

Bankruptcy Prediction Using Machine Learning Models: Firm Complexity Information Perspective

Tsung-Kang Chen *

Email: vocterchen@nycu.edu.tw

Department of Management Science, National Yang Ming Chiao Tung University
Center for Research in Econometric Theory and Applications, National Taiwan
University

Yijie Tseng

E-mail: 081647@mail.fju.edu.tw

Department of Accounting, Fu Jen Catholic University

Yun Hao

Email: peterhao.mg12@nycu.edu.tw

Department of Management Science, National Yang Ming Chiao Tung University

Chih-Hsuan Chen

Department of Management Science, National Yang Ming Chiao Tung University,

E-mail: sam0912149291@gmail.com

* Corresponding author. Address: No. 1001, Dasyue Rd., East Dist., Hsinchu City 300, Taiwan (R.O.C.). Tel: +886-3-5712121 ext 57124. The author gratefully acknowledge the financial support from the Center for Research in Econometric Theory and Applications (Grant no. 114L900203; 114L900201) from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

Bankruptcy Prediction Using Machine Learning Models: Firm Complexity Information Perspective

Abstract

Unlike previous bankruptcy prediction studies that primarily use financial information and annual report readability or tone measures as model input variables, we investigate whether additionally incorporating firm complexity information enhances bankruptcy prediction performance, using U.S. firm data from 1993 to 2022 and machine learning models such as Random Forest, XGBoost, CatBoost, and LightGBM. Based on the 53 complexity-related terms compiled by Loughran and McDonald (2023), we leverage GPT-4o to categorize them into five groups: financial instruments and accounting operations (FINACC), corporate financial distress (FD), business changes and organizational structure (BC), international and global operations (IGO), and debt and contractual terms (DC). Empirical results show that incorporating corporate complexity variables substantially improves bankruptcy prediction, especially for short-term forecasts. For one-year-ahead forecasts, the XGBoost model reaches an AUC of 0.9062 and an F1-score of 0.1529 while lowering both Type I and Type II errors, and it delivers the largest AUC and F1-score improvements across imbalance-handling methods. Moreover, we also find that BC- and DC-related firm complexity as well as the volatility of DC-related firm complexity, exhibit substantial feature importance. These results highlight that firm complexity provides distinct and incremental information for bankruptcy prediction beyond financial variables and traditional readability- or tone-based textual measures, with particularly strong effectiveness in identifying short-term financial distress.

Keywords: Bankruptcy prediction, Firm complexity information, Machine learning, Volatility of firm complexity, Bankruptcy risk term structure

1. Introduction

In recent years, corporate bankruptcy risk has intensified worldwide. Global markets have been disrupted by major events such as the subprime mortgage crisis, the European debt crisis, the U.S.–China trade war, and the COVID-19 pandemic, all of which have caused substantial economic turbulence. Many firms have gone bankrupt as their financial conditions deteriorated. According to the 2024 Global Insolvency Report released by Allianz, following a sharp increase in bankruptcies in 2023, global corporate insolvencies are expected to rise by an additional 11 percent in 2024 and continue rising through 2025 and 2026. In addition, S&P Global Market Intelligence further reports that in the first half of 2024, the United States recorded 346 corporate bankruptcy filings, marking the highest level for the same period in the past 13 years and surpassing the levels seen during the COVID-19 period. The World Economic Forum similarly identified debt crises and asset bubbles as major short- to medium-term global risks in 2022. As economic uncertainty continues to rise and bankruptcy filings reach new highs, corporate bankruptcy risk has become a central concern for financial institutions. For banks, accurately predicting corporate bankruptcy not only helps mitigate loan default losses but also improves the efficiency of lending decisions.

In prior bankruptcy prediction studies, most models relied on financial accounting variables and traditional statistical techniques. For example, Altman (1968) and Ohlson (1980) employed discriminant analysis and logistic regression, respectively, to predict corporate failure. However, Begley et al. (1996) argue that such statistical models are often constrained by the number of variables and by distributional assumptions, making them less effective in capturing the nonlinear and multifaceted nature of bankruptcy risk. In recent years, researchers have increasingly adopted machine learning models for bankruptcy prediction. Barboza et al. (2017), for instance, used eleven financial variables and compared the predictive performance of various machine learning methods, including logistic regression (LR), random forests (RF),

ensemble approaches such as bagging and boosting, and support vector machines (SVM). Their results show that machine learning models outperform traditional statistical models in predicting bankruptcy.

Beyond financial variables, a growing body of literature incorporates textual features from annual reports or earnings call transcripts into bankruptcy prediction models. Studies such as Mai et al. (2019), Kim and Yoon (2021), Chen et al. (2023), Chen et al. (2025), and Hao et al. (2025a) demonstrate that incorporating textual characteristics can enhance bankruptcy predictive accuracy. Mai et al. (2019) and Kim and Yoon (2021) find that word frequencies and sentiment in the Management’s Discussion and Analysis (MD&A) section of annual reports enhance bankruptcy prediction performance.¹ Chen et al. (2023) show that text-based communication values derived from annual reports help reduce the misclassification of non-bankrupt firms as bankrupt, particularly for short-term predictions, while Chen et al. (2025) find that the uncertainty in communication value decreases the misclassification of bankrupt firms as non-bankrupt.² Hao et al. (2025a) further demonstrate that text-based communication values extracted from earnings call transcripts improve the effectiveness of bankruptcy prediction models, and that their incremental contribution is more pronounced for longer-term predictions, consistent with the inherently forward-looking nature of earnings call transcripts.

However, most textual features used in prior research—such as readability metrics or sentiment measures—struggle to capture information that is directly value-relevant within financial disclosures, raising concerns about their explanatory power. To address this gap, Loughran and McDonald (2023) introduce the concept of firm complexity, defined as any

¹ Mai et al. (2019) use deep learning models to analyze word frequencies in the MD&A section of annual reports for bankruptcy prediction, while Kim and Yoon (2021) apply a BERT-based sentiment analysis model to predict corporate bankruptcy.

² Chen et al. (2023) use annual report readability and tone as input variables and employ four machine learning models, including random forest, SVM, logistic regression, and XGBoost, to predict bankruptcy. Chen et al. (2025) incorporate the uncertainty of text-based communication value from annual reports and combine it with random forest and XGBoost models for bankruptcy prediction.

aspect of a firm that makes its valuation more difficult or ambiguous and encompassing dimensions such as organizational structure, product logistics, financial reporting, information dissemination, or financial engineering. Using expert-curated terms combined with a Lasso selection process, they construct a lexicon of 53 informative complexity-related terms specifically designed for 10-K filings. This lexicon effectively captures the degree of complexity and underlying risk signals embedded in financial text. Specifically, the measure captures firm attributes reflected in 53 complexity-related terms—such as derivatives and restructuring provisions—that naturally increase the difficulty and ambiguity for investors to accurately assess the firm’s true value.

Moreover, unlike readability measures or semantic textual features that may be influenced by managerial strategic disclosure, firm complexity derives from underlying firm attributes reflected in the 53 complexity-related terms and is therefore less susceptible to managerial manipulation. These attributes inherently increase the difficulty and uncertainty investors face when assessing a firm’s true asset value. Loughran and McDonald (2023) document that firm complexity is significantly and positively associated with audit fees, unexpected earnings, and stock return volatility, consistent with the notion that complexity heightens valuation ambiguity. Given that information uncertainty affects external investors’ perceptions of asset value distributions (Lambert et al., 2007; Chen et al., 2025), firm complexity may reflect underlying firm attributes that elevate information uncertainty, thereby increasing the difficulty of accurately valuing the firm and reducing the communication value of information that is relevant to firm value. Hence, higher firm complexity may indicate that firms operate in environments characterized by greater uncertainty and risk exposure, thereby increasing their credit risk according to structural-form credit risk models (e.g. Duffie and Lando, 2001)³.

³ Hao et al. (2025b) apply the firm complexity measure developed by Loughran and McDonald (2023) to examine its relationship with corporate credit risk, measured by bond spreads, and find a positive association between the

Furthermore, building on the 53 complexity-related terms identified by Loughran and McDonald (2023), this study further classifies these terms into five dimensions—financial instruments and accounting operations (FINACC), corporate financial distress (FD), business changes and organizational structure (BC), international and global operations (IGO), and debt and contractual terms (DC)—and examines which category exhibits the strongest explanatory power for bankruptcy prediction.⁴ In addition, this study incorporates the volatility of each category as model inputs to capture incremental information related to bankruptcy risk, recognizing that firm complexity may serve as an important channel through which external investors encounter greater valuation difficulty. Overall, the objective of this research is to investigate whether incorporating firm complexity variables into bankruptcy prediction models offers incremental predictive power beyond traditional financial variables (e.g. Barboza et al., 2017) and annual report text-based communicative value (hereafter denoted TCV) features (e.g. Chen et al., 2023).

This study examines U.S. firms from 1993 to 2022 and employs four machine learning models—Random Forest, XGBoost, CatBoost, and LightGBM—combined with three data imbalance techniques, including BalancedBagging, EasyEnsemble, and Random Under Sampling, to construct bankruptcy prediction models. Using 2016 as the initial split year for training and testing sets, the study adopts a rolling-adjustment procedure to update the training and testing windows annually. In addition, to understand the role of firm complexity on bankruptcy risk term structure, the prediction periods are divided into one-year, two-year,

two. This is because a large presence of complexity-related terms in annual reports makes it harder for external investors to accurately assess a firm's true financial condition and asset value distribution, leading to incomplete accounting information and affecting credit risk (Duffie and Lando, 2001).

⁴ This study leverages GPT-4o to categorize terms based on their semantic meanings, guided by the conceptual domains outlined by Loughran and McDonald (2023), who highlight that firm complexity can arise from organizational structure, product logistics, financial reporting, information dissemination, and financial engineering. By mapping the 53 complexity-related terms onto these underlying sources of complexity, this study establishes an extended and structured framework for analyzing how different facets of firm complexity are predictable for bankruptcy risk.

three-year, and four-year horizons. To ensure the robustness of the empirical results, the study further analyzes the average performance of models run under ten random states.

This study adopts the 11 financial variables from Barboza et al. (2017) and the 11 annual report TCV variables from Chen et al. (2023) as the benchmark model, and subsequently incorporates 12 firm complexity variables to assess their incremental predictive contribution. The empirical results show that incorporating firm complexity variables significantly enhances the predictive performance of all models, particularly for short-term forecasts. For example, under the XGBoost model combined with Balanced Bagging, the one-year-ahead bankruptcy prediction achieves an AUC of 0.9017 and an F1-score of 0.1529, alongside improvements in both Type I and Type II errors.⁵ However, as the prediction horizon extends, the incremental contribution of firm complexity variables declines, indicating that such variables primarily capture managers' strategic narrative reporting intended to obscure value-relevant information over shorter horizons.

To further identify which dimensions of firm complexity contribute most to predictive performance, the study conducts feature engineering analyses. The results indicate that firm complexity arising from business changes and organizational structure, debt and contractual terms, the volatility of debt- and contract-related complexity, the volatility of international and global operations complexity, and the volatility of total complexity all exhibit substantial feature importance. Their predictive contributions exceed those of most of the 11 annual report TCV variables proposed by Chen et al. (2023). These findings indicate that firm complexity variables offer stronger predictive power and convey more distinct information for bankruptcy prediction than traditional annual report TCV measures, particularly in the dimensions of business and structural changes, debt and contractual arrangements, and firm complexity

⁵ Under this model specification, the Type I error decreases from 0.2789 to 0.2552, while the Type II error decreases from 0.1459 to 0.1337.

volatility. The results are consistent with the view that firm complexity stems from underlying firm attributes that make valuation more difficult or ambiguous, rather than from managerial manipulation (e.g. TCV).

The potential contributions of this study are as follows. First, unlike Chen et al. (2023), who rely on annual report TCV variables (e.g. readability and tone), this study incorporates the firm complexity measures developed by Loughran and McDonald (2023)—which capture underlying firm attributes that render valuation more difficult or ambiguous—into bankruptcy prediction models. In addition, this study further classifies the 53 firm complexity keywords into five distinct dimensions (FINACC, FD, BC, IGO, and DC) to provide a more nuanced understanding of the informational role of firm complexity. Second, the study identifies which dimensions of firm complexity are more relevant for bankruptcy prediction. Specifically, BC complexity, DC complexity, the volatility of DC complexity, the volatility of IGO complexity, and the volatility of total complexity exhibit substantial predictive importance. These findings offer meaningful economic insights into how firm complexity information contributes to bankruptcy prediction. Third, in contrast to Chen et al. (2023), whose enhancements mainly reduce Type II error, this study shows that firm complexity information effectively reduces both Type I error (i.e. misclassifying bankrupt firms as non-bankrupt) and Type II error (i.e. misclassifying non-bankrupt firms as bankrupt). Fourth, the study finds that firm complexity variables have stronger predictive power for shorter-term bankruptcy prediction. This result not only aligns with Chen et al. (2023), who document stronger short-term predictive power for the text-based communicative value of annual reports, but also accords with the theoretical implications of Duffie and Lando's (2001) credit risk model based on incomplete accounting information.⁶ Taken together, these findings demonstrate that firm complexity information

⁶ Duffie and Lando's (2001) incomplete accounting information-based credit risk model posits that incomplete accounting information primarily affects short-term credit risk.

provides novel and meaningful predictive features for bankruptcy risk assessment in banking practice and can help lenders reduce potential losses arising from borrower defaults. Finally, this study provides theoretical foundations and economic implications for understanding the impact of firm complexity information on corporate bankruptcy risk and its bankruptcy term structure.

The structure of this paper is organized as follows. Section 2 presents the literature review. Section 3 introduces the measures and research methodology. Section 4 presents empirical results. Section 5 further concludes the study.

2. Literature Review

The literature review of this study is organized into three parts. The first section discusses prior research on bankruptcy prediction, with particular emphasis on the input variables used in earlier studies. The second section reviews the literature on annual report readability and firm complexity. The final section examines studies that evaluate the predictive performance of machine learning models in forecasting bankruptcy events.

2.1. Bankruptcy Predictions Related Researches.

2.1.1. Researches Related to Financial Accounting Variables

Early research on bankruptcy prediction primarily relied on financial ratio variables combined with traditional statistical methods as the core modeling approach, as exemplified by Beaver (1966), Altman (1968), and Ohlson (1980).⁷ However, statistical models are often constrained by the number of variables and distributional assumptions, making them less

⁷ Beaver (1966) utilized univariate analysis alongside 14 financial ratios to assess whether a firm was experiencing financial distress. Building on this foundation, Altman (1968) applied Multiple Discriminant Analysis (MDA) and combined five key financial ratios—representing liquidity, profitability, productivity, financial leverage, and asset turnover—to construct the Z-Score model, which later became a cornerstone in bankruptcy prediction research. Ohlson (1980) further advanced the literature by employing logistic regression with nine financial accounting variables to estimate the likelihood of corporate bankruptcy.

effective in capturing complex, nonlinear characteristics inherent in corporate bankruptcy risk (Begley et al., 1996). In recent years, research has increasingly adopted machine learning models to enhance bankruptcy prediction. For instance, Barboza et al. (2017) employ several machine learning techniques—including Bagging, Boosting, and Random Forest—using eleven variables that capture firms’ financial and operational performance (Altman, 1968; Carton and Hofer, 2006). Their findings show that machine learning methods deliver superior predictive accuracy compared with traditional statistical models.

2.1.2. Researches Related to Textual Features from Annual Reports

In recent years, bankruptcy prediction research has expanded beyond traditional financial and accounting variables to incorporate textual features extracted from annual reports. Textual disclosures in annual reports convey firms’ narratives regarding their operating conditions and strategic objectives, providing valuable information to external stakeholders and facilitating the assessment of bankruptcy risk.

In the related literature, Mai et al. (2019) apply deep learning models to analyze word frequencies in MD&A disclosures for bankruptcy prediction. Kim and Yoon (2021) conduct sentiment analysis using a BERT-based model to forecast bankruptcy. Wang et al. (2018) propose the IST-RS (Incorporating Sentiment and Textual Information into a Random Subspace) framework, which integrates sentiment and textual information with traditional financial features, significantly enhancing the prediction of financial distress. Chen et al. (2023) utilize readability and tone features from annual reports as input variables and adopt four machine learning models—Random Forest, SVM, logistic regression, and XGBoost—to predict bankruptcy. Chen et al. (2025) employ the uncertainty in text-based communication value (U_TCV) derived from annual reports and integrate it with Random Forest and XGBoost models for bankruptcy prediction. Taken together, prior studies primarily rely on readability, semantic attributes, word frequency, and sentiment analysis as textual features in bankruptcy

prediction models. However, these traditional textual characteristics may be influenced by managerial strategic disclosure, which reduces their ability to capture the full range of value-relevant firm attributes and underlying economic fundamentals, thereby limiting their explanatory power in bankruptcy prediction.

2.2. Readability and Complexity of Annual Reports

2.2.1. Readability of Annual Reports

Readability refers to the extent to which a text can be easily read and understood, and has long been used as an indicator across various domains (Dale and Chall, 1948; Klare, 1963). Higher readability implies that a document is easier to comprehend and conveys greater communicative value (Seebeck and Kaya, 2023; Chen et al., 2023). In the context of annual reports, readability also reflects the transparency of corporate information disclosure. Because annual reports serve as a primary source of information for external stakeholders to understand a firm's operations and evaluate its risks, the readability of these reports is critical. When annual reports are written in a difficult or obscure style, the communication value of disclosures decreases, information transmission becomes slower and less efficient, and information asymmetry is exacerbated.

When financial disclosures are difficult to understand, information users must spend more time and effort extracting relevant information. Bloomfield (2002) further argues that when market participants struggle to comprehend public information, managers may exploit the situation by using obfuscating language to conceal poor performance. Ertugrul et al. (2017) document that firms with a higher proportion of uncertain and weak modal words in their 10-K filings face greater future stock price crash risk and more restrictive loan contract terms. This suggests that low readability and ambiguous tone may reflect managerial information hoarding behavior, making it harder for external stakeholders to grasp the firm's true condition, ultimately increasing the cost of debt financing and credit risk. In a related study, Chen and

Tseng (2021) find that the readability of financial statement footnotes is negatively associated with corporate bond yield spreads. Taken together, readability serves not only as an indicator of disclosure quality but may also have implications for a firm's bankruptcy risk (Chen et al., 2023; Chen et al., 2025).

2.2.2. Firm Complexity

Prior studies commonly use document size, sentence length, difficult word counts, and other readability-based features to proxy for textual complexity. However, as financial disclosures become increasingly intricate, scholars question whether traditional readability measures adequately capture the true complexity of a firm. Loughran and McDonald (2014) note that metrics such as the Fog Index, which rely heavily on syllable counts, often misclassify common business terms (e.g., financial, agreement, operating) as complex despite being easily understood by investors. As a result, readability measures may lack value relevance in financial contexts. To overcome these limitations, Loughran and McDonald (2023) propose the concept of firm complexity—defined as characteristics that make firm valuation more difficult or ambiguous—which offers a more meaningful, value-relevant text-based indicator than traditional readability metrics.

Firm complexity is measured using a dictionary of 53 complexity-related terms extracted from 10-K filings. The construction of this dictionary involves three steps. First, accounting experts identify an initial set of 374 terms that auditors typically pay close attention to during the audit process and that are believed to increase the difficulty of predicting future firm performance. This ensures the initial list captures terms relevant to both investors and auditors. Second, the list is narrowed to 198 terms that appear in at least 5% of 10-K filings. Third, a Lasso (Least Absolute Selection and Shrinkage Operator) regression model is employed to select the 53 most relevant terms from the 198 candidates. Firm complexity is then measured as the total frequency of these 53 terms in the annual report scaled by the total number of words

in the 10-K filing.

A higher frequency of complexity-related terms indicates that information users face greater difficulty interpreting the firm's true financial condition, thereby contributing to incomplete accounting information. Consistent with the theoretical framework of Duffie and Lando's (2001) incomplete accounting information-based credit risk model, firms with more incomplete disclosures exhibit higher credit risk. For example, Hao et al. (2025b) apply the Loughran and McDonald (2023) complexity measure to examine its relation to corporate credit risk, proxied by bond yield spreads, and find that firms with higher complexity face significantly higher credit risk. Because firm complexity obscures value-relevant information and influences external investors' perceptions of asset value, firms whose annual reports contain more pronounced complexity-related terms make it more difficult for investors to understand the true distribution of asset values, financial condition, and debt-related information. Therefore, firm complexity reflects the extent to which a firm's value and risk assessments are obscured, thereby capturing the assessed variance in firm value (Lambert et al., 2007) and the associated credit risk (Duffie and Lando, 2001).

Based on the above discussions, unlike readability measures or semantic textual features that may be influenced by managerial strategic disclosure, firm complexity (Loughran and McDonald, 2023) originates from underlying firm attributes reflected in the 53 complexity-related terms and is therefore less susceptible to managerial manipulation. By capturing characteristics that render firm valuation more difficult or ambiguous, firm complexity provides distinct and value-relevant information that differs from traditional annual report TCV variables, which may be subject to managerial manipulation aimed at influencing the speed or clarity of information transmission.

2.3. Machine Learning Models in Bankruptcy Prediction Research

Early studies on bankruptcy risk prediction largely relied on statistical models and focused

on the relationship between financial ratios and corporate default. Although statistical models offer interpretability, they are constrained by assumptions regarding variable distributions and the number of variables that can be included. These assumptions are often violated in real-world datasets, limiting the models' ability to capture the multifaceted determinants of bankruptcy risk. As a result, recent research increasingly turns to machine learning models for bankruptcy prediction. Machine learning techniques are capable of handling high-dimensional and large-scale data and are less restricted by distributional assumptions. For example, Nanni and Lumini (2009) show that machine learning models outperform traditional statistical approaches in classification and prediction tasks. Barboza et al. (2017) employ Random Forest and Boosting algorithms to predict corporate bankruptcy and document superior predictive performance relative to statistical models. Jabeur et al. (2021) utilize the CatBoost model to forecast bankruptcy among French firms and highlight its stable predictive accuracy. Wang et al. (2022) apply the Light Gradient Boosting Machine (LGBM) to predict firm financing risk and compare its performance with KNN, decision trees, and Random Forest, finding that LGBM consistently outperforms alternative models across evaluation metrics.

Despite their strong predictive power and flexibility, machine learning models are often criticized for their “black box” nature and limited interpretability. To address this concern, this study employs Random Forest and XGBoost to generate feature importance rankings, providing both interpretability and economic insights. In summary, this study adopts four machine learning models—Random Forest, XGBoost, CatBoost, and LGBM—to perform bankruptcy prediction and further examines whether firm complexity can offer incremental predictive power beyond traditional financial and annual report TCV predictors.

3. Research Methodology

3.1. Research Process

The research process of this study consists of seven steps. First, bankruptcy and non-bankruptcy firms from 1993 to 2022 are collected from the Compustat database, where bankruptcy status is identified based on a DLRSN code equal to 02 or 03. Second, financial data for the sampled firms are gathered from Compustat to construct the 11 financial variables developed by Barboza et al. (2017). Third, annual report TCV variables used in Chen et al. (2023) are obtained from the WRDS SEC Analytics Suite. Fourth, all 10-K filings are retrieved from the SEC EDGAR system, and firm complexity measures (Loughran and McDonald, 2023; Hao et al., 2025b) are computed through text mining techniques. Fifth, data preprocessing—including merging datasets, removing missing values, standardizing variables, splitting datasets, and addressing data imbalance—is conducted. Sixth, four machine learning models are constructed and trained using annually partitioned datasets to generate year-by-year predictions. Finally, predictive performance is evaluated using various performance metrics.

3.2. Data Sources and Variables

3.2.1. Dependent Variable

The sample comprises all U.S. bankruptcy and non-bankruptcy firms from 1993 to 2022, with financial data obtained from the Compustat database. Bankruptcy status is determined using Compustat's DLRSN code: firms with DLRSN codes of 02 or 03 are classified as bankrupt, and all others as non-bankrupt. Bankruptcy firms are coded as 1 (Bankruptcy=1), whereas non-bankruptcy firms are coded as 0 (Bankruptcy=0). After data preprocessing, the final sample consists of 64,259 non-bankruptcy observations and 1,983 bankruptcy observations.

3.2.2. Financial Information Variables

Following Barboza et al. (2017), this study employs 11 financial ratios as financial input variables. These variables include the five financial ratios originally introduced in Altman's

(1968) Z-score model and the six performance-variation measures developed by Carton and Hofer (2006). The variables are as follows: net working capital to total assets (NWC_TA), retained earnings to total assets (RE_TA), EBIT to total assets (EBIT_TA), equity market value to total debt (EMV_Debt), sales to total assets (Sales_TA), EBIT to sales (EBIT_Sales), asset growth (TA_Growth), sales growth (Sales_Growth), employee growth (EMP_Growth), change in ROE (ROE_Chg), and change in market-to-book ratio (PB_Chg). Detailed definitions of these 11 financial variables (FIN) are provided in Table 1.

[Insert Table 1 here]

3.2.3. *Firm Complexity Variables*

Following Loughran and McDonald (2023), firm complexity is defined as any aspect of a firm that makes its valuation more difficult or ambiguous, encompassing contexts such as organizational structure, product logistics, financial reporting, information dissemination, and financial engineering. Loughran and McDonald (2023) develop a 53-word complexity lexicon that captures a subset of terms to which auditors typically pay close attention during the audit process and that are identified through penalized regression as the most reliable predictors of firm complexity. These terms are also associated with audit fees, unexpected earnings, and post-filing stock return volatility.

This study further groups the 53 complexity-related terms into five categories: (1) financial instruments and accounting operations (FINACC; 17 terms), (2) corporate financial distress (FD; 5 terms), (3) business changes and organizational structure (BC; 13 terms), (4) international and global operations (IGO; 6 terms), and (5) debt and contractual terms (DC; 12 terms). The categorization primarily follows the semantic meaning of each term and draws on the conceptual domains identified by Loughran and McDonald (2023), who emphasize that firm complexity may arise from organizational structure, product logistics, financial reporting,

information dissemination, and financial engineering.⁸ By mapping the 53 complexity-related terms onto these underlying sources of firm complexity, this study develops an extended and structured framework that reflects the multifaceted firm complexity arising from a firm's financial operations, corporate financial distress, organizational structure, international activities, and contractual arrangements, as shown in Table 2. This framework further facilitates an examination of how different facets of firm complexity contribute to the prediction of bankruptcy risk.

[Insert Table 2 here]

The economic meanings of the five dimensions of firm complexity are described as follows. First, financial instruments and accounting operations (FINACC) encompass terms related to derivatives, hedging, securitizations, collateral, accrues and similar concepts, representing the complexity inherent in a firm's financial strategies and accounting practices. Firms may employ these instruments for legitimate purposes such as risk management or capital allocation, but they may also use structured financial products, accounting operations, or risk transfers to obscure underlying risks from external investors. Second, corporate financial distress (FD) includes terms such as bankruptcy, litigation, and infringement, reflecting the operational risks and legal disputes faced by the firm. Frequent disclosure of such terms suggests that the firm may be experiencing financial pressure or operational distress, thereby increasing its likelihood of bankruptcy. Third, business changes and organizational structure (BC) cover terms such as mergers, restructuring, affiliates, entities, and segments, indicating adjustments in the firm's business scope or organizational structure. While organizational changes may arise from routine business expansion, excessively frequent

⁸ In this study, GPT-4o is employed to assist in reclassifying the 53 complexity-related terms according to their semantic meanings and the conceptual domains identified by Loughran and McDonald (2023), who emphasize that firm complexity may arise from organizational structure, product logistics, financial reporting, information dissemination, and financial engineering.

restructuring may signal internal governance instability or unclear strategic direction, contributing to greater firm structure complexity. Fourth, international and global operations (IGO) involve terms related to international activities, global operations, and fund repatriation, reflecting the firm's engagement in multiple foreign markets. Multinational operations introduce additional layers of firm structure complexity, regulatory compliance, foreign exchange exposure, and geopolitical risks, thereby increasing disclosure and managerial challenges. Finally, debt and contractual terms (DC) include convertible instruments, covenants, leases, licensing agreements, and various contractual clauses, representing the complexity of a firm's financing structure and contractual arrangements. Such complexity may arise from regulatory requirements or financing needs; however, an abundance of contracts and intricate provisions may heighten information opacity and hinder external stakeholders' understanding of the firm's debt structure and contractual obligations.

Overall, firms with high levels of complexity produce disclosures that are more difficult for information users to interpret, thereby increasing information asymmetry. Such opacity may arise from organizational changes, business expansion, diversification strategies, multinational operations, regulatory-driven structural design, or the use of structured financial products and accounting techniques for risk management purposes.

With respect to the construction of firm complexity variables, this study follows Hao et al. (2025b) by computing complexity measures for each of the five categories—as well as total firm complexity—using term frequency. In addition, the five-year volatility (standard deviation) of each complexity measure is computed. Accordingly, this study constructs 12 complexity variables: five category-level firm complexity measures, their corresponding five-year volatilities, and the total firm complexity measure along with its five-year volatility. Detailed definitions of these 12 firm complexity variables are provided in Table 3.

[Insert Table 3 here]

3.2.4. Annual Report TCV Variable

With respect to annual report TCV variables, this study follows Chen et al. (2023) and employs their proposed 11 text-based communicative value (TCV) measures extracted from firms' 10-K filings, which may reflect managerial attempts to influence the speed and clarity of information transmission. These variables include file size, the Flesch Reading Ease Index, Coleman–Liau Readability Index, average words per paragraph, proportion of negative words, proportion of positive words, proportion of weak modal words, proportion of litigation-related terms, proportion of uncertainty words, and proportion of strong modal words. Detailed definitions of these 11 annual report TCV variables are presented in Table 4.

[Insert Table 4 here]

3.3. Data Preprocessing

The data preprocessing procedures in this study include merging the above variables, handling missing values, normalizing the data, partitioning the dataset into annual training and testing sets, and addressing data imbalance. Due to the imbalanced nature of the bankruptcy dataset, data preprocessing can make subsequent machine learning model predictions more robust. This section provides a detailed description of each step and the methods employed.

3.3.1 Missing Value Treatment

Because certain accounting items in the Compustat database contain missing values, the 11 financial ratio variables of Barboza et al. (2017) also inevitably exhibit missing observations. Most machine learning algorithms require complete input data; although various imputation methods are available in practice, this study prioritizes preserving the integrity of firms' financial information and textual disclosures. Considering that the sample size remains sufficiently large even after removing observations with missing values, this study excludes all firm-year observations with any missing value in the input variables to ensure that all variables

in the final dataset are complete. After removing missing observations, the final dataset consists of 66,242 firm-year observations.

3.3.2 Data Standardization

The variables used in this study include the financial variables of Barboza et al. (2017), the firm complexity variables, and the annual report TCV variables from Chen et al. (2023). Descriptive statistics for the original data are presented in Panels A, B, and C of Table 5. Because these variables differ substantially in scale, such magnitude disparities could distort the training and predictive performance of machine learning models. To mitigate this issue, this study employs Z-score standardization, transforming each input variable into a standardized normal distribution with a mean of 0 and a standard deviation of 1.

[Insert Table 5 here]

3.3.3 Data Splitting

This study partitions the sample into training and testing sets based on calendar years. Because firms' financial structures change annually, using sequential yearly data helps the model better capture temporal patterns. The full sample period spans 1993 to 2022, and bankruptcy predictions are performed for one-year-ahead, two-year-ahead, three-year-ahead, and four-year-ahead horizons.

The year 2016 is selected as the initial cutoff point, and a rolling-adjustment approach is adopted to update the training and testing windows annually. The training set always consists of all years prior to the first year of the corresponding testing period. For example, for the one-year-ahead prediction, training data for 1993–2016 (1993–2017; 1993–2018; 1993–2019; 1993–2020; 1993–2021) are used to predict bankruptcy in 2017 (2018; 2019; 2020; 2021; 2022). Performance metrics from all rolling predictions are averaged to obtain the final one-year-ahead predictive performance. For the two-year-ahead prediction, training data from 1993–2016 (1993–2017; 1993–2018; 1993–2019; 1993–2020) are used to predict bankruptcy

over the 2017–2018 (2018–2019; 2019–2020; 2020–2021; 2021–2022) periods. The same rolling procedure is applied to three-year-ahead and four-year-ahead predictions. This rolling-split design allows for a comprehensive assessment of predictive performance across different forecasting horizons.

3.3.4 Handling Data Imbalance

In bankruptcy prediction, the sample distribution is typically highly imbalanced. That is, the number of bankrupt firms is substantially smaller than that of non-bankrupt firms. When the data are imbalanced, machine learning models tend to favor the majority class during training, which reduces their ability to correctly identify the minority class and ultimately weakens overall predictive performance.

The annual sample distribution of bankrupt and non-bankrupt firms used in this study is presented in Table 6. As shown, the imbalance is substantial: the dataset contains 64,259 non-bankrupt observations and only 1,983 bankrupt observations, with bankrupt firms representing approximately 3% of the total sample. Therefore, data imbalance must be addressed. Following Chen et al. (2023) and Chen et al. (2025), this study applies three imbalance-handling techniques: Balanced Bagging, EasyEnsemble, and Random Under Sampling (RUS). Balanced Bagging combines the Bagging framework with an undersampling strategy, where each base classifier is trained on a randomly undersampled balanced subset of the data, and final predictions are obtained through majority voting. EasyEnsemble constructs multiple undersampled subsets, trains a boosting classifier on each subset, and then aggregates the models, thereby enhancing the model’s ability to recognize minority-class observations. RUS directly reduces the size of the majority class by randomly removing observations until the class sizes are approximately equal, achieving balanced class representation.⁹

⁹ This study also experimented with other imbalance-handling techniques, such as oversampling approaches. However, the results indicate that undersampling-based methods yield superior predictive performance. Therefore, only the results based on undersampling approaches are reported in this study. Moreover, unlike oversampling approaches that require the use of synthetic data, undersampling approaches retain only actual firm observations.

[Insert Table 6 here]

3.4. Machine Learning Models

Following Chen et al. (2023), Barboza et al. (2017), Jabeur et al. (2021), and Wang et al. (2022), this study adopts four machine learning models for bankruptcy prediction: Random Forest, XGBoost (Extreme Gradient Boosting), CatBoost (Categorical Boosting), and LGBM (Light Gradient Boosting Machine). The primary objective is to examine whether incorporating firm complexity variables into these models provides incremental predictive power beyond the benchmark models. In addition, this study compares the predictive performance across the four machine learning methods. The conceptual foundations of each model are introduced below.

3.4.1. Random Forest (RF)

Random Forest, proposed by Breiman (2001), is an ensemble-based machine learning algorithm grounded in the Bagging framework. It constructs multiple decision trees using bootstrap samples generated through sampling with replacement and aggregates their predictions to reduce overfitting and enhance model stability. For classification tasks, the final prediction is determined by majority voting across all trees. Empirically, Barboza et al. (2017) compare various machine learning models for bankruptcy prediction and find that Random Forest outperforms other approaches, highlighting its strong potential in predicting corporate bankruptcy.

3.4.2. XGBoost (Extreme Gradient Boosting)

XGBoost (Chen and Guestrin, 2016) is a boosting-based model built upon the Gradient Boosted Decision Tree (GBDT) framework. It constructs an ensemble of weak learners by iteratively optimizing the loss function. During training, XGBoost increases the weights of

This enhances the credibility of the analysis when examining the association between firm complexity and bankruptcy.

misclassified observations, allowing subsequent learners to focus more on difficult-to-classify samples, thereby improving predictive performance.

3.4.3. CatBoost (Categorical Boosting)

CatBoost, developed by Yandex in 2017, is also based on gradient boosting but introduces significant improvements in handling categorical features, missing values, and overfitting issues. Jabeur et al. (2021) apply CatBoost to bankruptcy prediction for French firms and document that CatBoost exhibits stable predictive performance and superior ability to identify potential bankruptcies compared with traditional models.

3.4.4. LGBM (Light Gradient Boosting Machine)

Developed by Microsoft in 2016, LGBM is another gradient boosting model that accelerates training and reduces memory usage through techniques such as leaf-wise tree growth and histogram-based algorithms. Wang et al. (2022) apply LGBM to predict corporate financing risk and compare its performance with KNN, decision trees, and Random Forest, finding that LGBM delivers strong predictive accuracy.

3.5. Performance Evaluation Metrics

This study uses the confusion matrix to compare model predictions with actual bankruptcy outcomes and assess model performance, as shown in Table 7. In the matrix, the positive class represents bankrupt firms (Bankruptcy=1), and the negative class represents non-bankrupt firms (Bankruptcy=0). The confusion matrix consists of four outcomes: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). TP indicates the number of firms that actually went bankrupt and were correctly predicted as bankrupt; FN represents bankrupt firms incorrectly predicted as non-bankrupt; FP refers to non-bankrupt firms incorrectly predicted as bankrupt; and TN indicates non-bankrupt firms correctly predicted as such. It is important to note that Type I error is typically considered more severe in bankruptcy

prediction, corresponding to misclassifying an actual bankrupt firm as non-bankrupt (i.e., False Negative). Type II error corresponds to misclassifying a non-bankrupt firm as bankrupt (i.e., False Positive). Specifically, Type I Error is defined as the proportion of bankrupt firms incorrectly classified as non-bankrupt, calculated as $\frac{FN}{TP+FN}$. Conversely, Type II Error represents the proportion of non-bankrupt firms mistakenly classified as bankrupt, computed as $\frac{FP}{TN+FP}$.

[Insert Table 7 here]

Using the confusion matrix, several performance metrics can be computed, including accuracy, precision, recall, F1-score, MCC, and AUC-ROC. Each metric is described in detail below. First, Accuracy is defined as the proportion of correctly predicted observations relative to the total number of observations, as shown in Eq. (1). Second, Precision is defined as the proportion of predicted positive (bankrupt) observations that are correctly classified as positive, while Recall is defined as the proportion of actual positive (bankrupt) observations that are correctly identified, as shown in Eqs. (2) and (3), respectively. Third, the F1-score, presented in Eq. (4), is the harmonic mean of Precision and Recall and provides a balanced measure of both. Fourth, the Matthews Correlation Coefficient (MCC), shown in Eq. (5), is a balanced performance metric that incorporates TP, TN, FP, and FN, making it particularly suitable for evaluating highly imbalanced datasets. The MCC ranges from -1 to 1 , where 1 indicates perfect prediction, -1 indicates total misclassification, and 0 represents random prediction. A value closer to 1 implies better model performance. Finally, AUC (Area Under the ROC Curve) represents the area under the Receiver Operating Characteristic (ROC) curve, where the x-axis is the false positive rate and the y-axis is the true positive rate. A higher AUC indicates stronger discrimination between positive and negative classes. AUC is particularly suitable for imbalanced data and provides a robust overall assessment of model performance.

$$Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)} \quad (1)$$

$$Precision = \frac{TP}{(TP+FP)} \quad (2)$$

$$Recall = \frac{TP}{(TP+FN)} \quad (3)$$

$$F1_score = \frac{2 \times Precision \times Recall}{(Precision + Recall)} \quad (4)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (5)$$

Given that bankruptcy prediction is a binary classification problem distinguishing bankrupt from non-bankrupt firms, the dataset exhibits a highly imbalanced structure. Because the majority class (non-bankrupt firms) dominates the sample, machine learning models tend to generate overly optimistic accuracy results by favoring the majority class. Therefore, this study does not rely on accuracy as the primary evaluation metric. Instead, due to the imbalanced nature of the data, greater emphasis is placed on the F1-score and AUC. The F1-score, as the harmonic mean of recall and precision, is well suited for bankruptcy prediction. Furthermore, because the primary concern of this study is whether bankrupt firms can be correctly identified, AUC is also adopted as a key performance metric. Finally, additional indicators—including MCC, Type I error, and Type II error—are incorporated to provide a comprehensive evaluation of each model's performance in predicting bankruptcy.

4. Empirical Analyses

This study examines U.S. bankrupt and non-bankrupt firms from 1993 to 2022 and evaluates whether incorporating firm complexity (Complexity) into bankruptcy prediction models enhances predictive accuracy relative to two benchmark models: (1) the 11 financial variables (FIN) proposed by Barboza et al. (2017) and (2) the combined FIN variables and the 11 text-based communicative value (TCV) variables proposed by Chen et al. (2023). The

analysis investigates whether the inclusion of Complexity variables improves prediction performance beyond these traditional indicators.

In addition, since prior literature consistently finds that Random Forest (RF) and XGBoost deliver superior predictive performance compared with logistic regression (LR), support vector machines (SVM), and other conventional approaches (e.g., Chen et al., 2023; Chen et al., 2025), this study adopts RF and XGBoost as the primary predictive models and additionally incorporates CatBoost and LightGBM (LGBM) for comparison. Meanwhile, three data imbalance techniques—Balanced Bagging, EasyEnsemble, and Random Under Sampling (RUS)—are applied to address severe class imbalance in the bankruptcy dataset. Moreover, bankruptcy predictions are generated for one- to four-year horizons using a rolling-window approach, and all models are trained using ten different random states to ensure robustness. The final results are the averages across the ten iterations.¹⁰

4.1. Prediction Results of Machine Learning Models

This section presents the performance of the four machine learning models under three data imbalance treatment methods for predicting corporate bankruptcy one to four years ahead. The models' predictive accuracy is evaluated and compared based on multiple performance metrics.

4.1.1 Predictive Performance After Adding Firm Complexity (Using FIN as the Benchmark Model)

The analysis first uses the 11 financial variables (FIN) from Barboza et al. (2017) as the benchmark model and then examines whether adding the firm complexity variables

¹⁰ This study sets 2016 as the initial cutoff year and adopts a rolling-adjustment procedure to update the training and testing windows annually. Specifically, the training data span 1993–2016 (1993–2017; 1993–2018; 1993–2019; 1993–2020; 1993–2021), which are used to predict bankruptcy outcomes for 2017 (2018; 2019; 2020; 2021; 2022). For each rolling window, the model is trained using 10 different random states, and the average of the evaluation metrics across the 10 runs is reported as the final one-year-ahead prediction result.

(Complexity) improves bankruptcy prediction. This approach evaluates the incremental contribution and economic significance of firm complexity relative to traditional financial predictors.

The left panel of Table 8 reports the AUC values before and after including Complexity in models constructed under the three imbalance handling methods (Balanced Bagging, EasyEnsemble, and RUS). The results show that adding firm complexity substantially improves model performance, particularly for CatBoost and XGBoost. For example, after incorporating the Complexity variables, the AUC of the XGBoost model combined with Balanced Bagging (EasyEnsemble) increases substantially across all prediction horizons. For the one-year-ahead forecast, the AUC rises from 0.8157 (0.8139) to 0.8609 (0.8664). Similar improvements are observed for longer horizons: the two-year-ahead AUC increases from 0.8134 (0.8079) to 0.8470 (0.8508), the three-year-ahead AUC increases from 0.8075 (0.8026) to 0.8342 (0.8376), and the four-year-ahead AUC increases from 0.8039 (0.7998) to 0.8273 (0.8292). These findings indicate that complexity significantly improves bankruptcy prediction, especially when combined with XGBoost and CatBoost.

Moreover, the improvement in AUC diminishes as the prediction horizon lengthens, implying that firm complexity is more relevant for short-term bankruptcy prediction. This aligns with the theoretical implication of Duffie and Lando's (2001) incomplete accounting information model, which posits that information incompleteness primarily affects short-term credit risk.¹¹

[Insert Table 8 here]

Tables 9–12 (left panels) further present the changes in F1-score, MCC, Type I error, and Type II error before and after incorporating the Complexity variables. The results reveal

¹¹ Duffie and Lando's (2001) incomplete accounting information-based credit risk model argues that incomplete accounting information has greater explanatory power for short-term credit spreads.

consistent improvements in both F1-score and MCC, with the largest gains observed under the configuration that combines XGBoost with the Balanced Bagging method. In this setting, the F1-score increases from 0.0857, 0.0847, 0.0830, and 0.0828 to 0.1141, 0.1062, 0.0994, and 0.0961 for the one- to four-year prediction horizons, respectively. Similarly, the MCC improves from 0.1346, 0.1335, 0.1306, and 0.1298 to 0.1704, 0.1591, 0.1490, and 0.1444. With respect to Type II error, the inclusion of the Complexity variables leads to substantial reductions across all models and prediction windows; for instance, in the configuration that combines XGBoost with the Balanced Bagging method, Type II error declines from 0.2418–0.2448 to 0.1716–0.1966. Although Type I error slightly increases under certain imbalance treatments, the configuration that combines XGBoost with the Random Under Sampling (RUS) method achieves reductions for shorter-horizon predictions. Overall, these findings demonstrate that firm complexity meaningfully enhances bankruptcy prediction performance, particularly by lowering the incidence of misclassifying bankrupt firms as non-bankrupt (Type II error), a factor that is especially critical in credit risk management.

[Insert Tables 9–12 here]

4.1.2. Predictive Performance After Adding Firm Complexity (Using FIN and TCV as the Benchmark Model)

The predictive performance after adding firm complexity is also evaluated using the second benchmark model, which combines the financial variables (FIN) from Barboza et al. (2017) and the annual report TCV variables from Chen et al. (2023). Because the 53-word complexity lexicon is developed through expert screening and reflects financially meaningful terminology associated with valuation difficulty, firm complexity captures a broader and more substantive range of value-relevant firm attributes and underlying economic fundamentals than traditional annual report TCV variables, which primarily reflect managerial attempts to manipulate the speed and clarity of information transmission in annual reports. Accordingly,

this analysis examines whether firm complexity provides incremental predictive information beyond conventional readability- and tone-based textual features (namely TCV variables) in annual reports.

As shown in the right panel of Table 8, the inclusion of the Complexity variables improves AUC values across all prediction horizons. For instance, when XGBoost is combined with the Balanced Bagging method (and similarly with EasyEnsemble), the one-year-ahead AUC increases from 0.8835 (0.8895) to 0.9017 (0.9062), the two-year-ahead AUC rises from 0.8775 (0.8827) to 0.8905 (0.8953), the three-year-ahead AUC increases from 0.8729 (0.8782) to 0.8823 (0.8866), and the four-year-ahead AUC improves from 0.8693 (0.8756) to 0.8765 (0.8809). These results indicate that firm complexity captures bankruptcy-relevant information not contained in TCV variables, thereby further enhancing model performance. Consistent with earlier findings, the magnitude of improvement diminishes as the forecasting horizon extends, underscoring that firm complexity operates primarily as a short-horizon signal reflecting near-term informational obfuscation and disclosure opacity.

Tables 9–12 (right panels) also show that the Complexity variables substantially improve F1-score and MCC for one- and two-year-ahead predictions, particularly when XGBoost is combined with the Balanced Bagging method. However, for longer prediction horizons—three or four years—the improvements in F1-score and MCC become smaller or even negative in certain configurations, further confirming that complexity primarily operates as a short-horizon predictor of bankruptcy risk. Moreover, in one-year-ahead predictions, the inclusion of the Complexity variables reduces both Type I and Type II errors in several model specifications. For example, in the configuration combining XGBoost with the Balanced Bagging method, Type I error decreases from 0.2789 to 0.2552 and Type II error decreases from 0.1459 to 0.1337. For longer horizons, reductions occur mainly in Type II error, while Type I error shows less consistent improvement. Overall, these results demonstrate that even after accounting for

annual report TCV variables, firm complexity continues to provide meaningful, economically interpretable information that strengthens short-term bankruptcy prediction accuracy.

Table 13 summarizes the impact of firm complexity on bankruptcy across different prediction horizons, effectively illustrating its contribution to the term structure of bankruptcy risk. The results reveal that improvements in AUC and F1-score systematically decline as the prediction window increases, indicating that firm complexity has greater relevance for short-term bankruptcy risk. This is consistent with the theoretical argument of Duffie and Lando (2001) that incomplete accounting information plays a more significant role in short-term credit risk. Overall, the empirical evidence suggests that complexity-related textual features capture valuation-difficulty characteristics arising from organizational structure, product logistics, financial reporting, information dissemination, and financial engineering, all of which materially influence short-term bankruptcy assessments. Over longer horizons, firms may adjust their reporting practices, restructure operations, or modify financial policies in ways that gradually reduce the relevance of previously observed complexity. In addition, as external stakeholders gradually obtain new information through subsequent disclosures and market signals, the initial informational disadvantage created by complexity is reduced. As a result, the complexity signal becomes less persistent and less informative over longer horizons, naturally weakening its predictive power for long-term bankruptcy risk.

[Insert Table 13 here]

4.2 Feature Importance Analysis

To further investigate the relative contribution of firm complexity to bankruptcy prediction, this study computes feature importance scores using Random Forest and XGBoost. The scores are averaged across all rolling training windows (1993–2016, 1993–2017, 1993–2018, 1993–2019, and 1993–2020) for one-year-ahead predictions. The feature set includes the

11 financial variables (FIN), the 11 annual report text-based communicative value variables (TCV), and the 12 firm complexity variables developed in this study.

Table 14 reports the feature importance rankings for both models. Across all model specifications, several complexity-related variables consistently appear among the top-ranked predictors, such as business changes and organizational structure complexity (COMPLEXITY_BC), debt and contractual terms complexity (COMPLEXITY_DC), the volatility of contractual complexity (SD_DC), the volatility of international and global operations complexity (SD_IGO), and the volatility of total complexity (SD_TOTAL). These variables frequently occupy top positions within the first 15 most important features in both Random Forest and XGBoost models. Notably, the proportion of firm complexity variables within the top 15 predictors is substantially higher—46.67% in Random Forest and 40% in XGBoost—compared with only 13.33% and 20% for the TCV variables, respectively. This pattern suggests that firm complexity variables provide richer economic interpretation and exhibit stronger bankruptcy relevance than traditional readability- or tone-based TCV measures. Moreover, within the Random Forest model, most of volatility-based firm complexity variables appear more prominently than the raw frequency-based firm complexity measures, indicating that fluctuations in firm complexity over time carry additional bankruptcy-related information.¹² This finding highlights the importance of temporal variation in firm complexity as an informative predictor of financial distress.

[Insert Table 14 here]

5. Conclusions

This study employs the annual report-based firm complexity lexicon developed by

¹² The volatility-based firm complexity variables with higher feature importance include SD_FINACC (volatility of financial instruments and accounting operations complexity), SD_BC (volatility of business changes and organizational structure complexity), SD_DC (volatility of debt and contractual terms complexity), SD_IGO (volatility of international and global operations complexity), and SD_TOTAL (volatility of total complexity).

Loughran and McDonald (2023) and, drawing on GPT-4o, the semantic meaning of each term, and the conceptual domains identified in their framework, classifies the 53 complexity-related terms into five categories: financial instruments and accounting operations, corporate financial distress, business changes and organizational structure, international and global operations, and debt and contractual terms. Building on the theoretical foundations of Duffie and Lando (2001) and Chen et al. (2025), this study further develops 12 firm complexity variables that capture both the level and the volatility of firm complexity. Using Barboza et al.'s (2017) financial variables (FIN) and Chen et al.'s (2023) annual report text-based communicative value (TCV) variables as benchmark models, this study examines whether the additional inclusion of the 12 firm complexity features enhances bankruptcy prediction performance.

The empirical results show that firm complexity variables provide bankruptcy-relevant information beyond what is captured by annual report TCV measures, thereby improving prediction accuracy—particularly for short-horizon forecasts. For example, under the XGBoost model combined with the Balanced Bagging classifier, incorporating firm complexity yields the strongest and most consistent improvements in AUC, F1-score, MCC, and reductions in both Type I and Type II errors for one-year-ahead predictions. However, for longer horizons (two to four years), complexity variables generally do not simultaneously reduce both Type I and Type II errors across model specifications. These findings suggest that the firm complexity signals embedded in financial-reporting contexts primarily capture contemporaneous or near-term firm attributes—including prevailing economic conditions, operational and financial risks, organizational structures, and multinational activities—whose predictive value is inherently concentrated in the short run. Although such complexity increases short-term information asymmetry and valuation difficulty, the impact of incomplete information naturally diminishes over time as additional disclosures, market signals, and firm adjustments emerge. This explains why the predictive power of firm complexity diminishes over longer horizons and is consistent

with Duffie and Lando's (2001) argument that incomplete accounting information plays a more prominent role in short-term credit risk.

Furthermore, the feature importance analysis shows that variables such as business changes and organizational structure complexity, debt and contractual terms complexity, the volatility of contractual complexity, the volatility of international and global operations complexity, and the volatility of total complexity consistently rank among the most influential predictors. These variables exhibit higher importance than most of the annual report TCV measures proposed by Chen et al. (2023), indicating that firm complexity features not only provide richer economic interpretability but also convey distinctive bankruptcy-relevant informational content that is not captured by traditional annual report TCV indicators. Taken together, these results demonstrate that incorporating firm complexity variables into bankruptcy prediction models can substantially reduce short-term credit losses and improve lending decision efficiency, highlighting their practical value for credit risk management. Moreover, firm complexity variables complement traditional financial and annual report TCV measures by capturing distinctive bankruptcy-related information and enhancing short-term predictive performance.

References

- Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589–609.
- Barboza, F., Kimura, H., & Altman, E. I. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83, 405–417.
- Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 4, 71–111.
- Begley, J., Ming, J., & Watts, S. (1996). Bankruptcy classification errors in the 1980s: An empirical analysis of Altman's and Ohlson's models. *Review of Accounting Studies*, 1(4), 267–284.
- Bloomfield, R. (2002). The "incomplete revelation hypothesis" and financial reporting. *Accounting Horizons*, 16(3), 233–243.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Carton, R. B., & Hofer, C. W. (2006). *Measuring organizational performance*. Edward Elgar Publishing.
- Chen, T.K., Hao, Y., Chang, T.R., & Lin, Y.C. (2025). Foreseeing bankruptcy: A novel perspective based on the uncertainty of text-based communicative value of annual reports. *Finance Research Letters*, 84, 107771.
- Chen, T.K., Liao, H.H., Chen, G.D., Kang, W.H., & Lin, Y.C. (2023). Bankruptcy prediction using machine learning models with the text-based communicative value of annual reports. *Expert Systems with Applications*, 233, 120714.
- Chen, T.K., & Tseng, Y. (2021). Readability of notes to consolidated financial statements and corporate bond yield spread. *European Accounting Review*, 30(1), 83–113.
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794
- Dale, E., & Chall, J. S. (1948). A formula for predicting readability: Instructions. *Educational Research Bulletin*, 37-54
- Duffie, D., & Lando, D. (2001). Term structures of credit spreads with incomplete accounting information. *Econometrica*, 69(3), 633–664.
- Ertugrul, M., Lei, J., Qiu, J., & Wan, C. (2017). Annual report readability, tone ambiguity, and the cost of corporate borrowing. *The Journal of Financial and Quantitative Analysis*, 52(2), 811–836.
- Hao, Y., Chen, T. K. & Lin, Y.C. (2025a). Bankruptcy prediction using the text-based communicative value of earnings call transcripts. *Review of Quantitative Finance and Accounting*, forthcoming.
- Hao, Y., Chen, T. K. & Tseng, Y. (2025b). Firm complexity and corporate credit risk. *Working*

- paper, National Yang Ming Chiao Tung University.
- Jabeur, S., Gharib, C., Mefteh-Wali, S., & Ben Arfi, W. (2021). CatBoost model and artificial intelligence techniques for corporate failure prediction. *Technological Forecasting and Social Change*, 166, 120658.
- Kim, A., & Yoon, S. (2021). Corporate bankruptcy prediction with domain-adapted bert. In *Proceedings of the Third Workshop on Economics and Natural Language Processing*, pages 26–36, Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Klare, G. R. (1963). *The measurement of readability*. Iowa State University Press.
- Lambert, R., Leuz, C., & Verrecchia, R. E. (2007). Accounting information, disclosure, and the cost of capital. *Journal of Accounting Research*, 45(2), 385–420.
- 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*, 59(6), 2487 – 2514.
- Mai, F., Tian, S., Lee, C., & Ma, L. (2019). Deep learning models for bankruptcy prediction using textual disclosures. *European Journal of Operational Research*, 274(2), 743–758.
- Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, 29(2), 449–470.
- Nanni, L., & Lumini, A. (2009). An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, 36, 3028–3033.
- Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18(1), 109–131.
- Seebeck, A., & Kaya, D. (2023). The power of words: An empirical analysis of the communicative value of extended auditor reports. *European Accounting Review*, 32(5), 1185–1215.
- Vapnik, V. (1963). Pattern recognition using generalized portrait method. *Automation and Remote Control*, 24(6), 774–780.
- Wang, D.-n., Li, L., & Zhao, D. (2022). Corporate finance risk prediction based on LightGBM. *Information Sciences*, 602,10
- Wang, G., Chen, G., & Chu, Y. (2018). A new random subspace method incorporating sentiment and textual information for financial distress prediction. *Electronic Commerce Research and Applications*, 29, 30–49.

Table 1. Barboza et al. (2017)'s Financial Variables (FIN)

Variable	Variable Definitions
NWC_TA	Net Working Capital / Total Assets
RE_TA	Retained Earnings / Total Assets
EBIT_TA	Earnings Before Interest And Taxes / Total Assets
EMV_Debt	(Market Value of Share * Number of Shares) / Total Debt
Sales_TA	Sales / Total Assets
EBIT_Sales	Earnings Before Interest And Taxes / Sales
TA_Growth	$(\text{Total Assets}_t - \text{Total Assets}_{t-1}) / \text{Total Assets}_{t-1}$
Sales_Growth	$(\text{Sales}_t - \text{Sales}_{t-1}) / \text{Sales}_{t-1}$
EMP_Growth	$(\text{Number of Employees}_t - \text{Number of Employees}_{t-1}) / \text{Number of Employees}_{t-1}$
ROE_Chg	$\text{ROE}_t - \text{ROE}_{t-1}$
PB_Chg	$\text{Price to Book}_t - \text{Price to Book}_{t-1}$

Table 2. Firm Complexity Dictionary (Loughran and McDonald, 2023)

Financial instruments and accounting operations (FINACC; 17 terms)	Business changes and organizational structure (BC; 13 terms)
ACCRUES	AFFILIATES
CARRYBACK	ENTITIES
CARRYFORWARD	MERGERS
CARRYFORWARDS	MERGING
COLLATERAL	OUTSOURCE
COLLATERIZATION	PARTNERING
DERIVATIVE	RESTRUCTURE
DERIVATIVES	RESTRUCTURED
FLOATING	RESTRUCTURING
HEDGED	SEGMENTS
HEDGES	SUBLEASES
RECLASSIFIED	SUBSIDY
SECURITIZATIONS	VENTURES
SECURITIZED	Debt and contractual terms (DC; 12 terms)
SWAPS	COMPLEX
INTANGIBLES	CONVERTIBLE
REVALUATION	COUNTERPARTIES
Corporate financial distress (FD; 5 terms)	COUNTERPARTY
BANKRUPTCIES	COVENANT
INFRINGEMENT	COVENANTS
LAWSUIT	LEASEHOLD
LAWSUITS	LEASES
LITIGATION	LESSORS
International and global operations (IGO; 6 terms)	LICENSING
GLOBAL	REVOCATION
INTERNATIONAL	SOVEREIGN
REPATRIATE	
REPATRIATED	
REPATRIATION	
WORLDWIDE	

Note: Table 2 presents the 53 firm complexity terms developed by Loughran and McDonald (2023). This study further classifies these terms into five categories based on their semantic meanings and the conceptual domains identified by Loughran and McDonald (2023). First, Financial instruments and accounting operations (FINACC) includes terms associated with derivatives, securitizations, collateralization, intangible assets, revaluation, and other financial engineering or accounting processes that require specialized financial expertise and increase information-processing demands. Second, Corporate financial distress (FD) captures terms related to legal disputes, bankruptcy proceedings, infringement, and litigation, which reflect heightened uncertainty regarding future cash flows and elevate valuation ambiguity. Third, Business changes and organizational structure (BC) includes terms describing mergers, restructuring, business segments, affiliates, partnerships, and other organizational adjustments that increase the scope, diversity, or structural complexity of firm operations. Fourth, International and global operations (IGO) contains terms reflecting multinational activities such as global expansion, foreign presence, repatriation, and sovereign exposure, all of which introduce additional layers of regulatory, currency, and geopolitical uncertainty. Finally, Debt and contractual terms (DC) includes terms directly tied to debt covenants, counterparties, leases, licensing, convertibility provisions, and other legally binding contractual arrangements that increase the complexity of firms' financial obligations.

Table 2. Firm Complexity Variables (COMPLEXITY)

Variable	Variable Definition
COMPLEXITY_TOTAL	The ratio of total firm complexity terms to total annual report words
COMPLEXITY_FINACC	The ratio of financial instruments and accounting operations terms to total annual report words
COMPLEXITY_FD	The ratio of corporate financial distress terms to total annual report words
COMPLEXITY_BC	The ratio of business changes and organizational structure terms to total annual report words
COMPLEXITY_IGO	The ratio of international and global operations terms to total annual report words
COMPLEXITY_DC	The ratio of debt and contractual terms to total annual report words
SD_TOTAL	The five-year standard deviation of TOTAL_COMPLEXITY
SD_FINACC	The five-year standard deviation of COMPLEXITY_FINACC
SD_FD	The five-year standard deviation of COMPLEXITY_FD
SD_BC	The five-year standard deviation of COMPLEXITY_BC
SD_IGO	The five-year standard deviation of COMPLEXITY_IGO
SD_DC	The five-year standard deviation of COMPLEXITY_DC

Table 3. Chen et al. (2023)’s Annual Report Text-based Communicative Value Variables (TCV)

Variable	Variable Definition
FSIZE	The total file size (in bytes) of the firm’s 10-K filing
FK_Ease	Flesch Reading Ease score of the 10-K filing: $206.835 - 1.015 \times (\text{total words} \div \text{total sentences}) - 84.6 \times (\text{total syllables} \div \text{total words})$
CL	Coleman–Liau Readability Index of the 10-K filing: $5.88 \times (\text{characters} \div \text{words}) - 29.6 \times (\text{sentences} \div \text{words}) - 15.8$
A_WPP	Average words per paragraph in the 10-K: $\text{Total number of words} \div \text{total number of paragraphs}$
FT_NEG	Negative word proportion in the 10-K (LM dictionary): The number of Loughran–McDonald (LM) negative financial words appearing in the 10-K filing divided by the total number of words in the LM master dictionary that appear in the text.
FT_MWeak_C	Count of LM weak-modal words in the 10-K: The total number of weak-modal terms defined by the Loughran–McDonald financial dictionary that appear in the 10-K filing.
FT_POS	Positive word proportion in the 10-K (LM dictionary): The number of Loughran–McDonald (LM) positive financial words appearing in the 10-K filing divided by the total number of words in the LM master dictionary that appear in the text.
Harv_NEG	Proportion of negative words in the 10-K (Harvard General Inquirer): The number of negative terms defined by the Harvard General Inquirer dictionary that appear in the 10-K filing, divided by the total number of words matched to the master dictionary.
FT_LITI	Proportion of litigious words in the 10-K (LM dictionary): The number of litigious terms defined by the Loughran–McDonald financial dictionary that appear in the 10-K filing, divided by the total number of words matched to the master dictionary.
FT_UNC	Proportion of uncertain words in the 10-K (LM dictionary): The number of uncertain terms defined by the Loughran–McDonald financial dictionary that appear in the 10-K filing, divided by the total number of words matched to the master dictionary.
FT_MStr	Proportion of strong modal words in the 10-K (LM dictionary): The number of strong modal terms defined by the Loughran–McDonald financial dictionary that appear in the 10-K filing, divided by the total number of words matched to the master dictionary.

Table 4. Summary Statistics of Major Variables

Variable	Obs	Mean	Std	1st percentile	25th percentile	50th percentile	75th percentile	99th percentile
Panel A. Barboza et al.'s (2017) financial variables (FIN)								
NWC_TA	66,242	-0.1442	17.1839	-2.3678	0.0537	0.2140	0.4089	0.8557
RE_TA	66,242	-4.2855	129.3280	-44.4161	-0.6291	0.0552	0.3154	0.9986
EBIT_TA	66,242	-0.1151	2.3527	-2.3588	-0.0344	0.0565	0.1103	0.3445
EMV_Debt	66,242	10.3157	797.6167	0.0303	0.8663	2.1076	5.2284	59.5348
Sales_TA	66,242	1.1339	1.4239	0.0139	0.5142	0.9184	1.4615	4.4938
EBIT_Sales	66,242	-4.8685	128.0708	-41.1403	-0.0360	0.0538	0.1233	0.4451
TA_Growth	66,242	0.3723	41.1255	-0.6174	-0.0577	0.0412	0.1659	2.4550
Sales_Growth	66,242	0.7421	57.0465	-0.7770	-0.0431	0.0646	0.1985	3.4248
EMP_Growth	66,242	0.0994	1.6424	-0.6445	-0.0556	0.0189	0.1207	1.4807
ROE_Chg	66,242	0.1821	134.4144	-10.9381	-0.1088	-0.0043	0.0680	12.2051
PB_Chg	66,242	0.2481	457.6858	-49.7929	-0.6447	-0.0145	0.5501	47.8383
Panel B. Firm Complexity variables (Complexity) (unit:%)								
COMPLEXITY_TOTAL	66,242	0.2731	0.1434	0.0545	0.1700	0.2489	0.3481	0.7193
COMPLEXITY_FIO	66,242	0.0655	0.0593	0.0035	0.0270	0.0479	0.0859	0.2752
COMPLEXITY_FD	66,242	0.0232	0.0286	0.0000	0.0061	0.0137	0.0294	0.1374
COMPLEXITY_BC	66,242	0.0608	0.0509	0.0065	0.0268	0.0465	0.0779	0.2477
COMPLEXITY_IGO	66,242	0.0441	0.0629	0.0000	0.0086	0.0219	0.0545	0.3088
COMPLEXITY_DC	66,242	0.0795	0.0572	0.0090	0.0404	0.0657	0.1033	0.2813
SD_TOTAL	66,242	0.0639	0.0493	0.0034	0.0289	0.0519	0.0852	0.2322
SD_FIO	66,242	0.0256	0.0322	0.0006	0.0082	0.0162	0.0311	0.1581
SD_FD	66,242	0.0098	0.0119	0.0000	0.0024	0.0060	0.0127	0.0558
SD_BC	66,242	0.0218	0.0213	0.0007	0.0078	0.0154	0.0284	0.1008
SD_IGO	66,242	0.0169	0.0237	0.0000	0.0033	0.0086	0.0211	0.1134
SD_DC	66,242	0.0234	0.0221	0.0009	0.0095	0.0175	0.0301	0.1094
Panel C. Chen et al.'s (2023) annual report text-based communicative value variables (TCV)								
FSIZE	66,242	6062056	9937007	1871	597280	1737324	8865258	37600000
FK_Ease	66,242	13.7901	18.6801	-27.9411	-5.3503	23.1331	28.0989	38.4827
CL	66,242	18.6359	3.5419	14.0939	15.6314	17.1892	22.1131	26.3730
A_WPP	66,242	92.7447	54.7750	23.0000	62.3950	83.1498	116.2759	221.0255
FT_NEG	66,242	0.0171	0.0161	0.0000	0.0128	0.0154	0.0183	0.0854
FT_MWeak_C	66,242	376.5544	315.9097	0.0000	149.0000	300.0000	526.0000	1404.0000
FT_POS	66,242	0.0077	0.0020	0.0000	0.0065	0.0076	0.0088	0.0129
Harv_NEG	66,242	0.0416	0.0169	0.0000	0.0366	0.0403	0.0439	0.1116
FT_LITI	66,242	0.0165	0.0075	0.0000	0.0113	0.0159	0.0207	0.0377
FT_UNC	66,242	0.0105	0.0030	0.0000	0.0085	0.0102	0.0123	0.0181
FT_MStr	66,242	0.0024	0.0012	0.0000	0.0016	0.0023	0.0031	0.0058

Note: The variable definitions in Panels A, B, and C correspond to Tables 1, 3, and 4, respectively.

Table 6. Yearly Sample Distribution of Bankrupt and Non-Bankrupt Firms

Year	Non-Bankrupt Bankruptcy=0	Bankrupt Bankruptcy=1	Total	The proportion of bankrupt firms
1993	5	0	5	0.0000%
1994	393	18	411	4.3796%
1995	900	36	936	3.8462%
1996	1,425	58	1,483	3.9110%
1997	2,014	79	2,903	2.7213%
1998	2,147	78	2,225	3.5056%
1999	2,109	78	2,187	3.5665%
2000	2,124	78	2,202	3.5422%
2001	2,350	93	2,443	3.8068%
2002	3,018	120	3,138	3.8241%
2003	3,132	114	3,246	3.5120%
2004	2,939	106	3,045	3.4811%
2005	2,738	95	2,833	3.3533%
2006	2,609	99	2,708	3.6558%
2007	2,525	97	2,622	3.6995%
2008	2,528	91	2,619	3.4746%
2009	2,757	82	2,839	2.8883%
2010	2,579	77	2,656	2.8991%
2011	2,366	70	2,436	2.8736%
2012	2,300	68	2,368	2.8716%
2013	2,266	64	2,330	2.7468%
2014	2,283	57	2,340	2.4359%
2015	2,228	59	2,287	2.5798%
2016	2,191	60	2,251	2.6655%
2017	2,165	52	2,217	2.3455%
2018	2,138	43	2,181	1.9716%
2019	2,106	36	2,142	1.6807%
2020	1,789	25	1,814	1.3782%
2021	2,001	26	2,027	1.2827%
2022	2,134	24	2,158	1.1121%
Total	64,259	1,983	66,242	2.9936%

Table 7. Confusion Matrix of Bankruptcy Prediction

Actual \ Predict	Positive	Negative
	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

Note: Positive denotes bankrupt firms (Bankruptcy = 1), whereas Negative denotes non-bankrupt firms (Bankruptcy = 0). The confusion matrix includes the following outcomes: True Positive (TP), which represents firms predicted to be bankrupt that are actually bankrupt; False Positive (FP), which represents firms that are actually non-bankrupt but are incorrectly predicted as bankrupt; True Negative (TN), which represents firms predicted to be non-bankrupt that are actually non-bankrupt; and False Negative (FN), which represents firms that are actually bankrupt but are incorrectly predicted as non-bankrupt.

**Table 8. Comparative Analysis of Bankruptcy Prediction Performance After
Incorporating Firm Complexity: Measured by AUC**

Panel A. One-Year-Ahead Bankruptcy Prediction Performance (AUC)						
ML model	FIN	FIN+ COMPLEXITY	Difference	FIN+TCV	FIN+ TCV +COMPLEXITY	Difference
Panel A.1. BalancedBaggingClassifier						
CatBoost	0.8310	0.8625	0.0315	0.8828	0.8964	0.0136
LGBM	0.8242	0.8590	0.0348	0.8780	0.8937	0.0157
RF	0.8289	0.8543	0.0254	0.8719	0.8849	0.0130
XGBoost	0.8157	0.8609	0.0452	0.8835	0.9017	0.0182
Panel A.2. EasyEnsemble						
CatBoost	0.8347	0.8648	0.0301	0.8846	0.8982	0.0136
LGBM	0.8232	0.8620	0.0388	0.8832	0.8969	0.0137
RF	0.8299	0.8575	0.0276	0.8748	0.8902	0.0154
XGBoost	0.8139	0.8664	0.0525	0.8895	0.9062	0.0167
Panel A.3. RandomUnderSampling						
CatBoost	0.8095	0.8454	0.0359	0.8709	0.8826	0.0117
LGBM	0.7806	0.8204	0.0398	0.8509	0.8620	0.0111
RF	0.7998	0.8337	0.0339	0.8548	0.8698	0.0150
XGBoost	0.7581	0.8137	0.0556	0.8459	0.8618	0.0159
Panel B. Two-Year-Ahead Bankruptcy Prediction Performance (AUC)						
Panel B.1. BalancedBaggingClassifier						
CatBoost	0.8267	0.8491	0.0224	0.8763	0.8850	0.0087
LGBM	0.8201	0.8460	0.0259	0.8716	0.8815	0.0099
RF	0.8254	0.8415	0.0161	0.8626	0.8717	0.0091
XGBoost	0.8134	0.8470	0.0336	0.8775	0.8905	0.0130
Panel B.2. EasyEnsemble						
CatBoost	0.8301	0.8522	0.0221	0.8777	0.8863	0.0086
LGBM	0.8166	0.8490	0.0324	0.8756	0.8856	0.0100
RF	0.8245	0.8447	0.0202	0.8654	0.8767	0.0113
XGBoost	0.8079	0.8508	0.0429	0.8827	0.8953	0.0126
Panel B.3. RandomUnderSampling						
CatBoost	0.8058	0.8326	0.0268	0.8634	0.8710	0.0076
LGBM	0.7752	0.8085	0.0333	0.8433	0.8516	0.0083
RF	0.7971	0.8200	0.0229	0.8461	0.8557	0.0096
XGBoost	0.7519	0.7977	0.0458	0.8367	0.8489	0.0122
Panel C. Three-Year-Ahead Bankruptcy Prediction Performance (AUC)						

Panel C.1. BalancedBaggingClassifier						
CatBoost	0.8226	0.8374	0.0148	0.8728	0.8767	0.0039
LGBM	0.8143	0.8346	0.0203	0.8676	0.8725	0.0049
RF	0.8207	0.8298	0.0091	0.8596	0.8644	0.0048
XGBoost	0.8075	0.8342	0.0267	0.8729	0.8823	0.0094
Panel C.2. EasyEnsemble						
CatBoost	0.8260	0.8413	0.0153	0.8741	0.8776	0.0035
LGBM	0.8120	0.8377	0.0257	0.8711	0.8768	0.0057
RF	0.8196	0.8324	0.0128	0.8618	0.8692	0.0074
XGBoost	0.8026	0.8376	0.0350	0.8782	0.8866	0.0084
Panel C.3. RandomUnderSampling						
CatBoost	0.8009	0.8221	0.0212	0.8597	0.8626	0.0029
LGBM	0.7701	0.7978	0.0277	0.8402	0.8452	0.0050
RF	0.7921	0.8087	0.0166	0.8436	0.8485	0.0049
XGBoost	0.7482	0.7849	0.0367	0.8333	0.8406	0.0073
Panel D. Four-Year-Ahead Bankruptcy Prediction Performance (AUC)						
Panel D.1. BalancedBaggingClassifier						
CatBoost	0.8211	0.8307	0.0096	0.8707	0.8704	-0.0003
LGBM	0.8118	0.8279	0.0161	0.8647	0.8672	0.0025
RF	0.8181	0.8226	0.0045	0.8567	0.8595	0.0028
XGBoost	0.8039	0.8273	0.0234	0.8693	0.8765	0.0072
Panel D.2. EasyEnsemble						
CatBoost	0.8233	0.8340	0.0107	0.8710	0.8707	-0.0003
LGBM	0.8101	0.8310	0.0209	0.8680	0.8708	0.0028
RF	0.8176	0.8250	0.0074	0.8587	0.8637	0.0050
XGBoost	0.7998	0.8292	0.0294	0.8756	0.8809	0.0053
Panel D.3. RandomUnderSampling						
CatBoost	0.7988	0.8165	0.0177	0.8549	0.8558	0.0009
LGBM	0.7670	0.7901	0.0231	0.8354	0.8388	0.0034
RF	0.7903	0.8015	0.0112	0.8404	0.8427	0.0023
XGBoost	0.7466	0.7789	0.0323	0.8295	0.8340	0.0045

Note: This table reports the predictive performance (measured by AUC) for bankruptcy risk across future horizons before and after incorporating firm complexity variables, based on 10 random-state replications. For each prediction horizon, model performance is computed as the average across all combinations of rolling training-testing windows and random states. Specifically, the one-year-ahead prediction includes six rolling forecasts—using 1993–2016 to predict 2017, 1993–2017 to predict 2018, 1993–2018 to predict 2019, 1993–2019 to predict 2020, 1993–2020 to predict 2021, and 1993–2021 to predict 2022. Considering 10 random states, the one-year-ahead result reflects the average of 60 prediction outcomes. The same procedure is applied to two-, three-, and four-year-ahead bankruptcy predictions.

Table 9. Comparative Analysis of Bankruptcy Prediction Performance After Incorporating Firm Complexity: Measured by F1-Score

Panel A. One-Year-Ahead Bankruptcy Prediction Performance (F1-score)						
ML model	FIN	FIN+ COMPLEXITY	Difference FIN+TCV		FIN+ TCV +COMPLEXITY	Difference
Panel A.1. BalancedBaggingClassifier						
CatBoost	0.0796	0.1028	0.0232	0.1222	0.1293	0.0071
LGBM	0.0816	0.1043	0.0227	0.1224	0.1321	0.0097
RF	0.0887	0.1116	0.0229	0.1282	0.1449	0.0167
XGBoost	0.0857	0.1141	0.0284	0.1375	0.1529	0.0154
Panel A.2. EasyEnsemble						
CatBoost	0.0688	0.0840	0.0152	0.1009	0.1041	0.0032
LGBM	0.0680	0.0834	0.0154	0.1025	0.1059	0.0034
RF	0.0686	0.0811	0.0125	0.0963	0.0990	0.0027
XGBoost	0.0687	0.0886	0.0199	0.1089	0.1152	0.0063
Panel A.3. RandomUnderSampling						
CatBoost	0.0678	0.0814	0.0136	0.0976	0.1013	0.0037
LGBM	0.0641	0.0760	0.0119	0.0924	0.0942	0.0018
RF	0.0673	0.0801	0.0128	0.0928	0.0974	0.0046
XGBoost	0.0617	0.0770	0.0153	0.0912	0.0976	0.0064
Panel B. Two-Year-Ahead Bankruptcy Prediction Performance (F1-score)						
Panel B.1. BalancedBaggingClassifier						
CatBoost	0.0790	0.0963	0.0173	0.1169	0.1191	0.0022
LGBM	0.0804	0.0972	0.0168	0.1172	0.1212	0.0040
RF	0.0880	0.1030	0.0150	0.1241	0.1313	0.0072
XGBoost	0.0847	0.1062	0.0215	0.1294	0.1387	0.0093
Panel B.2. EasyEnsemble						
CatBoost	0.0681	0.0796	0.0115	0.0961	0.0968	0.0007
LGBM	0.0667	0.0794	0.0127	0.0976	0.0986	0.0010
RF	0.0672	0.0773	0.0101	0.0931	0.0926	-0.0005
XGBoost	0.0669	0.0837	0.0168	0.1028	0.1068	0.0040
Panel B.3. RandomUnderSampling						
CatBoost	0.0673	0.0774	0.0101	0.0936	0.0948	0.0012
LGBM	0.0637	0.0724	0.0087	0.0883	0.0889	0.0006
RF	0.0664	0.0754	0.0090	0.0903	0.0918	0.0015
XGBoost	0.0604	0.0728	0.0124	0.0874	0.0919	0.0045
Panel C. Three-Year-Ahead Bankruptcy Prediction Performance (F1-score)						

Panel C.1. BalancedBaggingClassifier						
CatBoost	0.0781	0.0911	0.0130	0.1132	0.1110	-0.0022
LGBM	0.0790	0.0914	0.0124	0.1132	0.1135	0.0003
RF	0.0861	0.0965	0.0104	0.1220	0.1231	0.0011
XGBoost	0.0830	0.0994	0.0164	0.1232	0.1289	0.0057
Panel C.2. EasyEnsemble						
CatBoost	0.0675	0.0756	0.0081	0.0924	0.0910	-0.0014
LGBM	0.0659	0.0754	0.0095	0.0935	0.0927	-0.0008
RF	0.0663	0.0736	0.0073	0.0905	0.0881	-0.0024
XGBoost	0.0660	0.0794	0.0134	0.0982	0.1000	0.0018
Panel C.3. RandomUnderSampling						
CatBoost	0.0667	0.0735	0.0068	0.0901	0.0889	-0.0012
LGBM	0.0627	0.0690	0.0063	0.0856	0.0843	-0.0013
RF	0.0656	0.0715	0.0059	0.0874	0.0870	-0.0004
XGBoost	0.0597	0.0696	0.0099	0.0848	0.0869	0.0021
Panel D. Four-Year-Ahead Bankruptcy Prediction Performance (F1-score)						
Panel D.1. BalancedBaggingClassifier						
CatBoost	0.0783	0.0881	0.0098	0.1103	0.1053	-0.0050
LGBM	0.0790	0.0884	0.0094	0.1103	0.1081	-0.0022
RF	0.0860	0.0931	0.0071	0.1205	0.1185	-0.0020
XGBoost	0.0828	0.0961	0.0133	0.1201	0.1225	0.0024
Panel D.2. EasyEnsemble						
CatBoost	0.0673	0.0731	0.0058	0.0895	0.0867	-0.0028
LGBM	0.0657	0.0734	0.0077	0.0910	0.0883	-0.0027
RF	0.0663	0.0717	0.0054	0.0879	0.0849	-0.0030
XGBoost	0.0658	0.0767	0.0109	0.0963	0.0951	-0.0012
Panel D.3. RandomUnderSampling						
CatBoost	0.0667	0.0719	0.0052	0.0876	0.0852	-0.0024
LGBM	0.0624	0.0675	0.0051	0.0837	0.0812	-0.0025
RF	0.0657	0.0701	0.0044	0.0852	0.0841	-0.0011
XGBoost	0.0598	0.0684	0.0086	0.0822	0.0835	0.0013

Note: This table reports the predictive performance (measured by F1-score) for bankruptcy risk across future horizons before and after incorporating firm complexity variables, based on 10 random-state replications. For each prediction horizon, model performance is computed as the average across all combinations of rolling training–testing windows and random states. Specifically, the one-year-ahead prediction includes six rolling forecasts—using 1993–2016 to predict 2017, 1993–2017 to predict 2018, 1993–2018 to predict 2019, 1993–2019 to predict 2020, 1993–2020 to predict 2021, and 1993–2021 to predict 2022. Considering 10 random states, the one-year-ahead result reflects the average of 60 prediction outcomes. The same procedure is applied to two-, three-, and four-year-ahead bankruptcy predictions.

**Table 10. Comparative Analysis of Bankruptcy Prediction Performance After
Incorporating Firm Complexity: Measured by MCC**

Panel A. One-Year-Ahead Bankruptcy Prediction Performance (MCC)						
ML model	FIN	FIN+ COMPLEXITY	Difference FIN+TCV		FIN+ TCV +COMPLEXITY	Difference
Panel A.1. BalancedBaggingClassifier						
CatBoost	0.1338	0.1640	0.0302	0.1884	0.1970	0.0086
LGBM	0.1349	0.1648	0.0299	0.1865	0.1983	0.0118
RF	0.1397	0.1652	0.0255	0.1822	0.2007	0.0185
XGBoost	0.1346	0.1704	0.0358	0.2003	0.2201	0.0198
Panel A.2. EasyEnsemble						
CatBoost	0.1238	0.1490	0.0252	0.1716	0.1779	0.0063
LGBM	0.1209	0.1483	0.0274	0.1752	0.1789	0.0037
RF	0.1246	0.1444	0.0198	0.1650	0.1688	0.0038
XGBoost	0.1215	0.1550	0.0335	0.1846	0.1920	0.0074
Panel A.3. RandomUnderSampling						
CatBoost	0.1188	0.1408	0.0220	0.1639	0.1705	0.0066
LGBM	0.1078	0.1291	0.0213	0.1548	0.1567	0.0019
RF	0.1164	0.1378	0.0214	0.1551	0.1619	0.0068
XGBoost	0.0987	0.1279	0.0292	0.1500	0.1592	0.0092
Panel B. Two-Year-Ahead Bankruptcy Prediction Performance (MCC)						
Panel B.1. BalancedBaggingClassifier						
CatBoost	0.1336	0.1544	0.0208	0.1845	0.1865	0.0020
LGBM	0.1329	0.1545	0.0216	0.1832	0.1869	0.0037
RF	0.1397	0.1526	0.0129	0.1807	0.1864	0.0057
XGBoost	0.1335	0.1591	0.0256	0.1937	0.2053	0.0116
Panel B.2. EasyEnsemble						
CatBoost	0.1235	0.1422	0.0187	0.1676	0.1693	0.0017
LGBM	0.1191	0.1427	0.0236	0.1708	0.1708	0.0000
RF	0.1227	0.1386	0.0159	0.1623	0.1607	-0.0016
XGBoost	0.1181	0.1471	0.0290	0.1790	0.1823	0.0033
Panel B.3. RandomUnderSampling						
CatBoost	0.1189	0.1350	0.0161	0.1608	0.1628	0.0020
LGBM	0.1076	0.1231	0.0155	0.1499	0.1506	0.0007
RF	0.1161	0.1298	0.0137	0.1533	0.1556	0.0023
XGBoost	0.0961	0.1199	0.0238	0.1454	0.1518	0.0064
Panel C. Three-Year-Ahead Bankruptcy Prediction Performance (MCC)						

Panel C.1. BalancedBaggingClassifier						
CatBoost	0.1322	0.1469	0.0147	0.1815	0.1775	-0.0040
LGBM	0.1305	0.1459	0.0154	0.1797	0.1785	-0.0012
RF	0.1366	0.1438	0.0072	0.1806	0.1774	-0.0032
XGBoost	0.1306	0.1490	0.0184	0.1875	0.1947	0.0072
Panel C.2. EasyEnsemble						
CatBoost	0.1232	0.1355	0.0123	0.1632	0.1617	-0.0015
LGBM	0.1176	0.1356	0.0180	0.1659	0.1631	-0.0028
RF	0.1213	0.1323	0.0110	0.1601	0.1556	-0.0045
XGBoost	0.1167	0.1395	0.0228	0.1737	0.1743	0.0006
Panel C.3. RandomUnderSampling						
CatBoost	0.1184	0.1285	0.0101	0.1572	0.1557	-0.0015
LGBM	0.1055	0.1174	0.0119	0.1473	0.1451	-0.0022
RF	0.1146	0.1235	0.0089	0.1502	0.1498	-0.0004
XGBoost	0.0950	0.1142	0.0192	0.1430	0.1459	0.0029
Panel D. Four-Year-Ahead Bankruptcy Prediction Performance (MCC)						
Panel D.1. BalancedBaggingClassifier						
CatBoost	0.1326	0.1422	0.0096	0.1781	0.1699	-0.0082
LGBM	0.1302	0.1415	0.0113	0.1764	0.1714	-0.0050
RF	0.1362	0.1384	0.0022	0.1803	0.1724	-0.0079
XGBoost	0.1298	0.1444	0.0146	0.1842	0.1869	0.0027
Panel D.2. EasyEnsemble						
CatBoost	0.1226	0.1308	0.0082	0.1586	0.1543	-0.0043
LGBM	0.1169	0.1318	0.0149	0.1621	0.1559	-0.0062
RF	0.1211	0.1280	0.0069	0.1557	0.1504	-0.0053
XGBoost	0.1156	0.1339	0.0183	0.1711	0.1671	-0.0040
Panel D.3. RandomUnderSampling						
CatBoost	0.1179	0.1252	0.0073	0.1529	0.1495	-0.0034
LGBM	0.1041	0.1143	0.0102	0.1445	0.1404	-0.0041
RF	0.1147	0.1202	0.0055	0.1469	0.1448	-0.0021
XGBoost	0.0948	0.1116	0.0168	0.1393	0.1409	0.0016

Note: This table reports the predictive performance (measured by MCC) for bankruptcy risk across future horizons before and after incorporating firm complexity variables, based on 10 random-state replications. For each prediction horizon, model performance is computed as the average across all combinations of rolling training-testing windows and random states. Specifically, the one-year-ahead prediction includes six rolling forecasts—using 1993–2016 to predict 2017, 1993–2017 to predict 2018, 1993–2018 to predict 2019, 1993–2019 to predict 2020, 1993–2020 to predict 2021, and 1993–2021 to predict 2022. Considering 10 random states, the one-year-ahead result reflects the average of 60 prediction outcomes. The same procedure is applied to two-, three-, and four-year-ahead bankruptcy predictions.

**Table 11. Comparative Analysis of Bankruptcy Prediction Performance After
Incorporating Firm Complexity: Measured by Type I Error**

Panel A. One-Year-Ahead Bankruptcy Prediction Performance (Type I error)						
ML model	FIN	FIN+ COMPLEXITY	Difference FIN+TCV		FIN+ TCV +COMPLEXITY	Difference
Panel A.1. BalancedBaggingClassifier						
CatBoost	0.2121	0.2388	0.0267	0.2412	0.2326	-0.0086
LGBM	0.2311	0.2477	0.0166	0.2547	0.2452	-0.0095
RF	0.2810	0.3167	0.0357	0.3260	0.3176	-0.0084
XGBoost	0.2871	0.3027	0.0156	0.2789	0.2552	-0.0237
Panel A.2. EasyEnsemble						
CatBoost	0.1317	0.1401	0.0084	0.1623	0.1423	-0.0200
LGBM	0.1449	0.1377	-0.0072	0.1493	0.1507	0.0014
RF	0.1216	0.1375	0.0159	0.1552	0.1519	-0.0033
XGBoost	0.1505	0.1514	0.0009	0.1451	0.1482	0.0031
Panel A.3. RandomUnderSampling						
CatBoost	0.1612	0.1733	0.0121	0.1859	0.1702	-0.0157
LGBM	0.1986	0.1997	0.0011	0.1996	0.1994	-0.0002
RF	0.1722	0.1788	0.0066	0.1964	0.1897	-0.0067
XGBoost	0.2394	0.2262	-0.0132	0.2290	0.2199	-0.0091
Panel B. Two-Year-Ahead Bankruptcy Prediction Performance (Type I error)						
Panel B.1. BalancedBaggingClassifier						
CatBoost	0.2134	0.2537	0.0403	0.2321	0.2355	0.0034
LGBM	0.2399	0.2631	0.0232	0.2446	0.2501	0.0055
RF	0.2804	0.3401	0.0597	0.3217	0.3368	0.0151
XGBoost	0.2923	0.3228	0.0305	0.2739	0.2641	-0.0098
Panel B.2. EasyEnsemble						
CatBoost	0.1309	0.1450	0.0141	0.1501	0.1415	-0.0086
LGBM	0.1488	0.1386	-0.0102	0.1410	0.1487	0.0077
RF	0.1240	0.1410	0.0170	0.1539	0.1601	0.0062
XGBoost	0.1608	0.1608	0.0000	0.1346	0.1481	0.0135
Panel B.3. RandomUnderSampling						
CatBoost	0.1591	0.1764	0.0173	0.1750	0.1701	-0.0049
LGBM	0.2010	0.2070	0.0060	0.2005	0.1990	-0.0015
RF	0.1704	0.1892	0.0188	0.1937	0.1924	-0.0013
XGBoost	0.2494	0.2424	-0.0070	0.2277	0.2280	0.0003
Panel C. Three-Year-Ahead Bankruptcy Prediction Performance (Type I error)						

Panel C.1. BalancedBaggingClassifier						
CatBoost	0.2159	0.2598	0.0439	0.2215	0.2303	0.0088
LGBM	0.2447	0.2709	0.0262	0.2356	0.2468	0.0112
RF	0.2859	0.3456	0.0597	0.3056	0.3379	0.0323
XGBoost	0.2984	0.3354	0.0370	0.2679	0.2635	-0.0044
Panel C.2. EasyEnsemble						
CatBoost	0.1282	0.1494	0.0212	0.1436	0.1381	-0.0055
LGBM	0.1510	0.1459	-0.0051	0.1349	0.1463	0.0114
RF	0.1260	0.1466	0.0206	0.1440	0.1508	0.0068
XGBoost	0.1622	0.1708	0.0086	0.1273	0.1423	0.0150
Panel C.3. RandomUnderSampling						
CatBoost	0.1574	0.1797	0.0223	0.1634	0.1606	-0.0028
LGBM	0.2072	0.2075	0.0003	0.1885	0.1900	0.0015
RF	0.1735	0.1918	0.0183	0.1860	0.1828	-0.0032
XGBoost	0.2507	0.2481	-0.0026	0.2153	0.2171	0.0018
Panel D. Four-Year-Ahead Bankruptcy Prediction Performance (Type I error)						
Panel D.1. BalancedBaggingClassifier						
CatBoost	0.2143	0.2616	0.0473	0.2182	0.2317	0.0135
LGBM	0.2472	0.2721	0.0249	0.2318	0.2467	0.0149
RF	0.2882	0.3535	0.0653	0.2953	0.3349	0.0396
XGBoost	0.3029	0.3368	0.0339	0.2640	0.2644	0.0004
Panel D.2. EasyEnsemble						
CatBoost	0.1302	0.1533	0.0231	0.1448	0.1436	-0.0012
LGBM	0.1542	0.1476	-0.0066	0.1336	0.1493	0.0157
RF	0.1271	0.1543	0.0272	0.1474	0.1513	0.0039
XGBoost	0.1670	0.1790	0.0120	0.1253	0.1424	0.0171
Panel D.3. RandomUnderSampling						
CatBoost	0.1607	0.1825	0.0218	0.1662	0.1636	-0.0026
LGBM	0.2132	0.2094	-0.0038	0.1868	0.1868	0.0000
RF	0.1748	0.1984	0.0236	0.1854	0.1862	0.0008
XGBoost	0.2519	0.2504	-0.0015	0.2115	0.2136	0.0021

Note: This table reports the predictive performance (measured by Type I error) for bankruptcy risk across future horizons before and after incorporating firm complexity variables, based on 10 random-state replications. For each prediction horizon, model performance is computed as the average across all combinations of rolling training–testing windows and random states. Specifically, the one-year-ahead prediction includes six rolling forecasts—using 1993–2016 to predict 2017, 1993–2017 to predict 2018, 1993–2018 to predict 2019, 1993–2019 to predict 2020, 1993–2020 to predict 2021, and 1993–2021 to predict 2022. Considering 10 random states, the one-year-ahead result reflects the average of 60 prediction outcomes. The same procedure is applied to two-, three-, and four-year-ahead bankruptcy predictions.

**Table 12. Comparative Analysis of Bankruptcy Prediction Performance After
Incorporating Firm Complexity: Measured by Type II Error**

Panel A. One-Year-Ahead Bankruptcy Prediction Performance (Type II error)						
ML model	FIN	FIN+ COMPLEXITY	Difference FIN+TCV		FIN+ TCV +COMPLEXITY	Difference
Panel A.1. BalancedBaggingClassifier						
CatBoost	0.2931	0.2143	-0.0788	0.1752	0.1697	-0.0055
LGBM	0.2755	0.2081	-0.0674	0.1719	0.1636	-0.0083
RF	0.2355	0.1726	-0.0629	0.1499	0.1349	-0.0150
XGBoost	0.2418	0.1716	-0.0702	0.1459	0.1337	-0.0122
Panel A.2. EasyEnsemble						
CatBoost	0.3829	0.3045	-0.0784	0.2415	0.2432	0.0017
LGBM	0.3813	0.3083	-0.0730	0.2435	0.2383	-0.0052
RF	0.3883	0.3184	-0.0699	0.2631	0.2592	-0.0039
XGBoost	0.3748	0.2831	-0.0917	0.2278	0.2155	-0.0123
Panel A.3. RandomUnderSampling						
CatBoost	0.3750	0.3033	-0.0717	0.2441	0.2428	-0.0013
LGBM	0.3782	0.3155	-0.0627	0.2549	0.2543	-0.0006
RF	0.3738	0.3095	-0.0643	0.2590	0.2513	-0.0077
XGBoost	0.3755	0.3003	-0.0752	0.2475	0.2374	-0.0101
Panel B. Two-Year-Ahead Bankruptcy Prediction Performance (Type II error)						
Panel B.1. BalancedBaggingClassifier						
CatBoost	0.2930	0.2231	-0.0699	0.1835	0.1790	-0.0045
LGBM	0.2759	0.2179	-0.0580	0.1790	0.1721	-0.0069
RF	0.2356	0.1798	-0.0558	0.1499	0.1369	-0.0130
XGBoost	0.2412	0.1784	-0.0628	0.1534	0.1435	-0.0099
Panel B.2. EasyEnsemble						
CatBoost	0.3849	0.3184	-0.0665	0.2554	0.2569	0.0015
LGBM	0.3839	0.3219	-0.0620	0.2540	0.2504	-0.0036
RF	0.3922	0.3311	-0.0611	0.2652	0.2659	0.0007
XGBoost	0.3773	0.2951	-0.0822	0.2413	0.2289	-0.0124
Panel B.3. RandomUnderSampling						
CatBoost	0.3762	0.3161	-0.0601	0.2556	0.2550	-0.0006
LGBM	0.3773	0.3268	-0.0505	0.2637	0.2638	0.0001
RF	0.3758	0.3221	-0.0537	0.2609	0.2578	-0.0031
XGBoost	0.3755	0.3094	-0.0661	0.2575	0.2450	-0.0125
Panel C. Three-Year-Ahead Bankruptcy Prediction Performance (Type II error)						

Panel C.1. BalancedBaggingClassifier						
CatBoost	0.2957	0.2350	-0.0607	0.1930	0.1948	0.0018
LGBM	0.2795	0.2305	-0.0490	0.1889	0.1858	-0.0031
RF	0.2399	0.1922	-0.0477	0.1569	0.1474	-0.0095
XGBoost	0.2445	0.1890	-0.0555	0.1636	0.1563	-0.0073
Panel C.2. EasyEnsemble						
CatBoost	0.3882	0.3336	-0.0546	0.2686	0.2752	0.0066
LGBM	0.3868	0.3366	-0.0502	0.2680	0.2674	-0.0006
RF	0.3956	0.3450	-0.0506	0.2757	0.2824	0.0067
XGBoost	0.3807	0.3083	-0.0724	0.2560	0.2466	-0.0094
Panel C.3. RandomUnderSampling						
CatBoost	0.3801	0.3319	-0.0482	0.2704	0.2759	0.0055
LGBM	0.3802	0.3432	-0.0370	0.2775	0.2820	0.0045
RF	0.3791	0.3381	-0.0410	0.2733	0.2765	0.0032
XGBoost	0.3780	0.3217	-0.0563	0.2704	0.2634	-0.0070
Panel D. Four-Year-Ahead Bankruptcy Prediction Performance (Type II error)						
Panel D.1. BalancedBaggingClassifier						
CatBoost	0.2967	0.2442	-0.0525	0.2010	0.2077	0.0067
LGBM	0.2797	0.2394	-0.0403	0.1966	0.1978	0.0012
RF	0.2404	0.1981	-0.0423	0.1617	0.1554	-0.0063
XGBoost	0.2448	0.1966	-0.0482	0.1705	0.1665	-0.0040
Panel D.2. EasyEnsemble						
CatBoost	0.3899	0.3461	-0.0438	0.2798	0.2907	0.0109
LGBM	0.3880	0.3476	-0.0404	0.2788	0.2837	0.0049
RF	0.3973	0.3532	-0.0441	0.2853	0.2961	0.0108
XGBoost	0.3815	0.3188	-0.0627	0.2642	0.2629	-0.0013
Panel D.3. RandomUnderSampling						
CatBoost	0.3801	0.3407	-0.0394	0.2796	0.2897	0.0101
LGBM	0.3813	0.3520	-0.0293	0.2862	0.2966	0.0104
RF	0.3789	0.3436	-0.0353	0.2818	0.2872	0.0054
XGBoost	0.3791	0.3281	-0.0510	0.2828	0.2783	-0.0045

Note: This table reports the predictive performance (measured by Type II error) for bankruptcy risk across future horizons before and after incorporating firm complexity variables, based on 10 random-state replications. For each prediction horizon, model performance is computed as the average across all combinations of rolling training–testing windows and random states. Specifically, the one-year-ahead prediction includes six rolling forecasts—using 1993–2016 to predict 2017, 1993–2017 to predict 2018, 1993–2018 to predict 2019, 1993–2019 to predict 2020, 1993–2020 to predict 2021, and 1993–2021 to predict 2022. Considering 10 random states, the one-year-ahead result reflects the average of 60 prediction outcomes. The same procedure is applied to two-, three-, and four-year-ahead bankruptcy predictions.

Table 13. Term Structure Analysis of Bankruptcy Prediction Performance: A Firm Complexity Perspective

		Benchmark model: FIN				Benchmark model: FIN & TCV			
		Inclusion of COMPLEXITY				Inclusion of COMPLEXITY			
		1 year	2 year	3 year	4 year	1 year	2 year	3 year	4 year
Panel A. AUC Difference After Including Firm Complexity Variables									
Panel A.1.	BalancedBaggingClassifier								
CatBoost		0.0315	0.0224	0.0148	0.0096	0.0136	0.0087	0.0039	-0.0003
LGBM		0.0348	0.0259	0.0203	0.0161	0.0157	0.0099	0.0049	0.0025
RF		0.0254	0.0161	0.0091	0.0045	0.0130	0.0091	0.0048	0.0028
XGBoost		0.0452	0.0336	0.0267	0.0234	0.0182	0.0130	0.0094	0.0072
Panel A.2.	EasyEnsemble								
CatBoost		0.0301	0.0221	0.0153	0.0107	0.0136	0.0086	0.0035	-0.0003
LGBM		0.0388	0.0324	0.0257	0.0209	0.0137	0.0100	0.0057	0.0028
RF		0.0276	0.0202	0.0128	0.0074	0.0154	0.0113	0.0074	0.0050
XGBoost		0.0525	0.0429	0.0350	0.0294	0.0167	0.0126	0.0084	0.0053
Panel A.3.	RandomUnderSampling								
CatBoost		0.0359	0.0268	0.0212	0.0177	0.0117	0.0076	0.0029	0.0009
LGBM		0.0398	0.0333	0.0277	0.0231	0.0111	0.0083	0.0050	0.0034
RF		0.0339	0.0229	0.0166	0.0112	0.0150	0.0096	0.0049	0.0023
XGBoost		0.0556	0.0458	0.0367	0.0323	0.0159	0.0122	0.0073	0.0045
Panel B. F1-Score Difference After Including Firm Complexity Variables									
Panel B.1.	BalancedBaggingClassifier								
CatBoost		0.0232	0.0173	0.0130	0.0098	0.0071	0.0022	-0.0022	-0.0050
LGBM		0.0227	0.0168	0.0124	0.0094	0.0097	0.0040	0.0003	-0.0022
RF		0.0229	0.0150	0.0104	0.0071	0.0167	0.0072	0.0011	-0.0020
XGBoost		0.0284	0.0215	0.0164	0.0133	0.0154	0.0093	0.0057	0.0024
Panel B.2.	EasyEnsemble								
CatBoost		0.0152	0.0115	0.0081	0.0058	0.0032	0.0007	-0.0014	-0.0028
LGBM		0.0154	0.0127	0.0095	0.0077	0.0034	0.0010	-0.0008	-0.0027
RF		0.0125	0.0101	0.0073	0.0054	0.0027	-0.0005	-0.0024	-0.0030
XGBoost		0.0199	0.0168	0.0134	0.0109	0.0063	0.0040	0.0018	-0.0012
Panel B.3.	RandomUnderSampling								
CatBoost		0.0136	0.0101	0.0068	0.0052	0.0037	0.0012	-0.0012	-0.0024
LGBM		0.0119	0.0087	0.0063	0.0051	0.0018	0.0006	-0.0013	-0.0025
RF		0.0128	0.0090	0.0059	0.0044	0.0046	0.0015	-0.0004	-0.0011
XGBoost		0.0153	0.0124	0.0099	0.0086	0.0064	0.0045	0.0021	0.0013

Table 14. Feature Importance Analysis: A Firm Complexity Perspective

Random Forest			XGBoost		
Rank	Feature	IMP_AVG	Rank	Feature	IMP_AVG
1	EMV_Debt	0.0357	1	EBIT_Sales	0.0580
2	NWC_TA	0.0354	2	EMV_Debt	0.0369
3	RE_TA	0.0334	3	RE_TA	0.0340
4	COMPLEXITY_BC	0.0327	4	NWC_TA	0.0325
5	EBIT_Sales	0.0321	5	EBIT_TA	0.0322
6	SD_FINACC	0.0316	6	FK_Ease	0.0321
7	SD_BC	0.0314	7	COMPLEXITY_IGO	0.0311
8	SD_DC	0.0304	8	FT_MWeak_C	0.0311
9	SD_TOTAL	0.0299	9	COMPLEXITY_BC	0.0310
10	SD_IGO	0.0297	10	COMPLEXITY_DC	0.0308
11	A_WPP	0.0297	11	Sales_TA	0.0306
12	COMPLEXITY_DC	0.0296	12	SD_DC	0.0304
13	EBIT_TA	0.0295	13	SD_IGO	0.0304
14	TA_Growth	0.0294	14	SD_TOTAL	0.0298
15	FSIZE	0.0292	15	FT_UNC	0.0290
16	ROE_Chg	0.0291	16	A_WPP	0.0286
17	Sales_TA	0.0289	17	SD_FINACC	0.0285
18	CL	0.0288	18	FSIZE	0.0285
19	FK_Ease	0.0288	19	FT_NEG	0.0280
20	Harv_NEG	0.0287	20	Harv_NEG	0.0278
21	COMPLEXITY_IGO	0.0287	21	CL	0.0276
22	FT_NEG	0.0283	22	COMPLEXITY_FD	0.0276
23	FT_UNC	0.0281	23	FT_POS	0.0273
24	FT_POS	0.0280	24	COMPLEXITY_FINACC	0.0269
25	Emp_Growth	0.0279	25	SD_BC	0.0265
26	FT_MWeak_C	0.0278	26	SD_FD	0.0265
27	Sale_Growth	0.0276	27	Emp_Growth	0.0261
28	PB_Chg	0.0274	28	TA_Growth	0.0258
29	FT_LITI	0.0274	29	FT_MStr	0.0250
30	SD_FD	0.0273	30	COMPLEXITY_TOTAL	0.0249
31	FT_MStr	0.0273	31	FT_LITI	0.0247
32	COMPLEXITY_FINACC	0.0271	32	Sale_Growth	0.0246
33	COMPLEXITY_FD	0.0267	33	PB_Chg	0.0227
34	COMPLEXITY_TOTAL	0.0262	34	ROE_Chg	0.0224

Note: IMP_AVG represents the average importance score estimated using feature-importance measures from Random Forest and XGBoost. The value is computed as the moving average of importance scores across the training samples used for one-year-ahead bankruptcy predictions (i.e., using 1993–2018 to predict 2019, 1993–2019 to predict 2020, 1993–2020 to predict 2021, and 1993–2021 to predict 2022). Definitions of all variables are provided in Tables 1, 3, and 4.