Journal of Information-Based Decision Making

Journal homepage: www.jibdm.journal-publishing.org

ISSN: XXXX-XXXX

Asymmetric Effects of Corporate Governance on Financial Distress: A Granular Computing and Explainable Machine Learning Approach in Borsa Istanbul

Abdullah Kürşat Merter[1,*]

[1] Department of Business Administration, Faculty of Business Administration, Gebze Technical University, Kocaeli, Türkiye

ARTICLE INFO

Article history:
Received 11 May 2026
Received in revised form 22 June 2026
Accepted 23 August 2026
Available online 30 August 2026

Keywords:
Corporate Governance; Financial Distress; Explainable Machine Learning; XGBoost; SHAP; Granular Computing

ABSTRACT

We examine whether corporate governance indicators predict financial distress among 200 non-financial firms listed on Borsa Istanbul, using governance data from 2023 and financial outcomes from 2024, a year of severe macroeconomic stress in Türkiye. Naive designs inflate performance through label leakage and impose linearity on governance effects. We address both problems with a leakage-free nested cross-validation protocol, two distinct outcomes, and explainable machine learning. The first outcome is a narrowly defined new operating-loss onset, a firm that is not distressed in 2023 and reports negative operating profit in 2024 under the composite distress rule. The second outcome is the 2024 distress state conditional on the 2023 state and serves as a persistence benchmark. SHAP attributions are paired with formal segmented-regression threshold tests. The results identify a clear onset-persistence asymmetry. The onset models are estimated on the 155 firms at risk of onset, and under internal validation the onset classifiers produce mean AUC values between 0.54 and 0.65. The parsimonious specifications display moderate discrimination, but most interval estimates span 0.5 and the evidence remains weak. The onset outcome is entirely composed of new operating-loss cases in this sample. The conditional state model performs well, with AUC between 0.77 and 0.81, but its performance is dominated by the baseline distress state, whose average marginal effect is 0.413. The full governance block attains a grouped SHAP attribution of 0.130, far below the financial block. Dependence plots suggest threshold-like patterns, yet no formal break is detected, with no likelihood-ratio test reaching significance at the five percent level. An earlier naive specification reaches an AUC of 0.727 but changes the label, clustering and validation at the same time, so it is retained only to illustrate how evaluation design inflates performance. The study contributes by showing that separating distress onset from persistence and enforcing time-consistent validation yields a more credible basis for assessing the predictive value of governance measures in emerging-market firms.

1. Introduction

The separation of ownership from control creates the enduring tension of the modern corporation. Managers command information and discretion while dispersed shareholders bear the

* *Corresponding author.*
E-mail address: akmerter@gtu.edu.tr

residual risk, and agency theory assigns the board the task of narrowing that gap through internal control [1]. The monitoring account links board composition to the intensity of oversight and the conservatism of financial policy [2]. Whether such oversight protects firms when the macroeconomic environment deteriorates is a harder question. Pandemic era evidence indicates that corporate resilience was largely written in the balance sheet before the shock arrived [3], and that corporate social responsibility did not shield crisis period returns [4]. Board level evidence, by contrast, ties female directorships to more conservative financing and lower distress exposure [5]. Whether internal governance protects under economy wide stress remains unsettled.

Distress prediction is among the oldest quantitative problems in corporate finance, and its canonical instruments impose linearity in an index of ratios [6]. The linearity habit flattens exactly the patterns theory leads one to expect, because a liquidity ratio may matter only below a floor and a governance share only beyond a critical mass, while a single slope averages these regimes into irrelevance. Machine learning relaxes the functional form, and applied literature has grown rapidly. Systematic reviews now count hundreds of studies reporting discrimination that classical methods never approached [7], with applications extending to transactional data for small and medium enterprises [8] and to the communicative content of annual report text [9]. Reported performance, however, is only as credible as the evaluation that produced it.

A quieter weakness concerns evaluation. Target leakage, the use of information that would not exist at prediction time, inflates measured skill whenever overlapping temporal labels let the outcome window bleed into the predictor window or model selection touches the test data. Reproducibility audits document leakage invalidating published claims across machine learning based science at scale [10], and consensus recommendations now isolate evaluation from every modeling choice [11]. Finance has issued a parallel warning from within, since multiple testing alone generates many published factors [12]. Evaluation metrics add a further illusion. Distress studies increasingly use class imbalanced samples and expanded feature systems [13], a regime in which accuracy flatters weak models, and rankings reverse once the socio-economic costs of errors enter the assessment [14]. The reasonable prior for any distress study is suspicion of its own headline number.

Türkiye offers a demanding setting for confronting these questions. The economy moved through 2023 and 2024 under an exceptionally tight monetary stance, with the policy rate lifted to 50 percent by March 2024 and held there for most of the year [15]. Consumer price inflation closed in 2024 at 44 percent [16], the lira continued its long depreciation, and interest expenses on lira debt reset at levels borrowers had never absorbed. Among listed non-financial firms the share meeting a standard distress definition rose from 22.5 percent to 31.5 percent across the two years. Turkish boards entered this shock after a decade of gradual change, with female representation and committee adoption strengthening only recently [17] and governance index dimensions relating to performance in heterogeneous ways [18]. Nonconformity to global governance norms is patterned rather than random, so attributes cluster where codes permit explanation in place of compliance [19], and governance attributes link to distress risk across Asian emerging markets [20]. A shock of this scale turns the listed sector into a live test of whether pre shock governance variation carries survival content.

The present study reads that test through a leakage free predictive design. We assemble a governance panel for 200 nonfinancial firms listed on Borsa Istanbul, rebuilding every governance ratio from raw board counts in compliance reports and matching them to financial statements for both years of stress. Prediction proceeds under nested cross validation, so that model selection occurs only in inner folds and evaluation only in untouched outer folds. An XGBoost gradient boosting classifier paired with Shapley value attribution serves as the explanatory engine [21], [22], benchmarked against penalized and spline logits and two further ensembles. We estimate two distinct tasks, a new operating-loss onset task among firms not classified as distressed in 2023, and a conditional state task that classifies 2024 distress with the 2023 state included as a predictor. The contrast between a naive pooled design and this clean protocol measures how apparent predictability the evaluation design itself manufactures.

Four contributions follow from this design. The study imports the leakage discipline of the reproducibility literature into governance and distress research and quantifies the skill a pooled design invents, which vanishes once evaluation is honest. It establishes an internal-validation performance baseline for new operating-loss onset in an emerging market under stress, showing that the observed 2023 governance and financial variables provide weak discrimination for this outcome. It moves the question of nonlinear thresholds from visual interpretation toward formal segmented and spline tests specified before inspection, so that apparent cliffs must survive a disciplined null before they inform rules. It shows how the comply-or-explain regime compresses observable governance variation and thereby limits the governance signal available to the classifiers; the explainability layer identifies which variables drive the persistence benchmark. Explainable machine learning is used as a measurement layer that decomposes the fitted persistence benchmark and separates stable financial signals from weak governance contributions.

2. Theoretical Background

2.1 Corporate Governance Mechanisms Under Macroeconomic Stress

Agency theory treats the board as the principal internal device for mitigating the conflict between owners and managers, and the case for monitoring rests on the discretion that the separation of ownership and control leaves in managerial hands [23]. Monitoring matters most when external finance turns hostile, because managers then face stronger incentives to shift risk to claimholders or to postpone the recognition of losses. Independence and gender diversity are the two board attributes most often expected to strengthen oversight. Board level evidence links female directorships to closer attendance, tougher questioning, and more conservative financial policies [2], while the average association between women on boards and firm performance is modest and dependent on institutional context [24], [25]. For distress specifically, female directorships connect to more conservative capital structures that lower distress exposure [5].

The crisis record qualifies any simple protective account. During the pandemic crash, stocks with strong environmental and social ratings held value relative to peers [26], yet a broad cross-country test finds that corporate social responsibility did not protect crisis period returns once the shock became systemic [4]. The most direct governance test remains Fich and Slezak [27], who asked whether governance could save distressed firms from bankruptcy and found protection wanting. The credible reading of this literature is that whatever protection internal governance offers is conditional and slow moving, entangled with the financial structure it is supposed to discipline. Governance may constrain discretion in normal times without possessing the power to rescue firms in abnormal ones.

A second complication is measurement. Where codes operate on a comply or explain basis, compliance clusters at the minimum, and measured attributes display little dispersion precisely where monitoring theories predict their largest effects. Code compliance of this kind can reflect subjection rather than commitment [28], and reviews of emerging market governance reach a similar conclusion, since concentrated ownership and weak enforcement leave formal structures loosely coupled to practice [17]. Turkish evidence fits this pattern. Insider dominated board structures have been associated with weaker cross-sectional performance [29], and governance quality is priced but converges only gradually [30]. The binding Turkish requirement is a floor of two independent members, while the one third share operates on a comply or explain basis, so observed independence ratios concentrate at one third. This compression is not a data defect to be cleaned away. It is the institutional reality any predictive design must confront, and it disciplines expectations about what cross sectional governance variation can deliver.

2.2 The Target Leakage Problem in Predictive Modeling

Target leakage is the use of predictors whose values would not be knowable at the moment a prediction must be made. Its mechanisms are quiet. Overlapping temporal labels let the outcome window bleed into the predictor window, random folds place near identical firms on both sides of a split and let sector risk stand in for foresight, tuning on the full sample peeks at the test distribution, and survivorship removes the failures a model claims to foresee. A census of machine learning based science finds leakage of exactly these kinds invalidating published claims at scale, and it locates the source in evaluation design rather than in any single algorithm [10]. The bankruptcy literature is exposed to the same diagnosis. Reported discrimination is high across hundreds of studies, yet evaluation practices are heterogeneous and frequently incomparable across papers, so the headline figures resist accumulation into a reliable record [7]. Benchmarks that hold the protocol constant tell a sobering story, because algorithm rankings compress once the evaluation design is standardized [31], [32]. Flexible methods do deliver genuine gains where protocols respect time order, as large scale equity evidence shows [33], [34], but the gains belong to the design as much as to the learner.

Evaluation metrics impose distortions of their own. Accuracy and related scores mislead under class imbalance, because a trivial classifier performs well at low prevalence [35], and the Matthews correlation coefficient addresses several of these failures more directly [36]. Ranking metrics built for return prediction mislead in analogous ways when applied to distress [37]. Calibration, the agreement between predicted probabilities and observed frequencies, is a separate requirement that discrimination measures do not capture [38]. Small event counts destabilize both estimation and validation, and familiar rules of thumb for events per variable lack justification [39]. Reporting standards now codify these requirements, asking for interval estimates, calibration assessment, and full protocol disclosure [40]. A protocol that honors this standard treat evaluation design as part of the econometric model rather than as packaging.

2.3 Explainable AI and Nonlinear Thresholds

Linear index models impose linearity in the log odds, and that restriction mechanically masks the two patterns most plausible in distress data, thresholds and saturation. A liquidity ratio may protect only until a floor is reached, and board independence may matter only beyond a critical mass of outside voices, while a single slope averages these regimes into a report of no effect. Gradient boosted trees relax the functional form by construction [21], and Shapley value attribution decomposes each prediction into input contributions with a consistency guarantee drawn from cooperative game theory [22]. The combination has become the working standard in credit risk management [41], with extensions that weight attributions for regulatory reporting [42], construct feature weighted counterfactual explanations for bankruptcy predictions [43], and assemble models, visualizations, and summary explanations into holistic lending systems [44]. Applications to distress prediction confirm that boosted trees recover structure that parametric models miss and that the recovered variable rankings are stable enough to interpret [45-48].

A persistent argument holds that explanation and estimation are distinct tasks, and that interpretable models can match black box performance in high-stakes decisions, so accuracy should not be purchased at the price of opacity [49]. What remains unexamined in most applications is the inferential status of the patterns the opened box displays. A dependence plot can suggest a cliff or a saturation point, but the visual pattern is not a test, and attribution methods inherit the correlations of the data they summarize. The present study treats explainability output as hypothesis generating and moves the threshold question to formal segmented and spline tests specified before inspection of results. Where the literature admires nonlinear patterns, we ask whether the interpreted structure survives a disciplined null. That shift, from displaying nonlinearity to testing it, is what allows explainable machine learning to inform regulation rather than decorating prediction.

Two gaps remain. First, almost no study examines governance variables through the attribution lens in an emerging market under acute macroeconomic stress. Second, the heterogeneity of governance effects across firm groups is unaddressed, because existing applications estimate a single global model. This paper fills both gaps by combining granular segmentation with SHAP-based dependence analysis, which leads to one average-effect hypothesis and a two-stage hypothesis on non-linearity.

H1. Board independence and female representation measured in 2023 are negatively associated with the probability of financial distress realized in 2024, conditional on financial controls.

The non-linearity hypothesis is tested in two stages. The first stage asks only whether the relations between governance variables and distress probability depart from linearity at all, without committing to where any departure occurs. The second stage asks whether such a departure concentrates at identifiable threshold points, so that a change in sign or slope can be located within the observed range of the variable. Separating the two stages matters because visual evidence of curvature does not by itself establish that a specific kink exists.

H2a. The relations between governance variables and distress probability contain non-linear components that a linear specification does not capture.

H2b. The non-linear components take the form of sign or slope changes at identifiable threshold points, so that the average linear effect understates the true effect in specific ranges of the variable.

3. Data and Methodology

3.1 Sample Selection and Data Sources

The empirical setting of this study is the Borsa Istanbul main market during a period of acute macroeconomic tightening. We apply the inclusion criteria shown in Table 1 and obtain a final sample of 200 non-financial firms. The initial population encompasses all publicly traded firms at the end of the 2023 fiscal year. We exclude financial institutions along with banks and insurance companies and real estate investment trusts consistent with standard conventions in empirical corporate finance [23]. These entities operate under fundamentally different regulatory capital constraints and statutory accounting rules that confound the operational distress metrics applicable to industrial firms. We require the remaining firms to possess complete and continuous financial statements spanning the 2022 to 2024 observation window alongside accessible corporate governance disclosures.

Table 1
Sample attrition process

Description	Number of Firms
Initial universe of firms listed on Borsa Istanbul (as of 2023)	585
Less: Financial institutions, banks, insurance companies, and REITs	343
Less: Firms with missing continuous financial data (2022-2024)	32
Less: Firms with missing corporate governance reports	10
Final cross-sectional sample of non-financial firms	200

The dataset combines firm-level accounting data with governance disclosures. Financial statements and ratios for fiscal years 2022 to 2024 were obtained from FINNET, while governance measures were coded from corporate governance reports filed on the Public Disclosure Platform.

Journal of Information-Based Decision Making
Volume 1, Issue 1 (2026) 1-30

Turkish listed firms report under TMS 29, the Turkish adoption of IAS 29 for financial reporting in hyperinflationary economies. Accordingly, the financial inputs used here are expressed in end-of-period purchasing power, placing the ratios on a common price basis for cross-sectional comparison.

Commercial databases often provide coarse or missing coding of internal governance mechanisms in emerging markets operating under mandatory comply or explain regimes. We relied on a proprietary and hand collected corporate governance dataset to overcome this aggregation bias and establish strict construct validity. Board characteristics and committee structures along with auditor information were extracted directly from the official annual reports and corporate governance compliance reports filed on the Public Disclosure Platform of Türkiye. This manual curation allows us to rebuild every governance ratio from raw and verified counts rather than secondary proxies. The approach ensures high fidelity to the true governance environment prevailing strictly before the 2024 macroeconomic shock.

3.2 Target Outcomes and Predictive Modeling Framework

We formulate a strict composite definition to capture financial distress for the predictive tasks. A firm is classified as distressed each year if it meets any of three criteria. The first criterion is an Altman Emerging Market Z score below the critical threshold of 1.1 [50]. The second criterion is the reporting of negative operating profit. The third criterion is the reporting of negative book equity. The original 1.1 Z score threshold was calibrated on non restated statements. The composite rule reduces reliance on any single accounting criterion.

We estimate two target outcomes to separate distress onset from persistence. The conditional-state model predicts 2024 distress using 2023 information and includes the 2023 distress state. The onset model is estimated only among firms that were not distressed in 2023. A pooled label that combines prior distress components with 2024 outcomes would not represent the same prediction task.

3.3 Variable Measurement

Table 2 defines the variables and their measurement periods. All explanatory variables are observable by the end of 2023 and thus available when a prediction for 2024 would be made. Sales growth enters only in its 2022 to 2023 form, because the later rate cannot be known at prediction time and would leak outcome period information.

Table 2

Definitions, measurement periods and transformations of distress outcomes and predictor variables

Variable	Definition	Period
State24	Binary. Equals 1 if EM Z score below 1.1, negative operating profit or negative equity in 2024	2024
Onset24	Binary. Equals 1 if distressed in 2024 and not distressed in 2023	2023 to 2024
State23	Binary baseline state. Equals 1 if distressed in 2023	2023
EMZ score 2024	Continuous score defined in equation 1	2024
Board independence	Share of independent members on the board	2023
Female directors	Share of female members on the board	2023
CEO duality	Binary. Equals 1 if the CEO also chairs the board	2023
Board size	Natural logarithm of the number of board members	2023
Foreign directors	Share of foreign national members on the board	2023
Executive directors	Share of executive members on the board	2023
Board meetings	Natural logarithm of the number of board meetings	2023

Table 2
Conitnued

Variable	Definition	Period
Auditor change	Binary. Equals 1 if the external auditor changed during the year	2023
Auditor tenure	Consecutive years served by the incumbent auditor, descriptive only	2023
Audit quality	Binary. Equals 1 if the auditor is a Big 4 affiliate, descriptive only	2023
Firm size	Natural logarithm of total assets	2023
Leverage	Total liabilities over total assets	2023
Liquidity	Current ratio, current assets over current liabilities	2023
ROA	Net income over total assets	2023
Sales growth	Sales growth from 2022 to 2023	2022 to 2023

Note: EMZ = Emerging Market Z score. Continuous variables are winsorized at the 1st and 99th percentiles.

The independence ratio equals the independent member count divided by board size, and recomputation from the raw counts matches them exactly. Every firm has at least two independent members, the binding floor under the CMB Corporate Governance Communiqué, while the one third share operates on a comply or explain basis. Ratios below one third occur only in large boards meeting the floor but not the share, compliant outcomes rather than coding errors.

The 2024 Altman Emerging Market Z score serves both as a distress criterion and as a continuous outcome.

$$EMZ_{i,2024} = 3.25 + 6.56X_{1i} + 3.26X_{2i} + 6.72X_{3i} + 1.05X_{4i} \tag{1}$$

In Eq. (1), X_1 is working capital over total assets, X2 retained earnings over total assets, X3 operating profit over total assets and X4 book equity over total liabilities, all measured in 2024.

3.4 Granular Segmentation

Firms are grouped into information granules to examine model behavior within homogeneous subsets. Granules are built with K means clustering on standardized 2023 size, leverage and liquidity. The silhouette coefficient and the Calinski Harabasz index both favor k equal to 4 over the grid from 2 to 5, but k equal to 3 is retained for interpretability. k equal to 4 sensitivity split leaves the risk ordering unchanged.

The granules play a descriptive role and are excluded from every predictor set. The onset machine-learning models use the 13 firm-level predictors defined in Table 2, whereas the state models use the same predictors plus State23, for 14 predictors in total. A reduced penalized logit with six prespecified predictors, a fixed L2 penalty and no inner tuning is evaluated on the same outer folds as a parsimony sensitivity. Granules appear only in the explainability stage, where pooled SHAP attributions are compared across granules as a heterogeneity diagnostic.

3.5 Baseline Econometric Models

The baseline binary model is a logistic regression estimated by maximum likelihood [51].

$$\Pr(D_i = 1) = \Lambda(\beta_0 + x_i^{\top}\beta) \tag{2}$$

In Eq. (2), Lambda is the logistic cumulative distribution function and the outcome D is either Onset24 or State24. The onset specification uses eight governance variables and four financial controls, yielding 12 predictors for 27 onset events. The state specification adds State23 to these variables, yielding 13 predictors for 63 state events. Pre-distress sales growth is included only in the machine-learning feature set, giving 13 onset predictors and 14 state predictors. A ridge penalized logit and a spline logit with restricted cubic splines of four degrees of freedom on ROA, leverage and liquidity accompany the plain logit.

Propensity score matching for new onset distress in 2024 supplies a comparability check. Matching is restricted to 155 at risk firms, because onset is mechanically impossible for firms already distressed in 2023. Two treatments are defined at the at-risk medians, board independence above 0.333 and female representation above 0.182. Propensity scores come from firm size, leverage, liquidity, ROA, board size and CEO duality, and matching is nearest neighbor without replacement under a caliper of 0.05. The average treatment effect on the treated carries a bootstrap 95 percent confidence interval, sensitivity to unobserved confounding is bounded by the Rosenbaum sign test at Gamma from 1.0 to 2.0, and balance is displayed in the Love plot.

The continuous outcome counterpart regresses the 2024 EM Z score by ordinary least squares, asking whether variables silent at the binary threshold inform the soundness level. Because the score is heavy tailed, Huber and median quantile regressions accompany least squares. The maximum Cook's distance is 0.138, with 15 firms above the conventional 4/n threshold.

3.6 Machine Learning Protocol and Explainability

Six classifiers are compared under nested cross validation, a stratified 5-fold outer loop repeated 10 times giving 50 out of fold partitions. Only XGBoost is tuned in the inner loop, through a 4-fold grid search over depth 2 or 3, tree counts of 150 or 300, learning rates of 0.03 or 0.05 and L2 penalties of 1 or 5 [21]. The remaining learners use fixed configurations, a random forest of 300 trees, depth 4, minimum leaf 8 and balanced subsample, CatBoost with 300 iterations, depth 4, learning rate 0.05 and balanced weights [52], a penalized logit with L2 penalty C equal to 0.5, and the spline logit of Section 3.4. All preprocessing is refit inside each outer training fold, so no validation fold information enters any transformation. This nesting matters because model selection inside the evaluation folds biases error estimates downward [53].

The 50 outer folds are not independent, since each firm appears in about 10 test partitions, so fold level dispersion understates sampling variability. We report fold-level percentile intervals as measures of between-partition dispersion and firm-level bootstrap intervals as uncertainty intervals computed from the final firm-level prediction unit. Performance is summarized by ROC and precision recall areas, F1 and the Brier score, with PR AUC read against the prevalence base rates of 0.135 for onset and 0.315 for the state, which a classifier without skill attains at prevalence.

Calibration of the out of fold probabilities uses a logistic recalibration, whose intercept and slope are reported, the Brier score against the variance of the baseline rate, and a split half isotonic check. A decision table reports sensitivity, specificity, precision and the flagged share at thresholds of 0.20, the state prevalence of 0.313 and 0.50. The calibration intercept and slope are interpreted from the probabilities delivered by the fitted pipeline, while the ranking metrics and threshold-specific operating points are reported separately. Reporting follows the TRIPOD+AI guidance [40].

Explanations use SHAP values [22].

$$g(x_i) = \phi_0 + \sum_{j=1}^{p} \phi_{ij} \tag{3}$$

In Eq. (3) the attributions sit on the log odds scale and come from the interventional TreeSHAP algorithm, so XGBoost serves as the explanation vehicle [54]. Global importance is measured by mean absolute SHAP values, grouped SHAP aggregates attributions into the baseline state, the financial block, sales growth and the governance block, and permutation importance records the AUC loss when each variable is shuffled, with rank stability over 100 bootstrap refits. Because predictors are correlated, grouped and permutation measures are read alongside the raw values [55-59].

Formal threshold tests complete the design. Segmented logits with a single knot are fitted on the onset outcome for seven prespecified predictors, ROA, leverage, liquidity, female directors, board independence, executive directors and board size, with 200 bootstrap resamples for each knot interval. Likelihood ratio tests compare each segmented specification against the linear one, and a second family compares spline against linear logits for the same predictors, testing the existence of non-linearity rather than its location. The event counts of 27 and 63 limit power throughout, and we do not appeal to a mechanical events per variable rule, whose rationale has been questioned [39].

4. Empirical Results

This section reports the findings in six steps, moving from descriptive evidence to baseline logits, matching diagnostics, machine learning performance, explainability and formal non-linearity tests. The analysis comprises 200 non-financial firms.

4.1 Descriptive evidence

The 2024 distress indicator combines three criteria. A firm is distressed if its Altman Emerging Market Z-score falls below 1.1, or if it reports negative operating profit in 2024, or if it reports negative equity in 2024. Only 8 firms fall below 1.1 by the Z-score alone, and no firm reports negative equity, so the composite definition carries the classification. The transition matrix describes the stress episode. Of the 200 firms, 128 are never distressed, 36 are persistently distressed in both years, 9 resolve their 2023 distress by 2024, and 27 enter distress for the first time. The distressed share rises from 22.5 percent in 2023 (45 firms) to 31.5 percent in 2024 (63 firms). The 27 cases, representing 13.5 percent of the sample, constitute the new operating-loss onset outcome examined by the onset models. Table 3 reports descriptive statistics for the full sample.

Table 3
Descriptive statistics

Variable	N	Mean	SD	Min	Median	Max
Board independence	200	0.346	0.062	0.167	0.333	0.500
Female directors	200	0.192	0.159	0.000	0.167	0.600
CEO duality	200	0.495	0.501	0.000	0.000	1.000
Board size	200	1.931	0.274	1.609	1.946	2.639
Foreign directors	200	0.070	0.153	0.000	0.000	0.600
Executive directors	200	0.224	0.199	0.000	0.200	0.667
Board meetings	200	2.774	0.817	0.000	2.890	4.511
Auditor change	200	0.230	0.422	0.000	0.000	1.000
Auditor tenure	200	3.675	2.178	1.000	4.000	8.000
Audit quality	200	0.455	0.499	0.000	0.000	1.000
State23	200	0.225	0.419	0.000	0.000	1.000
Firm size	200	22.451	1.777	18.699	22.146	26.460
Leverage	200	0.448	0.219	0.068	0.422	0.961
Liquidity	200	1.849	1.833	0.307	1.289	11.755
ROA	200	0.059	0.103	-0.281	0.063	0.291
Sales growth 2022-2023	200	0.905	0.625	-0.259	0.903	3.115
EM Z-score 2024	200	7.881	18.774	-30.091	6.130	262.083

Note: SD = standard deviation. Sales growth is measured over 2022–2023.

Table 3 conveys the structure of the data. Board independence clusters tightly around the regulatory recommendation, with a mean of 0.346, a median of 0.333 and a standard deviation of 0.062, so cross-sectional variation in this variable is narrow. Female representation averages 0.192, CEO duality is present in 49.5 percent of firms and the mean board has about 7.2 members on the raw scale. Leverage averages 0.448 and ROA averages 0.059, with a loss-making tail reaching -0.281. Pre-distress sales growth from 2022 to 2023 averages 0.905, consistent with nominal expansion under high inflation. The EM Z-score is strongly skewed, with a mean of 7.881 against a median of 6.130, a pattern typical of book-based distress scores in samples containing low-liability firms. Table 4 reports the profile of each information granule under the retained k=3 partition.

Table 4
Information-granule profiles

Granule	n	Description	Mean size	Mean leverage	New-onset rate
G0	8	Ultra-liquid micro firms, mean current ratio 9.35	21.11	0.147	0.125
G1	91	Small low-leverage firms	21.50	0.293	0.099
G2	101	Large high-leverage firms	23.42	0.611	0.168

Note: Granules are obtained from K-means clustering of standardized 2023 firm size, leverage, and liquidity. The retained solution contains three clusters.

Table 4 shows an economically readable partition. G0 isolates eight ultra-liquid micro firms with a mean current ratio of 9.35, a small set whose extreme liquidity drives its separation under any k. G1 contains 91 small, conservatively financed firms with mean log assets of 21.50 and mean leverage of 0.293, and its new-onset rate is 9.9 percent. G2 contains 101 large, highly leveraged firms with mean log assets of 23.42 and mean leverage of 0.611, and its new-onset rate rises to 16.8 percent. The ordering of the new-onset rates follows the leverage gradient across the granules, which supports the face validity of the partition. The granules remain a descriptive device.

Table 5 reports Welch mean comparisons between the 27 onset firms and the 128 never-distressed firms. The pattern in Table 4 runs against standard priors. Firms entering distress in 2024 carried significantly higher 2023 leverage than never-distressed firms, 0.542 against 0.433. Profitability does not separate the groups, with onset ROA of 0.043 against 0.062. The executive-director share is higher among onset firms, 0.293 against 0.213, and the gap is only suggestive. Firm size is essentially identical. Read together, 2023 observables discriminate weakly between firms that will newly fail and firms that will remain sound, which anticipates the difficulty the onset models face below.

Table 5
Univariate comparisons by new-onset distress status

Variable	Onset mean	Never-distressed mean	t	p
Leverage	0.542	0.433	-2.36	0.024
ROA	0.043	0.062	1.04	0.305
Executive directors	0.293	0.213	-1.66	0.107
Firm size	22.455	22.451	-0.01	0.989

Note: Welch's t tests compare 2023 characteristics between firms with and without new-onset distress. The onset group contains 27 firms and the comparison group contains 128 firms. Negative t values indicate higher means in the onset group.

Figure 1 shows that new-onset distress is spread across the real sectors rather than concentrated in one activity. Industrial and technology firms share an onset rate of 13.3 percent and services stand at 10.9 percent. The 30.0 percent rate in the residual category rests on a small base and should not be over-interpreted. The transition panel clarifies the composition of the 2024 pool. Persistent cases, 36 firms, outnumber new onsets, 27 firms, and resolved cases, 9 firms, are rare. The 2024 environment, with the policy rate at 50 percent and inflation near 44 percent, propagated existing distress more than it created new cases. A persistence benchmark that conditions on the 2023 state therefore mechanically captures the largest block of 2024 distress.

Journal of Information-Based Decision Making
Volume 1, Issue 1 (2026) 1-30

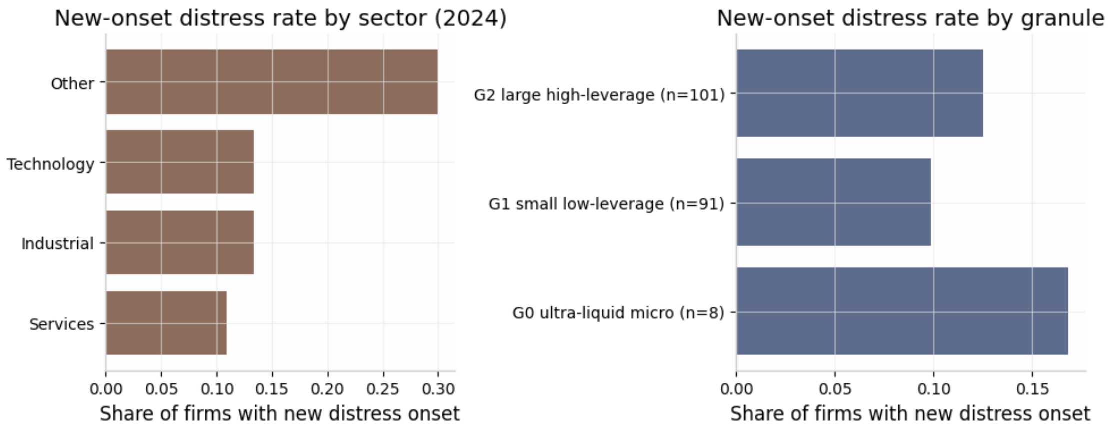

Fig. 1. Sector-specific new-onset distress rates and 2023–2024 distress transitions

The heatmap in Figure 2 shows moderate collinearity in a few pairs. Board size correlates with firm size at 0.48, as larger firms maintain larger boards. Board independence correlates with board size at -0.51, reflecting the two-member regulatory floor under the CMB Corporate Governance Communiqué. Large boards satisfy the floor with two independent members while missing the one-third comply-or-explain share, and the ratios of 0.143 to 0.182 correspond to two independent members on boards of 11 to 14. Leverage and the current ratio correlate at -0.45, and foreign directors with meeting frequency at -0.41. No correlation threatens estimation, but the independence-size link cautions that part of what an independence coefficient absorbs is board scale.

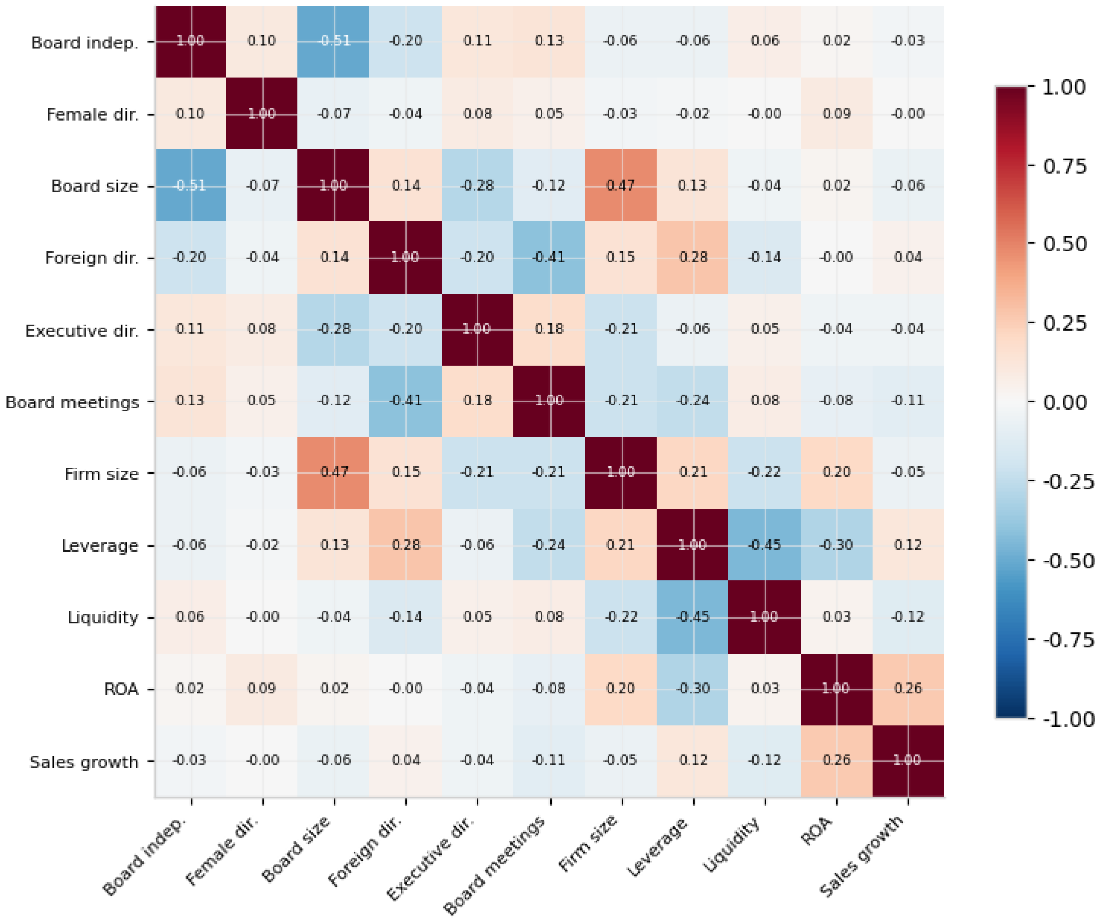

Fig. 2. Correlation matrix of governance and financial predictors

4.2 Baseline estimates

The onset model in Table 6 is estimated on the 155 firms at risk of onset and is weak by every summary measure. The pseudo R2 is 0.118, and the joint test does not reject an empty model, with a model p-value of 0.151. No coefficient reaches significance even at the 10 percent level. The two closest approaches are the executive-director share, with a coefficient of 1.739 and p = 0.156, and leverage, with a coefficient of 1.816 and p = 0.172. Profitability is unrelated to onset, with a p-value of 0.203 for ROA. The executive-director sign is adverse rather than protective, which contradicts the monitoring intuition behind H1, and the imprecision precludes any firmer conclusion. Hypothesis H1, which posits that stronger governance protects against distress, receives no support in the onset equation, and the at-risk estimation strengthens that null result.

Table 6
Logistic regression estimates for new-onset distress

Variable	Coefficient	SE	p-value	AME
Board independence	−1.664	4.861	0.732	−0.211
Female directors	0.290	1.480	0.845	0.037
CEO duality	−0.092	0.500	0.853	−0.012
Board size	−0.148	1.294	0.909	−0.019
Foreign directors	1.645	1.547	0.287	0.209
Executive directors	1.739	1.227	0.156	0.221
Board meetings	0.061	0.336	0.855	0.008
Auditor change	−0.569	0.651	0.382	−0.072
Firm size	−0.102	0.171	0.553	−0.013
Leverage	1.816	1.330	0.172	0.231
Liquidity	−0.145	0.242	0.550	−0.018
ROA	−4.032	3.165	0.203	−0.512
Observations	155			
Onset events	27			
McFadden pseudo-R^2	0.118			
LR $\chi^2(12)$	16.95			
Model p-value	0.151			
AIC	152.4			
BIC	192.0			

Note: AME = average marginal effect.

Table 7 reports the conditional state model, which explains the 2024 distress state using the 2023 state and the same covariates. Firms distressed in 2023 face a coefficient of 3.034, a marginal effect of 0.413 on the probability of remaining distressed. This single variable drives the fit, and the pseudo R2 of 0.303 triples that of the onset model. ROA enters with the expected negative sign at -4.044 but is imprecise, and its marginal effect of -0.550 should be read with caution. Every governance coefficient is insignificant, with p-values above 0.14 throughout. The contrast between Tables 5 and 6 frames is the paper’s central asymmetry. Distress is highly predictable from its own lag, while new onset is poorly predicted by the observed covariates, financial or governance related.

The continuous-outcome results in Table 8 are stronger than the binary onset results, because the Z-score exploits variation in the onset indicator discards. Leverage and ROA are highly significant under all three estimators, with leverage coefficients between -7.55 and -10.52 and ROA coefficients between 7.15 and 13.14. The governance coefficients are fragile. The executive-director share is significant under OLS and Huber but collapses with the median, so the association is carried by conditional means. Foreign directors follow the same pattern, significantly only at 10 percent under OLS and Huber. Liquidity turns significant only once outliers are downweighted. The governance signal in the continuous outcome is real but delicate, and it does not survive into the binary onset margin.

Table 7
Logistic regression estimates for conditional distress-state membership

Variable	Coefficient	SE	p-value	AME
Board independence	−3.581	3.967	0.367	−0.487
Female directors	0.206	1.236	0.867	0.028
CEO duality	0.335	0.424	0.429	0.046
Board size	0.064	1.062	0.952	0.009
Foreign directors	1.023	1.429	0.474	0.139
Executive directors	1.054	1.013	0.298	0.143
Board meetings	0.404	0.280	0.148	0.055
Auditor change	−0.666	0.527	0.206	−0.091
Firm size	0.038	0.142	0.791	0.005
Leverage	1.077	1.055	0.307	0.147
Liquidity	0.019	0.114	0.864	0.003
ROA	−4.044	2.464	0.101	−0.550
State23	3.034	0.524	<0.001	0.413
Observations	200			
Distressed firms	63			
McFadden pseudo-R^2	0.303			
LR $\chi^2(13)$	75.56			
Model p-value	<0.001			
AIC	201.7			
BIC	247.8			

Note: State23 denotes the 2023 distress-state indicator. AME = average marginal effect.

Table 8
Estimates for the 2024 Emerging Market Z score

Variable	OLS coefficient	OLS p-value	Huber coefficient	Huber p-value	Median coefficient	Median p-value
Leverage	−10.52	<0.001	−9.51	<0.001	−7.55	<0.001
ROA	13.14	<0.001	9.41	<0.001	7.15	<0.001
Executive directors	2.70	0.033	2.18	0.006	0.85	0.322
Foreign directors	2.72	0.060	2.03	0.076	0.52	0.675
Firm size	−0.17	0.290	−0.09	0.386	−0.15	0.199
Current ratio	0.20	0.194	0.25	0.007	0.31	0.002
Observations	200		200		200	
R^2	0.577					
Adjusted R^2	0.549					
Overall model test	$F(12,187) = 15.25$		Wald $\chi^2(12) = 343.12$		Wald $\chi^2(12) = 208.33$	
Model p-value	<0.001		<0.001		<0.001	
Scale or residual diagnostic	Max Cook's D = 0.138		Huber scale = 2.015		Pseudo R^2 = 0.371	
Governance controls	Yes		Yes		Yes	
Financial controls	Yes		Yes		Yes	

Note: OLS = ordinary least squares; Huber = Huber robust regression; Median = median quantile regression; R^2 = coefficient of determination.

4.3 Propensity score matching diagnostics

Table 9 reports matching diagnostics for two governance treatments, board independence above the at-risk median of 0.333 and female-director representation above the at-risk median of 0.182. The outcome in both exercises is new-onset distress in 2024. The matching sample is restricted to the 155 firms at risk of onset, namely the 27 firms that entered distress in 2024 and the 128 never-distressed firms. The restriction is necessary because an onset event is mechanically impossible for the 45 firms already distressed in 2023, and a full-sample design would match treated and control firms from a pool that does not correspond to the risk set. Matching is one-to-one nearest-neighbor matching without replacement on the propensity score, with a caliper of 0.05, and treatment thresholds are the at-risk-sample medians defined before outcome comparisons. The ATT bootstrap resamples matched-pair differences 2,000 times. Balance is judged against the conventional 0.10 criterion for absolute standardized mean differences, with 0.20 as a lenient bound. Panel A reports ATT estimates and matched-sample characteristics, Panels B and C report covariate balance separately for the two treatments, and Panel D reports Rosenbaum sensitivity results.

Table 9

Propensity-score matching estimates and diagnostics for 2024 new-onset distress in the at-risk sample

Panel A. Average treatment effects and matched-sample characteristics

Treatment	Treated N	Control N	Matched pairs	Off-support	ATT	95% CI	p-value
Board independence above median	68	87	31	9	−0.097	[−0.290, 0.065]	0.325
Female directors above median	77	78	39	5	0.051	[−0.154, 0.231]	0.600

Panel B. Covariate balance for the board-independence treatment

Covariate	Before SMD	After SMD
Firm size	−0.233	0.113
Leverage	−0.168	0.052
Current ratio	0.117	0.062
ROA	0.163	−0.152
Board size	−0.764	−0.010
CEO duality	0.083	−0.127

Panel C. Covariate balance for the female-director treatment

Covariate	Before SMD	After SMD
Firm size	−0.216	−0.063
Leverage	−0.093	−0.053
Current ratio	0.162	0.131
ROA	0.215	−0.092
Board size	−0.311	−0.223
CEO duality	0.247	−0.310

Panel D. Rosenbaum sensitivity analysis

Treatment	Γ = 1.00	Γ = 1.25	Γ = 1.50	Γ = 1.75	Γ = 2.00
Board independence above median	0.910	0.954	0.975	0.986	0.992
Female directors above median	0.395	0.563	0.693	0.786	0.851

Note: The at-risk sample comprises 155 firms. Treatment indicators equal one for firms above the at-risk median. ATT denotes the average treatment effect on the treated, SMD denotes the standardized mean difference, and Γ is Rosenbaum's sensitivity parameter. Confidence intervals are based on 2,000 firm-level bootstrap resamples. Matching uses one-to-one nearest-neighbor propensity scores without replacement with a caliper of 0.05.

Once the design is restricted to the appropriate risk set, matching substantially reduces the pre-treatment imbalance for the board-independence treatment, with the large board-size difference of −0.764 falling to −0.010. Residual imbalance remains for firm size, ROA, and CEO duality under the 0.10 criterion, although all post-matching differences for this treatment remain within the lenient 0.20 bound. The paired contrasts are therefore interpreted as adjusted associations rather than as evidence from a balanced quasi-experiment. The board-independence ATT is −0.097, with a bootstrap 95% confidence interval of [−0.290, 0.065] and p = 0.325. The point estimate is negative and therefore consistent with a protective direction, but it is statistically indistinguishable from zero. The female-director treatment yields an ATT of 0.051. For this treatment, current ratio, board size, and CEO duality remain above the 0.10 criterion, and board size and CEO duality also exceed the lenient 0.20 bound. The Rosenbaum sensitivity results in Panel D do not alter the inference that neither matched contrast provides evidence of a protective association. Hypothesis H1 receives no support from either matched contrast.

The Love plot in Figure 3 summarizes the covariate-balance results reported in Table 9. For the board-independence treatment, pre-matching differences are substantial, particularly for board size, whereas most post-matching differences lie within or close to the conventional tolerance band. For the female-director treatment, most post-matching differences are close to zero, although current ratio, board size, and CEO duality remain above the conventional absolute-difference threshold. The board-independence estimate is based on 31 matched pairs after 9 treated firms fall outside common support, while the female-director estimate is based on 39 matched pairs after 5 treated firms fall outside common support. Neither ATT is statistically distinguishable from zero. The matching results therefore provide no evidence in this sample of a protective association between the two governance treatments and new-onset distress, consistent with the onset logit and machine-learning results.

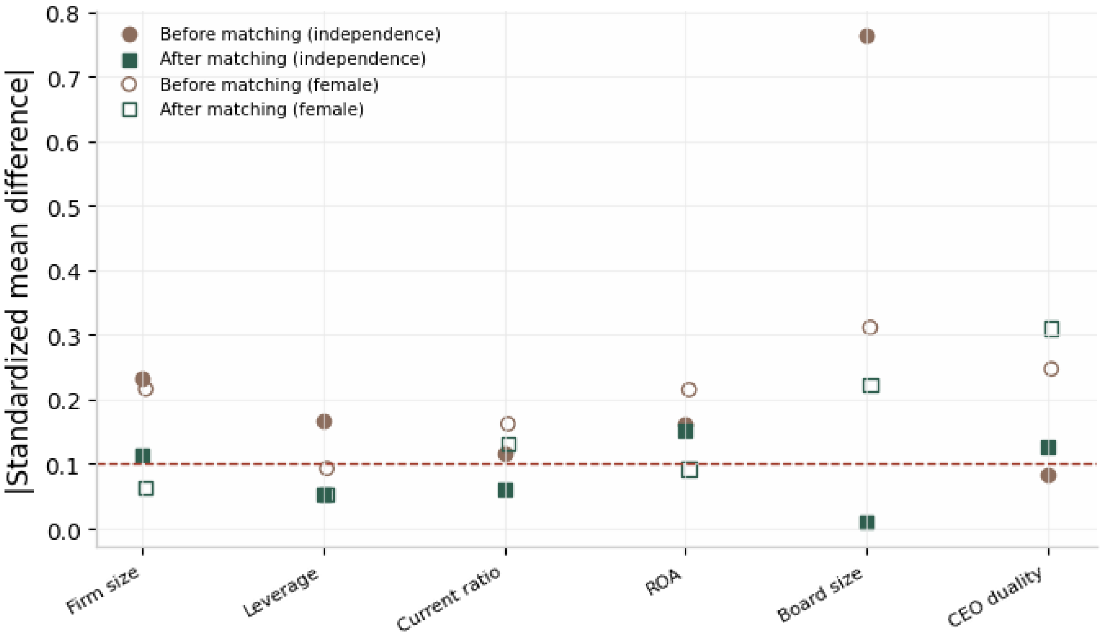

Fig. 3. Covariate balance before and after propensity-score matching

4.4 Machine Learning Performance

Table 10 reports out-of-sample discrimination for the onset outcome under nested cross-validation with 50 outer folds, for six classifiers that share the same 13 firm-level predictors and a reduced six-predictor penalized logit.

The onset results in Table 10 show weak and specification-dependent discrimination. Mean AUC ranges from 0.541 for CatBoost to 0.646 for the random forest. Every fold-level percentile interval includes 0.5, and the firm-level bootstrap intervals agree for five of the seven models, including the logit at [0.456, 0.709] and CatBoost at [0.418, 0.655]. Two intervals exclude 0.5, the random forest at [0.515, 0.757] and the reduced penalized logit at [0.513, 0.764]. The reduced penalized logit, built on leverage, the current ratio, ROA, firm size, board independence and female directors, attains a mean AUC of 0.641, so the narrowest specification is among the better performers. PR-AUC values range from 0.198 to 0.293 against a prevalence baseline of 0.174, so every model sits above the base rate while the absolute AUC levels remain weak. Because each firm sits in about 10 test partitions, the fold-level intervals are narrower than independent replication would imply. Parsimonious specifications display moderate discrimination for onset, but the evidence falls short of a strong result, and governance variables add no discernible lift, consistent with Table 5.

Table 10
Nested cross-validation performance for 2024 new-onset distress classification

Model	Predictors	ROC-AUC	PR-AUC	Brier score	F1 score	Fold-level 95% interval for ROC-AUC	Firm-bootstrap 95% CI for ROC-AUC
Logit	13	0.578	0.258	0.154	0.059	[0.374, 0.837]	[0.456, 0.709]
Penalized logit	13	0.580	0.268	0.238	0.361	[0.356, 0.814]	[0.448, 0.723]
Spline logit	13	0.563	0.214	0.160	0.000	[0.371, 0.733]	[0.439, 0.700]
Random forest	13	0.646	0.291	0.174	0.286	[0.461, 0.853]	[0.515, 0.757]
XGBoost	13	0.599	0.269	0.181	0.340	[0.462, 0.842]	[0.472, 0.721]
CatBoost	13	0.541	0.198	0.181	0.103	[0.346, 0.729]	[0.418, 0.655]
Reduced penalized logit	6	0.641	0.293	0.228	0.353	[0.460, 0.874]	[0.513, 0.764]
Outer validation design	5 folds × 10 repeats						
Held-out prediction unit	Firm-level out-of-fold probabilities						
Onset events	27						
Non-events	128						
Firm-bootstrap resamples	2,000						

Note: ROC-AUC = area under the receiver operating characteristic curve; PR-AUC = area under the precision-recall curve. The onset models are estimated on the 155 firms at risk of onset, namely the 27 onset firms and the 128 never-distressed firms. Mean ROC-AUC is averaged across the 50 outer validation folds. Fold-level intervals summarize dispersion across outer folds, whereas firm-bootstrap confidence intervals are based on 2,000 firm-cluster resamples of the aggregated out-of-fold predictions. PR-AUC, Brier score and F1 score are computed from the same aggregated firm-level out-of-fold probabilities. F1 is evaluated at the 0.5 threshold. The reduced penalized logit is reported as a prespecified parsimony specification.

A component decomposition qualifies the onset result and bounds its interpretation. The onset label is entirely driven by operating profitability. All 27 onset cases are firms with negative operating profit in 2024 after non-negative operating profit in 2023, new entries into the EMZ distress zone number zero, and negative-equity onset is zero, so the negative equity criterion is uninformative in this sample. The composite label therefore behaves as a new operating loss onset label in this sample, and the onset analysis speaks to profitability deterioration rather than to balance-sheet insolvency. We treat this as a limitation of the onset design.

Table 11 reports the corresponding results for the conditional state model, the persistence benchmark that includes the 2023 distress state among its predictors. Mean AUC ranges from 0.773 for the logit to 0.809 for the random forest, and all firm-level bootstrap intervals lie well above 0.5, for example [0.738, 0.877] for the random forest. The gains from flexible learners are real but small. XGBoost improves on the logit by +0.015, and the random forest improves on XGBoost by +0.022 (SE 0.005), increments of roughly two and four standard errors respectively that indicate modest non-linearity. The fold-level intervals remain wide, [0.631, 0.885] for XGBoost and [0.616, 0.917] for the logit, reflecting shared-fold dependence.

Table 11
Nested cross-validation performance for 2024 conditional distress-state classification

Model	Predictors	ROC-AUC	PR-AUC	Brier score	F1 score	Fold-level 95% interval for ROC-AUC	Firm-bootstrap 95% CI for ROC-AUC
Logit	14	0.773	0.642	0.159	0.636	[0.616, 0.917]	[0.706, 0.852]
Penalized logit	14	0.777	0.646	0.174	0.629	[0.603, 0.921]	[0.705, 0.854]
Random forest	14	0.809	0.665	0.171	0.651	[0.654, 0.950]	[0.738, 0.877]
XGBoost	14	0.787	0.659	0.168	0.661	[0.631, 0.885]	[0.705, 0.859]
Outer validation design	5 folds × 10 repeats						
Held-out prediction unit	Firm-level out-of-fold probabilities						
Observations	200						
Distressed firms	63						
State23 included	Yes						
Conditional-state PR-AUC benchmark	0.315						
Firm-bootstrap resamples	2,000						

Note: ROC-AUC = area under the receiver operating characteristic curve; PR-AUC = area under the precision–recall curve; State23 = 2023 distress-state indicator. Fold-level intervals summarize dispersion across the 50 outer validation folds, whereas firm-bootstrap confidence intervals are based on 2,000 firm-cluster resamples of the aggregated out-of-fold predictions. PR-AUC, Brier score and F1 score are computed from the same aggregated firm-level out-of-fold probabilities, with F1 evaluated at the 0.5 threshold.

Figure 4 plots the out-of-fold ROC curves for the state models. All four curves lie well above the diagonal across most of the false-positive range, indicating that the conditional state models rank firms by 2024 distress-state membership with useful consistency. The random forest curve sits outermost, in line with its mean AUC of 0.809, while the logit variants track each other near 0.77 and XGBoost lies in between. The curves are jagged because the out-of-fold evaluation uses 200 firms with 63 positives, and the separation between the best and worst curves remains modest relative to the interval widths reported for the state models in Table 11. Flexible models help, the margin over a plain logit is small, and most of the discrimination derives from persistence rather than governance

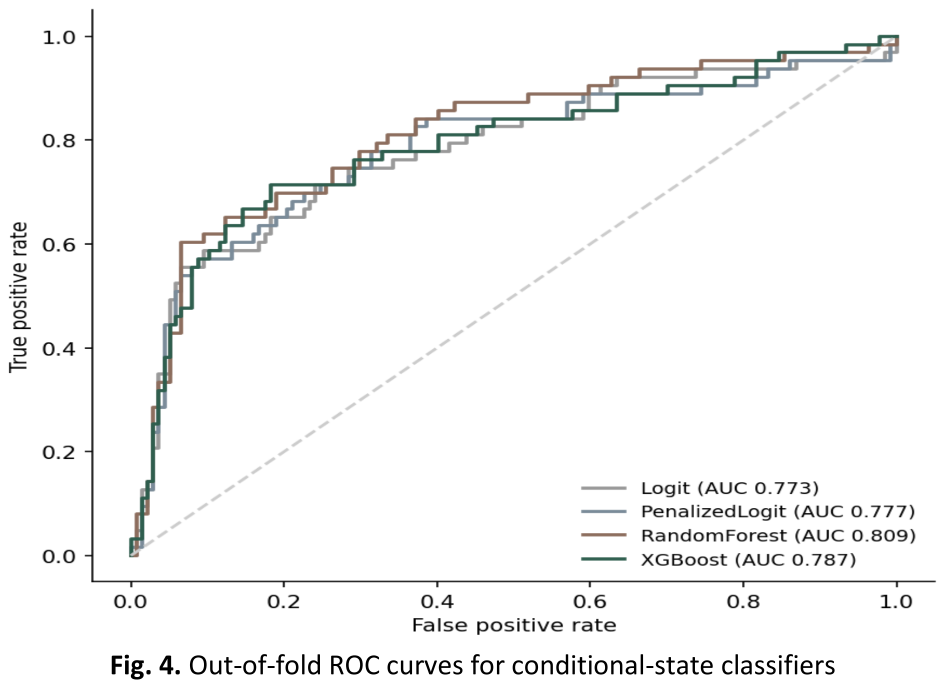

Fig. 4. Out-of-fold ROC curves for conditional-state classifiers

Table 12 reports calibration and decision-analytic metrics for the XGBoost conditional state model on out-of-fold predictions. The mean out-of-fold predicted probability is 0.408 against an observed distress rate of 0.315, so the negative intercept indicates that the model overestimates risk on average. The negative intercept indicates that the raw probabilities produced by the fitted pipeline are higher on average than the observed event rate; this calibration property is reported directly rather than attributed to a mechanism not documented in the model specification. The slope of 0.937 is close to one, and the Brier score of 0.168 improves on the baseline of 0.216, so the probabilities carry genuine information. Split-half isotonic recalibration did not improve the Brier score in this sample (0.173 versus 0.168). We therefore read the predicted probabilities as ranking scores and choose decision thresholds accordingly. The decision panel quantifies the screening trade-off. At 0.20 the model flags 71.0 percent of firms and catches 87.3 percent of distressed cases, with precision of 0.387. At the prevalence threshold of 0.313 the flag share falls to 54.0 percent with sensitivity of 0.810. At 0.50 precision rises to 0.672 while sensitivity falls to 0.651. Operating-point selection requires an explicit cost matrix; because that matrix is not estimated here, the table reports the sensitivity–specificity–precision trade-off without selecting a preferred threshold.

Table 12

Calibration and threshold-specific decision performance for 2024 conditional distress-state classification

Panel or metric	Value	Threshold	Sensitivity	Specificity	Precision	Flag share
Panel A. Calibration performance						
Calibration intercept	−0.570					
Calibration slope	0.937					
Brier score	0.168					
Baseline-rate variance	0.216					
Brier score after split-half isotonic recalibration	0.173					
Panel B. Threshold-specific decision performance						
Conditional-state classification		0.200	0.873	0.365	0.387	0.710
Conditional-state classification		0.313	0.810	0.599	0.481	0.530
Conditional-state classification		0.500	0.651	0.854	0.672	0.305
Prediction model	XGBoost conditional-state model					
Prediction unit	Firm-level out-of-fold probabilities					
Outer validation design	5 folds × 10 repeats, 50 outer folds					
Observations	200					
Distressed firms	63					
Firm-bootstrap resamples	2,000					

Note: The threshold of 0.313 is a prespecified operating point close to the observed conditional-state prevalence. Sensitivity, specificity, precision and flag share are calculated from firm-level out-of-fold probabilities. The split-half isotonic value is reported as a sensitivity analysis.

Figure 5 summarizes the calibration results reported in Table 12. The empirical points track the 45-degree line reasonably well through the middle of the probability range, where the slope near one originates. The average level of the predictions sits above the observed rate, 0.408 against 0.315, which produces the negative intercept of -0.570. The upper end of the curve is thin, since few firms receive high predicted probabilities. The split-half recalibration does not improve the Brier score in this sample, with 0.173 compared with 0.168; the data therefore does not demonstrate a systematic calibration improvement from this recalibration step. The out-of-fold probabilities therefore serve as ranks, and their levels overstate risk on average

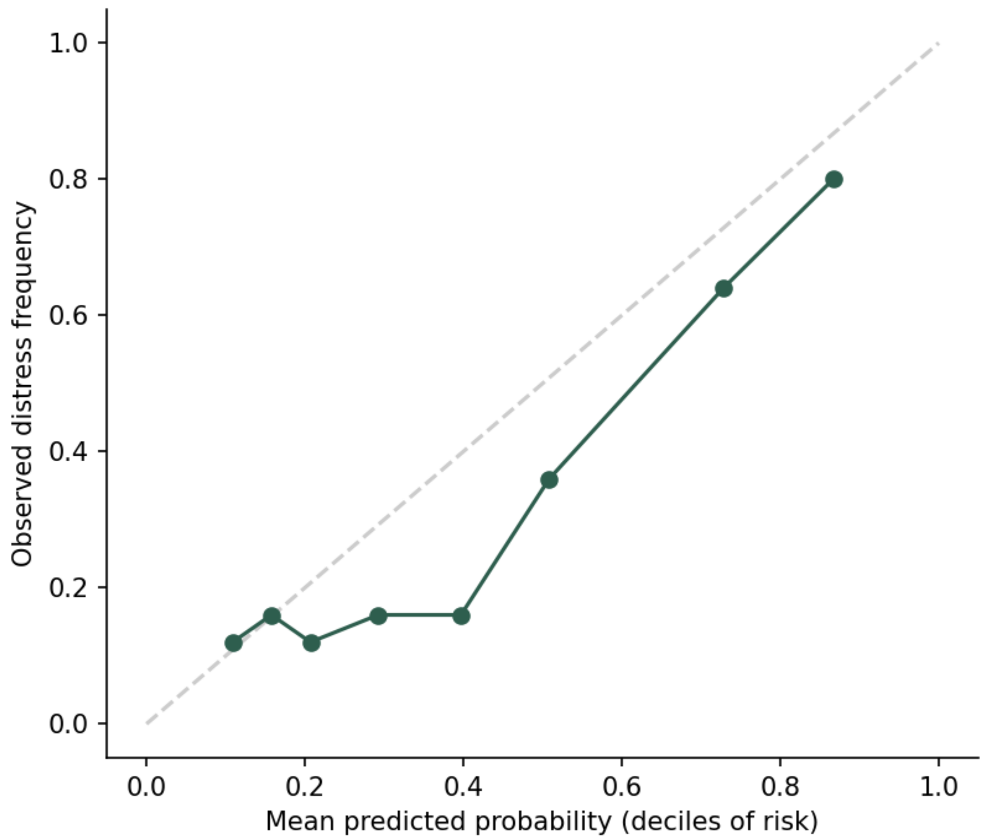

Fig. 5. Calibration curve for the conditional-state XGBoost model

4.5 Explainable AI Diagnostics

Table 13 reports global explainability results for the conditional state model, combining mean absolute SHAP values, bootstrap rank frequencies and permutation importance. The lagged distress state dominates by every measure, with mean absolute SHAP of 0.857, rank one in 100 percent of resamples and permutation importance of 0.191, an order of magnitude above every other predictor. The financial block occupies the next tier, at 0.625 against 0.130 for the entire governance block. Bootstrap frequencies separate stable signals from fragile ones. ROA, firm size, the current ratio and sales growth hold top five positions in 62 to 89 percent of resamples, while leverage and the executive-director share sit at 36 and 39 percent. The remaining governance variables fall below 5 percent, and independence, female directors, duality and foreign directors carry mean absolute SHAP at or below 0.014. Permutation importance confirms the ordering. These conditional-state explanations complement the econometric results by showing that the lagged state and financial variables dominate the fitted state classifier, whereas the governance block contributes comparatively little.

Table 13
Global explainability diagnostics for the XGBoost conditional-state model

Predictor	Mean absolute SHAP value	Bootstrap top-five frequency	Permutation importance, ΔROC-AUC
State23	0.857	100% at rank 1	0.191
ROA	0.352	89%	0.034
Firm size	0.239	77%	0.033
Current ratio	0.232	72%	0.021
Sales growth	0.180	62%	0.015
Leverage	0.144	36%	0.016

Table 13
Global explainability diagnostics for the XGBoost conditional-state model

Predictor	Mean absolute SHAP value	Bootstrap top-five frequency	Permutation importance, ΔROC-AUC
Executive directors	0.098	39%	0.012
Auditor change	0.042	<5%	≤0.003
Board meetings	0.042	15%	≤0.003
Board size	0.033	<5%	≤0.003
Foreign directors	0.014	<5%	≤0.003
CEO duality	0.011	<5%	≤0.003
Female directors	0.009	<5%	≤0.003
Board independence	0.008	<5%	≤0.003
Model	Conditional-state XGBoost		
Predictors	14		
Observations	200		
SHAP method	Interventional TreeSHAP		
SHAP scale	Log-odds		
Bootstrap refits	100 firm-level refits		

Note: State23 denotes the 2023 distress-state indicator. Mean absolute SHAP values summarize feature contributions on the log-odds scale. Bootstrap top-five frequency is the percentage of 100 firm-level bootstrap refits in which a predictor ranks among the five most important features. Permutation importance is the reduction in ROC-AUC after feature permutation.

Figure 6 presents the SHAP beeswarm plot, which shows the direction and dispersion of each predictor's contribution. High values of the 2023 distress indicator push predictions sharply toward distress, and its point cloud is separated from the rest by a wide margin. For ROA, high values shift predictions toward safety and low values toward distress, consistent with the conditional-state estimates in Table 7. The current ratio shows the reverse coloring, with low liquidity carrying positive contributions. The governance variables sit in a compressed band near zero, and their clouds show no strong color gradient, so their direction is not systematic. Firm size and sales growth display intermediate dispersion. The visual impression is consistent with the grouped attribution reported in Table 13, with governance at 0.130 of total attribution against 0.625 for financials

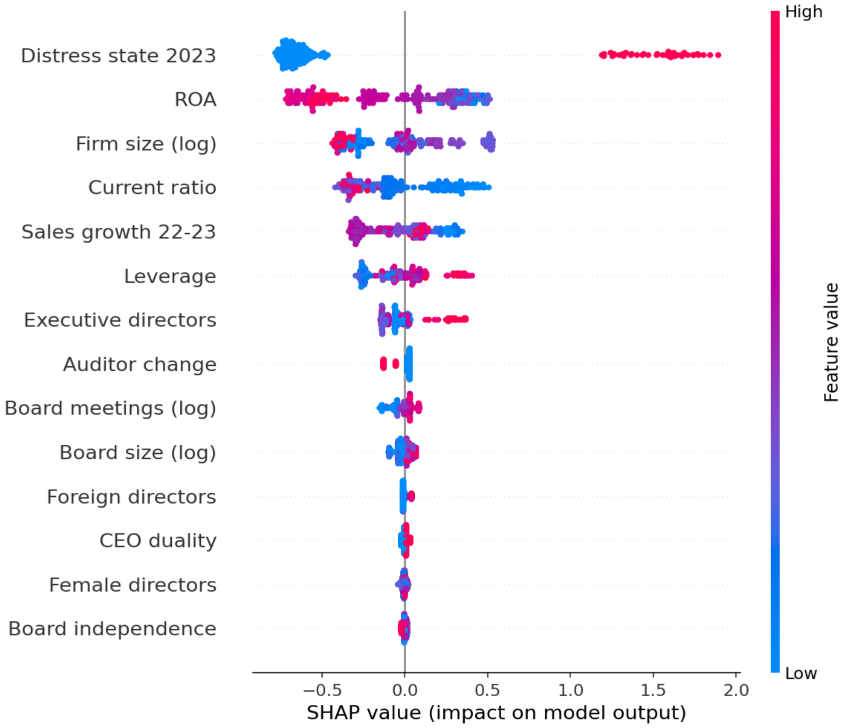

Fig. 6. SHAP beeswarm plot for the conditional-state XGBoost model

Figure 7 presents the mean absolute SHAP values as a ranked bar chart. The bar for the 2023 distress state dwarfs all others at 0.857, followed by ROA at 0.352, firm size at 0.239 and the current ratio at 0.232. From leverage at 0.144 downward the bars shrink quickly, and the five smallest, all governance variables, sit at 0.033 or less. Granule-level computation adds a qualification. The governance block's share of total attribution is 0.227 in granule G0 and 0.170 in each of G1 and G2. The G0 figure rests on only 8 firms and is suggestive at most. The ordering is stable across granules, since governance attribution never approaches the financial block in any group

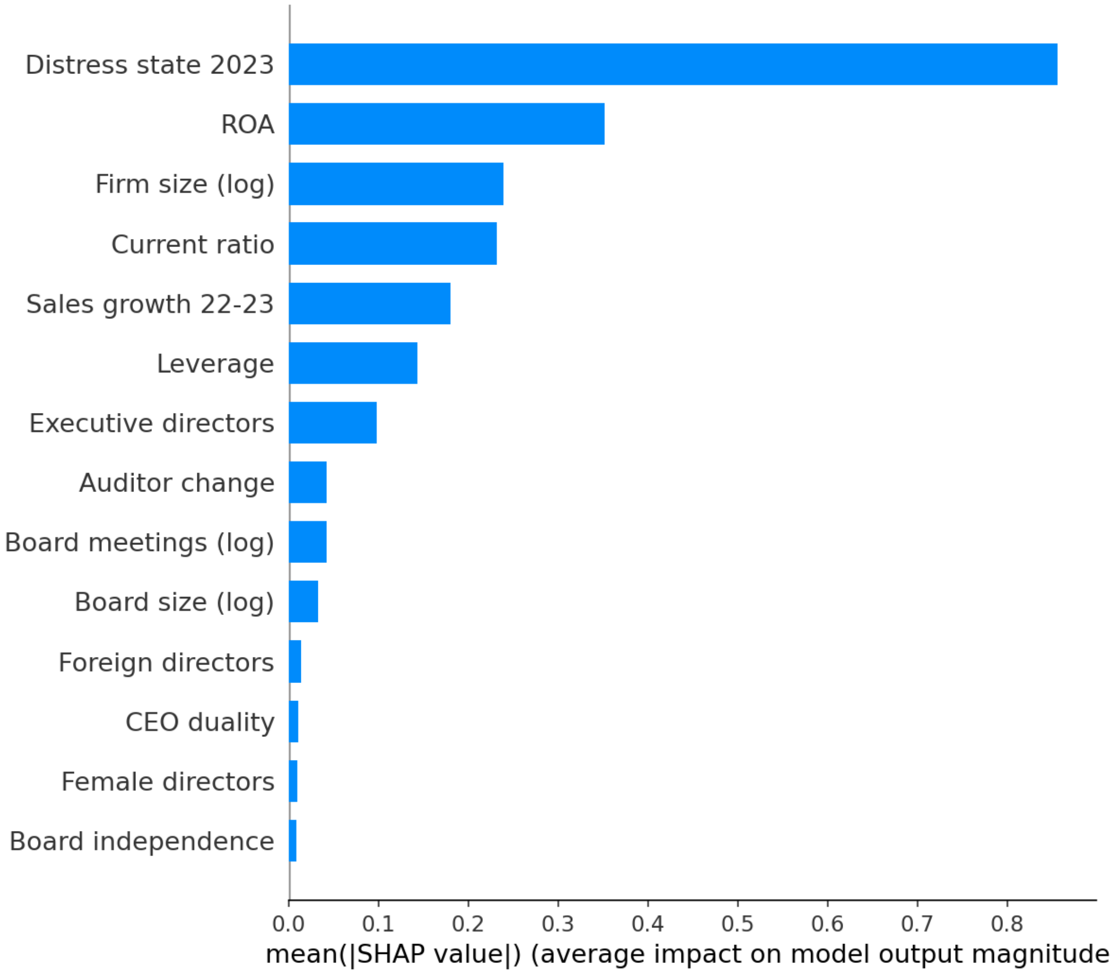

Fig. 7. Mean absolute SHAP importance for the conditional-state XGBoost model

4.6 Threshold Patterns and Formal Tests

Figure 8 presents SHAP dependence plots for the principal continuous predictors, with smoothed trends superimposed. The contribution of ROA falls monotonically and steepens around zero profitability before flattening. The current ratio shows a concentrated drop near 1.2 to 1.5, consistent with a liquidity cliff. Leverage contributions rise toward a ratio of about 0.40 and then level off. The executive-director share declines, with positive contributions at low insider shares turning negative at higher shares. These plots provide descriptive evidence about how the conditional state ensemble uses its predictors. They condition on the dominant lagged state and describe fitted-model behavior rather than structural effects. Table 14 reports separate formal nonlinearity diagnostics for the onset outcome; the state-model plots are not treated as confirmatory evidence for those onset tests

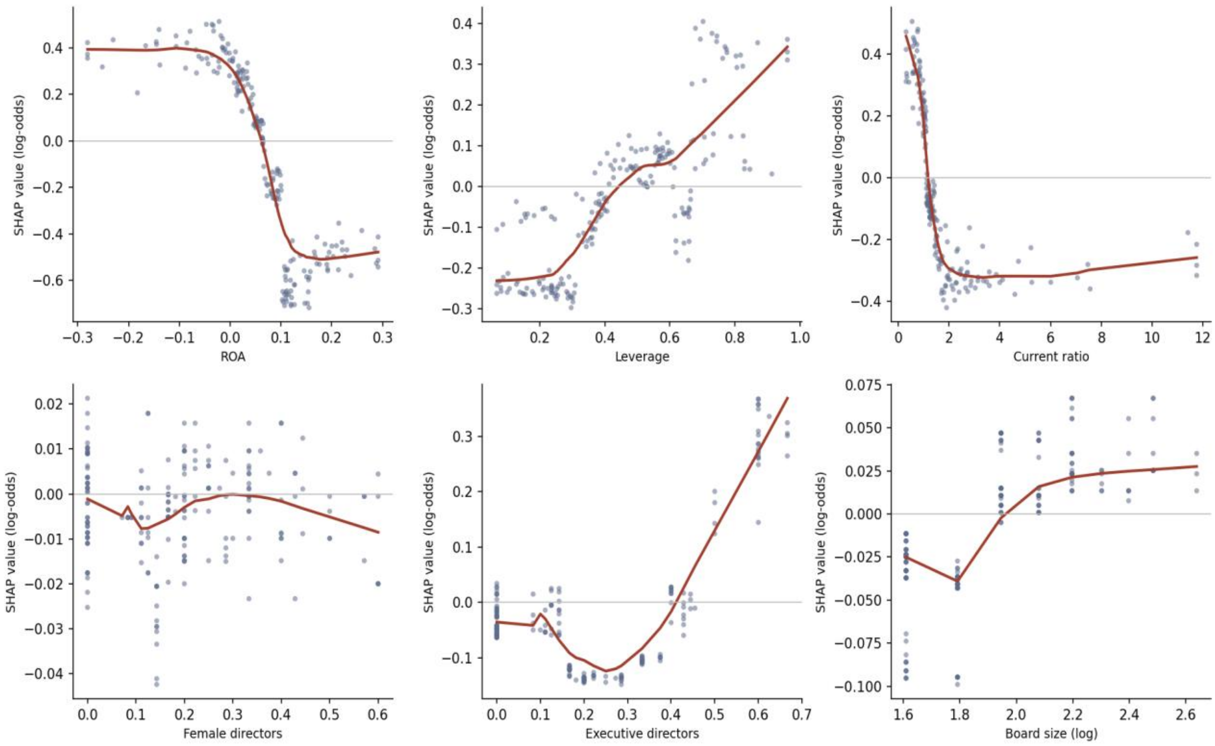

Fig. 8. SHAP dependence plots for selected predictors in the conditional-state XGBoost model

Table 14 reports on the formal tests, estimated on the 155 firms at risk of onset. Panel A gives segmented logit estimates at candidate knots with bootstrap confidence intervals on the onset outcome, and Panel B gives likelihood ratio tests of spline terms against linear terms on the onset outcome. In Panel A the estimated knots sit inside wide bootstrap intervals, and no likelihood ratio test reaches significance at the five percent level, with segmented p-values ranging from 0.061 for the current ratio to 0.777 for female directors. The knot intervals are themselves revealing. The current-ratio interval reaches down to 0.8, the board-independence interval covers the full compliance range from 0.20 to 0.45, and the ROA interval of [-0.05, 0.15] spans both sides of zero profitability, so the data cannot localize the candidate thresholds precisely. The closest approach, the current-ratio knot at 2.00 with p = 0.061, is marginal and does not survive a Bonferroni correction across the fourteen-test family. In Panel B the spline versus linear comparisons is uniformly insignificant, with p-values between 0.345 and 1.000. For the governance variables specified in H2, neither the spline comparisons nor the segmented tests support the hypothesized nonlinearity or threshold locations, so H2a and H2b receive no support.

Table 14
Formal nonlinearity tests for 2024 new-onset distress

Test family	Predictor	Candidate knot	Bootstrap 95% CI for knot	LR statistic	p-value
Panel A. Segmented-logit tests					
Segmented logit	Current ratio	2.00	[0.80, 2.00]	3.52	0.061
Segmented logit	Leverage	0.25	[0.25, 0.60]	0.59	0.444
Segmented logit	Female directors	0.25	[0.10, 0.40]	0.08	0.777
Segmented logit	ROA	0.125	[−0.05, 0.15]	0.77	0.382
Segmented logit	Board independence	0.40	[0.20, 0.45]	0.81	0.369
Segmented logit	Executive directors	0.50	[0.10, 0.50]	2.35	0.125
Segmented logit	Board size, log	2.40	[1.70, 2.40]	1.31	0.253

Table 14
Conituned

Test family	Predictor	Candidate knot	Bootstrap 95% CI for knot	LR statistic	p-value
Panel B. Spline-versus-linear LR tests					
Spline versus linear	ROA			0.16	0.922
Spline versus linear	Leverage			0.01	0.994
Spline versus linear	Current ratio			2.10	0.350
Spline versus linear	Board independence			0.00	1.000
Spline versus linear	Executive directors			2.13	0.345
Spline versus linear	Board size, log			0.41	0.816
Spline versus linear	Female directors			0.00	1.000
Outcome	Onset24				
Observations	155				
Onset events	27				
Segmented tests	7				
Spline-versus-linear tests	7				
Total tests	14				
Bonferroni-adjusted reference level	0.0036				
Knot selection	Profile-likelihood maximum				
Knot bootstrap resamples	200 firm-level resamples				

Note: LR = likelihood-ratio. All tests are estimated on the 155 firms at risk of onset. The candidate knot regions were set from the SHAP dependence patterns and fixed before estimation. Panel A reports the estimated knot and its bootstrap confidence interval; Panel B reports LR comparisons between spline and linear specifications. The p-values are unadjusted, and the Bonferroni-adjusted reference level is 0.0036 for the 14-test family.

5. Discussion and Conclusion

5.1 What Evidence Says About Governance Under Stress

Agency theory links board monitoring to the control of managerial discretion under separated ownership and control [1], while Fama and Jensen [23] place outside directors at the boundary between managers and residual claimants. The results do not support H1. None of the governance specifications provides reliable evidence that the 2023 board measures separated firms that newly entered distress in 2024. This finding does not reject monitoring as a mechanism. It indicates that formal board shares measured at one point do not provide enough variation to identify a short horizon protective association during a broad macroeconomic shock.

The result also clarifies how the present evidence relates to prior distress research. Fich and Slezak [27] show that governance characteristics affect the probability of bankruptcy among firms that are already distressed. Our design asks a different question and finds no useful governance signal for entry into distress among firms that were not distressed one year earlier. García and Herrero [5] associate female directorships with more conservative capital structures and lower distress exposure, yet the corresponding association is effectively zero in this sample. The difference is consistent with the institutional setting. Where a comply or explain code imposes binding floors, board attributes cluster near the legal minimum, measured ratios capture formal compliance more readily than monitoring intensity, and nonconformity is concentrated where firms can provide an explanation [28], [19]. When dispersion is compressed, a cross-sectional coefficient has little variation left to identify. The grouped SHAP attribution of 0.130 for governance, compared with 0.625 for the financial block, is consistent with that measurement constraint, although it is not a test that governance has no effect.

The timing evidence points in the same direction. Ding *et al.,* [3] show that firms entered the pandemic shock with resilience already reflected in their financial structure, and the onset results here likewise place the strongest information in the balance sheet position established before the shock rather than in board labels. This does not contradict the Turkish evidence of Ararat *et al.,* [30] or Ararat and Yurtoglu [17], which links governance to value and profitability over longer panels. It identifies a difference in horizon and outcome. Governance can be gradually reflected in firm value and profitability while one year of formal governance measures remain weak for forecasting an imminent transition into operating distress.

5.2 Persistence as Structural Propagation

The conditional state model contains most of the predictive performance in the study. The 2023 distress state has an average marginal effect of 0.413, and the explainability analysis places it well above the remaining predictors. Once that state is known, the additional contribution of governance and current financial ratios is limited. The model therefore functions primarily as a persistence benchmark. Its high discrimination should not be presented as evidence that governance forecasts new failures, because the leading predictor records whether distress was already present before the outcome year.

This result fits the view that distress develops through continuation and transition processes rather than through a single homogeneous event. Banerjee and Hofmann [60] describe a firm lifecycle in which weak firms persist under forbearance until conditions tighten, while de Moraes Souza *et al.,* [61] show that the transitions between distress stages contain much of the observable information. Costello [62] places an additional part of the transmission process outside the firm by showing how distress can travel through supply chains. The onset result adds a measurement implication. Internal governance and accounting variables observed one year ahead did not anticipate entry into distress, so a model designed to forecast entry may need variables that capture refinancing exposure, supply chain dependence, and sensitivity to the macroeconomic shock. The present evidence is consistent with that interpretation, but it does not identify the external carrier directly.

The distinction also refines the resilience literature. DesJardine *et al.,* [63] show that resilience practices shape which firms withstand a financial crisis, whereas this study asks when a firm crosses into a distress state. Resilience can support survival without predicting the exact timing of entry, and persistence can be highly predictable even when onset is not. These are different estimands and should not be combined in a single performance claim. Treating them separately prevents the strong classification of already distressed firms from being mistaken for evidence of successful early warning.

5.3 The Illusion of Predictability and the Discipline of Validation

The comparison between the contaminated benchmark and the leakage free design is the clearest validity result in the study. The naive specification reached an AUC of 0.727 after contemporaneous information and sector related structure entered the evaluation, whereas the clean protocol produced onset discrimination close to chance. The firms did not change between the two exercises. The change came from the information available to the model and from the way performance was measured. This result applies to the leakage taxonomy of Kapoor and Narayanan [10] and the associated reporting remedies of Kapoor *et al.,* [11] to distress prediction, where pooled samples and overlapping information windows can easily inflate apparent skill.

The validation result must be read together with the event count. The onset analysis contains 27 events in 200 firms, so the uncertainty around out of sample discrimination remains substantial, particularly for flexible learners. The evidence supports the statement that no useful onset discrimination was demonstrated under the stated information set. It does not establish that every moderate governance effect is absent. This distinction matters because events per variable is not a sufficient rule for judging prediction studies, yet the number of predictors, total sample size, and event fraction still shape out of sample performance [39]. Interval estimates, calibration measures, and transparent protocol reporting are therefore part of the evidence rather than supplementary presentation choices [40]. The contribution is a credible predictive baseline, not a claim that model complexity can overcome limited information.

5.4 Theoretical Contributions

The first theoretical contribution is to treat the distribution of governance measures as part of the governance distress relation. Where codes leave room for choice, a formal board structure may carry information about monitoring. Where legal floors bind, the same measure captures compliance more than implementation, and the identifying variation needed for a cross-sectional association is compressed. The finding gives the comply or explain discussion an empirical consequence for prediction. Institutional rules do not merely affect the level of governance; they shape the variation from which governance effects can be estimated.

The second contribution is the separation of onset from persistence. Many distress models pool entry into distress with continued membership in distress, even though the two outcomes have different information structures. By estimating them separately, the present design shows that predictive content is concentrated in continuation and is largely carried by the lagged state. This gives the propagation accounts of Banerjee and Hofmann [60] and Costello [62] a firm level statistical expression and establishes a boundary for interpreting high classification performance. A model that classifies persistence well should not be used to claim that it forecasts new distress.

The third contribution concerns the role of explainability in empirical prediction. SHAP dependence plots are useful for locating regions in which fitted contributions change, but a visible change is not by itself evidence of a structural threshold. The present analysis uses the plots to define candidate regions and then subjects those regions to segment and spline likelihood ratio tests. None of the formal tests support a threshold at the five percent level, and the result does not survive a stricter family wise reference. This sequence places explainability in an audit role. SHAP identifies what the fitted model uses and where a nonlinear pattern may warrant examination, while formal testing determines whether that pattern can support an inferential claim [22, 56]. It therefore avoids treating attractive visualization as a substitute for hypothesis testing.

The fourth contribution is a boundary condition for prediction in emerging market governance settings. When mandatory compliance compresses board shares into a narrow range, greater algorithmic flexibility cannot recover variation that the institutional arrangement does not provide. Single year governance ratios then have limited capacity to distinguish resilient firms from fragile ones, even when the classifier is able to represent nonlinear functions. Future governance research in such settings must either measure changes within firms, capture implementation quality and ownership structure, or examine a longer horizon over which governance practices can affect financial outcomes.

5.5 Practical Implications

For regulators, the evidence cautions against treating formal governance compliance as an early warning indicator for new distress. When issuers cluster near a mandated floor, a compliance score provides little information for ranking firms that will enter distress after a macroeconomic shock. Supervisory systems should keep governance monitoring and distress surveillance as distinct tasks. Financial condition, prior distress status, and probability calibration provide a more appropriate basis for an early warning screen than board composition alone.

For lenders and risk teams, the conditional state model is useful as a persistence benchmark rather than as a causal governance screen. A firm already classified as distressed should be evaluated with a different risk protocol from a firm entering distress for the first time. The calibration results also imply that predicted probabilities should be recalibrated for the institution and decision horizon, with thresholds selected from the costs of missed distress and unnecessary intervention rather than copied from the sample prevalence.

For boards and senior managers, the findings do not reduce governance to a box ticking exercise, but they do show that formal board ratios are not a substitute for financial resilience. Liquidity, leverage, profitability, and the capacity to absorb refinancing pressure remain central to stress readiness. Governance reforms should therefore be evaluated over a horizon long enough to capture implementation and financial consequences, not only through the presence of a legally required board structure in a single reporting year. Direct operational deployment of the present models without external validation would exceed the evidence.

5.6 Limitations and Future Research

The study has a defined empirical scope. It covers one market and one outcome year, and the onset task contains 27 events among 200 firms. The onset intervals are therefore wide, and the null result may reflect both weak governance information and limited power to detect moderate effects. No external validation has been conducted. The outcome is dominated by new operating losses, so the evidence concerns the onset of negative operating profit under the composite rule rather than every form of financial distress. IAS 29 inflation accounting may add measurement noise, and ownership structure, banking relationships, refinancing exposure, and supply chain dependence are not included among the predictors.

Future research can address these limits with longer panels and repeated stress episodes. Rolling within firm designs would allow researchers to examine whether governance changes precede changes in distress risk and whether the importance of governance stabilizes across regimes. External validation in subsequent years would test whether the persistence benchmark survives a change in macroeconomic conditions. Replication in other emerging markets with comply or explain codes would show whether variation compression is specific to the Turkish institutional setting. Models that include refinancing, supply chain, ownership, and enforcement variables could also test whether the weak onset signal reflects missing exposure measures rather than an absence of governance relevance.

The central result is that distress prediction changes when entry and persistence are separated and when evaluation is kept outside model construction. In this Borsa Istanbul sample, one year of governance data provides little information for forecasting new operating loss, while the 2023 distress state makes the continuation of distress highly predictable. The study therefore places a clear boundary around what the observed governance measures can support. It also shows why explainability should generate testable questions rather than replace formal inference.

6. Conclusion

This study examined whether corporate governance indicators can provide advance information about financial distress among non-financial firms listed on Borsa Istanbul, while distinguishing the arrival of new distress from the continuation of an existing condition. The evidence gives a qualified answer to this objective. Governance measures observed one year before the outcome did not produce a reliable early-warning signal for new-onset distress, and the regression, matching, and leakage-controlled machine-learning analyses offered no consistent support for a protective governance effect. In contrast, the conditional-state models predicted continuing distress much more effectively because the lagged distress condition carried most of the usable information, with financial characteristics contributing additional but smaller signals. The explainability analysis identified several plausible nonlinear patterns, yet the subsequent formal tests did not confirm statistically credible thresholds. The principal contribution is therefore a boundary around interpretation: formal board measures should not be treated as stand-alone predictors of imminent distress when institutional rules compress their variation and the forecasting horizon is short. Future applications should separate onset from persistence, enforce time-consistent validation, and combine governance indicators with financial and exposure measures before operational deployment.

Data Availability Statement

The financial statement items and governance disclosures underlying this study are publicly available from the Public Disclosure Platform.

Acknowledgement

This research was not funded by any grant.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm. Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4), 305–360. https://doi.org/10.1016/0304-405X(76)90026-X

[2] Adams, R. B., & Ferreira, D. (2009). Women in the boardroom and their impact on governance and performance. *Journal of Financial Economics*, 94(2), 291–309. https://doi.org/10.1016/j.jfineco.2008.10.007

[3] Ding, W., Levine, R., Lin, C., & Xie, W. (2021). Corporate immunity to the COVID-19 pandemic. *Journal of Financial Economics*, 141(2), 802–830. https://doi.org/10.1016/j.jfineco.2021.03.005

[4] Bae, K. H., El Ghoul, S., Gong, Z. J., & Guedhami, O. (2021). Does CSR matter in times of crisis? Evidence from the COVID-19 pandemic. *Journal of Corporate Finance*, 67, 101876. https://doi.org/10.1016/j.jcorpfin.2020.101876

[5] García, C. J., & Herrero, B. (2021). Female directors, capital structure, and financial distress. *Journal of Business Research*, 136, 592–601. https://doi.org/10.1016/j.jbusres.2021.07.061

[6] Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589–609. https://doi.org/10.2307/2978933

[7] Basilas, A., & Rigani, A. (2024). Machine learning techniques in bankruptcy prediction: A systematic literature review. *Expert Systems with Applications*, 255, 124761. https://doi.org/10.1016/j.eswa.2024.124761

[8] Kou, G., Xu, Y., Peng, Y., Shen, F., Chen, Y., Chang, K., & Kou, S. (2021). Bankruptcy prediction for SMEs using transactional data and two-stage multiobjective feature selection. *Decision Support Systems*, 140, 113429. https://doi.org/10.1016/j.dss.2020.113429

[9] Chen, C., Hanlon, D., Khedmati, M., & Wake, J. (2023). Annual report readability and equity mispricing. *Journal of Contemporary Accounting & Economics*, 19(3), 100368. https://doi.org/10.1016/j.jcae.2023.100368

[10] Kapoor, S., & Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*, 4(9), 100804. https://doi.org/10.1016/j.patter.2023.100804

[11] Kapoor, S., Cantrell, E. M., Peng, K., Pham, T. H., Bail, C. A., Gundersen, O. E., ... & Narayanan, A. (2024). REFORMS: Consensus-based recommendations for machine-learning-based science. *Science Advances*, 10(18), eadk3452. https://doi.org/10.1126/sciadv.adk3452

[12] Harvey, C. R., & Liu, Y. (2021). Lucky factors. *Journal of Financial Economics*, 141(2), 413–435. https://doi.org/10.1016/j.jfineco.2021.04.014

[13] Hong, Y., Jiang, M., Li, S., & Zhao, S. (2024). Class-imbalanced financial distress prediction with machine learning: Incorporating financial, management, textual, and social responsibility features into index system. *Journal of Forecasting*, 43(3), 593–614. https://doi.org/10.1002/for.3050

[14] Radovanovic, J., & Haas, C. (2023). The evaluation of bankruptcy prediction models based on socio-economic costs. *Expert Systems with Applications*, 227, 120275. https://doi.org/10.1016/j.eswa.2023.120275

[15] Central Bank of the Republic of Turkey. (2024). *Para politikasi kurulu toplanti karari: Faiz oranlarina iliskin basin duyurusu (2024-14)* [Press release]. https://www.tcmb.gov.tr/wps/wcm/connect/tr/tcmb+tr/main+menu/duyurular/basin/2024/duy2024-14

[16] Turkish Statistical Institute. (2025). *Tüketici fiyat endeksi, Aralık 2024* [Data set]. https://www.tuik.gov.tr
[17] Ararat, M., & Yurtoglu, B. B. (2021). Female directors, board committees, and firm performance: Time-series evidence from Turkey. *Emerging Markets Review*, 48, 100768. https://doi.org/10.1016/j.ememar.2020.100768
[18] Saygili, A. T., Saygili, E., & Taran, A. (2021). The effects of corporate governance practices on firm-level financial performance: Evidence from Borsa Istanbul Xkury companies. *Journal of Business Economics and Management*, 22(4), 884–904. https://doi.org/10.3846/jbem.2021.14440
[19] Witt, M. A., Fainshmidt, S., & Aguilera, R. V. (2022). Our board, our rules: Nonconformity to global corporate governance norms. *Administrative Science Quarterly*, 67(1), 131–166. https://doi.org/10.1177/00018392211022726
[20] Bounas, N., UdDin, S., Awan, T., & Khan, M. Y. (2021). Corporate governance and financial distress: Asian emerging market perspective. *Corporate Governance*, 21(4), 702–715. https://doi.org/10.1108/CG-04-2020-0119
[21] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785
[22] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *31st Conference on Neural Information Processing Systems (NIPS 2017)*.
[23] Fama, E. F., & Jensen, M. C. (1983). Separation of ownership and control. *The Journal of Law and Economics*, 26(2), 301–325. https://doi.org/10.2139/ssrn.94034
[24] Post, C., & Byron, K. (2015). Women on boards and firm financial performance: A meta-analysis. *Academy of Management Journal*, 58(5), 1546–1571. https://doi.org/10.5465/amj.2013.0319
[25] Terjesen, S., Couto, E. B., & Francisco, P. M. (2016). Does the presence of independent and female directors impact firm performance? A multi-country study of board diversity. *Journal of Management & Governance*, 20(3), 447–483. https://doi.org/10.1007/s10997-014-9307-8
[26] Albuquerque, R., Koskinen, Y., Yang, S., & Zhang, C. (2020). Resiliency of environmental and social stocks: An analysis of the exogenous COVID-19 market crash. *The Review of Corporate Finance Studies*, 9(3), 593–621. https://doi.org/10.1093/rcfs/cfaa011
[27] Rich, E. M., & Slezak, S. L. (2008). Can corporate governance save distressed firms from bankruptcy? An empirical analysis. *Review of Quantitative Finance and Accounting*, 30(2), 225–251. https://doi.org/10.1007/s11156-007-0048-5
[28] Roberts, J., Sanderson, P., Seidl, D., & Krivokapic, A. (2020). The UK corporate governance code principle of 'comply or explain': Understanding code compliance as 'subjection'. *Abacus*, 56(4), 602–626. https://doi.org/10.1111/abac.12208
[29] Sula, V. (2005). The impact of the roles, structure and process of boards on firm performance: Evidence from Turkey. *Corporate Governance: An International Review*, 13(2), 265–276. https://doi.org/10.1111/j.1467-8683.2005.00421.x
[30] Ararat, M., Black, B. S., & Yurtoglu, B. B. (2017). The effect of corporate governance on firm value and profitability: Time-series evidence from Turkey. *Emerging Markets Review*, 30, 113–132. https://doi.org/10.1016/j.ememar.2016.10.001
[31] Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 247(1), 124–136. https://doi.org/10.1016/j.ejor.2015.05.030
[32] Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83, 405–417. https://doi.org/10.1016/j.eswa.2017.04.006
[33] Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), 2223–2273. https://doi.org/10.1093/rfs/hhaa009
[34] Leippold, M., Wang, Q., & Zhou, W. (2022). Machine learning in the Chinese stock market. *Journal of Financial Economics*, 145(2), 64–82. https://doi.org/10.1016/j.jfineco.2021.08.017
[35] Luque, A., Carrasco, A., Martín, A., & de Las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, 91, 216–231. https://doi.org/10.1016/j.patcog.2019.02.023
[36] Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1), 6. https://doi.org/10.1186/s12864-019-6413-7
[37] Hessain, J. (2022). *Improving the prediction of asset returns with machine learning by using a custom loss function*. https://doi.org/10.2139/ssrn.3973086
[38] Van Calster, B., McLernon, D. J., van Smeden, M., Wynants, L., & Steyerberg, E. W. (2019). Calibration: The Achilles heel of predictive analytics. *BMC Medicine*, 17(1), Article 230. https://doi.org/10.1186/s12916-019-1466-7

[39] van Smeden, M., De Groot, J. A., Moons, K. G., Collins, G. S., Altman, D. G., Eijkemans, M. J., & Reitsma, J. B. (2016). No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. *BMC Medical Research Methodology*, 16(1), 163. https://doi.org/10.1186/s12874-016-0267-3

[40] Collins, G. S., Moons, K. G., Dhiman, P., Riley, R. D., Beam, A. L., Van Calster, B., ... & Logullo, P. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*, 385, e078378. https://doi.org/10.1136/bmj-2023-078378

[41] Lessmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable machine learning in credit risk management. *Computational Economics*, 57(1), 203–216. https://doi.org/10.1007/s10614-020-10042-0

[42] Giudici, P., & Raffinetti, E. (2021). Shapley-Lorenz eXplainable artificial intelligence. *Expert Systems with Applications*, 167, 114104. https://doi.org/10.1016/j.eswa.2020.114104

[43] Cho, S. H., & Shin, K. S. (2023). Feature-weighted counterfactual-based explanation for bankruptcy prediction. *Expert Systems with Applications*, 216, 119390. https://doi.org/10.1016/j.eswa.2022.119390

[44] Chen, C., Lin, K., Rudin, C., Shaposhnik, Y., Wang, S., & Wang, T. (2022). A holistic approach to interpretability in financial lending: Models, visualizations, and summary-explanations. *Decision Support Systems*, 152, 113647. https://doi.org/10.1016/j.dss.2021.113647

[45] Carmona, P., Dwekat, A., & Mardawi, Z. (2022). No more black boxes! Explaining the predictions of a machine learning XGBoost classifier algorithm in business failure. *Research in International Business and Finance*, 61, 101649. https://doi.org/10.1016/j.ribaf.2022.101649

[46] Zhang, Z., Wu, C., Qu, S., & Chen, X. (2022). An explainable artificial intelligence approach for financial distress prediction. *Information Processing & Management*, 59(4), 102988. https://doi.org/10.1016/j.ipm.2022.102988

[47] Liu, J., Li, C., Ouyang, P., Liu, J., & Wu, C. (2023). Interpreting the prediction results of the tree-based gradient boosting models for financial distress prediction with an explainable machine learning approach. *Journal of Forecasting*, 42(5), 1112–1137. https://doi.org/10.1002/for.2931

[48] Park, M. S., Son, H., Hyun, C., & Hwang, H. J. (2021). Explainability of machine learning models for bankruptcy prediction. *IEEE Access*, 9, 124887–124899. https://doi.org/10.1109/ACCESS.2021.3110270

[49] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

[50] Altman, E. I. (2005). An emerging market credit scoring system for corporate bonds. *Emerging Markets Review*, 6(4), 311–323. https://doi.org/10.1016/j.ememar.2005.09.007

[51] Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18(1), 109–131. https://doi.org/10.2307/2490395

[52] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)* (pp. 6638–6648). Curran Associates.

[53] Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, 7(1), 91. https://doi.org/10.1186/1471-2105-7-91

[54] Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., ... & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

[55] Aas, K., Jullum, M., & Løland, A. (2021). Explaining individual predictions when features are dependent: More accurate approximations to Shapley values. *Artificial Intelligence*, 298, 103502. https://doi.org/10.1016/j.artint.2021.103502

[56] Goldstein, A., Kapelner, A., Bleich, J., & Pitkin, E. (2015). Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *Journal of Computational and Graphical Statistics*, 24(1), 44–65. https://doi.org/10.1080/10618600.2014.907095

[57] Apley, D. W., & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(4), 1059–1086. https://doi.org/10.1111/rssb.12377

[58] Altman, E. I., Sabato, G., & Wilson, N. (2010). The value of non-financial information in SME risk management. *Journal of Credit Risk*, 6(2), 95–127. https://doi.org/10.21314/JCR.2010.110

[59] Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, 47(2), 427–465. https://doi.org/10.1111/j.1540-6261.1992.tb04398.x

[60] Banerjee, R., & Hofmann, B. (2022). Corporate zombies: Anatomy and life cycle. *Economic Policy*, 37(112), 757–803. https://doi.org/10.1093/epolic/eiac027

[61] de Moraes Souza, J. G., de Castro, D. T., Peng, Y., & Gartner, I. R. (2024). A machine learning-based analysis on the causality of financial stress in banking institutions. *Computational Economics*, 64(3), 1857–1890. https://doi.org/10.1007/s10614-023-10514-z

[62] Costello, A. M. (2020). Credit market disruptions and liquidity spillover effects in the supply chain. *Journal of Political Economy*, 128(9), 3434–3468. https://doi.org/10.1086/708736

[63] DesJardine, M., Bansal, P., & Yang, Y. (2019). Bouncing back: Building resilience through social and environmental practices in the context of the 2008 global financial crisis. *Journal of Management*, 45(4), 1434–1460. https://doi.org/10.1177/0149206317708854