audited by big four audit firms (*BIG4*) in a given year. Across both specifications, the *Bog Index* remains negatively and significantly associated with default risk (*DD*), reinforcing our central argument that lower readability is linked to heightened default risk, even after controlling for an extensive set of firm-specific and governance factors.[4]

[Insert Table 18 Here]

7.8 Holdout samples

To ensure the robustness of our initial findings, we conduct an additional set of regressions by partitioning the sample using two alternative thresholds: a 40%-60% split and a 25%-75% split based on the Bog Index distribution. This approach allows us to assess whether the effect of readability on default risk is consistent across different classifications of high- and low-readability

---

[4] Our (untabulated) results continue to hold if we include environmental factors, such as total carbon emissions, direct carbon emissions, and ESG scores, as additional control variables in our baseline regressions. As the inclusion of these environment-related variables significantly reduces the number of observations, the results remain untabulated.

firms. The results are presented in Appendix A4. Across all specifications, our results remain qualitatively unchanged, as we continue to find that a higher *Bog Index* (indicating lower readability) is associated with increased default risk.

8. Conclusion

This study investigates the association between the readability of narrative disclosures in 10-K reports and default risk. Less readable information in annual reports enhances information risk by deterring the information processing capability of debt investors. Hence, debt investors charge premiums and impose stricter loan covenants to compensate for their information risk, exposing the firms to greater default risk. To empirically test this hypothesis, we examine the association between the *Bog Index*, a measure of readability, and Merton's (1974) distance-to-default (*DD*) measure. Our results indicate a negative association between the *Bog Index* and Merton's (1974) *DD*, suggesting that an increase in the readability of 10-K reports corresponds to a decrease in corporate default risk. The association is also economically meaningful: a one standard deviation increase in the *BOG Index* is related to a 0.45 point drop in *DD*, which is a 10.78% decrease in *DD* relative to the mean of *DD*. We further explore the potential channels through which readability influences default risk. Our cross-sectional analyses reveal wide-range factors, such as level of monitoring, organizational capital and managerial ability, business strategy, debt concentration, firm performance, and overall risk, and the existence of the SEC 1998 Plain English Rule, which play important roles in the association between readability and default risk. This study further identifies information asymmetry, agency costs, corporate risk-taking strategies, cost of capital, and financial constraints as the underlying mechanisms for the intended association. Results remain robust to a battery of endogeneity and robustness tests.

This study relates to the existing literature by highlighting the importance of the readability of narrative disclosures as a valuable information source for investors (Rjiba et al., 2021). Thus, we extend the literature on how information risk, stemming from less readable reports, affects the debt market, specifically through their impact on corporate default risk (Bonsall & Miller, 2017; Chen et al., 2024). Our research further contributes to the prior research on the driving factors of default risk (Baghdadi et al., 2020; Kabir et al., 2021; Schultz et al., 2017). We identify poor readability as an accelerating factor of default risk, while strong monitoring mechanisms, greater managerial ability, and higher organizational capital within firms all emerge as potential mitigating factors. Finally, we build on the existing literature (Atif & Ali, 2021; Chava & Purnanandam, 2010), which considers better economic performance and a lower level of exposure to overall risk to be useful in the credit market. Our findings reveal that firms are likely to be exposed to a higher level of default risk despite having a better economic performance and a lower level of overall risk, further highlighting the importance of narrative disclosures to credit investors.

Our findings hold potential interest for various stakeholders. For corporate management, our research stresses the strategic significance of readily understandable annual reports in the debt market. By prioritizing clear and concise communication within these filings, companies can cultivate trust and potentially improve their access to capital at favorable terms. Financial analysts can also benefit from these findings by adopting a more proactive approach during Q&A sessions of earnings conference calls, particularly when dealing with firms historically known for less readable reports. By posing targeted questions that clarify specific aspects of firm performance and reporting, analysts can enhance their understanding and provide more informed investment recommendations.

References

Alam, M. S., Hasan, M. M., Alam, N., & Islam, M. S. (2024). Managerial Ability and Debt Choice. *Abacus*. https://doi.org/10.1111/abac.12334

Aldamen, H., & Duncan, K. (2013). Pricing of innate and discretionary accruals in Australian debt. *Accounting & Finance, 53*(1), 31-53.

Ali, S., Liu, B., & Su, J. J. (2018). Does corporate governance quality affect default risk? The role of growth opportunities and stock liquidity. *International Review of Economics & Finance, 58*, 422-448.

Anderson, R. C., Mansi, S. A., & Reeb, D. M. (2004). Board characteristics, accounting report integrity, and the cost of debt. *Journal of Accounting and Economics, 37*(3), 315-342.

Ang, J. S., Cole, R. A., & Lin, J. W. (2000, 2000/02). Agency Costs and Ownership Structure. *The Journal of Finance, 55*(1), 81-106.

Armstrong, C. S., Jagolinzer, A. D., & Larcker, D. F. (2010). Chief executive officer equity incentives and accounting irregularities. *Journal of Accounting Research, 48*(2), 225-271.

Ataullah, A., Vivian, A., & Xu, B. (2018). Optimistic Disclosure Tone and Conservative Debt Policy. *Abacus, 54*(4), 445-484.

Atif, M., & Ali, S. (2021). Environmental, social and governance disclosure and default risk. *Business Strategy and the Environment, 30*(8), 3937-3959.

Attig, N., & Cleary, S. (2014). Organizational capital and investment-cash flow sensitivity: The effect of management quality practices. *Financial Management, 43*(3), 473-504.

Baghdadi, G. A., Nguyen, L. H., & Podolski, E. J. (2020). Board co-option and default risk. *Journal of Corporate Finance, 64*, 101703.

Bagnoli, M., & Watts, S. G. (2010). Oligopoly, disclosure, and earnings management. *The Accounting Review, 85*(4), 1191-1214.

Bai, X., Dong, Y., & Hu, N. (2019). Financial report readability and stock return synchronicity. *Applied Economics, 51*(4), 346-363.

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

Ben-Nasr, H., Boubaker, S., & Sassi, S. (2021). Board reforms and debt choice. *Journal of Corporate Finance, 69*, 102009.

Bentley, K. A., Omer, T. C., & Sharp, N. Y. (2013). Business strategy, financial reporting irregularities, and audit effort. *Contemporary Accounting Research, 30*(2), 780-817.

Besuglov, E., & Crasselt, N. (2021). The effect of readability and language choice in management accounting reports on risk-taking: an experimental study. *Journal of Business Economics, 91*(1), 5-33.

Bharath, S. T., Dahiya, S., Saunders, A., & Srinivasan, A. (2011). Lending relationships and loan contract terms. *The Review of Financial Studies, 24*(4), 1141-1203.

Biddle, G. C., Hilary, G., & Verdi, R. S. (2009). How does financial reporting quality relate to investment efficiency? *Journal of Accounting and Economics, 48*(2-3), 112-131.

Blanco, B., Coram, P., Dhole, S., & Kent, P. (2021). How do auditors respond to low annual report readability? *Journal of Accounting and Public Policy, 40*(3), 106769.

Bloomfield, R., & Fischer, P. E. (2010). Disagreement and the Cost of Capital. *Journal of Accounting Research, 49*(1), 41-68.

Bloomfield, R. J. (2002). The 'incomplete revelation hypothesis' and financial reporting. *Accounting Horizons, 16*, 233-243.

Bogousslavsky, V., Fos, V., & Muravyev, D. (2024). Informed trading intensity. *The Journal of Finance, 79*(2), 903-948.

Bonsall, S. B., Leone, A. J., Miller, B. P., & Rennekamp, K. (2017). A plain English measure of financial reporting readability. *Journal of Accounting and Economics, 63*(2-3), 329-357.

Bonsall, S. B., & Miller, B. P. (2017). The impact of narrative disclosure readability on bond ratings and the cost of debt. *Review of Accounting Studies, 22*(2), 608-643.

Boot, A. W. A., Greenbaum, S. I., & Thakor, A. V. (1993). Reputation and Discretion in Financial Contracting. *The American Economic Review, 83*(5), 1165-1183.

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5-32.

Bris, A., Welch, I., & Zhu, N. (2006). The costs of bankruptcy: Chapter 7 liquidation versus Chapter 11 reorganization. *The Journal of Finance, 61*(3), 1253-1303.

Brogaard, J., Li, D., & Xia, Y. (2017). Stock liquidity and default risk. *Journal of Financial Economics, 124*(3), 486-502.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys, 22*(1), 31-72.

Campbell, J. L., Chen, H., Dhaliwal, D. S., Lu, H.-m., & Steele, L. B. (2014). The information content of mandatory risk factor disclosures in corporate filings. *Review of Accounting Studies, 19*, 396-455.

Carvajal, M., & Nadeem, M. (2023). Financially material sustainability reporting and firm performance in New Zealand. *Meditari Accountancy Research, 31*(4), 938-969.

Chava, S., & Purnanandam, A. (2010). Is default risk negatively related to stock returns? *The Review of Financial Studies, 23*(6), 2523-2559.

Chen, H., Ho, K.-C., Zhang, M., & Zhang, Q. (2023). Effect of managerial ability toward corporate social responsibility on enterprise default risk. *Finance Research Letters, 54*, 103700.

Chen, P.-C. (2016). Banks' acquisition of private information about financial misreporting. *The Accounting Review, 91*(3), 835-857.

Chen, T.-K., Tseng, Y., Lin, R.-C., & Hung, Y.-S. (2024). Readability of Pension Narrative Disclosures, Pension Regulatory Changes, and Corporate Credit Risk. *European Accounting Review*, 1-29.

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

Choi, Y., Levine, G., & Malone, S. W. (2020). The coronavirus (COVID-19) pandemic: Assessing the impact on corporate credit risk. *Moody's Analytics*, 1-17.

Claus, J., & Thomas, J. (2001). Equity premia as low as three percent? Evidence from analysts' earnings forecasts for domestic and international stock markets. *The Journal of Finance, 56*(5), 1629-1666.

Coles, J. L., Daniel, N. D., & Naveen, L. (2006). Managerial incentives and risk-taking. *Journal of Financial Economics, 79*(2), 431-468.

Contessi, S., De Pace, P., & Guidolin, M. (2014). How did the financial crisis alter the correlations of US yield spreads? *Journal of Empirical Finance, 28*, 362-385.

Corwin, S. A., & Schultz, P. (2012). A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices. *The Journal of Finance, 67*(2), 719-760.

Courtis, J. K. (2004). Corporate report obfuscation: artefact or phenomenon? *The British Accounting Review, 36*(3), 291-312.

Dalwai, T., Chinnasamy, G., & Mohammadi, S. S. (2021). Annual report readability, agency costs, firm performance: an investigation of Oman's financial sector. *Journal of Accounting in Emerging Economies, 11*(2), 247-277.

Demerjian, P., Lev, B., & McVay, S. (2012). Quantifying managerial ability: A new measure and validity tests. *Management Science, 58*(7), 1229-1248.

Demerjian, P. R., Lev, B., Lewis, M. F., & McVay, S. E. (2013). Managerial ability and earnings quality. *The Accounting Review, 88*(2), 463-498.

Dempsey, S. J., Harrison, D. M., Luchtenberg, K. F., & Seiler, M. J. (2012). Financial opacity and firm performance: the readability of REIT annual reports. *The Journal of Real Estate Finance and Economics, 45*(2), 450-470.

Denis, D. J., & Mihov, V. T. (2003, 2003/10). The choice among bank debt, non-bank private debt, and public debt: evidence from new corporate borrowings. *Journal of Financial Economics, 70*(1), 3-28.

Diamond, D. W. (1991). Debt maturity structure and liquidity risk. *The Quarterly Journal of Economics, 106*(3), 709-737.

Dieckmann, S., & Plank, T. (2012). Default risk of advanced economies: An empirical analysis of credit default swaps during the financial crisis. *Review of Finance, 16*(4), 903-934.

Duan, J.-C., Sun, J., & Wang, T. (2012). Multiperiod corporate default prediction—A forward intensity approach. *Journal of Econometrics, 170*(1), 191-209.

Easley, D., & O'Hara, M. (2004, 2004/08). Information and the Cost of Capital. *The Journal of Finance, 59*(4), 1553-1583.

Easterbrook, F. H. (1984). Two agency-cost explanations of dividends. *The American Economic Review, 74*(4), 650-659.

Easton, P. D. (2004). PE ratios, PEG ratios, and estimating the implied expected rate of return on equity capital. *The Accounting Review, 79*(1), 73-95.

Eisfeldt, A. L., & Papanikolaou, D. (2013). Organization capital and the cross-section of expected returns. *The Journal of Finance, 68*(4), 1365-1406.

Ertugrul, M., Lei, J., Qiu, J., & Wan, C. (2017). Annual report readability, tone ambiguity, and the cost of borrowing. *Journal of Financial and Quantitative Analysis, 52*(2), 811-836.

Fisher, R., Van Staden, C. J., & Richards, G. (2019). Watch that tone: An investigation of the use and stylistic consequences of tone in corporate accountability disclosures. *Accounting, Auditing & Accountability Journal, 33*(1), 77-105.

Fleming, G., Heaney, R., & McCosker, R. (2005). Agency costs and ownership structure in Australia. *Pacific-Basin Finance Journal, 13*(1), 29-52.

Francis, B., Hasan, I., Liu, L., Wu, Q., & Zhao, Y. (2021). Financial analysts' career concerns and the cost of private debt. *Journal of Corporate Finance, 67*, 101868.

Francis, J., LaFond, R., Olsson, P., & Schipper, K. (2005, 2005/06). The market pricing of accruals quality. *Journal of Accounting and Economics, 39*(2), 295-327.

Gebhardt, W. R., Lee, C. M., & Swaminathan, B. (2001). Toward an implied cost of capital. *Journal of Accounting Research, 39*(1), 135-176.

Goswami, R., Maji, S. G., & Hussain, F. (2023). Annual report readability and agency cost: the influence of firm size. *Business Perspectives and Research*, 1-16.

Guenther, D. A., Matsunaga, S. R., & Williams, B. M. (2017). Is tax avoidance related to firm risk? *The Accounting Review, 92*(1), 115-136.

Habib, A., & Hasan, M. M. (2020). Business strategies and annual report readability. *Accounting & Finance, 60*(3), 2513-2547.

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

Hasan, M. M. (2020). Readability of narrative disclosures in 10-K reports: does managerial ability matter? *European Accounting Review, 29*(1), 147-168.

Hassan, M. K., Abu Abbas, B., & Garas, S. N. (2019). Readability, governance and performance: a test of the obfuscation hypothesis in Qatari listed firms. *Corporate Governance: The International Journal of Business in Society, 19*(2), 270-298.

Healy, P. M., & Palepu, K. G. (2001, 2001/09). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics, 31*(1-3), 405-440.

Henry, D. (2010, 2010/01). Agency costs, ownership structure and corporate governance compliance: A private contracting perspective. *Pacific-Basin Finance Journal, 18*(1), 24-46.

Hilscher, J., & Wilson, M. (2017). Credit ratings and credit risk: Is one measure enough? *Management Science, 63*(10), 3414-3437.

Hoffmann, A. O., & Kleimeier, S. (2021). Financial disclosure readability and innovative firms' cost of debt. *International Review of Finance, 21*(2), 699-713.

Hoitash, R., & Hoitash, U. (2018). Measuring accounting reporting complexity with XBRL. *The Accounting Review, 93*(1), 259-287.

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

Hossain, M., Raghunandan, K., & Rama, D. V. (2020). Abnormal disclosure tone and going concern modified audit reports. *Journal of Accounting and Public Policy, 39*(4), 106764.

Hu, N., Liu, L., & Zhu, L. (2018). Credit default swap spreads and annual report readability. *Review of Quantitative Finance and Accounting, 50*(2), 591-621.

Jensen, M. C., & Meckling, W. H. (1976, 1976/10). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics, 3*(4), 305-360.

Jones, S., Moser, W. J., & Wieland, M. M. (2023). Machine learning and the prediction of changes in profitability. *Contemporary Accounting Research*.

Jurkus, A. F., Park, J. C., & Woodard, L. S. (2011, 2011/02). Women in top management and agency costs. *Journal of Business Research, 64*(2), 180-186.

Kabir, M. N., Rahman, S., Rahman, M. A., & Anwar, M. (2021). Carbon emissions and default risk: International evidence from firm-level data. *Economic Modelling, 103*, 105617.

Kaplan, S. N., & Zingales, L. (1997). Do investment-cash flow sensitivities provide useful measures of financing constraints? *The Quarterly Journal of Economics, 112*(1), 169-215.

Kim, C., Wang, K., & Zhang, L. (2019). Readability of 10-K reports and stock price crash risk. *Contemporary Accounting Research, 36*(2), 1184-1216.

Kim, D. S., Chung, C. Y., & Fard, A. (2023). Effects of 10-K readability on institutional blockholder monitoring of risk management. *Applied Economics Letters, 30*(18), 2657-2666.

Kim, J. H., Ahmed, K., & Ji, P. I. (2018). Significance Testing in Accounting Research: A Critical Evaluation Based on Evidence. *Abacus, 54*(4), 524-546.

Kong, D., Shi, L., & Zhang, F. (2021). Explain or conceal? Causal language intensity in annual report and stock price crash risk. *Economic Modelling, 94*, 715-725.

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

Lambert, R., Leuz, C., & Verrecchia, R. E. (2007). Accounting Information, Disclosure, and the Cost of Capital. *Journal of Accounting Research, 45*(2), 385-420.

Lawrence, A. (2013). Individual investors and financial disclosure. *Journal of Accounting and Economics, 56*(1), 130-147.

Lee, Y. J. (2012). The effect of quarterly report readability on information efficiency of stock prices. *Contemporary Accounting Research, 29*(4), 1137-1170.

Lehavy, R., Li, F., & Merkley, K. (2011). The effect of annual report readability on analyst following and the properties of their earnings forecasts. *The Accounting Review, 86*(3), 1087-1115.

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

Liao, H.-H., Chen, T.-K., & Lu, C.-W. (2009). Bank credit risk and structural credit models: Agency and information asymmetry perspectives. *Journal of Banking & Finance, 33*(8), 1520-1530.

Liao, S. (2015). Outside blockholders' monitoring of management and debt financing. *Contemporary Accounting Research, 32*(4), 1373-1404.

Lim, E. K., Chalmers, K., & Hanlon, D. (2018). The influence of business strategy on annual report readability. *Journal of Accounting and Public Policy, 37*(1), 65-81.

Liu, S., Wang, K. T., & Wu, Y. (2024). Corporate governance reforms and analyst forecasts: International evidence. *Abacus, 60*(2), 272-304.

Livingston, M., & Zhou, L. (2010). Split bond ratings and information opacity premiums. *Financial Management, 39*(2), 515-532.

Lo, K., Ramos, F., & Rogo, R. (2017). Earnings management and annual report readability. *Journal of Accounting and Economics, 63*(1), 1-25.

Loughran, T., & McDonald, B. (2014). Measuring readability in financial disclosures. *The Journal of Finance, 69*(4), 1643-1671.

Loughran, T., & McDonald, B. (2023). Measuring firm complexity. *Journal of Financial and Quantitative Analysis*, 1-28.

Luo, J.-h., Li, X., & Chen, H. (2018). Annual report readability and corporate agency costs. *China Journal of Accounting Research, 11*(3), 187-212.

McMullin, J. L., & Schonberger, B. (2020). Entropy-balanced accruals. *Review of Accounting Studies, 25*(1), 84-119.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance, 29*(2), 449-470.

Miles, R. E., Snow, C. C., Meyer, A. D., & Coleman Jr, H. J. (1978). Organizational strategy, structure, and process. *Academy of Management Review, 3*(3), 546-562.

Nadarajah, S., Duong, H. N., Ali, S., Liu, B., & Huang, A. (2021). Stock liquidity and default risk around the world. *Journal of Financial Markets, 55*, 100597.

Nadeem, M. (2022). Board gender diversity and managerial obfuscation: Evidence from the readability of narrative disclosure in 10-K reports. *Journal of Business Ethics, 179*(1), 153-177.

Obeng, V. A., Ahmed, K., & Cahan, S. F. (2021). Integrated reporting and agency costs: International evidence from voluntary adopters. *European Accounting Review, 30*(4), 645-674.

Ohlson, J. A. (2015). Accounting Research and Common Sense. *Abacus, 51*(4), 525-535.

Ohlson, J. A., & Juettner-Nauroth, B. E. (2005). Expected EPS and EPS growth as determinants of value. *Review of Accounting Studies, 10*, 349-365.

Panta, H., & Panta, A. (2023). Organizational capital and readability of financial reports. *Finance Research Letters, 55*, 103895.

Peters, R. H., & Taylor, L. A. (2017). Intangible capital and the investment-q relation. *Journal of Financial Economics, 123*(2), 251-272.

Plumlee, M. A. (2003). The effect of information complexity on analysts' use of that information. *The Accounting Review, 78*(1), 275-296.

Rahman, S. (2024). The importance of green patents for CDS pricing: The role of environmental disclosures. *Energy Economics, 139*, 107905.

Rahman, S., Sinnewe, E., & Chapple, L. (2024). Environment-specific political risk discourse and expected crash risk: The role of political activism. *International Review of Financial Analysis, 95*, 103494.

Rennekamp, K. (2012, 2012/06/06). Processing Fluency and Investors' Reactions to Disclosure Readability. *Journal of Accounting Research, 50*(5), 1319-1354.

Rjiba, H., Saadi, S., Boubaker, S., & Ding, X. S. (2021). Annual report readability and the cost of equity capital. *Journal of Corporate Finance, 67*, 101902.

Schrand, C. M., & Walther, B. R. (2000). Strategic benchmarks in earnings announcements: the selective disclosure of prior-period earnings components. *The Accounting Review, 75*(2), 151-177.

Schultz, E. L., Tan, D. T., & Walsh, K. D. (2017). Corporate governance and the probability of default. *Accounting & Finance, 57*, 235-253.

SEC Office of Investor Education and Advocacy. (2021). *How to read a 10-K/10-Q*

Switzer, L. N., Tu, Q., & Wang, J. (2018). Corporate governance and default risk in financial firms over the post-financial crisis period: International evidence. *Journal of International Financial Markets, Institutions and Money, 52*, 196-210.

Tunyi, A. A., Hussain, T., Areneke, G., & Agyemang, J. (2024). Co-opted Boards and the Obfuscation of Financial Reports. *Abacus*.

Valta, P. (2012). Competition and the cost of debt. *Journal of Financial Economics, 105*(3), 661-682.

Wang, A. W., & Zhang, G. (2009). Institutional ownership and credit spreads: An information asymmetry perspective. *Journal of Empirical Finance, 16*(4), 597-612.

Whited, T. M., & Wu, G. (2006). Financial constraints risk. *The Review of Financial Studies, 19*(2), 531-559.

Yin, J., Han, B., & Wong, H. Y. (2022). COVID-19 and credit risk: A long memory perspective. *Insurance: Mathematics and Economics, 104*, 15-34.

Zhang, X., Ouyang, R., Liu, D., & Xu, L. (2020). Determinants of corporate default risk in China: The role of financial constraints. *Economic Modelling, 92*, 87-98.

Appendix A1: Definition of variables

Variable	Definition
Dependent Variable	
DD	Distance-to-Default (DD) is an analytical metric utilized in finance to assess a firm's insolvency risk by quantifying the deviation between its current asset value and a default threshold, normalized by asset volatility. DD is an inverse measure of default risk, where a smaller value of DD indicates a higher default risk.
Independent Variable	
Bog Index	The BOG index, developed by Editor Software's StyleWriter, is a proprietary readability metric that assesses plain English issues in a document, including passive voice, redundant verbs, jargon, and sentence complexity (Bonsall et al., 2017). A higher BOG index score indicates lower readability.
Primary Control Variables	
Size	Natural logarithm of total assets in $ millions.
CAPEX	Amount of capital expenditures scaled by total assets at the beginning of the year.
Liquidity	The ratio of liquid assets to total assets.
Leverage	Long-term debt plus short-term debt, scaled by total assets.
MTB	The ratio of the book value to the market value of equity.
ROA	Profitability of the firm, measured as net income scaled by total assets.
Tangibility	The ratio of property, plant, and equipment to the total asset.
Channel variables	
R&D	The ratio of R&D expenditures to the total asset.
ROA Volatility	The standard deviation of operating earnings during the prior three years.
Spread	The natural logarithm of the annual quoted spread, measured over a firm's fiscal year multiplied by −1. The quoted spread is the average of the daily ratio between the closing ask-bid price and the mid-point price. The annual quoted spread is the average of the daily quoted spread. See Appendix A2 for the detailed calculation process.
Amihud	The natural logarithm of the annual Amihud illiquidity ratio, measured over a firm's fiscal year multiplied by −1. The daily Amihud illiquidity ratio is the average of the ratio of the daily absolute stock return to the trading volume on that day multiplied by 105. The annual Amihud illiquidity ratio is the average of the daily Amihud illiquidity ratio. See Appendix A2 for the detailed calculation process.
Agency Cost	A composite measure of agency costs, derived from free cash flow, expense ratio, dividend payout ratio, and asset turnover, following Ang et al. (2000). See Appendix A2 for the detailed calculation process.
KZ Index	KZ index is calculated following Kaplan and Zingales (1997).

WW Index	WW index is calculated following Whited and Wu (2006).
Moderating variables	
Institutional Ownership	The proportion of firm shares owned by institutional investors.
Analyst	The natural logarithm of the number of analysts providing one-year-ahead earnings forecasts.
Organization Capital	The organizational capital measure of Peters and Taylor (2017).
Managerial Ability	Managerial ability measure as in Demerjian et al. (2012).
WACC	The firm's overall cost of capital, calculated as the weighted average of the cost of equity and the after-tax cost of debt.
r_E	Cost of equity is calculated by taking the average of estimated implied cost of equity models, as in Claus and Thomas (2001), Gebhardt et al. (2001), Easton (2004), and Ohlson and Juettner-Nauroth (2005).
r_D	Cost of debt is calculated as total interest expense divided by total debt.
Prospector	A dummy variable coded 1 if the strategy score based on Bentley et al. (2013) is between 24 and 30 (both inclusive) and 0 otherwise
Defender	A dummy variable coded 1 if the strategy score based on Bentley et al. (2013) is between 6 and 12 (both inclusive) and 0 otherwise.
Tobin's Q	(Fair Market value + Total Liabilities)/Total Assets.
Overall Risk	Moving standard deviation of total stock return over the last three years.
Variables used in additional tests	
BusSegment	The natural logarithm of one plus the number of business segments at the end of the year
GeoSegment	The natural logarithm of one plus the number of geographic segments at the end of the year
Opaqueness	The moving sum of the absolute value of abnormal accruals in the prior three years (i.e., ABACCt + ABACCt–1 + ABACCt–2), where abnormal accruals are estimated using the modified Jones model
Special_Items	The special items at the end of the fiscal year t scaled by the book value of total assets at the beginning of the year
Age	The number of entire years since the firm's first appearance in COMPUSTAT
BIG4	A dummy variable is set to 1 if the firm hires one of the Big4 audit firms and 0 otherwise.
Loss	The indicator variable is set to 1 if net income before extraordinary items is negative in the current and previous year and 0 otherwise.
PD	The probability of default is built on the forward intensity mode of (Duan et al., 2012)
Altman Z score	Z-Score= 1.2A + 1.4B + 3.3C + 0.6D + 1.0E A = working Capital/Total Assets B= Retained Assets/Total Assets C = Earning Before Tax and Interest/Total Assets D = Market value of Equity/Total Liabilities E = Sales/Total Assets

Complexity	A measure of firm complexity based on textual analysis by Loughran and McDonald (2023).
ARC	Firm complexity is measured with accounting reporting complexity (ARC) and is based on the count of accounting items disclosed in eXtensible Business Reporting Language (XBRL) filings.
Gross File Size	The natural logarithm of the file size in megabytes of the SEC EDGAR "complete submission text file" for the 10-K filing.
Net File Size	The natural logarithm of the file size in megabytes of the SEC EDGAR "complete submission text file" for the 10-K filing, where only text content is included.
Bankruptcy	A dummy variable that equals 1 if a firm files for bankruptcy under Chapter 7 or Chapter 11 in a given year, and 0 otherwise.
Bog Index 3yr	The average of a firm's Bog Index over the prior three years.
Bog Index 5yr	The average of a firm's Bog Index over the prior five years.
Board_Size	The total number of directors serving on a firm's board.
Board_Independence	The proportion of independent directors on the board.
Director_Qualifications	Educational achievements of board members, including degrees earned and institutions attended.
Gender_Diversity	The proportion of female directors on the board.
Director_Networks	A measure of board interconnectivity based on shared directorships across firms.

Appendix A2: Calculation of Bid-Ask spread (*Spread*) and stock liquidity (*Amihud*)

This research measures *Spread* from daily high and low stock prices following Corwin and Schultz (2012). Similar to prior research (Nadarajah et al., 2021; Rahman, 2024), spread calculation is presented as follows:

$$\mathrm{Spread}_{i,t} = \mathrm{Average}[\frac{2(e^{x,t}-1)}{1+e^{x,t}}] \qquad \text{(i)}$$

$$x_t = \frac{\sqrt{2y_t}-\sqrt{y_t}}{3-2\sqrt{2}} - \sqrt{\frac{z_t}{3-2\sqrt{2}}}, \quad y_t = \sum_{k=0}^{1}\left[\ln\left(\frac{H^o{}_{t+k}}{L^o{}_{t+k}}\right)\right]^2, \quad z_t = \sum_{k=0}^{1}\left[\ln\left(\frac{H^o{}_{t,t+k}}{L^o{}_{t,t+k}}\right)\right]^2$$

where, $H^o{}_t$ represents the observed high price on day *t*, and $L^o{}_t$ represents the observed low price on day *t*. *y* is the sum over two days of the squared daily log (high/low), and *z* is the squared log (high/low) where the high (low) value is over two days.

Stock illiquidity (*Amihud*) is measured based on Amihud's (2002) stock illiquidity estimate, which refers to the daily ratio of absolute stock return to trading volume, averaged over a number of trading days in a given year. It reflects how much the absolute stock price changes with one dollar of trading volume. Amihud's stock illiquidity is computed as follows:

$$\mathrm{Illiquidity}_{i,t} = \frac{1}{D_{i,t}}\sum_{d=1}^{D_{i,t}}\frac{|\mathrm{Return}_{i,t,d}|}{\mathrm{Volume}_{i,t,d}} \qquad \text{(ii)}$$

where $|Return_{i,t,d}|$ denotes the absolute stock return of firm *i* on day *d* of year *t*, $Volume_{i,t,d}$ is the trading volume of firm *i* on day *d* of year *t*, and $D_{i,t}$ is the number of days with available data for firm *i* in year *t*. The higher the *Illiquidity*, the lower the stock liquidity. To indicate a positive relation between stock liquidity and the measure, *Illiquidity* is multiplied by −1, as follows:

$$\mathrm{Amihud}_{i,t} = \mathrm{Illiquidity}_{i,t} \times (-1) \qquad \text{(iii)}$$

where *Amihud*$_{i,t}$ is the stock liquidity of firm *i* in year *t*. The higher the *Amihud*, the higher the stock liquidity.

Similar to Obeng et al. (2021), we measure agency costs using four established proxies: free cash flow (*AC_FCF*), expense ratio (*AC_ER*), dividend payout ratio (*AC_DPR*), and asset turnover ratio (*AC_ATO*), following Ang et al. (2000), Easterbrook (1984); (Henry, 2010) and Jurkus et al. (2011). *AC_FCF* captures the inefficiencies associated with excess free cash flow, measured as the product of free cash flow and an indicator variable for low-growth firms (Tobin's Q < 1), where higher values indicate greater agency costs *AC_ER*, calculated as the ratio of operating expenses to sales, reflects excessive managerial spending, with higher values indicating higher agency costs (Ang et al., 2000; Fleming et al., 2005). *AC_DPR*, the ratio of dividends to net income, serves as a governance mechanism to prevent overinvestment, where lower values signify higher agency costs (Jurkus et al., 2011). Lastly, *AC_ATO*, measured as the asset turnover ratio, captures managerial inefficiency in generating revenue, where lower values indicate greater agency costs (Ang et al., 2000). For consistency, we multiply *AC_ATO* and *AC_DPR* by –1 so that higher values consistently reflect greater agency costs. To construct a comprehensive measure, we conduct principal component analysis (PCA) on these four proxies and use the first principal component score (*Agency Cost*) as our main agency cost variable.

Appendix A3: Role of board governance characteristics results

This Appendix provides the regression results for the role of key governance characteristics—board size, board independence, and gender diversity—on the association between the Bog Index and default risk. The results are displayed in the following table.

	(1)	(2)	(3)	(4)	(5)	(6)
	Board Size		Board Independence		Gender Diversity	
	Above-median	Below-median	Above-median	Below-median	Above-median	Below-median
Dependent variable	DD	DD	DD	DD	DD	DD
Bog Index	–0.050***	–0.052***	–0.049***	–0.052***	–0.050***	–0.052***
	(–15.417)	(–8.853)	(–15.634)	(–8.455)	(–15.138)	(–9.183)
Size	0.573***	0.467***	0.482***	0.545***	0.472***	0.554***
	(39.312)	(18.313)	(34.008)	(24.071)	(34.692)	(20.796)
CAPEX	3.612***	3.001***	3.428***	3.532***	3.084***	4.691***
	(11.803)	(5.920)	(12.077)	(5.936)	(11.248)	(6.739)
Liquidity	1.703***	1.984***	1.839***	1.919***	2.102***	1.413***
	(14.831)	(12.285)	(17.544)	(9.282)	(20.356)	(6.503)
Leverage	–0.160***	–0.150***	–0.139***	–0.208***	–0.135***	–0.206***
	(–21.040)	(–11.890)	(–20.233)	(–13.129)	(–19.954)	(–13.486)
MTB	0.051***	0.041***	0.042***	0.061***	0.039***	0.064***
	(21.164)	(11.372)	(19.932)	(13.084)	(18.267)	(14.450)
ROA	0.190***	0.812***	0.291***	0.717***	0.268***	1.440***
	(6.821)	(4.471)	(9.995)	(2.794)	(8.467)	(4.894)
Tangibility	–0.534***	–1.199***	–0.709***	–0.834***	–0.486***	–0.962***
	(–3.363)	(–4.774)	(–4.718)	(–2.837)	(–3.199)	(–3.236)
Constant	4.410***	5.871***	4.947***	5.296***	4.842***	5.545***
	(15.960)	(11.551)	(18.072)	(9.954)	(17.113)	(10.403)
Observations	16,150	24,181	19,526	20,942	18,954	21,512
R-squared	0.469	0.347	0.364	0.483	0.372	0.446
Industry effects	Yes	Yes	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes	Yes	Yes
Cluster by firm	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table presents the role of board governance characteristics, such *as Board Size*, *Board Independence*, and *Gender Diversity*, in the association between the Bog Index and default risk (DD). We split the sample based on the median values of these governance metrics, classifying firms as either Above-median or Below-median. The dependent variable is *DD*, an inverse measure of default risk, while the key independent variable is the *Bog Index*, an inverse measure of readability. Definitions of all variables are provided in Appendix A1. t-statistics are reported in parentheses. The superscript asterisks ***, **, and * denote two-tailed statistical significance at the 1%, 5%, and 10% levels, respectively.

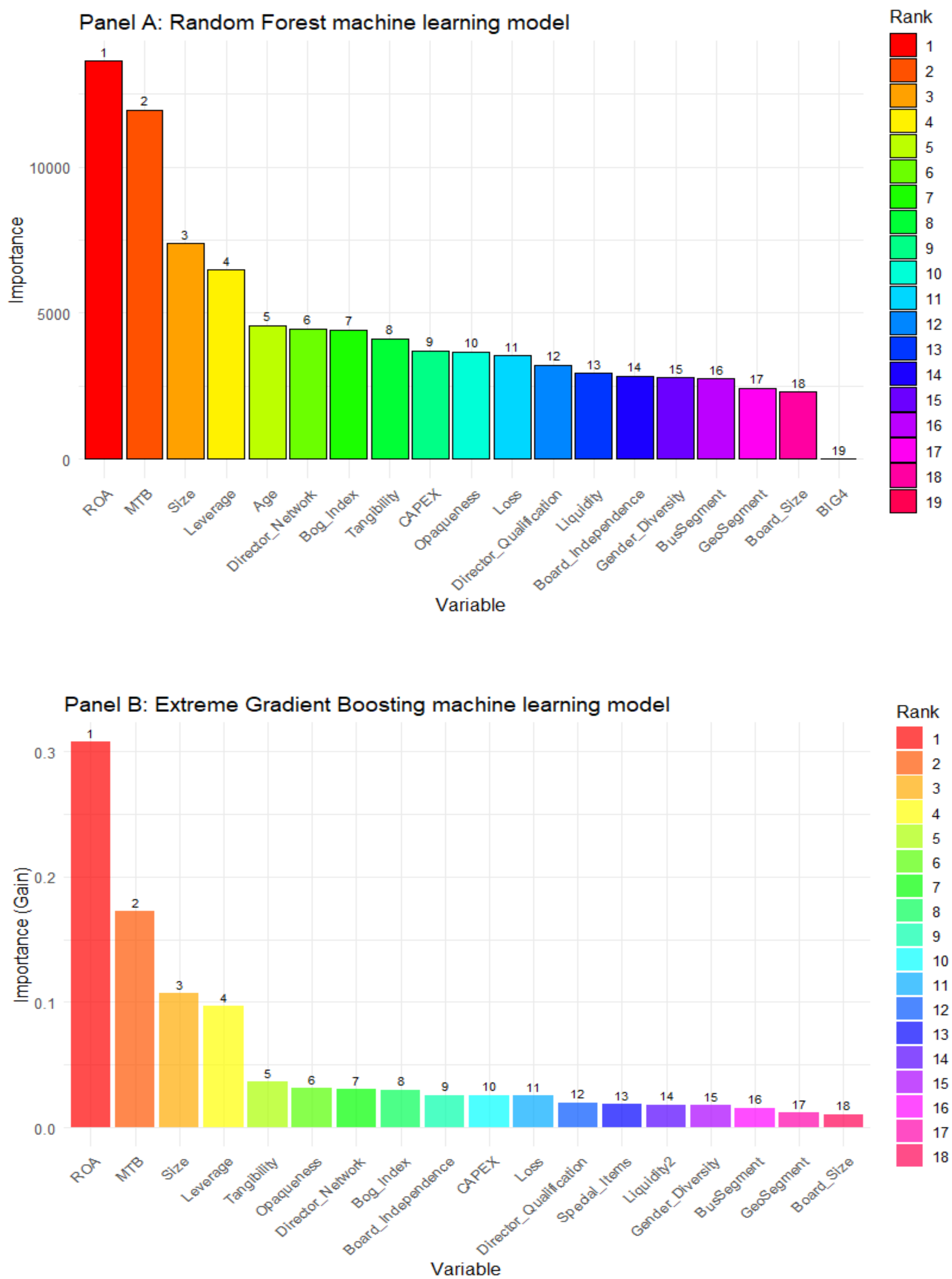
Appendix A4: Holdout sample results

This Appendix provides the regression results for the association between the *Bog Index* and default risk (*DD*) by partitioning the sample based on the Bog Index distribution. The results are presented in the following table.

	(1)	(2)	(3)	(4)
	60%	40%	75%	25%
	DD	DD	DD	DD
Bog Index	–0.042***	–0.064***	–0.044***	–0.069***
	(–9.624)	(–10.704)	(–11.046)	(–9.217)
Size	0.509***	0.570***	0.508***	0.603***
	(30.438)	(32.323)	(32.982)	(30.459)
CAPEX	3.422***	3.403***	3.686***	3.226***
	(8.229)	(10.261)	(10.351)	(8.186)
Liquidity	1.463***	2.555***	1.606***	2.643***
	(11.830)	(17.201)	(13.822)	(14.331)
Leverage	–0.132***	–0.198***	–0.140***	–0.206***
	(–16.962)	(–17.847)	(–18.995)	(–14.471)
MTB	0.042***	0.057***	0.045***	0.057***
	(17.844)	(15.535)	(19.864)	(12.256)
ROA	0.453***	0.173***	0.430***	0.113**
	(9.182)	(3.789)	(10.148)	(2.140)
Tangibility	–0.915***	–0.330*	–0.847***	–0.306
	(–4.830)	(–1.762)	(–4.863)	(–1.340)
Constant	4.632***	5.296***	4.707***	5.429***
	(11.752)	(11.152)	(13.150)	(9.355)
Observations	42,588	27,180	51,649	18,119
R-squared	0.391	0.491	0.400	0.512
Industry effects	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes
Cluster by firm	Yes	Yes	Yes	Yes

Notes: This table shows the association between readability and default risk by splitting the sample based on the Bog Index distributions, 25%, 40%, 60%, and 75%. The dependent variable is *DD*, an inverse measure of default risk. The key independent variable is *Bog Index*, an inverse measure of readability. Definitions of all variables are reported in Appendix A1. Robust standard errors adjusted for clustering by firm are reported in parentheses. The superscript asterisks ***, **, and * denote two–tailed statistical significance at the 1%, 5%, and 10% levels, respectively.

Figure 1: Variable importance plot using machine learning techniques



Notes: This figure displays the variable importance plot for predicting default risk (*DD*) using Random Forest (RF) and Extreme Gradient Boosting (XGBoost) machine learning techniques in Panels A and B, respectively. The figures are produced in R studio.

Table 1: Sample selection and summary statistics

PanelA: Sample selection process

	Firm-year Observation
Initial observations on Bog Index from Miller's website	189,665
Less: Merging data for Distance-to-default from CRI Database	96,260
Remaining Observations	93,405
Less: Missing data from control variables	15,558
Remaining observation	77,847
Less: Financial and Utility Companies	8,079
Remaining observation for baseline regression	69,768

Panel B: Descriptive statistics

	N	Mean	Std. Dev.	P25	Median	p75
DD	69,768	4.173	2.685	2.234	3.682	5.590
Bog Index	69,768	84.750	8.598	79.000	85.000	90.000
Size	69,768	5.772	2.188	4.180	5.699	7.290
CAPEX	69,768	0.053	0.061	0.016	0.033	0.065
Liquidity	69,768	0.216	0.245	0.031	0.115	0.319
Leverage	69,768	0.654	2.461	0.005	0.285	0.859
MTB	69,768	3.072	7.860	1.177	2.072	3.755
ROA	69,768	–0.109	0.613	–0.081	0.026	0.071
Tangibility	69,768	0.254	0.235	0.072	0.173	0.370

Panel C: Correlation matrix

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) DD	1.000								
(2) Bog Index	–0.042*	1.000							
(3) Size	0.392*	0.142*	1.000						
(4) CAPEX	0.001	–0.159*	0.046*	1.000					
(5) Liquidity	0.022*	0.283*	–0.327*	–0.201*	1.000				
(6) Leverage	–0.057*	0.000	0.129*	0.025*	–0.135*	1.000			
(7) MTB	0.118*	0.036*	0.006	–0.001	0.112*	0.470*	1.000		
(8) ROA	0.216*	–0.087*	0.339*	–0.003	–0.199*	0.046*	0.025*	1.000	
(9) Tangibility	0.009*	–0.164*	0.229*	0.598*	–0.432*	0.105*	–0.062*	0.083*	1.000

Notes: Panel A of this table reports the sample section process and Panel B shows descriptive statistics for variables used in this paper. Panel C shows the correlation matrix among the main variables used in this paper. * denotes statistical significance at the 5% level. All continuous variables are winsorized at the 1% and 99% levels, and definitions are provided in Appendix A1.

Table 2: Multivariate baseline OLS regression

	(1)	(2)	(3)	(4)	(5)
	Clustering By firm	Fama-MacBeth	Prais-Winsten	Firm fixed effects	Economic impact
Dependent Variable	DD	DD	DD	DD	
Bog Index	–0.052***	–0.039***	–0.014***	–0.040***	–0.572
	(–16.695)	(–11.436)	(–10.110)	(–12.490)	
Size	0.533***	0.482***	0.465***	0.489***	1.657
	(39.083)	(18.588)	(55.126)	(34.411)	
CAPEX	3.307***	1.443*	2.714***	0.987***	0.162
	(11.904)	(1.805)	(18.419)	(3.034)	
Liquidity	1.858***	1.980***	1.147***	1.625***	0.535
	(17.798)	(11.986)	(21.459)	(17.914)	
Leverage	–0.165***	–0.207***	–0.048***	–0.188***	–0.141
	(–23.829)	(–16.313)	(–17.188)	(–24.161)	
MTB	0.051***	0.064***	0.017***	0.059***	0.131
	(22.988)	(15.038)	(21.222)	(23.142)	
ROA	0.336***	0.860***	0.077***	0.382***	0.051
	(9.149)	(4.692)	(6.407)	(9.642)	
Tangibility	–0.595***	–0.118	–0.911***	–0.313**	0.177
	(–4.021)	(–0.520)	(–12.099)	(–2.381)	
Constant	5.029***	4.352***	2.412***	4.387***	
	(18.508)	(12.398)	(19.648)	(16.341)	
Observations	69,768	69,768	69,768	69,768	
R-squared	0.417	0.241	0.169	0.328	
Industry effects	Yes	Yes	Yes	No	
Year effects	Yes	No	Yes	Yes	
Cluster by firm	Yes	No	Yes	Yes	

Notes: This table shows the association between readability and default risk by employing alternative econometric methodologies. The dependent variable is *DD*, an inverse measure of default risk. The key independent variable is *Bog Index*, an inverse measure of readability. Definitions of all variables are reported in Appendix A1. Robust standard errors adjusted for clustering by firm are reported in parentheses. The superscript asterisks ***, **, and * denote two–tailed statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 3: 2SLS-Instrumental variable regression

	(1)	(2)
	1st stage regression	2nd stage regression
Dependent Variable	DD	DD
Ind_Avg_Bog	0.786***	
	(100.991)	
Bog Index		–0.074***
		(–18.085)
Size	0.655***	0.450***
	(23.397)	(37.609)
CAPEX	–0.843**	4.156***
	(–2.124)	(25.236)
Liquidity	–0.968***	1.829***
	(–6.971)	(31.729)
Leverage	0.018**	–0.081***
	(2.248)	(–24.494)
MTB	–0.004	0.028***
	(–1.608)	(27.892)
ROA	–0.325***	0.139***
	(–9.330)	(9.576)
Tangibility	–2.268***	–0.935***
	(–10.649)	(–10.533)
Constant	14.365***	7.721***
	(21.888)	(23.623)
Observations	69,768	69,768
R- squared	0.433	0.578
Industry effects	Yes	Yes
Year effects	Yes	Yes
Cluster by firm	Yes	Yes
F-statistics (P-value)	0.000	
Instrument Validity Tests for IV regression		
(i) F-test for excluded instrument in first stage		
Sanderson-Windmeijer F- test	450.80	
(ii) Under-identification test		
Anderson canon. Corr. LM statistic	59.37	
(iii) Weak identification test		
Cragg-Donald Wald F statistic	44.04	
Stock-Yogo weak ID test		
10% max IV size	16.38	
15% max IV size	13.45	
20% max IV size	9.34	
25% max IV size	7.23	

Notes: This table shows the effect of annual report readability on default risk using two-stage least square regression methods Our 1st stage regression results presented in Column (1) reveal the impact of the industry average of the Bog index (*Ind_Avg_Bog*) on *DD*, and 2nd stage regression results are presented in Column (2), showing the impact of the *Bog Index* on *DD*. The dependent variable is *DD*, an inverse measure of default risk. The key independent variable is the *Bog Index*, an inverse measure of readability. Definitions of all variables are reported in Appendix A1. Robust standard errors adjusted for clustering by firm are reported in parentheses. The superscript asterisks ***, **, and * denote two-tailed statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 4: BOG index and default risk: PSM analysis

Panel A: Pre-match propensity score regression and post-match diagnostic regression

	(1)	(2)
	Pre-Match	Post-Match
Dependent Variable	BOGDUMMY	
Size	0.123***	–0.085
	(30.033)	(–0.945)
CAPEX	–0.855***	0.702
	(–5.121)	(0.154)
Liquidity	1.935***	–1.752
	(48.012)	(–0.788)
Leverage	0.019***	–0.010
	(5.095)	(–0.634)
MTB	–0.004***	0.001
	(–3.560)	(0.359)
ROA	–0.341***	0.038
	(–17.232)	(1.290)
Tangibility	–0.524***	0.312
	(–11.229)	(0.465)
Constant	–1.086***	0.836
	(–35.878)	(0.877)
Observation	69,768	35,612
Pseudo R^2	0.050	0.022

Panel B: *Post-match d*ifference in firm characteristics

Variable	Treated	Control	Difference	t-test
Size	5.812	5.776	0.036	0.070
CAPEX	0.046	0.047	–0.001	–0.950
Liquidity	0.272	0.274	–0.002	–0.820
Leverage	0.632	0.602	0.030	1.610*
MTB	3.240	3.179	0.061	0.930
ROA	–0.150	–0.155	0.005	0.930
Tangibility	0.219	0.220	–0.001	–0.150

Panel C: Mean comparison test of Bog Index

Variable	Treated	Control	Difference	t-test
Bog Index	3.989	4.276	–0.287	12.039*

Panel D: Post-match regression analysis

Dependent Variable	Distance to default		
	(1)	(2)	(3)
	Neighbouring	Radius	Kernel
Bog Index	–0.014***	–0.014***	0.013***
	(–8.594)	(–7.510)	(7.307)
Size	0.548***	0.447***	0.447***
	(84.941)	(73.128)	(73.229)
CAPEX	0.055	0.341	0.337
	(0.227)	(1.325)	(1.311)

Liquidity	1.652***	1.349***	1.354***
	(27.458)	(25.479)	(25.609)
Leverage	–0.158***	–0.191***	–0.191***
	(–26.611)	(–38.606)	(–38.662)
MTB	0.045***	0.057***	0.058***
	(23.758)	(35.929)	(35.969)
ROA	0.606***	0.517***	0.517***
	(27.378)	(25.674)	(25.721)
Tangibility	–0.108	0.089	0.094
	(–1.492)	(1.269)	(1.352)
Constant	1.608***	–0.159	–0.121
	(11.524)	(–1.008)	(–0.774)
Observations	35612	34613	34513
R-squared	0.402	0.193	0.201
Industry effects	Yes	Yes	Yes
Year effects	Yes	Yes	Yes
Cluster by firm	Yes	Yes	Yes

Notes: This table reports the results of the propensity score matching procedure to investigate the effects of the Bog index on default risk. Panel A reports the parameter estimates from the logit model used to estimate propensity scores. The dependent variable BOGDUMMY in Columns (1) and (2) of Panel A is an indicator variable set to one if the firm has *Bog Index* above the median in a given year, zero otherwise. Panel A reports the pre-match propensity score regression and post-match diagnostic regression. Panel B reports the univariate comparisons of firm characteristics between treatment and control firms and the corresponding t statistics. Panel C reports the mean comparison test of *Bog Index* between treatment and control of group. Panel D reports multivariate results relating to default risk and the Bog index. Columns (1) – (3) of Panel C are estimated based on the neighboring matching method, radius matching method, and Kernel Matching methods, respectively. The dependent variable in Panel D is *DD*, an inverse measure of default risk. The key independent variable is the *Bog Index*, an inverse measure of readability. Definitions of all variables are reported in Appendix A1. Robust standard errors adjusted for clustering by firm are reported in parentheses. The superscript asterisks ***, **, and * denote two-tailed statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 5: Entropy balancing

Panel A: Before Balancing

	Treatment			Control		
	Mean	Variance	Skewness	Mean	Variance	Skewness
Size	5.805	4.875	0.21	5.742	4.709	0.108
CAPEX	0.046	0.003	3.149	0.058	0.004	2.573
Liquidity	0.273	0.078	1.005	0.165	0.039	1.744
Leverage	0.63	6.424	2.259	0.675	5.734	2.562
MTB	3.23	71.08	0.508	2.932	53.52	1.203
ROA	–0.161	0.407	–10.5	–0.062	0.344	–14.44
Tangibility	0.219	0.049	1.371	0.286	0.058	0.965

Panel B: After balancing

	Treatment			Control		
	Mean	Variance	Skewness	Mean	Variance	Skewness
Size	5.805	4.875	0.21	5.805	4.875	0.21
CAPEX	0.046	0.003	3.149	0.046	0.003	3.149
Liquidity	0.273	0.078	1.005	0.273	0.078	1.005
Leverage	0.63	6.424	2.259	0.63	6.424	2.259
MTB	3.23	71.08	0.508	3.23	71.08	0.508
ROA	–0.161	0.407	–10.5	–0.161	0.407	–10.5
Tangibility	0.219	0.049	1.371	0.219	0.049	1.371

Panel C: Regression after balancing

	Entropy balancing
Dependent Variable	DD
Bog Index	–0.026***
	(–5.353)
Size	0.433***
	(5.615)
CAPEX	4.034***
	(10.218)
Liquidity	1.538***
	(6.227)
Leverage	–0.060***
	(–7.557)
MTB	0.018***
	(6.273)
ROA	0.330***
	(6.056)
Tangibility	–0.729*
	(–1.716)

Constant	3.788***
	(6.733)
Observations	69,768
R-squared	0.711
Industry effects	Yes
Year effects	Yes
Cluster by firm	Yes

Notes: Panels A and B of this table report the means, variance, and skewness for the covariates of the treatment groups (firms with readability score above the median value) and the control groups, before and after balancing, respectively, as required for the entropy balancing estimates of Equation (3). Required balancing is achieved by using Hainmueller's Stata code, given that there is no mean, variance, or skewness difference between the treatment and control groups after the balancing. Panel C presents the regression based on the entropy balancing method. The dependent variable is DD, an inverse measure of default risk, while the key independent variable is the Bog Index, an inverse measure of readability. Definitions of all variables are provided in Appendix A1. t-statistics are reported in parentheses. The superscript asterisks ***, **, and * denote two-tailed statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 6: Channel mechanism: Information asymmetry and agency cost

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	DD	Spread	DD	Amihud	DD	Agency Cost	DD
Bog Index	–0.052***	0.001***	–0.030***	0.002***	–0.032***	0.007***	–0.048***
	(–16.695)	(15.801)	(–10.426)	(2.789)	(–19.711)	(4.908)	(–12.786)
Spread			–31.540***				
			(–31.821)				
Amihud					–0.156***		
					(–3.999)		
Agency cost							–0.395***
							(–11.028)
Size	0.533***	–0.009***	0.261***	–0.011***	0.405***	0.098***	0.494***
	(39.083)	(–45.617)	(15.953)	(–11.354)	(30.719)	(17.110)	(30.469)
CAPEX	3.307***	0.014**	3.523***	–0.123***	3.840***	–0.271*	1.106***
	(11.904)	(2.528)	(11.884)	(–3.765)	(19.544)	(–1.678)	(2.690)
Liquidity	1.858***	0.016***	2.177***	–0.050***	1.847***	0.970***	1.472***
	(17.798)	(9.950)	(20.914)	(–6.359)	(28.326)	(18.160)	(8.955)
Leverage	–0.165***	0.001***	–0.175***	0.001	–0.098***	0.026***	–0.299***
	(–23.829)	(4.623)	(–21.040)	(1.229)	(–24.120)	(7.390)	(–19.739)
MTB	0.051***	–0.000***	0.052***	–0.000**	0.032***	–0.007***	0.088***
	(22.988)	(–4.217)	(20.022)	(–2.143)	(26.691)	(–6.877)	(18.642)
ROA	0.336***	–0.004***	0.464***	–0.009***	0.229***	–1.735***	7.506***
	(9.149)	(–3.092)	(9.426)	(–2.606)	(11.138)	(–14.905)	(18.879)
Tangibility	–0.595***	0.006**	–0.378**	0.008	–0.587***	0.561***	0.157
	(–4.021)	(2.348)	(–2.520)	(0.866)	(–5.739)	(5.845)	(0.807)
Constant	5.029***	0.101***	8.415***	0.118***	4.408***	–1.584***	4.883***
	(18.508)	(26.046)	(29.079)	(8.054)	(26.703)	(–12.969)	(14.189)
Observations	69,768	69,768	69,768	69,768	69,768	69,768	69,768
R-squared	0.417	0.483	0.549	0.028	0.694	0.567	0.471
Industry effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster by firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sobel test (p-value)			<0.01		<0.01		<0.01

Notes: This table presents a potential channel—information asymmetry—through which the Bog Index may affect default risk. Column (1) reports the baseline regression results from Table 2 for comparison. *Spread* and *Amihud* serve as proxies for information asymmetry. *Agency Cost* is the measure of agency cost. Columns (2), (4), and (6) display the regression results of the impact of the Bog Index on *Spread*, *Amihud, and Agency Cost,* respectively. Columns (3), (5), and (7) present the regression results of the impact of the Bog Index on default risk (*DD*) after controlling for *Spread*, *Amihud, and Agency Cost,* respectively, along with other control variables. The dependent variable is *DD*, an inverse measure of default risk, while the key independent variable is the Bog Index, an inverse measure of readability. Definitions of all variables are provided in Appendix A1. t-statistics are reported in parentheses. The superscript asterisks ***, **, and * denote two-tailed statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 7: Channel mechanism: Risk taking strategy

	(1)	(2)	(3)	(4)	(5)
Dependent Variable	DD	R&D	DD	ROA Volatility	DD
Bog Index	–0.052***	0.002***	–0.028***	0.001**	–0.032***
	(–16.695)	(8.360)	(–13.680)	(2.262)	(–22.042)
R&D			–0.691***		
			(–9.052)		
ROA Volatility					–0.081***
					(–3.463)
Size	0.533***	–0.011***	0.510***	–0.011***	0.425***
	(39.083)	(–11.522)	(34.792)	(–9.717)	(36.702)
CAPEX	3.307***	0.108***	4.343***	0.157***	4.283***
	(11.904)	(3.297)	(18.520)	(4.345)	(26.196)
Liquidity	1.858***	0.116***	1.773***	0.128***	1.884***
	(17.798)	(14.432)	(27.019)	(8.885)	(32.808)
Leverage	–0.165***	–0.001***	–0.083***	–0.002***	–0.082***
	(–23.829)	(–2.972)	(–19.036)	(–3.525)	(–25.023)
MTB	0.051***	0.000	0.025***	0.001***	0.028***
	(22.988)	(0.846)	(22.129)	(3.039)	(28.301)
ROA	0.336***	–0.145***	0.192***	–0.447***	0.124***
	(9.149)	(–17.284)	(10.101)	(–43.329)	(7.395)
Tangibility	–0.595***	0.027**	–0.842***	–0.120***	–0.781***
	(–4.021)	(2.233)	(–6.823)	(–9.778)	(–8.869)
Constant	5.029***	–0.048**	3.054***	0.139***	4.130***
	(18.508)	(–2.451)	(16.753)	(5.023)	(28.436)
Observations	69,768	69,768	69,768	69,768	69,768
R-squared	0.417	0.573	0.712	0.459	0.698
Industry effects	Yes	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes	Yes
Cluster by firm	Yes	Yes	Yes	Yes	Yes
Sobel test (p-value)			<0.01		<0.01

Notes: This table presents a potential channel—risk-taking strategy—through which the Bog Index may affect default risk. Column (1) reports the baseline regression results from Table 2 for comparison. *R&D* and *ROA Volatility* serve as proxies for risk-taking strategy. Columns (2) and (4) display the regression results of the impact of the Bog Index on *R&D* and *ROA Volatility,* respectively. Columns (3) and (5) present the regression results of the impact of the Bog Index on default risk (DD) after controlling for *R&D* and *ROA Volatility*, respectively, along with other control variables. The dependent variable is *DD*, an inverse measure of default risk, while the key independent variable is the *Bog Index*, an inverse measure of readability. Definitions of all variables are provided in Appendix A1. t-statistics are reported in parentheses. The superscript asterisks ***, **, and * denote two-tailed statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 8: Channel mechanism: Cost of capital

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	DD	WACC	DD	r_E	DD	r_D	DD
Bog Index	–0.052***	–0.035***	–0.035***	0.050*	–0.042***	0.003**	–0.039***
	(–16.695)	(–3.241)	(–20.909)	(2.936)	(–16.841)	(2.629)	(–20.849)
WACC			–1.737***				
			(–7.828)				
Cost of Equity					–9.867***		
					(–21.987)		
Cost of Debt							–0.179***
							(–7.327)
Size	0.533***	–0.003***	0.426***	–0.001***	0.532***	–0.018***	0.433***
	(39.083)	(–13.763)	(32.944)	(–7.213)	(39.320)	(–14.889)	(33.497)
CAPEX	3.307***	0.015***	4.481***	0.003	3.568***	0.138***	4.440***
	(11.904)	(3.044)	(25.245)	(0.908)	(12.967)	(3.397)	(25.019)
Liquidity	1.858***	–0.010***	1.745***	–0.018***	1.667***	0.150***	1.738***
	(17.798)	(–5.363)	(26.023)	(–17.564)	(15.973)	(8.552)	(25.901)
Leverage	–0.165***	–0.000***	–0.079***	0.000***	–0.157***	–0.004***	–0.078***
	(–23.829)	(–3.595)	(–23.185)	(5.400)	(–23.214)	(–5.971)	(–23.030)
MTB	0.051***	–0.000	0.027***	–0.000***	0.046***	0.001***	0.027***
	(22.988)	(–0.926)	(24.758)	(–11.486)	(22.416)	(3.252)	(24.759)
ROA	0.336***	–0.009***	0.179***	0.001***	0.337***	–0.054***	0.189***
	(9.149)	(–3.896)	(10.818)	(3.122)	(9.170)	(–6.553)	(11.382)
Tangibility	–0.595***	–0.013***	–0.857***	–0.005***	–0.811***	–0.097***	–0.811***
	(–4.021)	(–6.086)	(–8.927)	(–3.174)	(–5.484)	(–6.212)	(–8.454)
Constant	5.029***	0.120***	4.221***	0.099***	6.028***	0.249***	3.989***
	(18.508)	(26.697)	(26.181)	(40.457)	(21.571)	(7.679)	(25.003)
Observations	69,768	69,768	69,768	69,768	69,768	69,768	69,768
R-squared	0.417	0.483	0.549	0.028	0.694	0.567	0.471
Industry effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster by firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sobel test (p-value)			<0.01		<0.01		<0.01

Notes: This table presents a potential channel—cost of capital—through which the Bog Index may affect default risk. Column (1) reports the baseline regression results from Table 2 for comparison. *WACC,* Cost of equity (r_E), and cost of debt (r_D) serve as proxies for the cost of capital. Columns (2), (4), and (6) display the regression results of the impact of the Bog Index on *WACC,* r_E, and r_D, respectively. Columns (3), (5), and (7) present the regression results of the impact of the Bog Index on default risk (*DD*) after controlling for *WACC,* r_E, and r_D, respectively, along with other control variables. The dependent variable is *DD*, an inverse measure of default risk, while the key independent variable is the Bog Index, an inverse measure of readability. Definitions of all variables are provided in Appendix A1. t-statistics are reported in parentheses. The superscript asterisks ***, **, and * denote two-tailed statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 9: Channel mechanism: Financial constraints

	(1)	(2)	(3)	(4)	(5)
Dependent Variable	DD	KZ Index	DD	WW Index	DD
Bog Index	–0.052***	0.184***	–0.035***	0.001***	–0.034***
	(–16.695)	(2.688)	(–22.193)	(9.655)	(–22.500)
KZ Index			–0.001***		
			(–4.699)		
WW Index					–1.548***
					(–11.239)
Size	3.307***	1.265***	0.427***	–0.048***	0.361***
	(11.904)	(5.777)	(36.804)	(–225.744)	(26.774)
CAPEX	1.858***	22.658***	4.381***	–0.036***	4.389***
	(17.798)	(5.421)	(26.868)	(–4.589)	(26.473)
Liquidity	–0.165***	–74.374***	1.802***	–0.008***	1.820***
	(–23.829)	(–21.436)	(31.129)	(–3.894)	(31.205)
Leverage	0.051***	–0.194	–0.082***	0.001***	–0.083***
	(22.988)	(–1.631)	(–25.100)	(6.658)	(–25.048)
MTB	0.336***	0.045	0.029***	–0.000***	0.029***
	(9.149)	(0.724)	(28.334)	(–6.992)	(27.780)
ROA	–0.595***	–9.458***	0.164***	–0.067***	0.100***
	(–4.021)	(–7.940)	(11.043)	(–47.689)	(5.297)
Tangibility	5.029***	34.948***	–0.732***	0.005*	–0.718***
	(18.508)	(14.129)	(–8.319)	(1.861)	(–8.033)
Constant	–0.052***	–32.209***	4.117***	–0.038***	4.137***
	(–16.695)	(–5.001)	(28.346)	(–8.926)	(28.211)
Observations	69,768	69,768	69,768	69,768	69,768
R-squared	0.417	0.158	0.700	0.856	0.704
Industry effects	Yes	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes	Yes
Cluster by firm	Yes	Yes	Yes	Yes	Yes
Sobel test (p-value)			<0.01		<0.01

Notes: This table presents potential channel — financial constraints —through which the Bog Index may affect default risk. Column (1) reports the baseline regression results from Table 2 for comparison. *KZ Index* and *WW Index* serve as proxies for risk-taking strategy. Columns (2) and (4) display the regression results of the impact of the Bog Index on *KZ Index* and *WW Index,* respectively. Columns (3) and (5) present the regression results of the impact of the Bog Index on default risk (*DD*) after controlling for *KZ Index* and *WW Index*, respectively, along with other control variables. The dependent variable is *DD*, an inverse measure of default risk, while the key independent variable is the Bog Index, an inverse measure of readability. Definitions of all variables are provided in Appendix A1. t-statistics are reported in parentheses. The superscript asterisks ***, **, and * denote two-tailed statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 10: Role of monitoring

	(1)	(2)	(3)	(4)
	Institutional Ownership		Analyst	
	Below-median	Above-median	Below-median	Above-median
Dependent Variable	DD	DD	DD	DD
Bog Index	–0.057***	–0.040***	–0.053***	–0.043***
	(–16.344)	(–8.211)	(–15.572)	(–8.426)
Size	0.533***	0.437***	0.551***	0.400***
	(33.809)	(17.683)	(39.811)	(11.552)
CAPEX	3.429***	2.898***	3.215***	3.348***
	(10.550)	(7.137)	(10.464)	(6.214)
Liquidity	1.749***	2.033***	1.889***	1.721***
	(13.945)	(14.368)	(16.565)	(10.165)
Leverage	–0.178***	–0.108***	–0.165***	–0.149***
	(–22.223)	(–10.527)	(–21.813)	(–12.126)
MTB	0.052***	0.041***	0.051***	0.044***
	(20.388)	(12.274)	(21.045)	(11.156)
ROA	0.355***	0.323***	0.259***	0.780***
	(8.860)	(5.632)	(8.533)	(4.326)
Tangibility	–0.640***	–0.326	–0.506***	–0.826***
	(–3.710)	(–1.606)	(–3.206)	(–3.261)
Constant	5.576***	4.322***	4.948***	5.304***
	(17.820)	(10.442)	(16.641)	(12.379)
Observations	51,255	18,513	50,696	19,072
R-squared	0.426	0.430	0.451	0.354
Chi-square	11.13***		18.93***	
P-value	0.000		0.000	
Industry effects	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes
Cluster by firm	Yes	Yes	Yes	Yes

Notes: This table examines the moderating role of monitoring in the association between the *Bog Index* and default risk (*DD*), with monitoring proxied by institutional ownership (*Institutional Ownership*) and number of analyst coverage (*Analyst*). Columns (1) and (2) display the impact of the *Bog Index* on *DD* for firms with below-median and above-median institutional ownership, respectively. Columns (3) and (4) present the regression results for the impact of the Bog Index on *DD* for firms with below-median and above-median analyst coverage, respectively. The dependent variable is *DD*, an inverse measure of default risk, while the key independent variable is the *Bog Index*, an inverse measure of readability. Definitions of all variables are provided in Appendix A1. t-statistics are reported in parentheses. The superscript asterisks ***, **, and * denote two-tailed statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 11: Role of organizational capital and managerial ability

	(1)	(2)	(3)	(4)
	Organizational Capital		Managerial Ability	
	Below-median	Above-median	Below-median	Above-median
Dependent Variable	DD	DD	DD	DD
Bog Index	–0.060***	–0.027***	–0.051***	–0.048***
	(–14.994)	(–5.006)	(–13.223)	(–13.280)
Size	0.583***	0.398***	0.555***	0.505***
	(32.607)	(15.133)	(34.595)	(30.351)
CAPEX	4.097***	2.371***	2.013***	3.846***
	(9.802)	(7.255)	(5.082)	(11.821)
Liquidity	2.046***	1.973***	1.764***	1.823***
	(15.244)	(14.074)	(14.274)	(14.279)
Leverage	–0.165***	–0.141***	–0.197***	–0.140***
	(–19.276)	(–13.433)	(–18.074)	(–18.583)
MTB	0.055***	0.033***	0.049***	0.053***
	(18.088)	(12.275)	(17.538)	(18.959)
ROA	0.624***	0.190***	0.247***	0.482***
	(5.889)	(6.626)	(6.223)	(7.862)
Tangibility	–0.525***	–0.326	–0.013	–0.797***
	(–2.724)	(–1.549)	(–0.064)	(–5.021)
Constant	5.251***	3.536***	4.966***	4.756***
	(15.187)	(7.509)	(14.584)	(15.217)
Observations	44,788	19,286	37,184	32,584
R-squared	0.464	0.388	0.430	0.428
Chi-square	15.89***		13.25***	
P-value	0.000		0.000	
Industry effects	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes
Cluster by firm	Yes	Yes	Yes	Yes

Notes: This table presents the moderating role of organizational capital and managerial ability on the relationship between the *Bog Index* and default risk (*DD*). Columns (1) and (2) display the impact of the *Bog Index* on *DD* for firms with below-median and above-median organizational capital, respectively. Columns (3) and (4) present the regression results for the impact of the *Bog Index* on *DD* for firms with below-median and above-median managerial ability, respectively. The dependent variable is *DD*, an inverse measure of default risk, while the key independent variable is the Bog Index, an inverse measure of readability. Definitions of all variables are provided in Appendix A1. t-statistics are reported in parentheses. The superscript asterisks ***, **, and * denote two-tailed statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 12: Role of business strategy and debt concentration

Panel A: Business strategy

	(1)		(2)
	Defender		Prospector
Dependent Variable	DD		DD
Bog Index	–0.032***		–0.054**
	(–6.973)		(–2.028)
Size	0.407***		0.624***
	(9.205)		(7.860)
CAPEX	2.925***		–4.710
	(3.025)		(–0.711)
Liquidity	2.332***		2.677***
	(6.530)		(4.708)
Leverage	–0.096***		–0.244***
	(–5.671)		(–3.649)
MTB	0.048***		0.038***
	(5.909)		(2.992)
ROA	0.619**		2.700***
	(2.141)		(5.748)
Tangibility	–0.413		5.601
	(–1.229)		(1.595)
Constant	5.550***		1.706
	(7.676)		(0.954)
Observations	5,537		846
R–squared	0.425		0.510
Chi-square		101.01**	
P-value		0.000	
Industry effects	Yes		Yes
Year effects	Yes		Yes
Cluster by firm	Yes		Yes

Panel B: Debt concentration

	(1)	(2)
	High debt concentration	Low debt concentration
Dependent Variable	DD	DD
Bog Index	–0.049***	–0.058***
	(–14.476)	(–10.999)
Size	0.531***	0.636***
	(37.240)	(29.462)
CAPEX	3.167***	4.594***
	(11.710)	(7.024)
Liquidity	1.901***	1.296***
	(17.617)	(5.581)
Leverage	–0.168***	–0.143***
	(–21.082)	(–12.732)
MTB	0.045***	0.053***
	(20.180)	(12.197)
ROA	0.301***	0.421***
	(8.153)	(3.937)

Tangibility	–0.442***		–1.106***
	(–2.826)		(–4.465)
Constant	4.755***		4.949***
	(16.168)		(10.690)
Observations	50,501		16,766
R-squared	0.390		0.513
Chi-square		31.43***	
P-value		0.000	
Industry effects	Yes		Yes
Year effects	Yes		Yes
Cluster by firm	Yes		Yes

Notes: This table presents the moderating role of Business strategy and Debt concentration on the association between the *Bog Index* and default risk (*DD*). Panel A displays the impact of business strategy on the association between *Bog Index* and *DD*, where *Prospector* refers to the firms that have strategy score between 24 and 3, and *Defender* refers to the firms that have strategy score between 6 and 12. Panel B displays the impact of debt concentration on the association between *the Bog Index and DD*, where high debt concentration refers to firms with above-median bank debt, and low debt concentration refers to firms with below-median bank debt concentration. The dependent variable is *DD*, an inverse measure of default risk, while the key independent variable is the *Bog Index*, an inverse measure of readability. Definitions of all variables are provided in Appendix A1. t-statistics are reported in parentheses. The superscript asterisks ***, **, and * denote two-tailed statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 13: Role of *SEC 1998 Plain English Rule*

	(1)	(2)
	1994–1997	1999–2002
	DD	DD
Bog Index	–0.057***	–0.033***
	(–11.624)	(–9.533)
Size	0.673***	0.379***
	(30.453)	(27.806)
CAPEX	2.604***	2.943***
	(5.721)	(8.941)
Liquidity	2.576***	1.660***
	(12.708)	(15.858)
Leverage	–0.222***	–0.131***
	(–12.049)	(–15.112)
MTB	0.055***	0.039***
	(10.229)	(15.204)
ROA	0.143**	0.190***
	(2.045)	(6.734)
Tangibility	0.137	–0.096
	(0.527)	(–0.627)
Constant	4.523***	3.108***
	(11.461)	(10.883)
Observations	8,119	15,294
R-squared	0.472	0.384
Chi-square	65.32***	
P-value	0.000	
Industry effects	Yes	Yes
Year effects	Yes	Yes
Cluster by firm	Yes	Yes

Notes: This table presents the role of the event of the SEC's 1998 Plain English Rule in the association between the Bog Index and default risk (DD), using the period 1994–1997 as the pre-event period and 1998–2002 as the post-event period. The dependent variable is *DD*, an inverse measure of default risk, while the key independent variable is the *Bog Index*, an inverse measure of readability. Definitions of all variables are provided in Appendix A1. t-statistics are reported in parentheses. The superscript asterisks ***, **, and * denote two-tailed statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 14: Role of firm performance and overall firm risk

	(1)	(2)	(3)	(4)
	Tobin's Q		Overall Risk	
	Above-median	Below-median	Above-median	Below-median
	DD	DD	DD	DD
Bog Index	–0.053***	–0.038***	–0.035***	–0.056***
	(–12.618)	(–12.586)	(–13.257)	(–8.445)
Size	0.598***	0.374***	0.428***	0.484***
	(35.417)	(29.504)	(36.269)	(15.734)
CAPEX	0.865**	2.411***	3.850***	3.624***
	(2.295)	(8.080)	(14.968)	(4.026)
Liquidity	0.955***	2.340***	1.847***	2.845***
	(7.660)	(18.111)	(21.095)	(9.188)
Leverage	–0.119***	–0.124***	–0.125***	–0.373***
	(–13.121)	(–16.194)	(–20.980)	(–11.040)
MTB	0.027***	0.063***	0.040***	0.106***
	(13.606)	(12.993)	(21.551)	(11.063)
ROA	0.227***	0.963***	0.296***	2.058**
	(6.627)	(9.246)	(9.488)	(2.127)
Tangibility	–0.080	–0.456***	–0.894***	–0.317
	(–0.367)	(–3.170)	(–6.647)	(–0.870)
Constant	5.827***	3.960***	3.866***	6.586***
	(15.426)	(15.284)	(16.944)	(11.056)
Observations	34,758	35,010	52,897	16,871
R-squared	0.455	0.384	0.359	0.504
Chi-square	8.15***		4.64**	
p-value	0.000		0.031	
Industry effects	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes
Cluster by firm	Yes	Yes	Yes	Yes

Notes: This table presents the role of firm performance (*Tobin's Q*) and overall firm risk (*Overall Risk*) in the association between the Bog Index and default risk (DD). Columns (1) and (2) display the impact of the *Bog Index* on *DD* for firms with above-median and below-median Tobin's Q, respectively. Columns (3) and (4) display the impact of the *Bog Index* on *DD* for firms with above-median and below-median Overall Risk, respectively. The dependent variable is *DD*, an inverse measure of default risk, while the key independent variable is the *Bog Index*, an inverse measure of readability. Definitions of all variables are provided in Appendix A1. t-statistics are reported in parentheses. The superscript asterisks ***, **, and * denote two-tailed statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 15: The association between readability and actual bankruptcy

	(1)	(2)	(3)
Dependent Variable	Bankruptcy	Bankruptcy	Bankruptcy
Bog Index	-0.005**		
	(-2.147)		
Bog Index 3yr		-0.006**	
		(-2.142)	
Bog Index 5yr			-0.006*
			(-2.485)
Size	-0.023**	-0.009	-0.008
	(-2.221)	(-0.732)	(-0.580)
CAPEX	1.959***	1.691***	1.780***
	(6.300)	(4.333)	(3.758)
Liquidity	-0.828***	-1.109***	-1.096***
	(-9.785)	(-10.374)	(-8.371)
Leverage	0.053***	0.054***	0.055***
	(7.814)	(6.809)	(5.903)
MTB	-0.016***	-0.020***	-0.022***
	(-6.379)	(-6.384)	(-5.751)
ROA	-0.502***	-0.842***	-0.898***
	(-11.234)	(-13.122)	(-11.110)
Tangibility	0.777***	0.885***	0.976***
	(7.933)	(7.725)	(7.267)
Constant	0.397**	0.372	0.262
	(2.251)	(1.634)	(0.926)
Observations	15,285	11,285	8,247
Pseudo R^2	0.040	0.052	0.054
Industry effects	Yes	Yes	Yes
Year effects	Yes	Yes	Yes
Cluster by firm	Yes	Yes	Yes

Notes: This table shows the association between readability and bankruptcy. The dependent variable is *Bankruptcy,* which is a dummy variable that equals 1 if a firm files for bankruptcy under Chapter 7 or Chapter 11 in a given year, and 0 otherwise. The key independent variable is *Bog Index*, an inverse measure of readability. Definitions of all variables are reported in Appendix A1. Robust standard errors adjusted for clustering by firm are reported in parentheses. The superscript asterisks ***, **, and * denote two-tailed statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 16: Alternative proxy of readability and default risk

Panel A: Alternative proxy of readability

	(1)	(2)	(3)	(4)
Dependent Variable	DD	DD	DD	DD
Complexity	–12.512***			
	(–26.084)			
ARC		–0.006***		
		(–18.282)		
Gross File Size			–0.126***	
			(–8.758)	
Net File Size				–0.334***
				(–18.168)
Size	0.660***	0.771***	0.522***	0.541***
	(38.711)	(31.278)	(36.669)	(38.212)
CAPEX	4.101***	3.573***	3.447***	3.509***
	(10.030)	(5.905)	(11.762)	(12.076)
Liquidity	1.727***	1.062***	1.745***	1.748***
	(13.128)	(5.634)	(15.664)	(15.894)
Leverage	–0.166***	–0.168***	–0.169***	–0.168***
	(–18.160)	(–13.840)	(–23.194)	(–23.223)
MTB	0.054***	0.054***	0.052***	0.052***
	(17.131)	(14.025)	(22.440)	(22.434)
ROA	0.583***	0.642***	0.351***	0.331***
	(7.238)	(5.514)	(8.791)	(8.500)
Tangibility	–1.109***	–1.266***	–0.465***	–0.490***
	(–5.664)	(–5.160)	(–3.013)	(–3.205)
Constant	2.952***	1.909***	2.494***	4.767***
	(20.020)	(10.958)	(11.283)	(19.826)
Observations	42,524	23,095	69,768	69,768
R-squared	0.444	0.408	0.411	0.417
Industry effects	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes
Cluster by firm	Yes	Yes	Yes	Yes

Panel B: Alternative proxy of default risk

	(1)	(2)
Dependent Variable	PD	Altman Z
Bog Index	0.001***	–0.067***
	(10.864)	(–6.038)
Size	–0.001***	–0.164***
	(– 11.472)	(–2.966)
CAPEX	–0.043***	14.351***
	(–10.399)	(9.583)
Liquidity	–0.012***	13.770***
	(–16.539)	(21.698)
Leverage	0.000***	–0.545***

	(3.762)	(–16.624)
MTB	–0.000***	0.274***
	(–10.601)	(16.068)
ROA	–0.008***	13.478***
	(–13.259)	(22.098)
Tangibility	0.007***	–3.553***
	(4.602)	(–5.903)
Constant	–0.004**	9.123***
	(–2.052)	(10.033)
Observations	69,768	69,768
R-squared	0.109	0.430
Industry effects	Yes	Yes
Year effects	Yes	Yes
Cluster by firm	Yes	Yes

Notes: This table presents the regression results using alternative proxies for readability and default. In Panel A, the proxies for readability include *Complexity*, Annual Report Complexity (*ARC*), *Gross File Size*, and *Net File Size*. In Panel B, the proxies for default risk include Probability of Default (*PD*) and *Altman Z-score*. The dependent variable is *DD*, an inverse measure of default risk. Definitions of all variables are provided in Appendix A1. t-statistics are reported in parentheses. The superscript asterisks ***, **, and * denote two-tailed statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 17: Change regression and lagged independent variable

Panel A: Change regression		Panel B: Lagged independent variable	
	(1)		(2)
Dependent Variable	Δ DD		DD
Δ Bog Index	–0.009***	Bog Index$_{t-1}$	–0.053***
	(–5.236)		(–15.835)
Δ Size	–0.028	Size$_{t-1}$	0.514***
	(– 1.310)		(34.466)
Δ CAPEX	1.654***	CAPEX$_{t-1}$	1.371***
	(11.601)		(4.681)
Δ Liquidity	0.732***	Liquidity$_{t-1}$	1.861***
	(10.986)		(16.358)
Δ Leverage	–0.021***	Leverage$_{t-1}$	–0.180***
	(–8.233)		(–22.546)
Δ MTB	0.009***	MTB$_{t-1}$	0.048***
	(10.708)		(20.089)
Δ ROA	0.094***	ROA$_{t-1}$	0.796***
	(7.105)		(10.021)
Δ Tangibility	–0.736***	Tangibility$_{t-1}$	–0.254
	(–5.738)		(–1.573)
Constant	–0.064***	Constant	5.297***
	(–15.873)		(17.935)
Observations	61,005	Observations	61,005
R-squared	0.24	R-squared	0.437
Industry effects	Yes	Industry effects	Yes
Year effects	Yes	Year effects	Yes
Cluster by firm	Yes	Cluster by firm	Yes

Notes: This table presents additional robustness tests to assess the relationship between the *Bog Index* and default risk (*DD*). Panel A shows the regression results using change regression. Panel B shows the regression results using lead-lag analysis, where t – 1 refers to one-year lag of independent variables. The dependent variable is *DD*, an inverse measure of default risk, while the key independent variable is the Bog Index, an inverse measure of readability. Δ indicates the change of a respective variable compared to the previous year. Definitions of all variables are provided in Appendix A1. t-statistics are reported in parentheses. The superscript asterisks ***, **, and * denote two-tailed statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 18: Inclusion of additional controls

	(1)	(2)
Dependent Variable	DD	DD
Bog Index	–0.040***	–0.034***
	(–11.380)	(–15.472)
Board_Size	–0.034**	
	(–2.322)	
Board_Independence	1.834***	
	(10.718)	
Director_Qualification	–0.053	
	(–0.938)	
Gender_Diversity	1.168***	
	(4.570)	
Director_Network	–0.000	
	(–1.378)	
BusSegment		0.104**
		(2.309)
GeoSegment		–0.009
		(–0.339)
Opaqueness		–0.012***
		(-6.940)
Special_Items		0.001
		(1.52)
Age		0.621***
		(16.954)
BIG4		0.237***
		(7.128)
Loss		–1.307***
		(–42.411)
Constant	8.393***	2.638***
	(16.407)	(11.428)
Observations	40,463	23,590
R-squared	0.406	0.471
Other controls	Yes	Yes
Industry effects	Yes	Yes
Year effects	Yes	Yes
Cluster by firm	Yes	Yes

Notes: This table presents the association between the *Bog Index* and default risk (*DD*), incorporating additional control variables beyond those included in the baseline regressions. The dependent variable is *DD*, an inverse measure of default risk, while the key independent variable is the *Bog Index*, an inverse measure of readability. Other controls include *Size*, *CAPEX*, *Liquidity*, *Leverage*, *MTB*, *ROA*, and *Tangibility*. Definitions of all variables are provided in Appendix A1. t-statistics are reported in parentheses. The superscript asterisks ***, **, and * denote two-tailed statistical significance at the 1%, 5%, and 10% levels, respectively.