

Accounting Warning Consensus and Next-Year Corporate Accounting Stress in Bangladesh

Gourav Roy^{1*} and Susmita Sarker²

¹Bangladesh Institute of Capital Market (BICM), Ministry of Finance, Government of Bangladesh

²Department of Accounting, National University, Bangladesh

ABSTRACT

Purpose – This study tests whether consensus across partially overlapping accounting-warning systems predicts next-year corporate accounting stress beyond raw fundamentals among provider-covered listed non-financial firms in Bangladesh.

Design/methodology/approach – Of 236 listed non-financial firms, 109 have sufficient Refinitiv/Worldscope coverage to construct the Distress Consensus Index (DCI). The DCI combines Zmijewski, Springate, Piotroski F-score, and Financial Health Composite (FHC) risk ranks for 2012-2024, with outcomes through 2025. Firm and year fixed effects, conditional fixed-effects logit, composition and selection diagnostics, Raw Fundamentals Risk Index (RFRI) benchmarks, severe accounting-stress endpoints, and leakage-controlled temporal validation are used.

Finding/Results – In the preferred provider-covered sample (N = 1,066), a 10-point DCI increase is associated with 6.86 percentage points more next-year accounting stress; the estimate is 7.25 points in complete four-score observations and 4.05 points for severe accounting stress. DCI remains incremental to RFRI, including when FHC is removed from the DCI. Holdout AUC is 0.780 under contemporaneous annual ranking and 0.766 under fully historical frozen scaling; adding DCI to raw fundamentals raises AUC from 0.777 to 0.797.

Originality/Value – DCI is a transparent relative screening feature, not a bankruptcy probability or transferable event-risk estimate. Generalization is limited to provider-covered listed non-financial firms with sufficient accounting data, while measurement composition, score overlap, peer ranking, temporal preprocessing, and calibration drift are explicit inferential boundaries.

ARTICLE INFO

Keywords:

Accounting Stress,
Accounting Warning
Systems,
Distress Consensus,
Frontier Market,
Bangladesh

Article Information:

Received: 20/08/2026

Revise: 10/09/2026

Accepted: 30/09/2026

ISSN:

2985-3168 (Online)

2985-3222 (Print)

*Corresponding Author at:

Bangladesh Institute of Capital Market (BICM), Ministry of Finance, Government of Bangladesh.

E-mail address: gouravroy.du@gmail.com (Gourav Roy), susmitasarkersuchi1@gmail.com (Susmita Sarker)

The work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International \(CC BY-SA 4.0\)](https://creativecommons.org/licenses/by-sa/4.0/)



1. Introduction

Financial distress usually emerges through combinations of weakening profitability, liquidity, leverage, cash generation, and operating efficiency rather than through one ratio. Classic warning systems translated these signals into statistical models, including Altman (1968), Ohlson (1980), Zmijewski (1984), and Piotroski's (2000) fundamental-strength score. The unresolved practical question is what to infer when credible systems disagree—and, more importantly, whether their agreement on weakness predicts future deterioration more reliably than reliance on one imported model.

The systems are heterogeneous but not statistically independent. Zmijewski emphasizes profitability, leverage, and liquidity; Piotroski combines discrete changes in fundamental strength; Springate applies a different discriminant structure; and the FHC summarizes six operating and balance-sheet dimensions. Several inputs overlap across these systems, so agreement can reflect both a common latent fragility component and repeated exposure to the same accounting deterioration. Beneish remains conceptually separate because it targets reporting behavior rather than general solvency (Beneish, 1999; Piotroski, 2000; Zmijewski, 1984). Their coefficients and historical cutoffs were developed in different settings, making universal transportability doubtful (Altman et al., 2017; Laitinen & Suvas, 2016; Sayari & Mugan, 2017).

Machine learning improves classification by combining features or learners, and recent distress studies report strong performance from bagging, boosting, stacking, random forests, and imbalance-aware ensembles (Barboza, Kimura, & Altman, 2017; Liang, Tsai, Lu, & Chang, 2020; Liu, Suzuki, & Du, 2025; Wang & Chi, 2024). Yet optimized ensembles can be tuning-dependent and less transparent. More importantly, when an accounting score generates the dependent class, high predictive accuracy may partly reproduce that score's algebra rather than forecast an independently realized adverse state.

This study therefore examines a transparent middle ground: the common relative-risk ranking that emerges when structurally different but partially overlapping accounting systems place the same firm in a weak position. The measurement argument is deliberately conditional rather than assuming independent errors. If the systems share a latent fragility component but also retain model-specific information, consensus may be more stable than reliance on a single imported specification. That proposition is tested directly through component correlations, principal-component decomposition, comparison with a simple Raw Fundamentals Risk Index (RFRI), and validation against future outcomes external to the DCI itself.

Bangladesh is a demanding setting for this test. Bangladesh Bank (2026) reports that the DSEX fell 6.73% in 2025 and average daily turnover declined 17.56%. World Bank (2026) data underscore the relatively small scale of the listed-company market, while Dhaka Stock Exchange information reported by The Business Standard (2026) identified going-concern warnings for 37 listed firms. These conditions make early-warning quality economically relevant and amplify concerns about mechanically transporting cutoffs calibrated in other countries and periods.

Local evidence also leaves a specific gap. Abdullah (2021) shows that machine learning can reproduce Altman-based solvency categories for Bangladeshi listed firms, while Rafi (2026) documents sectoral variation in Beneish manipulation signals. Neither establishes whether agreement across heterogeneous accounting-warning systems predicts a separately realized future deterioration event.

The study uses an unbalanced panel of listed non-financial Bangladeshi firms, with 2012-2024 as the principal predictor period and forward outcomes through 2025. Among 236 non-financial firms in the source universe, 109 have at least one computable core warning signal and 109 have a usable DCI observation; 127 firms never enter the construct because the supplied Refinitiv/Worldscope workbook contains insufficient financial-statement coverage for them. Firm-level selection diagnostics therefore precede the usual lead-outcome attrition analysis, and the study's population inference is explicitly limited to Bangladesh-listed non-financial firms with sufficiently complete provider coverage. The DCI harmonizes FHC, Zmijewski, Springate, and Piotroski signals to annual risk percentiles and averages at least three observed components. The resulting measure is explicitly relative: it identifies firms that are weaker than contemporaneous listed non-financial peers and is not an aggregate measure of system-wide deterioration.

The primary outcome is next-year accounting stress, defined as negative net income or negative operating cash flow. This is an adverse accounting-condition screen rather than bankruptcy, default, restructuring, or insolvency. Firm and year fixed effects test whether within-firm changes in consensus precede deterioration after controlling for current accounting stress, size, and strictly consecutive-year sales growth. Additional analyses distinguish new stress onset from persistence, construct a stricter severe accounting-stress endpoint, compare DCI directly with RFRI, diagnose both source-universe selection and future-data attrition, examine changing component availability and annual peer composition, and evaluate nonlinear and clustered-inference robustness. A 2022-2024 holdout contrasts contemporaneous annual ranking with fully historical frozen and expanding-window transformations.

The preceding literature and construct-validity considerations lead to three research questions:

RQ1. Does accounting-warning consensus predict independently realized next-year corporate accounting stress, including new stress onset among firms that are currently non-stressed and stricter adverse-accounting endpoints?

RQ2. Does the DCI retain incremental information after explicitly accounting for shared inputs, changing component availability, annual peer composition, source-universe selection, an RFRI benchmark, alternative aggregation rules, attrition sensitivity, and robust inference?

RQ3. Does the consensus retain useful out-of-time discrimination under contemporaneous, fully historical frozen, and expanding-window transformations, while remaining competitive rather than universally superior to raw-fundamental and machine-learning benchmarks?

The study makes two closely related contributions. First, it shifts the unit of analysis from choosing a single imported distress model to evaluating a transparent common risk ranking across heterogeneous but partially overlapping accounting systems. This contribution is falsifiable: pairwise correlations and PCA quantify commonality, while the RFRI benchmark tests whether DCI adds information beyond direct aggregation of the six underlying financial-health dimensions. Equal weighting is retained for transparency rather than optimized against the outcome.

Second, the design separates construct formation from realized future validation and from deployment assumptions. Annual percentile harmonization reduces dependence on historical absolute cutoffs but yields a relative within-year measure, not a macro-financial stress index. The annual holdout is therefore out-of-time with respect to outcomes and model fitting but partly transductive with respect to contemporaneous percentile scaling. Frozen pre-2022 and

expanding-window transformations test whether predictive usefulness survives when the reference distribution is historically anchored. The reduced M5 manipulation screen remains outside the core DCI as a construct-boundary comparator.

The evidence is economically substantial but deliberately bounded. In the preferred fixed-effects sample ($N = 1,066$), a 10-point DCI increase is associated with a 6.86-percentage-point higher probability of next-year accounting stress. The coefficient is 7.25 percentage points in complete four-score observations and 6.89 points after controlling for the numbers of available DCI and FHC components. A conditional fixed-effects logit, which conditions out time-invariant firm heterogeneity and uses only firms with outcome variation, yields an odds ratio of 1.75 per 10 DCI points. DCI remains incremental to RFRI in the preferred joint model (0.0448 per 10 points; $p = 0.0035$). More stringently, a DCI reconstructed without FHC remains significant when entered jointly with RFRI (0.0363 per 10 points; $p = 0.0041$). In the 2022-2024 holdout, DCI AUC is 0.780 under contemporaneous annual ranking, 0.766 under fully historical frozen scaling, and 0.768 under expanding historical scaling. Adding DCI to a raw-fundamentals logit raises AUC from 0.777 to 0.797, while nonlinear tree benchmarks remain competitive.

These results do not imply that DCI is a bankruptcy probability, an absolute measure of Bangladesh-wide corporate deterioration, a causal treatment, or the universally best predictive model. The defensible claim is narrower: among provider-covered Bangladesh-listed non-financial firms with sufficiently complete Refinitiv/Worldscope accounting data, a transparent relative consensus across partially overlapping warning systems contains incremental and temporally robust information about subsequent accounting stress and can serve as a screening feature that warrants deeper investigation.

2. Literature Review and Research Questions

2.1 From accounting ratios to dynamic financial-distress prediction

Corporate financial-distress prediction developed from a simple but durable premise: deterioration in a firm's financial statements is observable before the terminal event of bankruptcy. Beaver (1966) established the predictive content of individual financial ratios, showing that accounting measures can discriminate between failed and non-failed firms before failure becomes legally visible. Altman (1968) extended this logic through multiple discriminant analysis, combining profitability, liquidity, accumulated profitability, leverage, and asset-use measures into the Z-score. Ohlson (1980) subsequently replaced the distributional restrictions of discriminant analysis with a probabilistic logit framework, while Zmijewski (1984) used a probit specification and, importantly, demonstrated how choice-based sampling and complete-data selection can distort distress-model estimation. These studies created the central empirical architecture of the field: financial condition is multidimensional, no single ratio is sufficient, and model performance depends not only on variable choice but also on sampling and validation design.

The literature later moved from static classification toward explicitly dynamic prediction. Shumway (2001) showed that single-period bankruptcy models can be misspecified when failure risk changes over time, while Campbell, Hilscher, and Szilagyi (2008) demonstrated the value of dynamic failure probabilities that combine accounting and market information. This evolution matters because an early-warning system is useful only if it predicts a future adverse state rather than merely describing contemporaneous weakness.

Recent evidence confirms that accounting variables remain central even as prediction methods become more sophisticated. Zhao, Ouenniche, and De Smedt (2024) and Dasilas and Rigani (2024) document a field spanning statistical models, machine learning, preprocessing, feature selection, and out-of-sample evaluation, while Nugroho and Dewayanto (2025) emphasize the continuing value of interpretable statistical models alongside hybrid approaches. Firm-level studies similarly show that modern machine learning can improve distress prediction when the information set and validation design are disciplined (Zhao, Xu, & Ji, 2023; Rahman & Zhu, 2024). The contemporary question is therefore how accounting information should be combined, transported across contexts, and validated against genuinely future outcomes.

2.2 Distinct accounting-warning systems capture different dimensions of corporate condition

A major limitation of treating financial-distress scores as interchangeable is that they were developed for different objectives and encode different economic concepts. Zmijewski's model emphasizes profitability, leverage, and liquidity and belongs directly to the statistical failure-prediction tradition (Zmijewski, 1984). Piotroski's (2000) F-score, in contrast, was designed as a fundamental-strength screen using nine binary signals spanning profitability, cash flow, leverage/liquidity, operating efficiency, and changes in financial condition. A high F-score therefore represents stronger fundamentals rather than a calibrated probability of bankruptcy. These distinctions are important because two firms can receive different risk classifications simply because alternative scores emphasize different dimensions of the same balance sheet. Beneish (1999) illustrates the construct-boundary problem even more clearly. The M-score was developed to identify financial-statement patterns associated with earnings manipulation, not to estimate general insolvency or cash-flow distress. Its variables are designed to capture unusual receivables, margin deterioration, asset quality, sales growth, depreciation choices, and other conditions that may accompany manipulation. Treating a Beneish signal as another bankruptcy score therefore conflates reporting-quality risk with financial fragility. The appropriate empirical question is whether manipulation-oriented signals overlap with, or remain distinct from, solvency and fundamental-strength signals.

The Financial Health Composite follows a complementary strategy: rather than reproducing the coefficients of one legacy model, it summarizes six recurrent dimensions of operating and financial condition profitability, liquidity, leverage, operating cash generation, interest-servicing capacity, and asset-use efficiency. These dimensions recur in contemporary distress research. Ashraf, Félix, and Serrasqueiro (2020) show that accounting, market, and reporting-quality information jointly improve prediction across developed and emerging settings, while Liu, Suzuki, and Du (2025) identify profitability, cash flow, solvency, and structural ratios as important features in ensemble models. The literature therefore supports a multidimensional view of financial health without requiring one historical set of externally estimated weights.

2.3 Model transportability, industry heterogeneity, and the frontier-market problem

The widespread use of established scores can obscure a transportability problem: parameters and cutoffs estimated in one country, period, or industry may not transfer unchanged to another setting. Altman et al. (2017), using firms from 31 European and three non-European countries, show that a general Z-score specification performs reasonably on average but that country-specific estimation materially improves classification accuracy. Established scores may therefore contain useful information even when their original calibration is context dependent.

Industry heterogeneity creates a related problem. Sayari and Mugan (2017) show that the information content of financial ratios differs across industries and argue for industry-specific distress modeling. Cross-country evidence reaches the same broad conclusion. Laitinen and Suvas (2016) find that cultural context moderates the effects of financial predictors in international failure models, while ElBannan (2021) shows that firm-level predictors of distress in 11 emerging markets operate alongside differences in financial-market development and institutional quality. Ashraf et al. (2020) further document that the predictive role of financial-reporting-quality measures is relevant in both the United Kingdom and Pakistan, but the broader modeling environment differs across developed and emerging markets.

These findings have direct implications for a frontier-market study. An imported absolute threshold embeds assumptions about accounting regimes, capital structures, inflation, firm size distributions, market institutions, and the original estimation sample. A within-year relative-risk transformation offers a different objective: it asks where a firm stands relative to its contemporaneous domestic peers rather than whether it crosses a threshold calibrated decades earlier in another market. Such a transformation does not eliminate contextual dependence, but it reduces reliance on the external calibration of any one model. This is especially relevant when the research objective is to extract the common information across several warning systems rather than to defend the universal validity of one cutoff.

2.4 From single scores to ensembles: why agreement may contain information

The machine-learning literature provides a useful but limited analogy for combining warning systems. Ensemble methods can improve generalization when component learners make partly non-overlapping errors (Opitz & Maclin, 1999), and distress research increasingly uses bagging, boosting, stacking, and imbalance-aware learning (Barboza et al., 2017; Liang et al., 2020; Wang & Chi, 2024; Liu et al., 2025). That analogy does not imply that the accounting scores used here are independent classifiers: several rely on common profitability, liquidity, leverage, cash-flow, or efficiency information.

Algorithmic ensemble learning and accounting-score consensus are therefore distinct empirical objects. Trained ensembles optimize base learners and combination rules against a labeled outcome; the DCI does neither. It applies established score structures, converts their outputs to a common ordinal risk scale, and averages them without outcome-driven weights. The relevant construct-validity question is not whether component inputs are independent, but whether the systems retain enough non-redundant structure for their common ranking to add information beyond a direct aggregation of the underlying fundamentals.

The measurement interpretation is accordingly framed in terms of a common latent fragility component plus correlated and model-specific information. Correlated score errors limit any simple error-cancellation claim. The design therefore quantifies pairwise Pearson and Spearman dependence, decomposes the four component risk ranks through PCA, and introduces an RFRI constructed directly from the six FHC fundamentals. If DCI remains informative after RFRI is controlled, the evidence supports incremental structure in cross-model synthesis rather than a claim of statistical independence.

2.5 Prediction design: realized outcomes, temporal validation, and incremental information

A second methodological issue concerns the target variable. Some distress studies use legal bankruptcy, delisting, restructuring, or researcher-defined events; others use the output of an established accounting score as the class label. The latter approach creates a circular validation risk because a high-performing algorithm may learn the algebra of the benchmark score without demonstrating that it predicts an independently realized future adverse state.

This distinction is visible in Bangladesh. Abdullah (2021) reports strong machine-learning performance for 244 listed firms, but the training and testing classes are generated from Altman Z-score categories. That design demonstrates reproduction of an Altman-based solvency classification rather than direct prediction of a separately observed future deterioration event.

Contemporary research increasingly emphasizes chronological data separation, probability calibration, class imbalance, and multiple performance metrics rather than accuracy alone (Zhao et al., 2024; Dasilas & Rigani, 2024; Liu et al., 2025). A credible early-warning design should therefore assess rank discrimination, precision-recall behavior, probabilistic loss, and performance in a genuinely later test period.

The present study validates consensus against realized next-year accounting stress negative earnings and/or negative operating cash flow rather than against another score. This endpoint intentionally captures an adverse accounting condition and is broader than legal or solvency distress. A growing or investment-intensive firm can temporarily generate negative operating cash flow, and a cyclical firm can report a loss without approaching default. Construct validity is therefore assessed through stress onset, dual stress, repeated loss, persistence, two-year horizons, and a stricter severe accounting-stress endpoint based on multiple contemporaneous red flags. Estimation and evaluation are chronological, and deployment sensitivity is assessed with fully historical frozen and expanding-window transformations.

2.6 Evidence from Bangladesh and the unresolved empirical gap

Bangladesh-specific distress research remains fragmented across models and sectors. Many applications focus on Altman-type scores or regulated financial institutions, whose capital and liability structures are not directly comparable with non-financial firms. Abdullah (2021) broadens the setting to listed companies but uses Altman-derived outcome classes, while Rafi (2026) studies Beneish manipulation risk rather than subsequent solvency or cash-flow stress. The unresolved issue is therefore whether a transparent consensus across heterogeneous but partially overlapping accounting-warning systems contains incremental information about a realized future adverse accounting state in a broad non-financial frontier-market panel. The design addresses that gap through component-dependence tests, a direct raw-fundamentals benchmark, within-firm estimation, onset/persistence decomposition, attrition and inference sensitivity, alternative consensus constructions, and leakage-controlled temporal validation.

3. Methodology

3.1 Research design and analytical sample

The study uses an unbalanced panel of listed non-financial firms in Bangladesh. The source workbook contains 354 company records, including 118 firms classified as financial institutions. Excluding banks, insurers, non-bank financial institutions, and other financial firms leaves 236 non-financial firms. Of these, 109 have at least one computable core warning signal and 109 have at least one usable DCI observation during 2012-2024, whereas 127 have no usable DCI because the supplied provider workbook contains insufficient accounting coverage. This initial construct-availability selection is analytically distinct from later loss of next-year outcomes. Firm-level diagnostics compare the 109 included and 127 excluded firms on all source-universe metadata and market variables that are actually observable. The excluded group has no usable market-capitalization, enterprise-value, or shares-outstanding history in the supplied workbook and only 7.1% have a populated fiscal-year-end field, so accounting size, reporting quality, listing age, governance, and survival cannot be credibly

compared. Accordingly, all inference is restricted to the provider-covered subset rather than the full listed non-financial market. The year 2011 is retained only for lag construction, while 2025 is used only to realize forward outcomes for 2024 predictors.

The panel is kept naturally unbalanced rather than restricted to firms with complete histories. For every lead or lag, calendar continuity is enforced: an observation at t is paired with $t+1$ or $t-1$ only when the relevant calendar year is immediately consecutive. Sales growth is computed only from immediately consecutive fiscal-year observations; values spanning multi-year reporting gaps are treated as missing. Applying this rule yields the preferred fixed-effects sample of 1,066 firm-years from 107 firms. All tables and interpretations use this continuity-consistent sample.

Continuous variables entering the FHC and the main continuous controls are winsorized within calendar year at the 1st and 99th percentiles. Refinitiv/Worldscope retrieval strings such as #ERROR and NO WORLDScope DATA are treated as missing, never as numeric values. The design is predictive and associational. Firm fixed effects absorb time-invariant firm heterogeneity, but no exogenous treatment or instrument is available; coefficients are not interpreted as causal effects.

3.2 Construction of the Financial Health Composite

A formative Financial Health Composite (FHC) is constructed to summarize six economically distinct dimensions of corporate condition: profitability (return on assets), liquidity (current ratio), leverage (total debt to total assets), cash generation (operating cash flow to total assets), debt-service capacity (interest coverage), and operating efficiency (total asset turnover). Leverage is reverse-oriented so that every component has the same interpretation: higher values indicate stronger financial health.

For the contemporaneous annual DCI, each FHC component k is winsorized within calendar year at the 1st and 99th percentiles and standardized using that year's cross-sectional mean and standard deviation. Leverage is reverse-oriented so that higher standardized values uniformly indicate stronger financial condition. Equation (1) describes this annual construction. This contemporaneous version is intentionally peer-relative and uses information from the year- t cross-section.

$$Z_{k,it} = \frac{H_{k,it} - \bar{H}_{k,t}}{S_{H_{k,t}}} \quad (1)$$

In Equation (1), k indexes the health component, i the firm, and t the year. The annual FHC is the equal-weight mean of the available standardized components, as shown in Equation (2), with at least four of six components required. For deployment validation, however, the FHC is reconstructed from raw inputs under fully historical scaling. The frozen specification estimates each component's 1st/99th-percentile bounds, mean, and standard deviation exclusively from 2012-2021 and applies those parameters unchanged to 2022-2024. The expanding specification re-estimates the same parameters using information available only through $t-1$. Thus no holdout-year cross-sectional location or scale enters the frozen or expanding FHC.

$$FHC_{it} = \frac{1}{K_{it}} \sum_{k \in \mathcal{A}_{it}} Z_{k,it}, \quad K_{it} \geq 4 \quad (2)$$

The equal-weight specification is preferred because it is transparent and avoids estimating sample-specific weights that could overfit the same panel used for subsequent prediction. Construct robustness is assessed in two ways: first, the FHC is converted to an annual

percentile rank; second, principal-component analysis is applied to complete standardized component observations as a diagnostic alternative to equal weighting. The PCA exercise is treated as a robustness assessment rather than as a re-estimation of the primary construct.

3.3 Benchmark accounting warning systems

The consensus framework combines three established accounting-warning systems with the newly constructed FHC. The models were selected because their required inputs are fully observable in the non-financial panel and because they capture different aspects of corporate condition. This cross-model design is preferable to forcing canonical Altman or Ohlson scores when essential inputs are absent.

The Zmijewski (1984) score is calculated from profitability, leverage, and liquidity using Equation (3). Higher values indicate greater financial-distress risk.

$$X_{it}^{ZM} = -4.3 - 4.5 \frac{NI_{it}}{TA_{it}} + 5.7 \frac{TL_{it}}{TA_{it}} - 0.004 \frac{CA_{it}}{CL_{it}} \quad (3)$$

In Equation (3), NI denotes net income, TA total assets, TL total liabilities, CA current assets, and CL current liabilities. The formula is applied directly to firm-year accounting values; no distress-event oversampling is used in constructing the score.

The Springate score is computed using the four-ratio discriminant specification in Equation (4) (Springate, 1978). Unlike the Zmijewski score, a higher Springate value indicates stronger condition and is therefore reverse-oriented when the common risk scale is constructed.

$$S_{it}^{SP} = 1.03 \frac{WC_{it}}{TA_{it}} + 3.07 \frac{EBIT_{it}}{TA_{it}} + 0.66 \frac{EBT_{it}}{CL_{it}} + 0.40 \frac{Sales_{it}}{TA_{it}} \quad (4)$$

Here, WC is working capital (current assets minus current liabilities), EBIT is operating income, and EBT is pretax income. Sales and the balance-sheet denominators are measured contemporaneously.

Piotroski's F-score is implemented as the sum of nine binary financial-statement signals, as represented in Equation (5) (Piotroski, 2000). The signals span profitability, positive operating cash flow, improvement in return on assets, accrual quality, leverage, liquidity, equity issuance, gross-margin improvement, and asset-turnover improvement. A larger F-score represents stronger fundamentals and is therefore reverse-oriented for the common risk scale.

$$F_{it}^P = \sum_{m=1}^9 I_{m,it}, \quad F_{it}^P \in \{0,1, \dots, 9\} \quad (5)$$

The canonical eight-variable Beneish model is not calculable because selling, general, and administrative expense is unavailable. The manuscript therefore does not label a five-variable equation as the Beneish (1999) model. Instead, it uses the reduced five-variable M-score reported by Roxas (2011), which retains the days-sales-in-receivables index (DSRI), the gross-margin index (GMI), the asset-quality index (AQI), the sales-growth index (SGI), and the depreciation index (DEPI), as specified in Equation (6). Each index is the ratio of the current-year to the prior-year value of the underlying accounting relation, so a value above one denotes deterioration relative to the previous year, and higher M5 values denote a greater likelihood of earnings manipulation. The five component indices are winsorized within year at the 1st and 99th percentiles before aggregation. This reduced manipulation-screening score is retained only as a construct-boundary comparator and is not treated as equivalent to Beneish's canonical eight-variable specification.

$$M5_{it} = -6.065 + 0.823 DSRI_{it} + 0.906 GMI_{it} + 0.593 AQI_{it} + 0.717 SGI_{it} + 0.107 DEPI_{it} \quad (6)$$

Reduced M5 is not treated as a solvency model. Its role is to test construct specificity: if a manipulation-oriented warning signal behaves differently from the four financial-health/distress measures, that separation supports rather than weakens the interpretation of the consensus construct.

Canonical Altman Z and Ohlson O scores are intentionally not used in the empirical models. The workbook does not contain retained earnings, which is required for the public-firm Altman Z specification, and it lacks the price-level-normalized size term and direct funds-from-operations measure required for a canonical Ohlson O-score. No accounting substitute is inserted for these missing variables. This restriction prevents a non-standard proxy from being mislabeled as an established score.

3.4 Annual risk harmonization and the Distress Consensus Index

The component scores have different units, directions, and historical thresholds. In the contemporaneous construction, each score is transformed into a within-year empirical risk percentile using average ranks for ties. For transparency, the rank rule is the empirical mid-rank percentile: tied observations receive their average position; higher-risk direction is retained for Zmijewski and reversed for Springate, Piotroski, and FHC. Values closer to 100 therefore indicate greater relative accounting weakness within the relevant reference set.

$$R_{it}^{(j)} = 50(1 - d_j) + 100d_j P_{j,t}(S_{it}^{(j)}), \quad d_j \in \{-1, +1\} \quad (7)$$

In Equation (7), S is the raw score produced by warning system j , P is the within-year empirical mid-rank cumulative proportion of that score in the year- t reference set, and d equals $+1$ for systems in which higher values indicate greater risk (Zmijewski) and -1 for systems in which higher values indicate greater strength (Springate, Piotroski, and the FHC). R is therefore bounded between 0 and 100, with higher values denoting greater relative accounting weakness. Annual ranking is intentionally cross-sectional and therefore depends on the size and composition of the year- t peer set. The number of usable DCI firms varies from 70 in 2012 to a maximum of 102 in 2020-2021, before declining to 96 in 2024. Piotroski is discrete and exhibits material ties: only 7-9 distinct F-score values are observed per year and the largest annual tie contains approximately 20.8%-29.6% of usable observations. Main-result sensitivity therefore includes a stable 44-firm peer subset and sector-year percentile ranks. These tests distinguish movement in a firm's accounting condition from changes induced purely by the reference universe.

The primary DCI is the equal-weight mean of available annual risk percentiles from Zmijewski, Springate, Piotroski, and FHC, with at least three of four signals required, as defined in Equation (8), in which the index set contains the component systems observed for firm i in year t and J is the number of those available signals. Component availability is recorded explicitly. Of 1,189 usable DCI observations, 1,084 contain all four systems and 105 contain three; every three-system DCI is the Zmijewski-Springate-FHC combination because Piotroski is the missing signal. For FHC, 1,053 observations use all six dimensions, 119 use five, and 17 use four. Robustness models control directly for the number of observed DCI and FHC components and re-estimate the preferred specification on complete four-score observations.

$$DCI_{it} = \frac{1}{J_{it}} \sum_{j \in J_{it}} R_{it}^{(j)}, \quad J_{it} \geq 3 \quad (8)$$

No data-driven weights are estimated for the primary DCI. Alternative median, complete-case, PCA, and leave-one-score-out constructions are treated as robustness checks rather than ex post model searches. To address shared inputs directly, the study also constructs a Raw Fundamentals Risk Index (RFRI) from annual risk percentiles of ROA, current ratio, debt/assets, CFO/assets, interest coverage, and asset turnover. The same directional logic is used so that higher RFRI indicates greater risk; at least four of six components are required. RFRI is deliberately simple and is not tuned against the holdout outcome.

Cross-model disagreement is measured as the sample standard deviation of the available component risk percentiles, as shown in Equation (9). Pairwise Pearson and Spearman correlations quantify component dependence, and PCA on complete four-score observations decomposes common versus residual variation. These diagnostics explicitly test the degree to which DCI reflects common information rather than presuming independent model errors.

$$DIS_{it} = \left[\frac{1}{J_{it} - 1} \sum_{j \in J_{it}} \left(R_{it}^{(j)} - DCI_{it} \right)^2 \right]^{1/2} \quad (9)$$

3.5 Realized future accounting-stress outcomes

The primary dependent variable is an observed future accounting state rather than a distress label generated by any component model. Next-year accounting stress equals one when the firm reports negative net income or negative operating cash flow in the immediately following year. The outcome is determined when next-year net income is available; a missing CFO observation is not itself coded as stress. Equation (10) formalizes the binary screen.

$$HS_{i,t+1} = \mathbf{1}(\text{NI}_{i,t+1} < 0 \vee \text{CFOA}_{i,t+1} < 0) \quad (10)$$

In Equation (10), NI denotes net income and CFOA denotes operating cash flow scaled by total assets, so the indicator is unaffected by the scaling of the cash-flow term. The primary endpoint captures earnings loss and/or failure to generate positive operating cash flow, but it is intentionally broader than conventional financial distress. It does not establish bankruptcy, default, restructuring, insolvency, or going-concern failure. The paper therefore uses 'accounting stress' or 'adverse accounting condition' for the realized outcome and reserves 'financial distress' for the established prediction literature and the historical labels of component models.

Secondary outcomes assess construct breadth and severity: next-year loss; next-year negative operating cash flow; dual stress in which both occur; any accounting stress within t+1 or t+2; persistent accounting stress in both years; repeated loss; and a severe accounting-stress endpoint. The severe endpoint equals one when at least two of four red flags are observed at t+1 net loss, negative operating cash flow, interest coverage below one, and non-positive book equity with at least three of the four indicators observed. This endpoint remains accounting-based rather than legal distress but has a substantially lower event rate. Separate samples define stress onset among currently non-stressed firms and persistence/recovery among currently stressed firms.

3.6 Panel econometric specification

The main explanatory model is a firm- and year-fixed-effects linear probability model. Equation (11) is the preferred specification. DCI is divided by 10 so that beta has a direct probability interpretation for a 10-point movement on the 0-100 consensus scale.

$$HS_{i,t+1} = \beta \frac{DCI_{it}}{10} + \rho HS_{it} + \gamma' X_{it} + \alpha_i + \lambda_t + \varepsilon_{i,t+1} \quad (11)$$

In Equation (11), HS is the future accounting-stress indicator, X contains current accounting stress, log total assets, and annual sales growth, α_i denotes firm fixed effects, and λ_t denotes year fixed effects. Sales growth is computed only when sales at t and t-1 are both observed in immediately consecutive calendar years; non-consecutive last-observation growth is not permitted. Log assets and sales growth are winsorized within year at the 1st and 99th percentiles. In the onset subsample, current accounting stress is omitted because it is identically zero.

Standard errors are clustered by firm in the primary linear probability specification, allowing arbitrary within-firm heteroskedasticity and serial dependence across the 107 firm clusters. Two-way firm-and-year clustering is reported as a sensitivity check for common year shocks, but the year dimension contains only 13 clusters and its smaller standard error is not treated as stronger evidence. A null-imposed year-cluster wild bootstrap-t with Webb six-point weights and 9,999 replications (Cameron, Gelbach, & Miller, 2008; MacKinnon & Webb, 2017) is retained solely as a small-year-cluster sensitivity diagnostic; because it clusters by year only, it does not reproduce the full dependence structure of two-way clustering. Sector-by-year fixed effects provide a separate control for sector-specific common shocks.

The linear probability model remains primary because its coefficient has a direct probability-point interpretation and its fixed-effects transformation is transparent. The nonlinear robustness model is now a conditional fixed-effects logit rather than a pooled sector/year logit. Conditioning on firm-specific sufficient statistics removes time-invariant firm heterogeneity; as expected, firms with no within-firm outcome variation do not contribute to the conditional likelihood. Year indicators and the same contemporaneous controls are included. Equation (12) states the corresponding logistic form, in which the logistic cumulative distribution function maps the linear index to a probability and the firm effect is conditioned out of the likelihood rather than estimated. The conditional logit is used to evaluate functional-form robustness, not to replace the firm-clustered LPM as the primary inferential model.

$$\Pr(HS_{i,t+1} = 1 | X_{it}) = \Lambda(\eta_{it}), \quad \eta_{it} = \beta \frac{DCI_{it}}{10} + \rho HS_{it} + \gamma' X_{it} + \alpha_i + \lambda_t \quad (12)$$

The logistic model is not used to replace the fixed-effects linear probability estimates; it is used to verify that the direction and statistical content of the result are not artifacts of the linear functional form. Across all econometric models, emphasis is placed on coefficient magnitude, confidence intervals, and stability across specifications rather than on statistical significance alone.

3.7 Component dependence, incremental information, attrition, and robustness

Construct dependence is examined before predictive claims are made. In addition to pooled Pearson/Spearman correlations and PCA, DCI and RFRI are residualized with respect to firm and year fixed effects and their within-transformed correlation is reported together with a variance-inflation diagnostic. Incremental information is then tested by entering DCI and RFRI separately and jointly, with and without preferred controls, under complete-score and sector-by-year specifications. A stricter overlap test reconstructs DCI without FHC, using only the Zmijewski, Springate, and Piotroski annual risk ranks, and enters this reduced consensus jointly with RFRI in the preferred complete-score fixed-effects sample. This specification removes the direct inclusion of FHC the component constructed from the same six fundamentals used by RFRI from the DCI side of the joint model. Measurement-composition

sensitivity also includes controls for DCI/FHC component counts, complete four-score observations, leave-one-score-out indices, a stable peer subset, and sector-adjusted percentiles. Selection is analyzed at two stages. First, the 109 firms entering DCI construction are compared with the 127 non-financial firms excluded before construct formation on all observable source-universe metadata and market-data availability. Second, among usable DCI firm-years, ObservedLead equals one when an immediately consecutive t+1 accounting record permits the outcome to be determined. Retained and missing-lead observations are compared using standardized mean differences, a pooled observation logit, extreme missing-outcome bounds, and inverse-probability-of-observation weighting. These later tests do not resolve the earlier provider-coverage selection; the manuscript therefore restricts generalizability explicitly.

3.8 Temporal out-of-sample validation

Predictive validity is evaluated using 2012-2021 predictor observations for model development and 2022-2024 for temporal holdout evaluation, with 2024 outcomes realized from 2025 statements. The contemporaneous annual DCI uses the year-t peer cross-section and is therefore out-of-time with respect to outcomes and predictive-model fitting but partly transductive with respect to scaling. It represents an annual screening setting in which peer accounts for year t are jointly observable.

Two deployment-oriented alternatives are fully historical from the raw accounting inputs onward. For the frozen version, 2012-2021 data determine FHC component winsorization bounds, means, standard deviations, the resulting FHC reference distribution, and the empirical distributions of Zmijewski, Springate, and Piotroski. Holdout values are mapped by an empirical mid-CDF rule: $\text{percentile} = 100 \times [\text{number of historical values below } x + 0.5 \times \text{number equal to } x] / N$, with values below the historical minimum assigned 0 and values above the maximum assigned 100 before direction reversal where required. For the expanding version, 2022 uses 2012-2021, 2023 uses 2012-2022, and 2024 uses 2012-2023. No transformation parameter or outcome from year t or t+1 enters the score assigned to year t. Table A1 sets out the preprocessing order for the three transformations and identifies, at each stage, whether holdout-year information is used.

Table 1. Leakage-control preprocessing pipeline

Stage	Contemporaneous annual DCI	Frozen fully historical DCI	Expanding fully historical DCI
1. Raw inputs	Year-t accounting inputs	Year-t accounting inputs	Year-t accounting inputs
2. FHC winsorization + scaling	Bounds, mean, SD estimated from year t peers	All parameters estimated once from 2012-2021	All parameters estimated from 2012 through t-1
3. FHC construction	Mean of ≥ 4 standardized dimensions	Mean of ≥ 4 historically standardized dimensions	Mean of ≥ 4 historically standardized dimensions
4. Score-to-risk mapping	Year-t average-rank percentiles	Empirical mid-CDF of 2012-2021 reference scores	Empirical mid-CDF of data available through t-1

Stage	Contemporaneous annual DCI	Frozen fully historical DCI	Expanding fully historical DCI
5. DCI	Mean of ≥ 3 component risk percentiles	Mean of ≥ 3 component risk percentiles	Mean of ≥ 3 component risk percentiles
Holdout-year cross-section used?	Yes, for scaling/ranking only	No	No

Note: Historical empirical mid-CDF assigns 0 to values below the reference minimum and 100 to values above the reference maximum before direction reversal. For 2022, 2023, and 2024, the expanding references are 2012-2021, 2012-2022, and 2012-2023, respectively.

$$BS = \frac{1}{N} \sum_{n=1}^N (y_n - \hat{p}_n)^2 \quad (13)$$

In Equation (13), y is the realized binary outcome and $\hat{p}$ is the predicted probability. Logarithmic loss, shown in Equation (14), more heavily penalizes highly confident incorrect predictions and provides a complementary assessment of probabilistic accuracy.

$$LL = -\frac{1}{N} \sum_{n=1}^N [y_n \ln(\hat{p}_n) + (1 - y_n) \ln(1 - \hat{p}_n)] \quad (14)$$

Holdout performance is summarized by AUC, AUPRC, Brier score, and log loss. A univariate logistic calibration model is estimated on 2012-2021 annual DCI only, yielding $\text{logit}(p) = -3.419 + 0.4146 \times (\text{DCI}/10)$; the fitted mapping is frozen for holdout Brier, log-loss, calibration, and threshold analysis. Firm-block bootstrap resampling with 1,000 replications provides both paired confidence intervals for metric differences and individual 95% confidence intervals for each principal model's AUC, AUPRC, Brier score, and log loss. Operating thresholds are selected only from training probabilities and frozen before holdout evaluation.

3.9 Machine-learning benchmark design

Machine learning is used as a predictive benchmark rather than as a causal identification strategy. The raw-fundamentals feature set contains return on assets, current ratio, debt-to-assets, operating cash flow to total assets, interest coverage, asset turnover, sales growth, log total assets, operating margin, net margin, and sector. These variables are intentionally more granular than the DCI so that the incremental-value test is demanding: the DCI must add information beyond a model that already observes the underlying accounting fundamentals. Four benchmark algorithms are estimated on the pre-2022 training sample: L2-regularized logistic regression, elastic-net logistic regression, random forest, and gradient boosting. Hyperparameters were specified ex ante as deliberately low-complexity benchmark settings and were not selected or tuned using the 2022-2024 holdout. The L2 logistic model uses $C = 0.01$; elastic net uses $C = 0.01$ and L1 ratio 0.25; random forest uses 600 trees, maximum depth 5, minimum leaf size 5, square-root feature sampling, and class-balanced bootstrap weighting; gradient boosting uses 100 depth-1 learners, learning rate 0.10, and minimum leaf size 10. This protocol gives each benchmark a fixed development budget and prevents post-hoc tuning advantages against DCI.

For linear models, missing continuous predictors are replaced with the training-sample median and then standardized using training-sample means and standard deviations. Sector is imputed using the training-sample mode and one-hot encoded with unseen categories handled without failure. Tree models use training-sample medians for missing numeric values

and fixed dummy-variable columns. All preprocessing is therefore learned from the estimation sample only, which prevents information leakage from the holdout.

Incremental predictive value is evaluated by comparing a raw-fundamentals L2-regularized logistic model with an otherwise identical model that additionally includes DCI. The comparison uses common holdout observations and firm-block bootstrap confidence intervals for changes in AUC, AUPRC, and Brier score. Machine-learning results are interpreted strictly as forecasting evidence: a more accurate algorithm does not establish a causal mechanism, and DCI is not expected to dominate every tuned nonlinear learner.

3.10 Statistical inference, software, and reproducibility

All reported tests are two-sided, with 95% confidence intervals emphasized alongside economic magnitude. The empirical workflow was implemented in Python using NumPy, statsmodels, scikit-learn, SciPy, and matplotlib, with deterministic seeds for resampling and stochastic learners. The authors retain the licensed raw workbook, reconstruction logic, variable dictionary, transformation specifications, model outputs, and analysis code. The licensed Refinitiv/Worldscope raw data cannot be redistributed, but code, variable definitions, and non-proprietary derived results can be provided to editors and reviewers for reproducibility checks. No holdout outcome is used to estimate transformations, preprocessing parameters, predictive coefficients, calibration functions, or operating thresholds.

4. Results and Discussion

4.1 Descriptive evidence and the cross-model risk gradient

The analytical audit starts from 354 company records. After excluding 118 financial institutions, 236 non-financial firms remain. Only 109 firms have sufficient provider coverage to construct a core warning signal; the remaining 127 are excluded before consensus formation. This is the dominant selection boundary in the study, not the later loss of 42 forward outcomes. Within the 109 provider-covered firms, at least one core warning signal is computable for 1,191 firm-years and DCI for 1,189. A consecutive t+1 accounting-stress outcome is observed for 1,147 DCI observations; after current stress, log assets, and strictly consecutive-year sales growth are imposed, the preferred sample contains 1,066 firm-years from 107 firms. Table 1A shows that identifier, exchange, industry, and sector metadata are complete in both source-universe groups, but provider financial/market coverage is sharply different: fiscal-year-end is populated for all 109 included firms but only 7.1% of excluded firms, while market capitalization, enterprise value, and common shares are observed for included firms but for none of the 127 excluded firms in the supplied workbook. Sector composition differs moderately but does not reject equality at conventional 5% levels (chi-square $p = 0.079$; Cramer's $V = 0.256$). Consequently, the evidence generalizes to provider-covered listed non-financial firms, not automatically to the full Bangladesh non-financial market.

Table 2. Source-universe construct-availability selection diagnostics

Characteristic	Included firms (N=109)	Excluded firms (N=127)	Interpretation
PermID, primary RIC, ISIN, CUSIP, SEDOL, ticker, exchange, TRBC industry, GICS sector	100%	100%	Identifier/classification metadata complete in both groups

Characteristic	Included firms (N=109)	Excluded firms (N=127)	Interpretation
Fiscal-year-end field populated	100.0%	7.1%	Strong provider-coverage difference
Any market capitalization, 2012-2024	97.2%	0.0%	No size comparison is possible for excluded firms
Any enterprise value, 2012-2024	97.2%	0.0%	No EV comparison is possible for excluded firms
Any common shares outstanding, 2012-2024	100.0%	0.0%	No market-history comparison is possible for excluded firms
Communication Services	3.7%	3.9%	GICS sector share
Consumer Discretionary	30.3%	43.3%	GICS sector share
Consumer Staples	11.0%	12.6%	GICS sector share
Energy	5.5%	1.6%	GICS sector share
Health Care	11.9%	7.1%	GICS sector share
Industrials	10.1%	9.4%	GICS sector share
Information Technology	0.9%	2.4%	GICS sector share
Materials	18.3%	18.9%	GICS sector share
Real Estate	0.9%	0.0%	GICS sector share
Utilities	7.3%	0.8%	GICS sector share
Sector-distribution test	chi-square p = 0.079	Cramer's V = 0.256	Moderate composition difference; not significant at 5%

Note: 'Included' means at least one usable DCI observation during 2012-2024. The supplied workbook contains no usable market-capitalization, enterprise-value, or shares-outstanding history for the 127 excluded non-financial firms; therefore size, listing-age, governance, reporting-quality, and survival differences cannot be estimated from these data. This is a population-coverage limitation, not evidence that the groups are economically identical.

Source: Authors' calculations based on the study data.

Table 3. Descriptive statistics for principal risk signals and realized outcomes

Variable	N	Mean	SD	P25	Median	P75
Distress Consensus Index	1,189	49.758	22.095	32.123	50.327	66.921
Score disagreement	1,189	18.986	9.731	11.475	17.789	26.024
Zmijewski risk	1,191	50.546	28.878	25.546	50.515	75.490
Springate risk	1,191	49.454	28.878	24.510	49.485	74.454
Piotroski risk	1,084	49.400	28.363	22.464	47.396	73.469
FHC risk	1,189	49.453	28.878	24.510	49.485	74.286
Raw Fundamentals Risk Index	1,189	49.558	19.067	35.404	51.609	64.019
Next-year accounting stress	1,147	0.261	0.439	0.000	0.000	1.000
Next-year loss	1,147	0.104	0.305	0.000	0.000	0.000
Next-year negative CFO	1,143	0.205	0.404	0.000	0.000	0.000

Note: Risk measures are annual within-year percentiles, with higher values indicating greater relative accounting risk. Outcome descriptives are reported among observations with a usable DCI; sample sizes vary with outcome availability.

Source: Authors' calculations based on the study data.

The annual DCI has a mean of 49.76 and standard deviation of 22.10. The economic ordering is pronounced before covariate adjustment. Across DCI quintiles, next-year accounting-stress incidence rises monotonically from 5.65% in Q1 to 52.17% in Q5, a difference of approximately 46.5 percentage points. Loss, negative operating cash flow, and two-year persistent accounting stress also rise with the consensus ranking, as shown in Table 2 and Figure 1. Because DCI is reranked within each year, this gradient is a cross-sectional ordering of firms rather than evidence that aggregate corporate stress rises with the level of the index over calendar time.

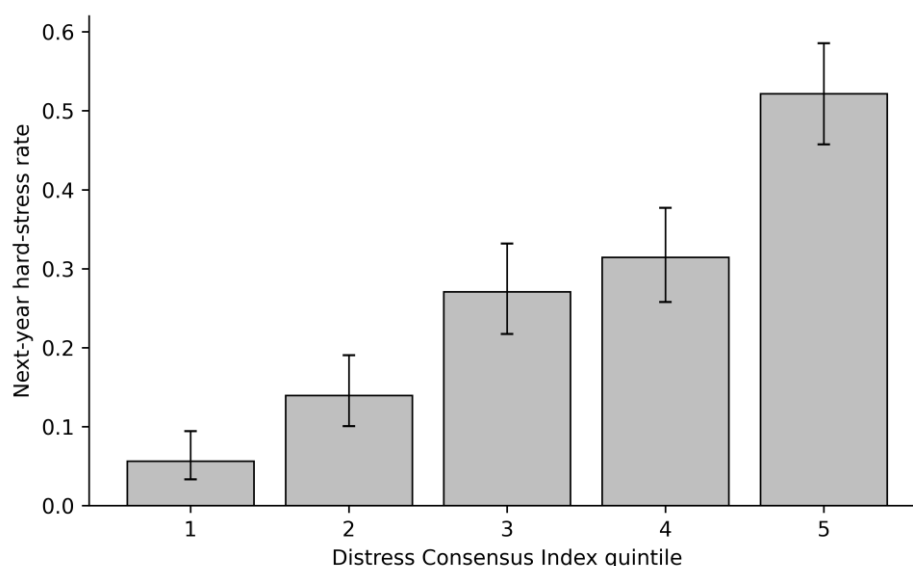
Table 4. Realized future accounting stress across Distress Consensus Index quintiles

Quintile	N	Mean DCI	Accounting stress	Loss	Neg. CFO	Persistent 2y
Q1	230	18.6	5.7%	0.9%	5.7%	1.9%
Q2	229	35.6	14.0%	5.2%	9.6%	7.3%
Q3	229	50.1	27.1%	9.6%	21.4%	11.4%
Q4	229	63.6	31.4%	13.1%	25.0%	20.7%
Q5	230	80.2	52.2%	23.0%	40.6%	31.7%

Note: Percentages are realized future event rates. Persistent 2y equals one when accounting stress occurs in both $t+1$ and $t+2$.

Source: Authors' calculations based on the study data.

Figure 1. Next-year accounting-stress incidence across DCI quintiles



Source: Authors' calculations; error bars are Wilson 95% confidence intervals.

4.2 Dependence across warning systems and the raw-fundamentals benchmark

The component-dependence diagnostics reject any description of the four DCI inputs as statistically independent. Pearson correlations range from 0.148 between Zmijewski and Piotroski risk to 0.796 between Springate and FHC risk. Spearman correlations are closely similar. The overlap is therefore economically meaningful, especially for Springate and FHC, while Piotroski retains substantially more distinct variation.

Table 5. Dependence and principal-component structure of DCI components

Panel A: Pearson correlations	Zmijewski	Springate	Piotroski	FHC
Zmijewski	—			
Springate	0.485	—		
Piotroski	0.148	0.262	—	
FHC	0.563	0.795	0.342	—

Note: Pairwise Pearson correlations are shown below the diagonal. The maximum pairwise correlation is Springate-FHC (0.796). Spearman correlations and pairwise counts are reported in the supplementary analysis.

Source: Authors' calculations based on the study data.

Table 6. (continued). PCA variance decomposition and loadings

Component / loading	Eigenvalue / PC1	Variance / PC2	Cumulative
PC1	2.386	59.6%	59.6%
PC2	0.884	22.1%	81.7%
PC3	0.548	13.7%	95.4%
PC4	0.185	4.6%	100.0%
Loading: Zmijewski	0.474	-0.371	
Loading: Springate	0.570	-0.123	
Loading: Piotroski	0.306	0.919	
Loading: FHC	0.598	-0.059	

Note: PCA uses 1,084 complete four-score observations after standardization. PC1 is oriented so that higher scores correspond to greater average component risk. The first component explains 59.6% of standardized component variance and the first two components explain 81.7%.

Source: Authors' calculations based on the study data.

The PCA confirms a substantial common factor without implying complete redundancy. PC1 explains 59.6% of standardized component variance, while PC2 adds 22.1%; the first two components jointly explain 81.7%. PC1 loads positively on all four component risks, with the smallest loading on Piotroski. The appropriate interpretation is a common accounting-weakness component embedded in partially overlapping systems, not cancellation of independent errors.

The direct RFRI benchmark is intentionally demanding because it aggregates the same six fundamentals used in FHC. The pooled DCI-RFRI Pearson correlation is 0.889, but after residualizing both variables with firm and year fixed effects the correlation declines to 0.840, corresponding to a two-variable VIF of 3.40 rather than near-singularity. In the preferred joint model, DCI remains positive and statistically significant at 0.0448 per 10 points (SE = 0.0154; 95% CI [0.0148, 0.0749]; $p = 0.0035$), while RFRI is weaker (0.0429; SE = 0.0237; 95% CI [-0.0036, 0.0892]; $p = 0.070$). The stricter test removes FHC from DCI before the RFRI comparison. In the resulting complete-score sample, the firm/year-residualized DCI(-FHC)-RFRI correlation is 0.752 (VIF = 2.30). When both are entered jointly with the preferred controls, DCI(-FHC) remains significant at 0.0363 per 10 points (SE = 0.0126; 95% CI [0.0115, 0.0611]; $p = 0.0041$; $N = 1,045$), and RFRI is also significant at 0.0615 (SE = 0.0220; 95% CI [0.0183, 0.1046]; $p = 0.0053$). The full DCI also remains significant in complete four-score observations and with sector-by-year effects, and all leave-one-score-out DCI coefficients remain positive. The incremental

claim is therefore that cross-model synthesis retains information beyond direct duplication of the FHC raw-fundamental component, not that the score systems are statistically independent.

Table 7. Measurement composition and overlap-robustness diagnostics

Panel / specification	N	DCI coefficient per 10	SE / diagnostic	p / result
DCI with all four warning systems	1,084	—	91.2% of DCI observations	—
DCI with three systems: Zmijewski + Springate + FHC	105	—	8.8% of DCI observations	Piotroski missing
FHC with 6 dimensions	1,053	—	88.6% of FHC observations	—
FHC with 5 dimensions	119	—	10.0% of FHC observations	—
FHC with 4 dimensions	17	—	1.4% of FHC observations	—
Preferred + DCI/FHC component-count controls	1,066	0.0689	0.0104	<0.001
Complete four-score DCI only	1,045	0.0725	0.0102	<0.001
Stable 44-firm peer subset	557	0.0677	0.0175	<0.001
Sector-adjusted percentiles (≥ 5 peers per component)	1,002	0.0640	0.0109	<0.001
Leave out Zmijewski	1,045	0.0645	0.0087	<0.001
Leave out Springate	1,045	0.0606	0.0096	<0.001
Leave out Piotroski	1,066	0.0634	0.0112	<0.001
Leave out FHC	1,045	0.0637	0.0094	<0.001
DCI-RFRI pooled correlation	1,189	—	$r = 0.889$	—
DCI-RFRI firm/year residualized correlation	1,066	—	$r = 0.840$; VIF = 3.40	—

Note: Regression rows use firm and year fixed effects, current accounting stress, log assets, strict consecutive-year sales growth, and firm-clustered standard errors unless otherwise stated. Component-count controls are themselves statistically insignificant.

Source: Authors' calculations based on the study data.

Table 8. Annual reference-universe size and Piotroski tie structure

Year	Usable DCI firms	Usable Piotroski firms	Distinct Piotroski scores	Largest tie share
2012	70	45	8	26.7%
2013	80	64	7	23.4%
2014	85	76	7	26.3%
2015	72	68	7	25.0%
2016	89	69	7	21.7%
2017	97	88	9	23.9%
2018	97	93	8	28.0%
2019	100	95	8	23.2%
2020	102	98	9	26.5%
2021	102	99	9	23.2%
2022	101	98	8	29.6%
2023	98	95	8	24.2%
2024	96	96	9	20.8%

Note: Percentiles use average ranks for ties. The stable-peer sensitivity restricts the ranking universe to 44 firms with usable DCI observations in every year from 2012 through 2024.

Source: Authors' calculations based on the study data.

4.3 Baseline within-firm association and incremental information

Table 4 reports the preferred within-firm association and the strengthened incremental-information tests. With firm and year fixed effects, current accounting stress, log assets, and strict sales growth, a 10-point DCI increase is associated with a 6.86-percentage-point increase in next-year accounting-stress probability (SE = 0.0104; 95% CI [0.0483, 0.0889]; $p < 0.001$). The estimate is stable under sector-by-year effects, component-count controls, complete four-score observations, and alternative peer definitions. In the joint DCI-RFRI model, the DCI coefficient is 0.0448 ($p = 0.0035$). When FHC is removed from DCI, the reduced three-system consensus remains significant jointly with RFRI (0.0363; $p = 0.0041$), directly addressing partial duplication between FHC and RFRI. A conditional fixed-effects logit provides a nonlinear analogue: after 38 invariant-outcome firms are removed by conditioning, 702 observations from 69 firms remain and the DCI log-odds coefficient is 0.560 (SE = 0.090; $p < 0.001$), equivalent to an odds ratio of 1.75 (95% CI [1.47, 2.09]) per 10 DCI points.

Table 9. Baseline panel estimates and incremental tests against RFRI

Specification / predictor	Coef./10	SE	95% CI	N	R ²
Sector + year FE	0.0694***	0.0069	[0.0556, 0.0831]	1147	0.202
Firm + year FE	0.0594***	0.0089	[0.0418, 0.0770]	1147	0.405
+ current stress	0.0670***	0.0111	[0.0451, 0.0889]	1147	0.407
Preferred: + strict growth	0.0686***	0.0104	[0.0483, 0.0889]	1066	0.430
Firm + sector x year FE	0.0705***	0.0120	[0.0468, 0.0942]	1066	0.498
Preferred, two-way clustered SE	0.0686***	0.0091	[0.0488, 0.0885]	1066	0.430
Preferred + RFRI: DCI	0.0448**	0.0154	[0.0148, 0.0749]	1066	0.433
Preferred + DCI: RFRI	0.0429	0.0237	[-0.0036, 0.0892]	1066	0.433
Preferred + RFRI: DCI(-FHC)	0.0363**	0.0126	[0.0115, 0.0611]	1045	0.429
Preferred + DCI(-FHC): RFRI	0.0615**	0.0220	[0.0183, 0.1046]	1045	0.429

Note: Dependent variable = next-year accounting stress. Preferred model includes firm and year fixed effects, current accounting stress, log assets, and strictly consecutive-year winsorized sales growth. Except the two-way row, SEs are clustered by firm. RFRI and DCI are both scaled per 10 points. DCI(-FHC) is the equal-weight mean of Zmijewski, Springate, and Piotroski annual risk percentiles in complete-score observations. ***, **, and * denote $p < 0.001$, $p < 0.01$, and $p < 0.05$, respectively.

Source: Authors' calculations based on the study data.

Table 10. Nonlinear and composition robustness of the primary DCI association

Specification	N / firms	Effect per 10 DCI points	95% CI	p
Preferred firm/year FE LPM	1,066 / 107	+0.0686 probability	[0.0483, 0.0889]	<0.001
Conditional fixed-effects logit	702 / 69	log-odds = 0.560; OR = 1.75	OR [1.47, 2.09]	<0.001
Complete four-score LPM	1,045 / 107	+0.0725 probability	[0.0525, 0.0925]	<0.001
Component-count-controlled LPM	1,066 / 107	+0.0689 probability	—	<0.001

Note: Conditional logit conditions out firm fixed effects and necessarily excludes firms with no within-firm variation in the binary outcome. Its model-based standard error is used only for nonlinear robustness. The firm-clustered LPM remains the primary inferential specification.

Source: Authors' calculations based on the study data.

4.4 New stress onset, persistence, recovery, and alternative outcomes

The onset analysis directly addresses the possibility that the baseline association merely reflects persistence in earnings or cash-flow weakness. Among 803 currently non-stressed observations, 127 enter accounting stress in the following year. The DCI coefficient is 0.0610 per 10 points (SE = 0.0128; 95% CI [0.0356, 0.0864]; $p < 0.001$). Hence the early-warning association is present even when current accounting stress is absent by construction.

Among 263 currently stressed observations, higher DCI predicts persistence and lower recovery. DCI is also positively associated with dual stress, repeated loss, and the stricter severe accounting-stress endpoint. Severe accounting stress occurs in 107 of 1,065 preferred-control observations (10.0%) and rises by 4.05 percentage points per 10 DCI points (SE = 0.0077; 95% CI [0.0255, 0.0556]; $p < 0.001$). Repeated loss occurs in 57 of 1,066 observations (5.35%) and rises by 2.64 percentage points per 10 DCI points ($p < 0.001$). These lower-frequency endpoints strengthen construct validity for adverse accounting conditions without converting the outcome into a legal bankruptcy measure.

Table 11. Stress transitions, stricter endpoints, and horizon robustness

Outcome	Events	Coef./10	SE	95% CI	p	N
Stress onset among currently non-stressed	127	0.0610***	0.0128	[0.0356, 0.0864]	<0.001	803
Persistent stress among currently stressed	151	0.0814*	0.0327	[0.0162, 0.1465]	0.015	263
Recovery among currently stressed	112	-0.0814*	0.0327	[-0.1465, -0.0162]	0.015	263
Next-year dual stress (loss AND negative CFO)	52	0.0192**	0.0071	[0.0051, 0.0334]	0.008	1064

Outcome	Events	Coef./10	SE	95% CI	p	N
Next-year loss	114	0.0402***	0.0080	[0.0243, 0.0560]	<0.001	1066
Next-year negative CFO	216	0.0475***	0.0109	[0.0258, 0.0691]	<0.001	1064
Any accounting stress within next 2 years	352	0.0529***	0.0147	[0.0238, 0.0820]	<0.001	944
Persistent accounting stress over t+1,t+2	142	0.0363**	0.0123	[0.0119, 0.0608]	0.004	944
Severe accounting stress (>=2 of 4 red flags)	107	0.0405***	0.0077	[0.0255, 0.0556]	<0.001	1065
Repeated loss (loss at t and t+1)	57	0.0264***	0.0066	[0.0134, 0.0394]	<0.001	1066

Note: All coefficients are from firm/year fixed-effects linear probability models with firm-clustered SEs. Onset is estimated only among currently non-stressed firms and omits current stress. Persistence and recovery are estimated only among currently stressed firms and are complements. Other outcomes use the preferred controls. Next-year weak FHC is reported only in the supplementary analysis because it is construct-consistent rather than independent validation. ***, **, and * denote $p < 0.001$, $p < 0.01$, and $p < 0.05$, respectively.

Source: Authors' calculations based on the study data.

4.5 Attrition, missing future outcomes, and inference robustness

The later lead-outcome design loses relatively few DCI observations: 42 of 1,189 usable DCI firm-years lack a consecutive t+1 accounting-stress outcome. This should not be confused with the earlier source-universe selection in which 127 of 236 non-financial firms never enter DCI construction. Among DCI observations, missing-lead records are somewhat weaker on several dimensions: mean DCI is 53.59 versus 49.62 among retained observations, RFRI is 55.15 versus 49.35, current accounting-stress incidence is higher, and log assets are lower. These patterns justify sensitivity analysis rather than a missing-at-random claim.

In the multivariable observation model, DCI does not significantly predict next-year data availability after current accounting stress, firm size, strict sales growth, sector effects, and a year trend are controlled (odds ratio = 0.964 per 10 DCI points; $p = 0.652$). This lead-availability result does not resolve source-universe coverage selection. Missing-lead observations are not labeled as delistings because the data cannot distinguish delisting or suspension from merger, restructuring, temporary non-reporting, or provider coverage gaps.

Table 12. Sample flow, missing-outcome bounds, and covariance robustness

Panel A: sample flow	Observations	Firms
Potential non-financial predictor firm-years, 2012-2024	3,068	236
At least one computable core warning signal	1,191	109
Usable DCI (at least 3 of 4 signals)	1,189	109
Usable DCI and observed consecutive t+1 accounting-stress outcome	1,147	108
Plus current stress and log assets	1,147	108
Final preferred sample with strict consecutive-year sales growth	1,066	107

Note: Potential predictor firm-years cover 2012-2024. A usable DCI requires at least three of four core component scores. Strict sales growth requires immediately consecutive t and $t-1$ sales observations.

Source: Authors' calculations based on the study data.

Table 13. (continued). Attrition sensitivity

Scenario	DCI/10	SE	95% CI	N
Observed leads only	0.0686***	0.0104	[0.0481, 0.0892]	1066
Missing leads = stressed	0.0701***	0.0107	[0.0489, 0.0913]	1106
Missing leads = non-stressed	0.0622***	0.0099	[0.0426, 0.0818]	1106
Inverse-probability-of-observation weighted	0.0687***	0.0104	[0.0484, 0.0890]	1066

Note: The two missing-outcome scenarios are extreme bounds, not alternative preferred datasets. Forty otherwise eligible strict-control observations lack a consecutive $t+1$ outcome. Stabilized observation weights are tightly distributed (approximately 0.97-1.21), and the weighted estimate is essentially unchanged. *** denotes $p < 0.001$.

Source: Authors' calculations based on the study data.

Table 14. (continued). Inference sensitivity

Method	DCI/10	SE	p	Cluster information
Firm-clustered (primary)	0.0686	0.0104	<0.001	107 firm clusters
Two-way firm/year clustered	0.0686	0.0091	<0.001	firm + 13 year clusters
Year-cluster wild bootstrap-t (Webb, 9,999)	0.0686	0.0084	0.0002	13 year clusters

Note: The wild cluster bootstrap-t is null-imposed, clustered by year, uses Webb six-point weights, and uses 9,999 replications. The smaller two-way clustered SE is not interpreted as stronger evidence.

Source: Authors' calculations based on the study data.

The preferred association is robust to deliberately extreme assignments of missing future outcomes. The coefficient is 0.0686 using observed leads, 0.0701 when all otherwise eligible missing leads are classified as accounting stress, and 0.0622 when all are classified as non-stressed; all remain $p < 0.001$. Inverse-probability-of-observation weighting produces 0.0687 ($p < 0.001$). The year-cluster wild bootstrap-t yields $p = 0.0002$, but this procedure addresses a different dependence concern from firm clustering and is not presented as a substitute for the primary firm-clustered covariance estimator.

4.6 Temporal, frozen, and expanding-window validation

The chronological holdout contains 279 observations and 85 accounting-stress events. Under contemporaneous annual ranking, DCI AUC is 0.780 and AUPRC is 0.614. This version is out-of-time with respect to outcomes and predictive-model fitting but partly transductive with respect to year- t cross-sectional scaling and ranking. The fully historical frozen reconstruction — where FHC winsorization, location, scale, and all component empirical reference distributions are estimated only from 2012-2021 raw inputs — produces AUC 0.766 and AUPRC 0.595. The fully historical expanding version produces AUC 0.768 and AUPRC 0.598. Thus completely eliminating holdout-year cross-sectional information modestly lowers discrimination but leaves substantial out-of-time ranking power. Figures 2 and 3 present the corresponding receiver-operating-characteristic and precision-recall curves.

Paired firm-block bootstrap differences quantify the cost of removing contemporaneous peer information. Annual DCI exceeds the fully historical frozen DCI by 0.0136 AUC points (95% CI [0.0031, 0.0240]) and 0.0198 AUPRC points (95% CI [0.0036, 0.0335]); the Brier difference is only 0.0018 (95% CI [-0.0029, 0.0071]). Relative to the expanding version, the annual AUC

advantage is 0.0116 (95% CI [0.0024, 0.0217]). The interpretation is therefore not that scaling is immaterial, but that DCI remains useful after a deliberately stricter deployment-style leakage control.

Table 15. Temporal holdout predictive performance

Model	N	Events	AUC	AUPRC	Brier	Log loss
Annual contemporaneous DCI	279	85	0.780	0.614	0.176	0.522
Frozen fully historical DCI	279	85	0.766	0.595	0.175	0.519
Expanding fully historical DCI	279	85	0.768	0.598	0.174	0.519
RFRI	279	85	0.753	0.562	0.184	0.543
Raw fundamentals L2 logit	279	85	0.777	0.608	0.176	0.523
Raw fundamentals + DCI logit	279	85	0.797	0.624	0.170	0.507
Raw fundamentals ElasticNet	279	85	0.769	0.594	0.184	0.543
Random forest fundamentals	279	85	0.804	0.603	0.187	0.554
Gradient boosting fundamentals	279	85	0.804	0.614	0.162	0.488
Springate risk	279	85	0.750	0.562	0.187	0.549
FHC risk	279	85	0.730	0.549	0.188	0.556
Piotroski risk	274	83	0.696	0.510	0.194	0.575
Zmijewski risk	279	85	0.659	0.551	0.200	0.592

Note: Predictors are from 2022-2024 and outcomes extend through 2025. The annual contemporaneous DCI is partly transductive only with respect to year- t scaling/ranking; the frozen and expanding DCI variants are fully historical from the raw inputs onward. Calibration coefficients, predictive-model preprocessing parameters, and screening operating points are learned from pre-2022 data. Lower Brier and log loss are better. Piotroski has a smaller common sample because of missing score observations.

Source: Authors' calculations based on the study data.

Table 16. Firm-block bootstrap intervals and paired incremental-performance differences

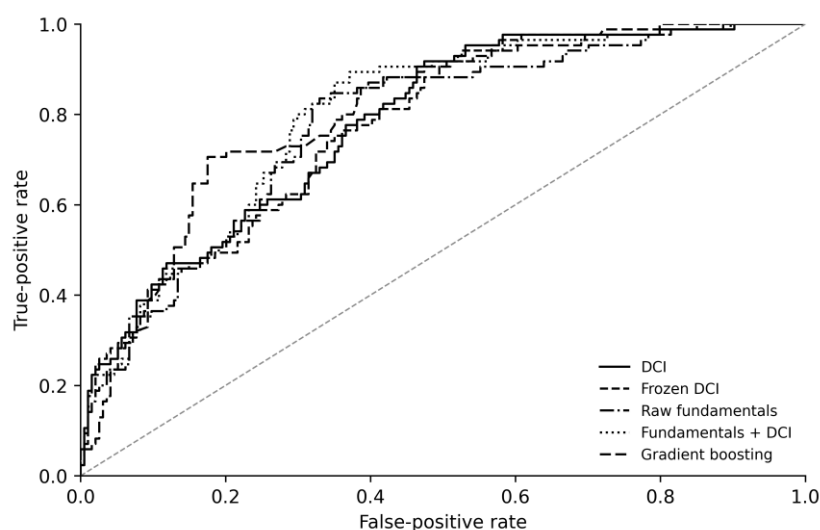
Model / comparison	AUC / 95% CI	AUPRC / 95% CI	Brier / 95% CI	Log-loss / 95% CI
Annual DCI	[0.718, 0.842]	[0.485, 0.740]	[0.148, 0.206]	[0.454, 0.596]
Frozen fully historical DCI	[0.702, 0.833]	[0.471, 0.724]	[0.147, 0.203]	[0.453, 0.586]
Expanding fully historical DCI	[0.702, 0.830]	[0.459, 0.726]	[0.147, 0.203]	[0.453, 0.586]
RFRI	[0.686, 0.813]	[0.448, 0.675]	[0.156, 0.216]	[0.474, 0.622]
Raw fundamentals L2	[0.712, 0.833]	[0.487, 0.727]	[0.152, 0.203]	[0.465, 0.585]
Raw fundamentals + DCI	[0.729, 0.849]	[0.490, 0.730]	[0.144, 0.199]	[0.445, 0.572]
Random forest fundamentals	[0.738, 0.860]	[0.474, 0.727]	[0.161, 0.214]	[0.490, 0.620]
Gradient boosting fundamentals	[0.742, 0.855]	[0.500, 0.723]	[0.135, 0.189]	[0.423, 0.553]

Model / comparison	AUC / 95% CI	AUPRC / 95% CI	Brier / 95% CI	Log-loss / 95% CI
Raw fundamentals + DCI minus raw fundamentals (paired delta)	+0.0196 [0.0034, 0.0380]	+0.0154 [-0.0084, 0.0418]	-0.0056 [-0.0087, -0.0029]	Not estimated in the paired bootstrap

Note: Model-specific intervals are percentile intervals from 1,000 firm-block bootstrap replications. In the final row, delta equals raw fundamentals + DCI minus raw fundamentals on identical holdout observations. Positive delta AUC/AUPRC and negative delta Brier indicate improvement. The paired AUC and Brier intervals exclude zero; the paired AUPRC interval includes zero. The prespecified paired comparison covered AUC, AUPRC, and Brier, not log loss.

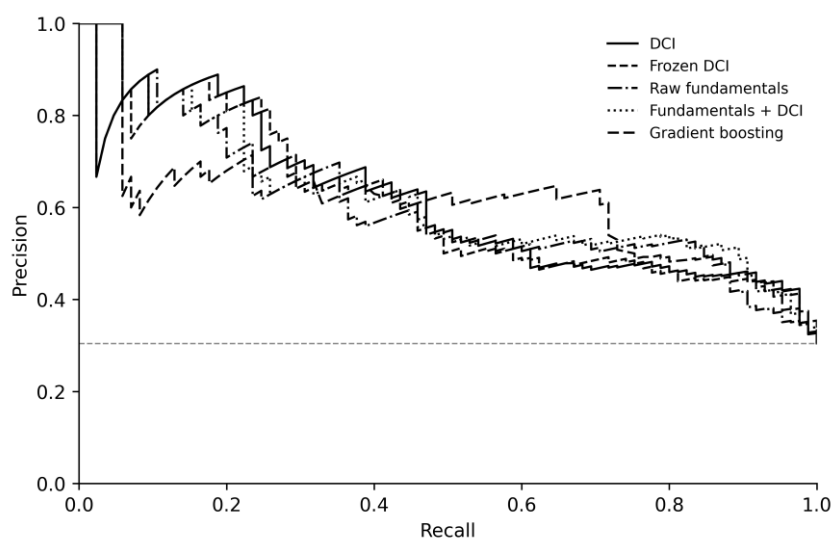
Source: Authors' calculations based on the study data.

Figure 2. ROC curves for selected temporal-holdout models



Source: Authors' calculations. All transformations and model parameters are learned without holdout outcomes.

Figure 3. Precision-recall curves for selected temporal-holdout models



Source: Authors' calculations. The horizontal reference corresponds to holdout event prevalence.

A supplementary five-fold GroupKFold analysis, which prevents any firm from appearing in both estimation and validation folds, yields mean AUC 0.736 (SD = 0.067) and mean AUPRC 0.486 (SD = 0.127). Performance is lower and more variable than in the chronological holdout. Accordingly, the main holdout establishes temporal generalization to later firm-year observations more strongly than generalization to entirely unseen firms.

4.7 Predictive benchmarks, operating thresholds, and calibration

The predictive comparison supports a competitive rather than dominant interpretation of DCI. Raw-fundamentals L2 logistic regression produces AUC 0.777. Adding DCI raises AUC to 0.797. The paired firm-block bootstrap difference is +0.0196 (95% CI [0.0034, 0.0380]). AUPRC rises by 0.0154, but its paired interval includes zero (95% CI [-0.0084, 0.0418]); Brier score improves by -0.0056 (95% CI [-0.0087, -0.0029]), where a negative change is favorable. Thus the incremental gain is supported for discrimination by AUC and for Brier loss, while the AUPRC increase is numerically positive but not statistically resolved at the 95% level. Gradient boosting remains slightly higher in AUC at about 0.804. DCI's comparative advantage is transparent synthesis and incremental feature value, not universal superiority over nonlinear prediction.

Table 17. Holdout operating characteristics at training-selected DCI thresholds

Operating point	Threshold	Sens.	Spec.	PPV	NPV	Bal. acc.	F1	Flagged
Youden J	0.333	47.1%	86.1%	59.7%	78.8%	0.666	0.526	67 (24.0%)
~80% sensitivity	0.178	83.5%	56.2%	45.5%	88.6%	0.699	0.589	156 (55.9%)
~80% specificity	0.345	44.7%	88.7%	63.3%	78.5%	0.667	0.524	60 (21.5%)

Note: These values are training-selected, sample-specific screening operating points based on the frozen pre-2022 calibration mapping and are evaluated unchanged in the holdout. They are not transferable event-risk cutoffs, regulatory thresholds, investment action thresholds, or estimates of a firm's probability of bankruptcy. The high-sensitivity rule detects 83.5% of holdout accounting-stress events but flags 55.9% of observations; the high-specificity rule attains 88.7% specificity and 63.3% PPV while detecting 44.7% of events. Firm-block bootstrap confidence intervals are reported in the supplementary analysis.

Source: Authors' calculations based on the study data.

Operating characteristics illustrate the trade-off relevant to screening. The training-selected Youden operating point (0.333) yields 47.1% sensitivity, 86.1% specificity, and 59.7% PPV in the holdout. The operating point selected to provide approximately 80% training sensitivity (0.178) yields 83.5% holdout sensitivity but only 56.2% specificity. Conversely, the high-specificity operating point (0.345) yields 88.7% specificity and 63.3% PPV but 44.7% sensitivity. These values are sample-specific triage settings produced by the training-period calibration model. They are not transferable risk cutoffs and should not be interpreted as regulatory, lending, audit, or investment action thresholds.

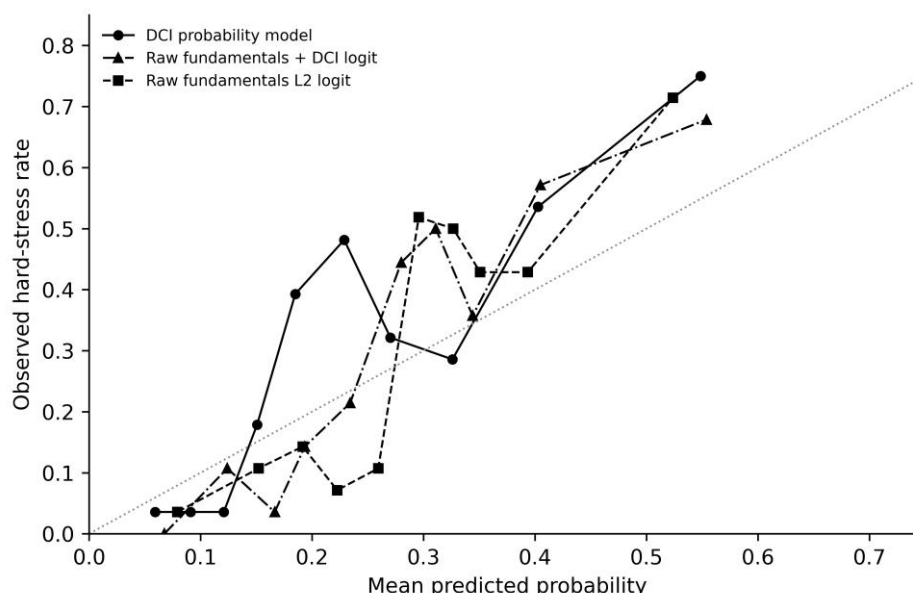
Table 18. (continued). Holdout calibration diagnostics

Model	Intercept	95% CI	Slope	95% CI	Brier	O/E
DCI probability model	0.794	[0.289, 1.299]	1.382	[0.985, 1.779]	0.176	1.278
Raw fundamentals L2 logit	0.974	[0.392, 1.557]	1.997	[1.366, 2.627]	0.176	1.090
Raw fundamentals + DCI logit	0.977	[0.427, 1.527]	1.867	[1.322, 2.413]	0.170	1.137

Note: O/E = observed events divided by the sum of predicted probabilities. Values above one indicate underprediction on average. For the DCI probability mapping, a positive calibration intercept, a slope above one, and O/E above one indicate underprediction in the later holdout and compression of predicted probabilities. Calibration parameters describe this sample and period; they do not convert DCI into a transferable event-risk probability.

Source: Authors' calculations based on the study data.

Figure 4. Holdout calibration by predicted-risk group



Source: Authors' calculations. Points show quantile-group mean predicted probabilities and observed accounting-stress rates; the diagonal indicates perfect calibration.

Calibration is imperfect, especially at higher predicted values, as Figure 4 shows. For the DCI probability model, the holdout calibration intercept is 0.794, the slope is 1.382, and the observed-to-expected ratio is 1.278. The positive intercept and O/E ratio above one indicate meaningful underprediction in the later holdout, while the slope above one indicates that predictions are too compressed. These results reinforce the use of DCI for relative ranking and triage rather than event-risk estimation. The numerical operating points in Table 8 therefore require prospective recalibration before any external deployment and must not be treated as bankruptcy probabilities or transferable regulatory or investment cutoffs.

4.8 Reduced M5 and construct specificity

The reduced five-variable manipulation comparator is attributed to Roxas (2011), not to Beneish's canonical eight-variable M-score. In the fixed-effects sample available for M5, its coefficient is 0.0072 per 10 risk-percentile points ($p = 0.194$). In the joint model, DCI remains strongly associated with next-year accounting stress (0.0703; $p < 0.001$), while M5 remains small and statistically insignificant (0.0034; $p = 0.549$). On an identical 2022-2024 common holdout sample of 219 observations and 67 events, M5 has AUC 0.500 with a firm-block-bootstrap 95% CI of [0.402, 0.590], whereas DCI has AUC 0.804 with 95% CI [0.734, 0.863]. The null M5 result is not a general test of Beneish effectiveness; it shows only that this reduced manipulation-oriented signal does not discriminate the paper's accounting-stress outcome in this dataset.

Table 19. Reduced M5 as a construct-specific comparator

Specification	Predictor	Coef./10	SE	p	N	R ²
Roxas M5 only	m5_risk10	0.0072	0.0056	0.194	711	0.432
Joint: DCI	dci10	0.0703***	0.0144	<0.001	711	0.457
Joint: Roxas M5	m5_risk10	0.0034	0.0057	0.549	711	0.457

Note: Dependent variable = next-year accounting stress. Models include firm and year fixed effects, current stress, log assets, and strictly consecutive-year sales growth. Reduced M5 and DCI are annual risk percentiles scaled per 10 points. *** denotes $p < 0.001$.

Source: Authors' calculations based on the study data.

Table 20. Common-sample temporal discrimination: reduced M5 versus DCI

Model	N	Events	AUC	95% CI
Roxas (2011) reduced M5 risk	219	67	0.500	[0.402, 0.590]
DCI on identical observations	219	67	0.804	[0.734, 0.863]

Note: Both metrics use exactly the same 2022-2024 holdout observations. Confidence intervals are firm-block bootstrap percentile intervals.

Source: Authors' calculations based on the study data.

5. Conclusion and Suggestions

5.1 Conclusions

This study evaluates whether a transparent consensus across partially overlapping but structurally heterogeneous accounting-warning systems contains incremental information about next-year corporate accounting stress. The evidence supports that proposition only for the provider-covered subset of Bangladesh-listed non-financial firms with sufficient Refinitiv/Worldscope accounting data. Of 236 identified non-financial firms, 109 enter DCI construction, and the preferred estimation sample contains 1,066 firm-years from 107 firms. Within that sample, a 10-point DCI increase is associated with a 6.86-percentage-point higher probability of next-year accounting stress; the result survives complete-score restrictions, component-count controls, stable-peer and sector-adjusted ranking, severe accounting-stress endpoints, and a conditional fixed-effects logit.

The dependence tests materially refine the construct claim. Component risk ranks share a substantial common factor, and full DCI and RFRI remain highly correlated after firm/year residualization. Nevertheless, DCI is significant when entered jointly with RFRI. More importantly, when FHC is removed from DCI so that the DCI no longer directly contains the six-fundamental composite underlying RFRI, the reduced three-system consensus remains significant jointly with RFRI (0.0363 per 10 points; 95% CI [0.0115, 0.0611]; $p = 0.0041$). This is stronger evidence of incremental cross-model information, although it does not imply component independence. Fully historical preprocessing lowers holdout AUC modestly from 0.780 to 0.766-0.768, showing that contemporaneous peer information contributes some discrimination but is not the sole source of temporal ranking power. Adding DCI to raw fundamentals increases holdout AUC by 0.0196 in paired bootstrap evaluation, while calibration diagnostics show that later-period event probabilities are underpredicted.

5.2 Suggestions and practical implications

For investors, analysts, lenders, auditors, and regulators, DCI is best used as a triage feature for adverse accounting conditions among provider-covered listed non-financial firms. Convergence of several warning systems can identify firms that merit deeper review of earnings quality, operating cash generation, leverage, liquidity, and financing conditions. DCI should not be used as a stand-alone bankruptcy classifier, an insolvency diagnosis, or an event-risk probability. The numerical operating points reported in Table 8 are sample-specific screening settings and do not constitute regulatory, investment, lending, or audit action thresholds.

For frontier-market research, relative harmonization remains useful when imported absolute cutoffs are questionable, but interpretation must be precise. The contemporaneous annual DCI identifies relative weakness within the observed year-*t* peer set and is partly transductive with respect to cross-sectional scaling and ranking. The frozen and expanding alternatives are fully historical from raw inputs onward. Stable-peer, sector-adjusted, and fully historical tests show that the core association survives alternative reference structures. The initial provider-coverage gap is more consequential for population inference than later lead attrition and should remain a central boundary in subsequent applications.

6. Limitations and Future Research

Several limitations define the scope of the evidence. First, and most importantly for population inference, 127 of 236 identified non-financial firms lack sufficient Refinitiv/Worldscope coverage to enter DCI construction. Because those excluded firms have virtually no usable financial or market series in the supplied workbook, the results cannot be generalized mechanically to the full Bangladesh listed non-financial market. Second, the primary endpoint captures negative earnings and/or negative operating cash flow, not legal bankruptcy, default, restructuring, insolvency, or a verified going-concern event. Severe accounting-stress and repeated-loss analyses narrow the outcome but remain accounting-based. Third, DCI components overlap in profitability, leverage, liquidity, and cash-flow information. The RFRI joint models, including the DCI(-FHC)-RFRI specification, and the leave-one-score-out tests reduce concerns about direct duplication but do not establish statistical independence.

Fourth, the naturally unbalanced panel cannot identify why some future accounting records are missing. Missingness may reflect delisting, suspension, merger, restructuring, temporary non-reporting, or provider coverage gaps. Fifth, annual percentiles depend on the available peer universe and Piotroski has substantial ties; stable-peer and sector-adjusted ranks provide sensitivity evidence but do not create an absolute distress scale. Sixth, conditional fixed-effects logit necessarily discards firms with invariant outcomes and is therefore a robustness model rather than the primary estimator.

Seventh, probability calibration is imperfect: the DCI holdout calibration intercept is positive, the slope exceeds one, and the observed-to-expected ratio is 1.278, indicating later-period underprediction. DCI and the Table 8 operating points should therefore be treated as screening tools rather than calibrated event-risk estimates. Eighth, fully historical scoring modestly reduces holdout discrimination relative to the partly transductive contemporaneous peer ranking. Future work should validate the framework with verified bankruptcy, default, restructuring, or going-concern events; richer coverage of excluded firms; external markets; prospectively frozen preprocessing pipelines; and prospective recalibration of any operating thresholds.

Data availability and reproducibility

The underlying firm-level accounting and market data were obtained from licensed Refinitiv/Worldscope sources and cannot be redistributed publicly. The authors retain the reconstruction code, variable dictionary, exact preprocessing rules, model specifications, deterministic random seeds, and non-proprietary derived outputs and will provide these materials to the editor or reviewers for reproducibility checks subject to journal procedures and the data licence.

References

- Abdullah, M. (2021). The implication of machine learning for financial solvency prediction: An empirical analysis on public listed companies of Bangladesh. *Journal of Asian Business and Economic Studies*, 28(4), 303–320. <https://doi.org/10.1108/JABES-11-2020-0128>
- 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/10.1111/j.1540-6261.1968.tb00843.x>
- Altman, E. I., Iwanicz-Drozowska, M., Laitinen, E. K., & Suvas, A. (2017). Financial distress prediction in an international context: A review and empirical analysis of Altman's Z-score model. *Journal of International Financial Management & Accounting*, 28(2), 131–171. <https://doi.org/10.1111/jifm.12053>
- Ashraf, S., Félix, E. G. S., & Serrasqueiro, Z. (2020). Development and testing of an augmented distress prediction model: A comparative study on a developed and an emerging market. *Journal of Multinational Financial Management*, 57–58, 100659. <https://doi.org/10.1016/j.mulfin.2020.100659>
- Bangladesh Bank. (2026). *Financial stability report 2025*. Retrieved from <https://www.bb.org.bd/pub/annual/fsr/financialstabilityreport2025.pdf>
- 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>
- Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 4, 71–111. <https://doi.org/10.2307/2490171>
- Beneish, M. D. (1999). The detection of earnings manipulation. *Financial Analysts Journal*, 55(5), 24–36. <https://doi.org/10.2469/faj.v55.n5.2296>
- The Business Standard. (2026, April 4). Auditors flag 'going concern' threat for 37 listed companies. *The Business Standard*. Retrieved from <https://www.tbsnews.net/economy/stocks/auditors-flag-going-concern-threat-37-listed-companies-1402596>
- Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3), 414–427. <https://doi.org/10.1162/rest.90.3.414>
- Campbell, J. Y., Hilscher, J., & Szilagyi, J. (2008). In search of distress risk. *The Journal of Finance*, 63(6), 2899–2939. <https://doi.org/10.1111/j.1540-6261.2008.01416.x>
- Dasilas, 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>
- ElBannan, M. A. (2021). On the prediction of financial distress in emerging markets: What matters more? Empirical evidence from Arab Spring countries. *Emerging Markets Review*, 47, 100806. <https://doi.org/10.1016/j.ememar.2021.100806>

- Laitinen, E. K., & Suvas, A. (2016). Financial distress prediction in an international context: Moderating effects of Hofstede's original cultural dimensions. *Journal of Behavioral and Experimental Finance*, 9, 98–118. <https://doi.org/10.1016/j.jbef.2015.11.003>
- Liang, D., Tsai, C.-F., Lu, H.-Y. R., & Chang, L.-S. (2020). Combining corporate governance indicators with stacking ensembles for financial distress prediction. *Journal of Business Research*, 120, 137–146. <https://doi.org/10.1016/j.jbusres.2020.07.052>
- Liu, W., Suzuki, Y., & Du, S. (2025). Ensemble learning algorithms based on EasyEnsemble sampling for financial distress prediction. *Annals of Operations Research*, 346, 2141–2172. <https://doi.org/10.1007/s10479-025-06494-y>
- MacKinnon, J. G., & Webb, M. D. (2017). Wild bootstrap inference for wildly different cluster sizes. *Journal of Applied Econometrics*, 32(2), 233–254. <https://doi.org/10.1002/jae.2508>
- Nugroho, D. S., & Dewayanto, T. (2025). Application of statistics and artificial intelligence for corporate financial distress prediction models: A systematic literature review. *Journal of Modelling in Management*, 20(6), 1999–2023. <https://doi.org/10.1108/JM2-12-2024-0412>
- 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>
- Opitz, D., & Maclin, R. (1999). Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research*, 11, 169–198. <https://doi.org/10.1613/jair.614>
- Piotroski, J. D. (2000). Value investing: The use of historical financial statement information to separate winners from losers. *Journal of Accounting Research*, 38(Supplement), 1–41. <https://doi.org/10.2307/2672906>
- Rafi, N. H. (2026). Earnings manipulation in Bangladeshi listed firms: Sectoral patterns, Covid-19 anomalies, and discriminant reliability of the Beneish M-Score. *Journal of Financial Markets and Governance*, 4(2), 63–82. <https://doi.org/10.54728/JFMG.202510.00118>
- Rahman, M. J., & Zhu, H. (2024). Predicting financial distress using machine learning approaches: Evidence China. *Journal of Contemporary Accounting & Economics*, 20(1), 100403. <https://doi.org/10.1016/j.jcae.2024.100403>
- Roxas, M. L. (2011). Financial statement fraud detection using ratio and digital analysis. *Journal of Leadership, Accountability and Ethics*, 8(4), 56–66.
- Sayari, N., & Mugan, C. S. (2017). Industry specific financial distress modeling. *BRQ Business Research Quarterly*, 20(1), 45–62. <https://doi.org/10.1016/j.brq.2016.03.003>
- Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business*, 74(1), 101–124. <https://doi.org/10.1086/209665>
- Springate, G. L. V. (1978). *Predicting the possibility of failure in a Canadian firm: A discriminant analysis* [Unpublished master's project]. Simon Fraser University.
- Wang, S., & Chi, G. (2024). Cost-sensitive stacking ensemble learning for company financial distress prediction. *Expert Systems with Applications*, 255, 124525. <https://doi.org/10.1016/j.eswa.2024.124525>
- World Bank. (2026). *Market capitalization of listed domestic companies (current US\$) - Bangladesh [Data set]*. World Development Indicators. Retrieved from <https://data.worldbank.org/indicator/CM.MKT.LCAP.CD?locations=BD>
- Zhao, J., Ouenniche, J., & De Smedt, J. (2024). Survey, classification and critical analysis of the literature on corporate bankruptcy and financial distress prediction. *Machine Learning with Applications*, 15, 100527. <https://doi.org/10.1016/j.mlwa.2024.100527>

- Zhao, Q., Xu, W., & Ji, Y. (2023). Predicting financial distress of Chinese listed companies using machine learning: To what extent does textual disclosure matter? *International Review of Financial Analysis*, 89, 102770. <https://doi.org/10.1016/j.irfa.2023.102770>
- Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 22, 59–82. <https://doi.org/10.2307/2490859>