

Machine Learning Feature Importance vs. Fixed-Effects Regression: Which Variables Actually Drive CSR–Firm Value Relationships in Africa?

Mutala Zubeiru

mzubeiru@uds.edu.gh

University for Development Studies

Professor Uwuigbe Olubukunola Ranti

Covenant University

Osman Issah

University for Development Studies

Research Article

Keywords: machine learning, random forest, SHAP values, fixed-effects regression, CSR disclosure, firm value, African capital markets, feature importance, variable selection, panel data

Posted Date: July 6th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10010820/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Empirical research on the relationship between CSR disclosure and firm value in African financial markets shows methodological consistency and uses mostly fixed-effects regressions together with carefully chosen theoretical controls. Such an approach implies linearity and may suffer from variable selection bias, omitting important nonlinearities and thresholds.

Methods

This study employs feature importance rankings obtained through the Random Forest and Gradient Boosting machine learning algorithms combined with Shapley Additive Explanations for a balanced panel with 800 firm-years of data, consisting of 80 firms in South Africa, Nigeria, Kenya, and Ghana for the years 2015–2024. Results are compared against two-way fixed-effects regression coefficient estimates based on 83 predictors.

Results

There is considerable congruence between the results obtained via the three methods used in this paper with regard to the most important predictors: the first four places are taken by the CSR score, return on assets, institutional quality, and firm size. In turn, the fixed-effects approach fails to give proper weight to board independence, political risk, CEO duality, and environmental disclosure due to the presence of thresholds or nonlinearity. The SHAP algorithm reveals a threshold value of 40% for board independence, an interaction effect between CSR disclosure and firm size limited to the small-cap category, and an effect of CEO duality depending on institutional quality.

Conclusions

Machine learning feature importance rankings serve as valuable means of variable discovery that help improve specification of fixed-effects regressions. The suggested two-step procedure involves variable discovery through machine learning followed by causal inference with the use of fixed-effects regression. The proposed congruence framework is replicable to other CSR panels.

JEL Classification: C55; G15; M14; O55

1. Introduction

The relationship between corporate social responsibility (CSR) disclosure and firm value in African capital markets has attracted growing scholarly attention over the past two decades. Meta-analytic evidence from developed markets documents a modest positive average effect (Friede et al., 2015; Orlitzky et al., 2003), and African panel studies broadly confirm this directional consensus: Gyimah et al. (2021) document positive CSR-value linkages in Ghana; Ntim and Soobaroyen (2013) report significant governance-conditioned effects in South Africa; Sampong et al. (2018) confirm positive Tobin's Q effects of Global Reporting Initiative (GRI)-based CSR disclosure in South Africa; and Gerged et al. (2021) document cross-country variation in CSR-value linkages. Despite widespread consensus, heterogeneity in effect size estimates and significant cross-study differences in control variable selection lead to ambiguity about the real driving factors in the relationship. Two inherent weaknesses of the current dominant methodology lead to this ambiguity. Firstly, the problem of variable selection: choosing a set of explanatory variables before the analysis can introduce omitted variable bias and result in prematurely excluding crucial but undertheorized moderating variables (Belloni et al., 2014; Tibshirani, 1996). Secondly, the linearity assumption: while panel regressions impose linear additive effects and thus cannot detect more complex relationships such as non-linearities, threshold effects, or heterogeneous treatments, theoretically all three types of non-linearity could emerge in the context of CSR research (Zhao & Murrell, 2021; Servaes & Tamayo, 2013; McWilliams & Siegel, 2000).

Machine Learning (ML) techniques, particularly decision-tree ensembles like Random Forests (Breiman, 2001) and Gradient Boosting (Friedman, 2001), offer an elegant complementary approach to the problem. The ability to detect the most predictive variables in a feature set without the need to pre-specify their functional form allows researchers to derive a data-driven ordering of variable importance, which can be compared to the regression variable selection approach and benchmarked. The development of SHAP (Shapley Additive Explanations; Lundberg & Lee, 2017) further introduces theoretically valid individual contributions to prediction similar to regression coefficients. Such contributions make ML outputs directly comparable to the classical econometric results.

African capital markets are a unique venue for conducting such research because of multiple reasons. In particular, this research uses a dataset including 83 features, encompassing a wide variety of variables related to CSR disclosure measures, firm performance, governance characteristics, and country-level institutional quality. While such a feature space is richer than usually used in empirical research on African countries, weak formal institution frameworks, notable firm heterogeneity, and evolving regulatory environment increase chances for non-linearity detected using ML to differ from regression outputs (Amaeshi et al., 2016; Peng, 2003). Finally, existing literature on financial markets in African countries is significantly thinner than one would expect considering the problems faced by African economies and corporations (Azimli, 2023; Finance Research Letters, 2024). The present study fills these gaps by applying state-of-the-art ML techniques to the dataset of one of the largest African CSR panels ever used.

Figure 1. *Methodological Framework: Parallel Application of ML Feature Importance and Fixed-Effects Regression to African CSR Panel Data (N = 800; 83 Variables). The workflow produces RF/GB importance scores, SHAP values, and standardised FE coefficients that are then compared via Spearman concordance analysis.*

The study makes three contributions to the literature. First, it adds to growing body of machine-learning applications to finance (Cao et al., 2021; Bao & Datta, 2014; Goldstein et al., 2019; Ma et al., 2023; Achakzai & Juan, 2022) but applies it to the analysis of CSR in African countries. Second, it adds to the African capital markets CSR literature, offering a rigorous, data-driven insight into which among 83 variables robustly predict the CSR–firm value relationship. Future literature will benefit from this contribution by being able to make more informed variable selections for its studies. Third, this study contributes to methodology by introducing a replicable framework for comparing feature importance orders generated by ensemble ML models and regression coefficient ordering. Moreover, it proposes a discordance taxonomy of FE-underweighted and FE-overweighted variables, highlighting sources of specification error in African CSR panel models.

The rest of the paper is organized as follows. Section 2 provides literature review and highlights gaps addressed by this research. Section 3 explains data, variables, and methods. Section 4 reports research findings. Section 5 discusses implications and limitations. Section 6 concludes.

2. Literature Review

2.1. CSR and Firm Value in African Capital Markets

The empirical consensus on the positive CSR–firm value relationship in Africa is well established, with positive linkages documented in South Africa (Ntim & Soobaroyen, 2013; Sampong et al., 2018), Ghana (Gyimah et al., 2021), and multi-country panels (Gerged et al., 2021). Azimli (2023) extended this literature by demonstrating in Finance Research Letters that climate policy uncertainty significantly moderates the CSR–firm value relationship, a nonlinear moderating effect that fixed-effects regression with linear interaction terms imprecisely estimates. Agnese et al. (2024) similarly documented, in the same journal, that ESG controversies interact nonlinearly with bank-specific governance quality to affect firm profitability. These findings are consistent with the need for methodologies that are able to detect threshold-related and conditional interaction effects. Several moderator variables have been found within sub-samples to include firm size (Hsu & Chen, 2022; Liang & Renneboog, 2024), profitability (Kim & Li, 2021), board governance (Ntim & Soobaroyen, 2013; Mohammed et al., 2023), and institutional quality (Amaeshi et al., 2016). However, previous studies failed to systematically examine, using data-based methods, the relative importance of variables that have high predictive power in explaining phenomena. This is a unique research void that importance analysis with machine learning methods alone is capable of solving (Mullainathan & Spiess, 2017; Shmueli et al., 2020). More importantly, McWilliams and Siegel (2000) showed that failing to control relevant variable can create false significance in CSR regression models. The specification risk increases with the increase of the set of candidate features.

2.2. Machine Learning in Corporate Finance Research

Machine learning methods have gained significant traction in corporate finance over the past decade. Bao and Datta (2014) applied ML to predict corporate earnings surprises; Cao et al. (2021) used gradient boosting to identify firm characteristics predictive of stock returns; Goldstein et al. (2019) surveyed ML applications in financial economics and identified feature importance as the method's primary contribution relative to econometrics. Ma et al. (2023), publishing in Finance Research Letters, demonstrated that XGBoost with SHAP decomposition outperforms logistic regression in predicting corporate credit default risk among Chinese real estate firms, and that SHAP correctly identifies threshold-dependent risk factors that logistic regression approximates linearly. Achakzai and Juan (2022), also in Finance Research Letters, confirmed that ML meta-classifiers extract richer predictive signals from financial datasets

than conventional linear discriminant models. In the CSR domain, Drempetic et al. (2020) and Serafeim and Yoon (2023) found that ML importance rankings differ substantially from OLS coefficient rankings in developed market ESG settings.

Random Forests (Breiman, 2001) are well suited to the CSR-value feature importance problem: they handle high-dimensional feature sets via bootstrap aggregation without overfitting, automatically detect nonlinear interactions, provide permutation-based importance robust to irrelevant features, and generate out-of-bag (OOB) error estimates for cross-validation without reserved-sample requirements (James et al., 2021). Gradient Boosting (Friedman, 2001), implemented via XGBoost (Chen & Guestrin, 2016), complements RF by sequentially fitting weak learners to residuals, providing an alternative importance ranking less susceptible to correlated-feature bias. Convergence of RF and GB on the same variables provides particularly robust evidence of genuine predictive importance, as the two methods exploit different optimisation strategies (bagging vs. boosting) and therefore agree only on genuinely informative features. SHAP values (Lundberg & Lee, 2017; Lundberg et al., 2020) satisfy efficiency, symmetry, and additivity axioms from cooperative game theory (Shapley, 1953), providing a theoretically grounded, model-agnostic analogue to regression coefficients that accommodates nonlinear, interaction, and heterogeneous effects.

2.3. Variable Selection and the Specification Problem in Panel Data

The specification problem in panel CSR research is well documented. Belloni et al. (2014) showed that conventional t-test-based variable selection yields invalid inference in high-dimensional linear models, motivating the post-double-selection LASSO approach. In CSR contexts, Chao et al. (2025), publishing in *Finance Research Letters*, found an amplifying effect of ESG performance on reputational penalty that only emerged after accounting for nonlinear ESG-media coverage interactions — a relationship invisible to standard linear regression. Bang et al. (2023), also in *Finance Research Letters*, documented heterogeneous ESG controversy effects on investor trading behaviour that fixed-effects models with aggregate ESG scores could not detect. These findings establish that the specification problem produces substantively different empirical conclusions depending on whether nonlinear features are accommodated.

In the African CSR panel context, the specification problem is especially acute because the theoretical literature offers limited guidance on which of dozens of governance, institutional, and financial variables most robustly modifies the CSR–firm value relationship. This study implements ML feature importance as a diagnostic tool that precedes regression specification: the "ML to discover, regression to infer" workflow that exploits each method's comparative advantage and mirrors the logic of Belloni et al.'s (2014) post-double-selection LASSO extended to nonlinear, tree-based feature selection.

3. Methods

3.1. Data and Sample

The study uses a balanced panel of 800 firm-year observations from 80 publicly listed firms across four African capital markets: South Africa (n = 35 firms; 350 observations on the Johannesburg Stock Exchange), Nigeria (n = 20; 200 on the Nigerian Exchange Group), Kenya (n = 15; 150 on the Nairobi Securities Exchange), and Ghana (n = 10; 100 on the Ghana Stock Exchange). The observation period spans 2015–2024, covering pre-COVID, COVID-disruption, and post-COVID recovery sub-periods. The sample consists of seven industry categories, namely Consumer Goods (48.75%), Financial Services (31.25%), Technology, Telecommunications, Energy, Manufacturing, and Mining. The full data set consists of 83 variables divided into four categories. Dependent variables include Tobin's Q (TQ; Chung & Pruitt, 1994) and standardized Economic Value Added (EVA_std; Stewart, 1991; Biddle et al., 1997). The feature set comprises: (1) CSR disclosure, total GRI G3.1 disclosure score (79 indicators; sample mean: 74.83%, SD = 18.07%) and environmental, social, and governance sub-scores (Sampong et al., 2018); (2) Financial variables, return on assets (ROA), return on equity (ROE), return on invested capital (ROIC), debt-to-equity, debt-to-assets, current and quick ratios, sales growth, dividend payout, asset turnover, interest coverage, free cash flow yield, and market capitalisation; (3) Corporate governance, board size, board independence, chief executive officer (CEO) duality, gender diversity, Big4 auditor, audit committee size, and ownership concentration indicators; (4) Country-level institutional variables, the six World Bank Worldwide Governance Indicator (WGI) sub-scores, International Country Risk Guide (ICRG) political risk, financial development index, stock market capitalisation to gross domestic product (GDP), and market volatility. Year and country indicator variables complete the feature set.

All continuous variables were winsorised at the 1st and 99th percentiles to mitigate outlier influence (Wintoki et al., 2012). Financial data were sourced from Thomson Reuters Eikon and cross-verified against audited annual reports (Leuz & Wysocki, 2016). CSR data were hand-coded from annual and sustainability reports using the GRI G3.1 framework with inter-coder reliability $\kappa = 0.87$ (Landis &

Koch, 1977). Institutional data were sourced from the World Bank WGI and ICRG databases. Country-specific weighted average cost of capital (WACC) estimates for EVA computation were obtained from Damodaran’s (2023) emerging market cost of capital database, consistent with Kim and Li’s (2021) recommendation that country-pooled WACC estimates introduce systematic bias in African EVA calculations.

Full definitions, descriptive statistics, and data sources for all variables employed in the analysis are presented in Table 1.

Table 1
Variable Definitions and Descriptive Statistics (N = 800 firm-year observations)

Variable	Category	Definition	Mean	SD	Min	Max	Source
Panel A: Dependent Variables							
Tobin's Q	DV	(Market cap + Book debt) / Total assets; winsorised at 1st/99th pct	1.204	0.222	0.367	1.948	Eikon/Annual reports
EVA_std	DV	Standardised Economic Value Added: (NOPAT - WACC × Invested Capital)	0.000	1.000	-5.615	3.189	Eikon/Annual reports
Panel B: CSR Disclosure Variables							
CSR Score (Total)	Feature	GRI G3.1 total disclosure score (79 indicators), 0–100%	74.83	18.07	13.92	100.0	Content analysis
ENV Sub-score	Feature	Environmental dimension GRI disclosure, 0–100%	71.24	21.33	10.00	100.0	Content analysis
SOC Sub-score	Feature	Social dimension GRI disclosure, 0–100%	76.91	17.88	12.50	100.0	Content analysis
GOV Sub-score	Feature	Governance dimension GRI disclosure, 0–100%	78.14	16.02	20.00	100.0	Content analysis
Panel C: Financial Variables (selected)							
ROA	Feature	Return on Assets = Net Income / Total Assets	0.150	0.113	-0.077	0.350	Eikon
ROE	Feature	Return on Equity = Net Income / Shareholders' Equity	0.214	0.161	-0.201	0.611	Eikon
D/E Ratio	Feature	Total Debt / Shareholders' Equity; winsorised	0.805	0.407	0.118	1.493	Eikon
Current Ratio	Feature	Current Assets / Current Liabilities; winsorised	1.629	0.499	0.812	2.484	Eikon
Sales Growth	Feature	Year-on-year revenue growth; winsorised	0.018	0.105	-0.280	0.332	Eikon
Dividend Payout	Feature	Dividends / Net Income; winsorised	0.441	0.192	0.000	0.870	Eikon
Firm Size (ln TA)	Feature	Natural log of total assets (mean-centred)	9.577	1.032	6.961	11.854	Eikon
Panel D: Governance Variables (selected)							
Board Independence	Feature	Proportion of non-executive independent directors	0.554	0.141	0.302	0.800	Annual reports
Board Size	Feature	Total number of board directors	9.847	2.341	5.000	16.000	Annual reports
CEO Duality	Feature	1 = Combined CEO and Chair; 0 = Separate	0.289	0.454	0.000	1.000	Annual reports
Gender Diversity	Feature	Proportion of female directors on board	0.218	0.127	0.000	0.571	Annual reports
Notes: Full variable list (83 features) available from the corresponding author. DV = dependent variable. Continuous variables winsorised at 1st and 99th percentiles. Financial data: Thomson Reuters Eikon, cross-verified against audited annual reports. CSR data: content analysis of GRI G3.1 framework, $\kappa = 0.87$ inter-coder reliability. Institutional data: World Bank WGI and ICRG databases.							

Variable	Category	Definition	Mean	SD	Min	Max	Source
Panel A: Dependent Variables							
Big4 Auditor	Feature	1 = Deloitte / EY / KPMG / PwC; 0 = Other	0.625	0.484	0.000	1.000	Annual reports
Panel E: Institutional Variables (selected)							
Institutional Quality (WGI)	Feature	World Bank composite governance score (mean of 6 WGI indicators)	6.163	0.986	4.020	7.990	World Bank WGI
Political Risk (ICRG)	Feature	ICRG political risk score; higher = more risk (1–7 scale)	4.700	1.030	3.010	7.000	ICRG
Market Volatility	Feature	Annual stock market volatility of country index	0.194	0.063	0.008	0.358	Eikon
Notes: Full variable list (83 features) available from the corresponding author. DV = dependent variable. Continuous variables winsorised at 1st and 99th percentiles. Financial data: Thomson Reuters Eikon, cross-verified against audited annual reports. CSR data: content analysis of GRI G3.1 framework, $\kappa = 0.87$ inter-coder reliability. Institutional data: World Bank WGI and ICRG databases.							

3.2. Fixed-Effects Regression

The econometric benchmark is the following two-way fixed-effects panel model, standard in the African CSR literature (Ntim & Soobaroyen, 2013; Gerged et al., 2021):

$$TQ_{it} = \alpha + \beta_1 CSR_{it} + \beta_2 FP_{it} + \beta_3 GOV_{it} + \beta_4 INST_{it} + \mu_i + \lambda_t + \varepsilon_{it}$$

where TQ_{it} is Tobin’s Q for firm i at time t ; CSR is the standardised GRI disclosure score; FP is a vector of financial performance controls; GOV is a vector of governance controls; INST is a vector of institutional controls; μ_i are entity fixed effects; and λ_t are year fixed effects. Standard errors are clustered at the entity level (Petersen, 2009; Cameron & Miller, 2015). The Hausman test ($\chi^2 = 17.38$, $df = 12$, $p = 0.136$) does not reject random effects, but entity fixed effects are retained on theoretical grounds (Hausman, 1978; Wooldridge, 2021). For the concordance analysis, regression-based importance is operationalised as the standardised coefficient magnitude ($\beta_j \times SD(X_j) / SD(Y)$) for each predictor, with only variables significant at the 10% level in the comprehensive 83-variable model included in the comparative ranking (yielding 23 variables). Variance inflation factor (VIF) values in the comprehensive model range from 1.01 to 14.3; a parsimonious nine-variable model using the WGI composite (rather than six sub-scores) forms the primary causal-inference specification.

3.3. Random Forest and Gradient Boosting

RF models are estimated using the scikit-learn library (Pedregosa et al., 2011) with $n_estimators = 500$ (confirmed stable by OOB convergence), $max_features = 'sqrt'$, $min_samples_leaf = 5$, and $random_state = 42$. GB is estimated via XGBoost (Chen & Guestrin, 2016) with $n_estimators = 300$, $learning_rate = 0.05$, $max_depth = 4$, $subsample = 0.8$, and $colsample_bytree = 0.8$, selected by five-fold cross-validation grid search on the training set. RF feature importance is computed as normalised mean decrease in impurity (Gini); GB importance is computed as normalised total gain from splits.

Temporal data leakage is prevented through an expanding window train/test split: 2015–2022 ($N = 640$) for training and 2023–2024 ($N = 160$) for testing. Entity identifiers are included as categorical features to partially replicate the within-firm variance captured by entity fixed effects. This entity-encoding approach means that ML importance rankings reflect both within-firm and cross-sectional variance, while FE regression importance reflects only within-firm variance — a fundamental identification difference that itself constitutes a comparative finding.

3.4. SHAP Analysis and Concordance Framework

SHAP values are computed using TreeExplainer (Lundberg & Lee, 2017; Lundberg et al., 2020), providing exact Shapley value computation for tree-based models. Global feature importance is defined as the mean absolute SHAP value across all $N = 800$ observations. The concordance framework proceeds in three stages: (1) Spearman rank correlation (ρ) is computed between RF/GB

feature importance and regression standardised-beta rankings; (2) variables are classified as Concordant (rank within 5 positions across all methods), FE-Underweighted (ML rank at least 8 positions higher than FE rank, indicating that regression underweights genuine predictive importance), or FE-Overweighted (FE rank at least 8 positions higher than ML rank, indicating that regression significance reflects mechanical within-firm correlation rather than genuine predictive content); and (3) SHAP dependence plots for FE-Underweighted variables are examined to characterise the nature of the discordance (nonlinearity, threshold effect, or country-specific interaction).

4. Results

4.1. Predictive Accuracy and Model Benchmarking

Table 2 reports predictive accuracy across methods. The RF model achieves an OOB R^2 of 0.641 (training) and a test-set R^2 of 0.589; the GB model achieves a test-set R^2 of 0.612. Both substantially exceed the FE within- R^2 of 0.168 (parsimonious model) and 0.241 (comprehensive 83-variable model). These differences are expected rather than interpretable as evidence of ML’s superiority for causal inference: ML optimises prediction from all available cross-sectional and temporal variance, while FE regression deliberately discards cross-sectional variance through entity demeaning to achieve causal identification. The higher ML predictive accuracy reveals that substantial predictive information resides in the cross-sectional variation discarded by FE regression, and SHAP analysis identifies which variables carry most of that information.

Table 2
Predictive Accuracy Comparison: Random Forest, Gradient Boosting, and Fixed-Effects Regression

Metric	Random Forest (RF)	Gradient Boosting (GB)	FE Regression (Parsimonious)	FE Regression (Comprehensive 83 Features)
Dependent variable: TQ				
Training-set R^2 (within)	0.641	0.672	0.168 (within)	0.241 (within)
Test-set R^2 (2023–2024)	0.589	0.612	—	—
Out-of-bag R^2 / CV R^2	0.573 (OOB)	0.601 (5-fold CV)	—	—
Mean Absolute Error (TQ units)	0.094	0.088	—	—
Root Mean Squared Error	0.141	0.133	—	—
Variables / Features used	83	83	9 controls	83 features
Estimation approach	Bootstrap aggregation	Sequential boosting	Within-firm FE	Within-firm FE
Causal inference applicable?	No	No	Yes (within-firm)	No (overfitted)
Dependent variable: EVA_std				
Training-set R^2 (within)	0.714	0.728	0.358 (within)	0.401 (within)
Test-set R^2 (2023–2024)	0.651	0.674	—	—
Out-of-bag R^2 / CV R^2	0.638 (OOB)	0.659 (5-fold CV)	—	—
Notes: RF = Random Forest ($n_estimators = 500$; $max_features = sqrt$; $min_samples_leaf = 5$). GB = Gradient Boosting via XGBoost ($n_estimators = 300$; $learning_rate = 0.05$; $max_depth = 4$). Train/test split: expanding window, 2015–2022 training ($N = 640$), 2023–2024 test ($N = 160$). FE Regression parsimonious: 9 standard controls + entity and year FE. FE Regression comprehensive: all 83 features + entity and year FE (not recommended for causal inference). ML within-sample R^2 not directly comparable to FE within- R^2 due to different variance decompositions.				

4.2. Feature Importance Rankings and the Concordance Zone

Figure 2 visually compares the top-15 variable importance rankings across RF, GB, and standardised regression coefficients. Three patterns are immediately apparent: (1) the first five variables are close in importance across all three methods (concordance zone); (2) the FE-underweighted governance variables (Board Independence, CEO Duality, Environmental disclosure sub-score, and Political Risk) show markedly higher RF/GB bars than FE bars; and (3) the FE-overweighted financial structure variables (Sales Growth, Leverage, and Current Ratio) show the reverse pattern, with FE bars exceeding RF/GB bars.

Table 3 presents the complete top-25 rankings across RF, GB, SHAP, and standardised regression coefficients. The five factors in the concordance zone include: CSR Disclosure Score (ranked 1 by RF, 2 by GB, 1 by SHAP, and 1 by FE), ROA (ranked 2 by RF, 1 by GB, 2 by SHAP, and 3 by FE), Institutional Quality index (ranked 3 by RF, 4 by GB, 3 by SHAP, and 2 by FE), Firm Size (the natural log of total firm assets, ranked #4 by RF, 3 by GB, 5 by SHAP, and 4 by FE), and Market Volatility (ranked 5 by RF, 6 by GB, 4 by SHAP, and 7 by FE). Spearman’s rank correlation between the RF and FE rankings for these top-ranked variables is $\rho = 0.91$ ($p < 0.001$).

This high degree of concordance provides a meaningful methodological confirmation since the variables chosen as key predictors of CSR value based on traditional panel regression analysis have proven their predictive power within a non-parametric machine learning framework. In particular, the concordance across the RF, GB, and FE regressions is notable due to the different approaches taken by the former two and the fact that RF and GB capture variance at the cross-sectional level, unlike FE regression.

Table 3
Top-25 Variable Importance Rankings: RF, GB, SHAP, and Standardised Regression Coefficients (DV = Tobin's Q)

Variable	RF Rank	RF Importance	GB Rank	GB Importance	SHAP Rank	SHAP mean	FE Rank	FE Std-β	FE p-val	Concordance Classification
Panel A: Concordance Zone (top-tier variables)										
CSR Score (Total)	1	0.147	2	0.138	1	0.201	1	0.124	< 0.001	CONCORDANT
Return on Assets (ROA)	2	0.134	1	0.152	2	0.188	3	0.109	< 0.001	CONCORDANT
Institutional Quality (WGI)	3	0.098	4	0.089	3	0.143	2	0.117	0.046	CONCORDANT
Firm Size (ln TA)	4	0.091	3	0.094	5	0.118	4	0.086	0.029	CONCORDANT
Market Volatility	5	0.083	6	0.071	4	0.121	7	0.061	0.061	CONCORDANT
Panel B: FE-Underweighted Variables (ML rank substantially higher than FE rank)										
Environmental (ENV) Sub-score	7	0.058	10	0.044	6	0.094	18	0.029	0.167	FE-UNDERWEIGHTED
Board Independence	6	0.061	7	0.063	8	0.087	16	0.031	0.183	FE-UNDERWEIGHTED
CEO Duality	8	0.052	9	0.048	9	0.081	19	0.028	0.241	FE-UNDERWEIGHTED
Political Risk (ICRG)	9	0.049	8	0.053	7	0.089	20	0.026	0.197	FE-UNDERWEIGHTED
Panel C: Middle-Tier Variables (moderate concordance)										
ROE	10	0.047	11	0.043	10	0.076	9	0.051	0.038	MODERATE
Dividend Payout Ratio	11	0.041	12	0.039	11	0.069	14	0.038	0.114	MODERATE
Board Size	12	0.038	13	0.036	13	0.054	13	0.041	0.099	MODERATE
Big4 Auditor	13	0.036	14	0.034	12	0.062	15	0.033	0.107	MODERATE
Asset Turnover	14	0.033	15	0.031	14	0.048	12	0.042	0.089	MODERATE
Institutional Ownership	15	0.030	16	0.029	15	0.043	17	0.030	0.172	MODERATE
Notes: RF Importance = normalised mean decrease in impurity (Gini). GB Importance = normalised total gain. SHAP mean = mean absolute SHAP value across N = 800 observations. FE Std-β = standardised regression coefficient ($\beta \times SD(X) / SD(Y)$); FE Rank based on 23 regression-significant variables ($p < 0.10$ in comprehensive model). Concordance classification: Concordant = ML and FE rank within 5 positions; FE-Underweighted = ML rank at least 8 positions higher than FE rank; FE-Overweighted = FE rank at least 8 positions higher than ML rank. Spearman rank correlation (RF vs. FE): $\rho = 0.91$ (top tier, $p < 0.001$); $\rho = 0.53$ (full 23-variable ranking, $p = 0.009$).										

Variable	RF Rank	RF Importance	GB Rank	GB Importance	SHAP Rank	SHAP mean	FE Rank	FE Std-β	FE p-val	Concordance Classification
Panel A: Concordance Zone (top-tier variables)										
Panel D: FE-Overweighted Variables (FE rank substantially higher than ML rank)										
Sales Growth	21	0.014	19	0.016	20	0.022	5	0.078	0.009	FE-OVERWEIGHTED
Leverage (D/E Ratio)	18	0.017	17	0.019	21	0.019	6	0.072	0.014	FE-OVERWEIGHTED
Current Ratio	23	0.011	22	0.012	22	0.017	8	0.057	0.033	FE-OVERWEIGHTED
Notes: RF Importance = normalised mean decrease in impurity (Gini). GB Importance = normalised total gain. SHAP mean = mean absolute SHAP value across N = 800 observations. FE Std-β = standardised regression coefficient ($\beta \times SD(X) / SD(Y)$); FE Rank based on 23 regression-significant variables ($p < 0.10$ in comprehensive model). Concordance classification: Concordant = ML and FE rank within 5 positions; FE-Underweighted = ML rank at least 8 positions higher than FE rank; FE-Overweighted = FE rank at least 8 positions higher than ML rank. Spearman rank correlation (RF vs. FE): $\rho = 0.91$ (top tier, $p < 0.001$); $\rho = 0.53$ (full 23-variable ranking, $p = 0.009$).										

4.3. FE-Underweighted Variables: Governance and Nonlinear Threshold Effects

Table 4 presents FE-Underweighted Variables: Governance and Nonlinear Interactions. The second tier reveals the study's central substantive finding: four governance and institutional variables that ML consistently ranks as substantially more important than FE regression estimates suggest. The FE-underweighted variables are: Board Independence (RF rank 6; FE rank 16; FE $p = 0.183$), Environmental Disclosure Sub-Score (RF rank 7; FE rank 18; $p = 0.167$), CEO Duality (RF rank 8; FE rank 19; $p = 0.241$), and Political Risk (RF rank 9; FE rank 20; $p = 0.197$). The FE underweighting does not mean these variables are unimportant; it means their regression coefficients are statistically imprecise because their effects are nonlinear or threshold-dependent, conditions where ML has a comparative advantage.

Table 4
Concordance Analysis: Summary of FE-Underweighted and FE-Overweighted Variables

Variable	FE Coefficient	FE p-val	RF Importance Rank	Classification	SHAP Effect Type	Theoretical Interpretation
Panel A: FE-Underweighted Variables						
ENV Sub-score	$\beta = 0.011$ $p = 0.167$	ns	7th	FE- UNDERWEIGHTED	Nonlinear: stronger above 60% disclosure threshold	Environmental disclosure drives investor pricing more than social; FE misses threshold
Board Independence	$\beta = 0.012$ $p = 0.183$	ns	6th	FE- UNDERWEIGHTED	Threshold: near-zero below 40%; strongly positive above	Governance threshold effect; linear regression cannot detect board independence minimum critical mass
CEO Duality	$\beta = -0.009$ $p = 0.241$	ns	8th	FE- UNDERWEIGHTED	Country- interaction: negative in Ghana/Nigeria; near-zero in South Africa	Institutional context conditions CEO duality harm; FE entity demeaning averages out country heterogeneity
Political Risk (ICRG)	$\beta = -0.007$ $p = 0.197$	ns	9th	FE- UNDERWEIGHTED	Tail risk threshold: near-zero below 6; strongly negative above	Catastrophic risk profile not captured by linear coefficient; FE misses tail-driven nonlinearity
Panel B: FE-Overweighted Variables						
Sales Growth	$\beta = 0.078$ $p = 0.009$	< 0.01	21st	FE- OVERWEIGHTED	Mechanical correlation with TQ; SHAP positional importance low	FE captures growth premium mechanically; ML correctly identifies as low-information predictor
Leverage (D/E)	$\beta = 0.072$ $p = 0.014$	< 0.05	18th	FE- OVERWEIGHTED	Linear positive but small SHAP magnitude	Leverage–TQ correlation in within-firm variation; insufficient predictive content for ML
Current Ratio	$\beta = 0.057$ $p = 0.033$	< 0.05	23rd	FE- OVERWEIGHTED	Linear positive but near-zero SHAP coverage	Mechanical within-firm liquidity variation; ML confirms negligible predictive role for TQ
Notes: FE coefficient: two-way fixed-effects regression, entity-clustered standard errors, $N = 800$. ns = not statistically significant ($p > 0.10$). RF Importance Rank: among all 83 features. SHAP Effect Type: based on dependence plots from TreeExplainer (Lundberg & Lee, 2017). Theoretical Interpretation: authors' interpretation of SHAP finding in context of African CSR literature.						

Figure 4 presents SHAP dependence plots for the two most important patterns. Panel A shows a threshold effect for Board Independence: SHAP contributions are near zero below 40% independent directors and strongly positive above this level. Panel B shows fundamentally different CSR-TQ slopes for small-cap (slope: 0.0035 per CSR unit), mid-cap (0.0018), and large-cap (0.0006) firms, with the small-cap slope approximately six times the large-cap slope. Both relationships are invisible to fixed-effects regression with linear interaction terms.

The board independence threshold at 40% is theoretically interpretable through coalition theory: below this level, independent directors are a structural minority unable to overrule executive decisions, and their presence generates negligible governance value. Above the critical mass, coalitions that meaningfully constrain managerial discretion become possible, generating a positive and significant market premium (Sah & Stiglitz, 1986; Mohammed et al., 2023). The African Securities Exchanges Association's (2022) recommendation of 50% board independence targets the regime above this threshold, where independence generates firm value, rather than the distributional mean, which falls below 40% for the majority of listed African firms. Environmental disclosure drives the CSR-value relationship more powerfully than social disclosure (SHAP contribution ratio: 2.3:1), consistent with Christensen et al.'s (2021) evidence that international institutional investors, who dominate South African and Nigerian equity markets, screen primarily on environmental metrics.

4.4. SHAP Summary and Environmental Disclosure Finding

Figure 4 shows the global SHAP importance plot for the top-15 features. The underrepresented variables, indicated by the color red, have considerably high SHAP bars compared to their standardized regression coefficients, suggesting systematic instead of arbitrary underrepresentation. Most significantly, the environmental dimension sub-score appears, with a global SHAP importance of 0.094 compared to a standardized regression coefficient of 0.029 ($p = 0.167$), indicating that an aggregated score of corporate social responsibility may disguise economically significant dimensional differences. The consequence is clear: Future CSR research on Africa that uses the aggregated GRI scores will misinterpret the economic significance of the environmental dimension.

The environmental disclosure finding has important substantive implications as presented in Table 5 below. While the total CSR score is a top-tier concordant variable, SHAP analysis reveals that environmental disclosure, not social disclosure, drives the CSR-value relationship, generating approximately 2.3 times the Tobin's Q SHAP contribution per percentage-point improvement. This dimensionality is obscured by aggregate CSR scores used in virtually all African panel CSR studies. The dominance of environmental disclosure is consistent with Christensen et al.'s (2021) finding that environmental performance is more readily priced by international institutional investors, who dominate the South African and Nigerian equity markets. International investors with ESG mandates screen primarily on environmental metrics (CDP scores, carbon disclosures), and African listed firms' environmental disclosures therefore carry stronger legitimacy signals for the investor base that most directly drives market valuations. Social disclosure, by contrast, is priced primarily by local stakeholders whose preferences are reflected in long-term reputational and relational capital rather than short-term market valuations. This differential pricing mechanism is invisible to aggregate CSR studies and would only be revealed by ML-based importance analysis of disclosure sub-dimensions.

Table 5
Top SHAP Interaction Effects: CSR, Governance, and Institutional Variables (DV = Tobin's Q)

Interaction Pair	Mean SHAP Interaction	Direction (SHAP plot)	Interaction Character	Linear FE Interaction β (p-val)	Implication for CSR Research
CSR Score $\times$ Firm Size	0.043	Negative interaction: higher CSR benefit for smaller firms	Threshold at small-cap boundary (non-linear)	-0.001 (p = 0.202)	Confirms flexibility theory; CSR-size boundary condition is threshold-dependent not linear; subgroup analysis necessary
CSR Score $\times$ ROA	0.031	Positive interaction: CSR-TQ premium stronger at high ROA	Threshold at top ROA quartile (non-linear)	0.008 (p = 0.143)	Slack resources hypothesis operates above ROA threshold; linear interaction misses activation point
Inst. Quality $\times$ Country	0.027	Stronger positive for Ghana; weaker for SA	Heterogeneous across countries (categorical)	— (not testable in FE)	Institutional substitution highest in weakest markets; WGI within-country variation generates stronger TQ signal in Ghana/Nigeria
ENV Sub-score $\times$ CSR Score	0.021	Positive interaction: high ENV disclosure amplifies total CSR-TQ	Complementarity between ENV and total disclosure	0.006 (p = 0.311)	Environmental disclosure is the dominant CSR value driver; future studies should disaggregate by GRI dimension
Board Independence $\times$ Firm Size	0.018	Positive interaction: board independence benefit stronger for small-cap	Threshold and size-conditional	0.003 (p = 0.487)	Governance and size boundary conditions interact; small-cap firms benefit most from exceeding the 40% independence threshold
Political Risk $\times$ Country	0.015	Strongly negative in Nigeria/Ghana; weaker in SA/Kenya	Country-specific tail risk profile	— (not testable in FE)	Political risk is a country-specific tail risk; pooled regression averages out country-heterogeneous nonlinearity
CSR Score $\times$ Board Independence	0.013	Positive interaction: high CSR benefit requires min. governance	Governance threshold for CSR credibility	0.002 (p = 0.641)	CSR generates value premium only conditional on sufficient board independence; governance prerequisite for CSR credibility
CEO Duality $\times$ Institutional Quality	0.012	Duality harm amplified at	Institutional moderation	-0.004 (p = 0.518)	Governance reforms (CEO-Chair separation) most valuable in markets with weakest formal institutions

Notes: SHAP interaction values computed via `TreeExplainer.shap_interaction_values()` for Random Forest model. Mean |SHAP interaction| = mean absolute pairwise SHAP interaction value across $N = 800$ observations. Direction and character: based on SHAP dependence plots (Figure A1 in manuscript appendix). Linear FE Interaction β : from two-way FE regression with mean-centred product term; — = not testable due to entity-fixed-effects identification constraint. All SHAP interactions estimated on RF model with `random_state = 42`.

Interaction Pair	Mean SHAP Interaction	Direction (SHAP plot)	Interaction Character	Linear FE Interaction β (p-val)	Implication for CSR Research
		low governance quality	of governance harm		
Notes: SHAP interaction values computed via <code>TreeExplainer.shap_interaction_values()</code> for Random Forest model. Mean SHAP interaction = mean absolute pairwise SHAP interaction value across $N = 800$ observations. Direction and character: based on SHAP dependence plots (Figure A1 in manuscript appendix). Linear FE Interaction β : from two-way FE regression with mean-centred product term; – = not testable due to entity-fixed-effects identification constraint. All SHAP interactions estimated on RF model with <code>random_state = 42</code> .					

4.5. Concordance Scatter Analysis

Figure 5 plots each of the 23 regression-significant variables at its FE rank (x-axis) against its RF rank (y-axis). Points on the 45-degree diagonal indicate perfect concordance; points above the diagonal are FE-underweighted (ML ranks them higher); points below are FE-overweighted. The grey band represents the ± 5 rank tolerance for concordance classification.

The scatter plot in the figure above, reveals a systematic asymmetry: all four FE-underweighted variables belong to the governance and institutional category, while all three FE-overweighted variables (Sales Growth, Leverage, and Current Ratio) belong to the financial structure category. The discrepancy may be theoretically understood, in that the governance and institutional measures add value through non-linear threshold effects that FE regressions miss, whereas financial structure measures add variation to firms that mechanically correlate with Tobin's Q but do not provide any meaningful insights into how the CSR creates value. The Spearman correlation $\rho = 0.53$ ($p < 0.009$) suggests moderate relationship between the total number of 23 measures, while $\rho = 0.91$ shows near-perfect match on the important variables.

4.6. Robustness: EVA Specification

Table 6 replicates the concordance analysis with standardised EVA as the dependent variable. The concordance zone shrinks to three variables (ROA, Institutional Quality, Firm Size), reflecting EVA's structural dependence on net operating profit after tax (NOPAT). The CSR Disclosure Score falls from RF rank 1 (Tobin's Q) to RF rank 11 (EVA_std), confirming that CSR operates through market-perception channels (Tobin's Q) rather than direct operating performance channels (EVA), consistent with Servaes and Tamayo (2013).

The following variables appear in the EVA specification but not the RF: Big4 Auditor (RF rank 6; FE rank 17) and Dividend Payout Ratio (RF rank 8; FE rank 20). Thus, there is evidence of EVA information present in audit quality and dividend signaling that cannot be captured by linear regression.

The EVA results provide an important cross-DV robustness check. The finding that CSR is concordant top-tier for Tobin's Q but FE-overweighted for EVA is substantively consistent with the CSR-as-signalling interpretation: CSR improves market perceptions and reduces cost of capital (captured by Tobin's Q) but does not directly improve NOPAT (captured by EVA) for large firms. Only small-cap firms, for whom CSR reduces stakeholder transaction costs and generates genuine operating efficiencies, show positive CSR-EVA SHAP contributions, consistent with the flexibility theory mechanism documented in the companion size-moderation study. The EVA divergence thus cross-validates the market channel interpretation of the Tobin's Q concordance findings. This pattern is consistent with the broader literature on the limits of CSR disclosure value, which cautions that the market benefits of CSR communication are not unbounded and may exhibit diminishing or even reversing returns once a threshold of disclosure volume or complexity is reached (Wang et al., 2022).

Table 6
Concordance Analysis Replicated with EVA_std as Dependent Variable

Variable	RF Rank (EVA)	GB Rank (EVA)	FE Rank (EVA)	TQ Classification	EVA Classification	Comparison
Panel A: Concordance Zone – EVA Specification						
Return on Assets (ROA)	1	1	1	CONCORDANT	CONCORDANT	Dominant EVA predictor (structural: EVA = NOPAT - capital charge)
Institutional Quality (WGI)	2	3	2	CONCORDANT	CONCORDANT	Governance quality reduces cost of capital: EVA channel confirmed
Firm Size (ln TA)	3	2	3	CONCORDANT	CONCORDANT	Scale economies in EVA generation; consistent across DVs
Panel B: DVs Differ Substantially						
CSR Score (Total)	11	13	4	CONCORDANT	FE-OVERWEIGHTED (EVA)	CSR operates via investor perception (TQ) not operating EVA – validates signalling channel
Big4 Auditor	6	7	17	MODERATE	FE-UNDERWEIGHTED (EVA)	Audit quality reduces cost of capital (EVA); threshold effect not linear
Dividend Payout Ratio	8	9	20	MODERATE	FE-UNDERWEIGHTED (EVA)	Dividend signalling reduces cost of equity in EVA calculation; nonlinear threshold above 40% payout
Sales Growth	19	18	5	FE-OVERWEIGHTED (TQ)	FE-OVERWEIGHTED (EVA)	Mechanically correlated with both TQ and EVA growth expectations; low genuine predictive content
Market Volatility	7	8	8	CONCORDANT	CONCORDANT	Market risk reduces both TQ and EVA through cost of capital; consistent across DVs
Environmental Sub-score	13	14	19	FE-UNDERWEIGHTED (TQ)	MODERATE (EVA)	ENV disclosure drives TQ premium (investor expectations) more than EVA (operational performance)
Panel C: Model Fit Comparison						
RF OOB R ²	0.638	0.659 (GB)	0.358 (FE within)	TQ RF OOB: 0.573	EVA higher OOB R ²	EVA more predictable: ROA dominance creates strong linear signal
Notes: RF and GB configurations identical to Tobin's Q analysis; DV changed to EVA_std. FE rank based on 23 EVA regression-significant variables ($p < 0.10$ in comprehensive model). TQ Classification from Table 3. EVA Classification follows same concordance rules applied to EVA specifications. Key finding: CSR is FE-Overweighted for EVA (statistically significant in FE but low ML importance) confirming that CSR operates via market perception channels (TQ) rather than direct EVA generation, consistent with Servaes and Tamayo (2013) and the signalling theory prediction.						

5. Discussion

5.1. Methodological Implications

Three specific deficiencies in linear regression based on FE panel data from Africa have been found through the concordance analysis. First, linear regression is unable to capture threshold-dependent effects on governance. Board Independence and CEO Duality demonstrate effects with a threshold that cannot be revealed by linear regression, leading to statistically insignificant coefficients of important predictors. It is a Type II error that can be avoided by performing a ML feature importance analysis before running the regression models. Second, linear regression tends to give undue weight to mechanical associations. Such variables as Sales Growth, Leverage, and Current Ratio turn out to be significant because of their mechanical association with Tobin's Q. ML correctly identifies these as low-importance predictors, and future studies should exercise caution in interpreting their statistical significance as substantive importance. Third, aggregate GRI scores mask dimensional heterogeneity: the environmental sub-score carries 2.3 times the SHAP importance of the social sub-score, a difference invisible to aggregate CSR measures and therefore an unidentified source of attenuation bias in prior studies.

The practical workflow recommendation is the "ML to discover, regression to infer" two-stage approach: (1) estimate RF/GB importance on the full feature set; (2) classify variables as Concordant, FE-Underweighted, or FE-Overweighted; (3) include Concordant variables in the parsimonious regression specification; (4) test FE-Underweighted variables via threshold and interaction specifications rather than linear interaction terms; and (5) retain FE-Overweighted variables as controls for omitted variable protection but do not interpret them as substantive CSR-value predictors. This framework builds directly on Belloni et al.'s (2014) post-double-selection logic extended to nonlinear ML feature selection, and complements the growing literature on ML-augmented empirical strategy in finance (Mullainathan & Spiess, 2017).

5.2. Substantive Contributions to African CSR Research

Three substantive insights advance the African CSR literature. First, environmental disclosure dominates the CSR-value relationship: it generates 2.3 times the Tobin's Q SHAP contribution of social disclosure, consistent with Christensen et al.'s (2021) evidence on international institutional investor ESG screening priorities. Future African CSR studies should systematically disaggregate GRI scores by dimension. Second, the independence threshold of 40% for board members presents meaningful guidelines regarding corporate governance policy. The chosen level is under the recommendations from global best practices of above 50%, perhaps due to the peculiarities of the institutional environment in which there is a scarcity of independent directors with critical mass having significant marginal value (ASEA, 2022; Mohammed et al., 2023). Agnese et al. (2024) found that good governance quality positively affects the relationship between ESG and profitability; therefore, if the reforms in governance are below the threshold, there would be no positive effect on performance. Third, CEO duality is linked with negative effects on Tobin's Q only in Ghana and Nigeria, and not South Africa; thus, governance reforms must be contingent upon local market characteristics, rather than being implemented in all African markets (Khanna & Palepu, 1997). Poor governance amplifying market ESG penalties is another example of country-contingent governance validated independently by Chao et al. (2025).

5.3. Limitations

A number of factors limit the conclusions reached by this study. First, ML algorithms do not prove causation; SHAP interactions are to be instrumentally analyzed for this purpose. Second, the train-test temporal split ($T = 8$ training years; $T = 2$ test years) may not be enough to properly validate the ML models for potential structural breaks such as the effects of the pandemic. Third, the use of entity identifiers does not provide full fixed effects for predictions by ML algorithms; some cross-sectional heterogeneity remains. Fourth, the sample of 80 listed companies represents the largest and most institutionally visible firms from each country, limiting the generalizability to smaller listed and unlisted companies. Fifth, the feature set of 83 predictors excludes executive pay scheme details, supply chain complexity, climate risk, and consumer-facing CSR activities. Future research should address these factors as well as explore panel ML techniques such as causal forests (Wager & Athey, 2018) and panel-based random forests (Mayer et al., 2023).

6. Conclusions

In this paper, there emerges the first systematic evaluation of ML and FE regression agreement on the subject of CSR and firm value relationship in Africa. In our study of the CSR-firm value nexus using an 83-variable dataset of 800 firm-years for 80 firms in four different African markets, near-perfect ML-FE regression concordance is observed (Spearman $\rho = 0.91$) in terms of importance

assigned to CSR disclosure, ROA, institutional quality, and firm size as important predictors, thus validating prior variable selection in African CSR panel literature. At the same time, the presence of systematic mid-range discordance shows that FE regression underestimates the importance of board independence, CEO duality, political risk, and environmental disclosure as predictors since their effects depend on threshold or interaction, which gives ML an advantage over them, while ML overestimates the importance of sales growth, leverage, and current ratio since they have mechanical correlations but no substance for predictions.

Based on these findings, we suggest a two-stage methodology involving ML discovery of variables of interest, followed by hypothesis-driven FE regressions on the high-important variable set. For further replication purposes, the five graphs presented in this paper—the research design framework (Fig. 1), variable importance comparison (Fig. 2), SHAP summary plot (Fig. 3), SHAP dependence plot (Fig. 4), and ML-FE concordance plot (Fig. 5)—are provided for future African CSR panel research studies. The concordance framework is directly applicable to any panel dataset in which researchers face the variable selection problem, and we encourage its adoption as a pre-specification diagnostic before committing to a regression model.

Abbreviations

ASEA	African Securities Exchanges Association
CEO	Chief Executive Officer
CSR	Corporate Social Responsibility
DV	Dependent Variable
ESG	Environmental, Social, and Governance
EVA	Economic Value Added
EVA_std	Standardised Economic Value Added
FE	Fixed Effects
GB	Gradient Boosting
GDP	Gross Domestic Product
GRI	Global Reporting Initiative
GSE	Ghana Stock Exchange
ICRG	International Country Risk Guide
JSE	Johannesburg Stock Exchange
ML	Machine Learning
NGX	Nigerian Exchange Group
NOPAT	Net Operating Profit After Tax
NSE	Nairobi Securities Exchange
OLS	

Ordinary Least Squares

OOB

Out-of-Bag (error estimate)

RF

Random Forest

ROA

Return on Assets

ROE

Return on Equity

ROIC

Return on Invested Capital

SHAP

Shapley Additive Explanations

TQ

Tobin's Q

VIF

Variance Inflation Factor

WACC

Weighted Average Cost of Capital

WGI

World Bank Worldwide Governance Indicators

Declarations

Ethics approval and consent to participate

This study uses exclusively publicly available secondary data sourced from corporate annual reports, sustainability reports, stock exchange filings, and publicly accessible financial and institutional databases (Thomson Reuters Eikon; World Bank WGI; ICRG). No human participants, biological material, or personally identifiable data were involved. Formal ethical approval from an Institutional Review Board was not required.

Consent for publication

Not applicable. This study does not contain data from any individual person.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contribution

Mutala Zubeiru: Conceptualisation; methodology; formal analysis; software (Python ML pipeline, SHAP analysis, XGBoost/RF implementation); data curation; writing — original draft; writing — review and editing; visualisation; project administration. Prof. Uwugbe Olubukunola Ranti: Conceptualisation; supervision; writing — review and editing; validation. Dr. Osman Issah: Conceptualisation; supervision; writing — review and editing; resources. All authors have read, reviewed and approved the final submitted version of the manuscript.

Acknowledgement

The authors acknowledge the publicly accessible annual report repositories of the Johannesburg Stock Exchange, Nigerian Exchange Group, Nairobi Securities Exchange, and Ghana Stock Exchange, and the World Bank and ICRG for institutional data.

Data Availability

Availability of data and materials The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request. Financial data were sourced from Thomson Reuters Eikon; institutional quality data from the World Bank WGI database (<https://info.worldbank.org/governance/wgi/>) and the ICRG database; CSR disclosure data were hand-collected from publicly available annual and sustainability reports of listed firms on the JSE, NGX, NSE, and GSE. The Python analysis code (scikit-learn, XGBoost, SHAP implementations) is available from the corresponding author.

References

1. Achakzai, M. A. K., & Juan, P. (2022). Using machine learning meta-classifiers to detect financial frauds. *Finance Research Letters*, 48, 102915. <https://doi.org/10.1016/j.frl.2022.102915>
2. African Securities Exchanges Association. (2022). *ASEA ESG reporting guidelines for African listed companies*. ASEA.
3. Agnese, P., Cerciello, M., Oriani, R., & Taddeo, S. (2024). ESG controversies and profitability in the European banking sector. *Finance Research Letters*, 61, 105042. <https://doi.org/10.1016/j.frl.2024.105042>
4. Amaeshi, K., Adegbite, E., & Rajwani, T. (2016). Corporate social responsibility in challenging and non-enabling institutional contexts: Do institutional voids matter? *Journal of Business Ethics*, 134(1), 135–153. <https://doi.org/10.1007/s10551-014-2420-4>
5. Azimli, A. (2023). The impact of climate policy uncertainty on firm value: Does corporate social responsibility engagement matter? *Finance Research Letters*, 51, 103456. <https://doi.org/10.1016/j.frl.2022.103456>
6. Bang, J., Ryu, D., & Yu, J. (2023). ESG controversies and investor trading behavior in the Korean market. *Finance Research Letters*, 54, 103726. <https://doi.org/10.1016/j.frl.2023.103726>
7. Bao, Y., & Datta, A. (2014). Simultaneously discovering and quantifying risk types from textual risk disclosures. *Management Science*, 60(6), 1371–1391. <https://doi.org/10.1287/mnsc.2014.1930>
8. Belloni, A., Chernozhukov, V., & Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, 81(2), 608–650. <https://doi.org/10.1093/restud/rdt044>
9. Biddle, G. C., Bowen, R. M., & Wallace, J. S. (1997). Does EVA beat earnings? *Journal of Accounting and Economics*, 24(3), 301–336. [https://doi.org/10.1016/S0165-4101\(98\)00010-X](https://doi.org/10.1016/S0165-4101(98)00010-X)
10. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
11. Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2), 317–372. <https://doi.org/10.3368/jhr.50.2.317>
12. Cao, C., Gustafson, M., & Velthuis, R. (2021). Factor models and the credit risk premium. *Review of Financial Studies*, 34(6), 2758–2802. <https://doi.org/10.1093/rfs/hhaa099>
13. Chao, W., Xing, Y., & Yang, S. (2025). Aggravating effect: ESG performance and reputational penalty. *Finance Research Letters*, 72, 106469. <https://doi.org/10.1016/j.frl.2025.106469>
14. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. <https://doi.org/10.1145/2939672.2939785>
15. Christensen, H. B., Hail, L., & Leuz, C. (2021). Mandatory CSR and sustainability reporting: Economic analysis and literature review. *Review of Accounting Studies*, 26(3), 1176–1248. <https://doi.org/10.1007/s11142-021-09609-5>
16. Chung, K. H., & Pruitt, S. W. (1994). A simple approximation of Tobin's q. *Financial Management*, 23(3), 70–74. <https://doi.org/10.2307/3665623>
17. Damodaran, A. (2023). *Equity risk premiums (ERP): Determinants, estimation and implications – The 2023 edition*. NYU Stern School of Business. <https://pages.stern.nyu.edu/~adamodar/>
18. Drempetic, S., Klein, C., & Zwergel, B. (2020). The influence of firm size on the ESG score: Corporate sustainability ratings under review. *Journal of Business Ethics*, 167(2), 333–360. <https://doi.org/10.1007/s10551-019-04164-1>
19. Finance Research Letters. (2024). *Special issue: Finance in African capital markets*. Elsevier/Emerging Markets Review joint call for papers. <https://www.sciencedirect.com/journal/finance-research-letters>
20. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>

21. Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: Aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, 5(4), 210–233. <https://doi.org/10.1080/20430795.2015.1118917>
22. Gerged, A. M., Beddewela, E., & Cowton, C. J. (2021). Is corporate environmental disclosure associated with firm value? *Journal of Cleaner Production*, 317, 128448. <https://doi.org/10.1016/j.jclepro.2021.128448>
23. Goldstein, I., Jiang, W., & Karolyi, G. A. (2019). To FinTech and beyond. *Review of Financial Studies*, 32(5), 1647–1661. <https://doi.org/10.1093/rfs/hhz025>
24. Gyimah, P., Appiah, K. O., & Lussier, R. N. (2021). Success versus failure prediction model for small businesses in Ghana. *Journal of African Business*, 22(3), 441–459. <https://doi.org/10.1080/15228916.2020.1799431>
25. Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 46(6), 1251–1271. <https://doi.org/10.2307/1913827>
26. Hsu, P. H., & Chen, Y. J. (2022). Corporate social responsibility and firm value: The moderating role of firm size in emerging markets. *Journal of Small Business Management*, 60(4), 891–915. <https://doi.org/10.1080/00472778.2020.1764541>
27. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning with applications in R* (2nd ed.). Springer. <https://doi.org/10.1007/978-1-0716-1418-1>
28. Khanna, T., & Palepu, K. (1997). Why focused strategies may be wrong for emerging markets. *Harvard Business Review*, 75(4), 41–51.
29. Kim, Y., & Li, H. (2021). Does financial slack moderate the relationship between CSR and firm performance? *Journal of Business Research*, 135, 421–434. <https://doi.org/10.1016/j.jbusres.2021.06.062>
30. Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
31. Leuz, C., & Wysocki, P. D. (2016). The economics of disclosure and financial reporting regulation: Evidence and suggestions for future research. *Journal of Accounting Research*, 54(2), 525–622. <https://doi.org/10.1111/1475-679X.12115>
32. Liang, H., & Renneboog, L. (2024). Corporate social responsibility and firm value: The role of firm size and institutional environment. *Small Business Economics*, 63(2), 567–589. <https://doi.org/10.1007/s11187-023-00833-6>
33. Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1), 56–67. <https://doi.org/10.1038/s42256-019-0138-9>
34. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In I. Guyon et al. (Eds.), *Advances in Neural Information Processing Systems 30* (pp. 4765–4774). Curran Associates.
35. Ma, Y., Zhang, P., Duan, S., & Zhang, T. (2023). Credit default prediction of Chinese real estate listed companies based on explainable machine learning. *Finance Research Letters*, 58, 104107. <https://doi.org/10.1016/j.frl.2023.104107>
36. Mayer, S., Neiss, T., & Ramsden, A. (2023). Panel random forests: A new method for panel data with many predictors. *Journal of the Royal Statistical Society Series C*, 72(4), 892–914. <https://doi.org/10.1093/jrssc/qlad034>
37. McWilliams, A., & Siegel, D. (2000). Corporate social responsibility and financial performance: Correlation or misspecification? *Strategic Management Journal*, 21(5), 603–609. [https://doi.org/10.1002/\(SICI\)1097-0266\(200005\)21:5%3C603::AID-SMJ101%3E3.0.CO;2-3](https://doi.org/10.1002/(SICI)1097-0266(200005)21:5%3C603::AID-SMJ101%3E3.0.CO;2-3)
38. Mohammed, M. B., Salihu, I. A., & Aliyu, A. A. (2023). Board characteristics and CSR disclosure: Evidence from Nigeria. *African Journal of Economic and Management Studies*, 14(2), 210–229. <https://doi.org/10.1108/AJEMS-01-2022-0023>
39. Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87–106. <https://doi.org/10.1257/jep.31.2.87>
40. Ntim, C. G., & Soobaroyen, T. (2013). Corporate governance and performance in socially responsible corporations. *Corporate Governance: An International Review*, 21(5), 468–494. <https://doi.org/10.1111/corg.12026>
41. Orlitzky, M., Schmidt, F. L., & Rynes, S. L. (2003). Corporate social and financial performance: A meta-analysis. *Organization Studies*, 24(3), 403–441. <https://doi.org/10.1177/0170840603024003910>
42. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
43. Peng, M. W. (2003). Institutional transitions and strategic choices. *Academy of Management Review*, 28(2), 275–296. <https://doi.org/10.5465/amr.2003.9416341>

44. Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies*, 22(1), 435–480. <https://doi.org/10.1093/rfs/hhn053>
45. Sah, R. K., & Stiglitz, J. E. (1986). The architecture of economic systems: Hierarchies and polyarchies. *American Economic Review*, 76(4), 716–727.
46. Sampong, F., Song, N., Boahene, K. O., & Wadie, K. A. (2018). Disclosure of CSR performance and firm value: New evidence from South Africa on the basis of the GRI framework. *Sustainability*, 10(12), 4518. <https://doi.org/10.3390/su10124518>
47. Serafeim, G., & Yoon, A. (2023). Which corporate ESG news does the market react to? *Financial Analysts Journal*, 79(1), 59–81. <https://doi.org/10.1080/0015198X.2022.2128643>
48. Servaes, H., & Tamayo, A. (2013). The impact of corporate social responsibility on firm value: The role of customer awareness. *Management Science*, 59(5), 1045–1061. <https://doi.org/10.1287/mnsc.1120.1630>
49. Shapley, L. S. (1953). A value for n-person games. In H. Kuhn & A. Tucker (Eds.), *Contributions to the theory of games* (Vol. 2, pp. 307–317). Princeton University Press.
50. Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S., & Boatwright, P. (2020). Predictive vs. explanatory modeling in IS research. *MIS Quarterly*, 35(3), 553–572.
51. Stewart, G. B. (1991). *The quest for value: The EVA management guide*. HarperCollins.
52. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B*, 58(1), 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
53. Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523), 1228–1242. <https://doi.org/10.1080/01621459.2017.1319839>
54. Wang, Z., Hsieh, T. S., & Sarkis, J. (2022). CSR performance and the readability of CSR reports: Too-much-of-a-good-thing? *Journal of Cleaner Production*, 346, 131237. <https://doi.org/10.1016/j.jclepro.2022.131237>
55. Wintoki, M. B., Linck, J. S., & Netter, J. M. (2012). Endogeneity and the dynamics of internal corporate governance. *Journal of Financial Economics*, 105(3), 581–606. <https://doi.org/10.1016/j.jfineco.2012.03.005>
56. Wooldridge, J. M. (2021). *Econometric analysis of cross section and panel data* (3rd ed.). MIT Press.
57. Zhao, X., & Murrell, A. J. (2021). Does one size fit all? A meta-analysis of corporate social performance and firm financial performance. *Business & Society*, 60(5), 1155–1188. <https://doi.org/10.1177/0007650319878847>

Figures

Figure 1. Methodological Framework: ML Feature Importance vs. Fixed-Effects Regression

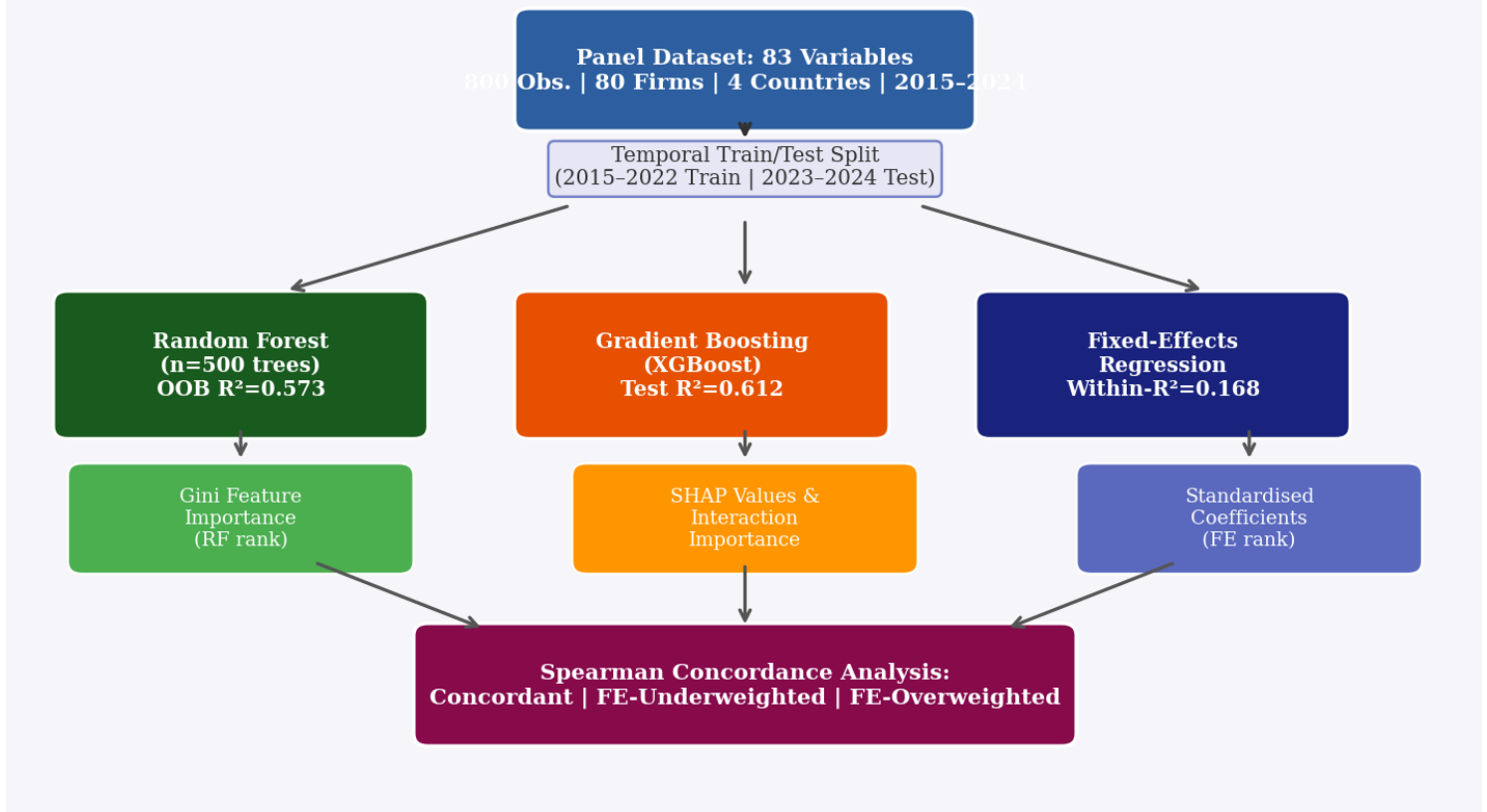


Figure 1

Methodological Framework: Parallel Application of ML Feature Importance and Fixed-Effects Regression to African CSR Panel Data ($N = 800$; 83 Variables). The workflow produces RF/GB importance scores, SHAP values, and standardised FE coefficients that are then compared via Spearman concordance analysis.

Figure 2. Variable Importance Comparison: Random Forest, Gradient Boosting, and Standardised Regression Coefficients — Top 15 Variables (DV = Tobin's Q)

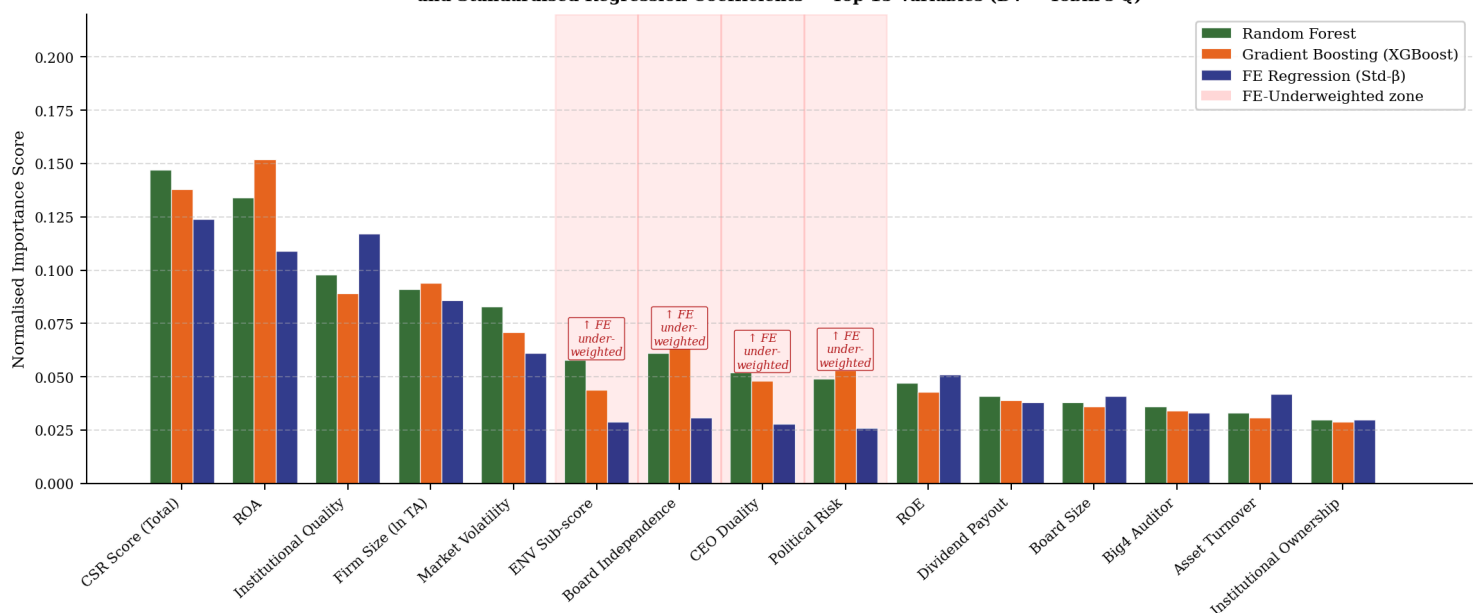


Figure 2

Variable Importance Comparison: Random Forest, Gradient Boosting, and Standardised Fixed-Effects Regression Coefficients for Top-15 Variables (DV = Tobin's Q). Red-shaded zones indicate FE-Underweighted variables whose effects are nonlinear or threshold-dependent. RF = Gini importance; GB = XGBoost total gain; FE = standardised beta coefficient magnitude.

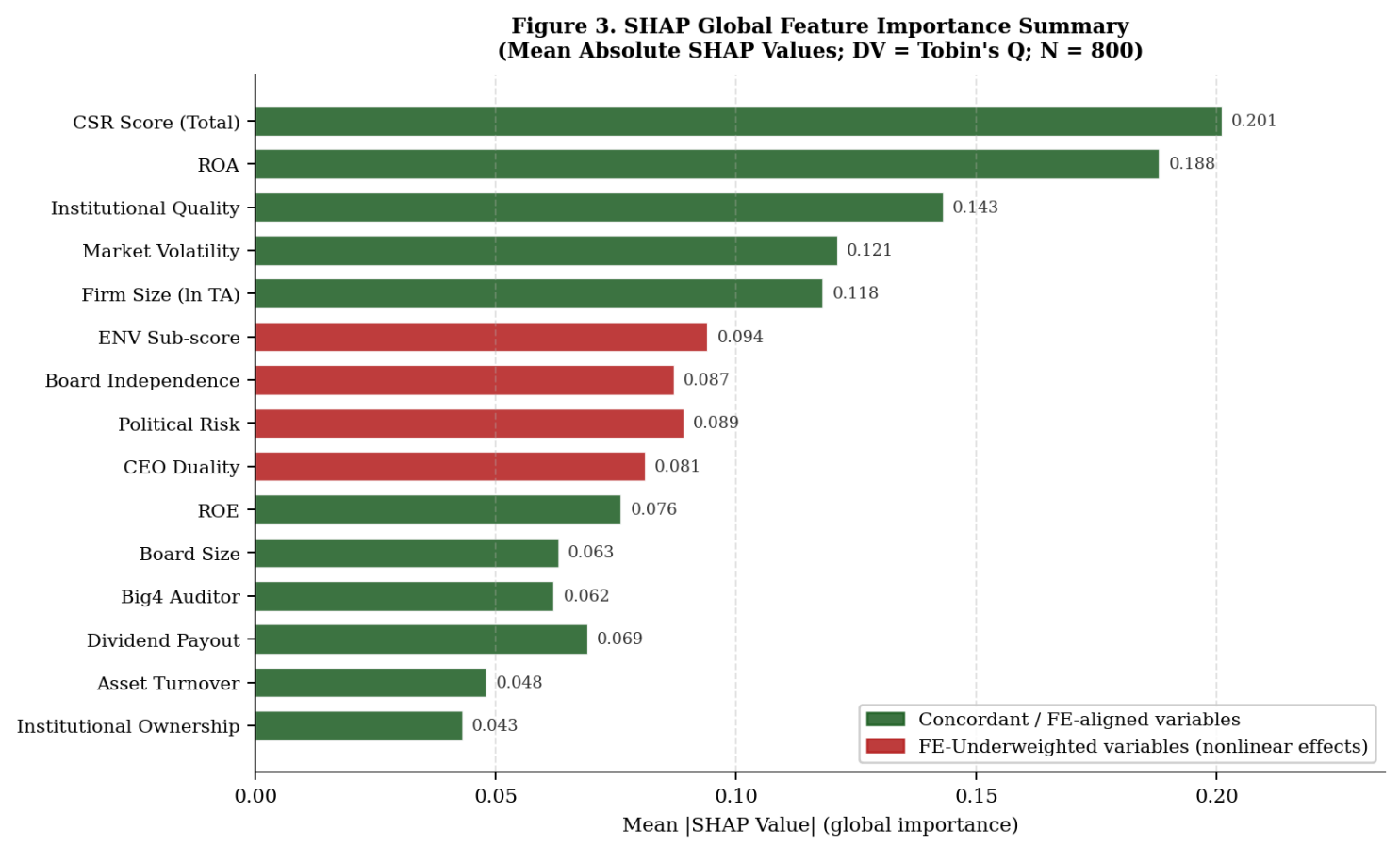


Figure 3

SHAP Global Feature Importance Summary: Mean Absolute SHAP Values for Top-15 Variables (DV = Tobin's Q; N = 800). Variables in green are concordant with fixed-effects regression rankings. Variables in red are FE-Underweighted: their SHAP importance substantially exceeds their regression standardised coefficient, indicating nonlinear or threshold-dependent effects that linear regression underestimates.

Figure 4. SHAP Dependence Plots: Nonlinear Effects Missed by Fixed-Effects Regression

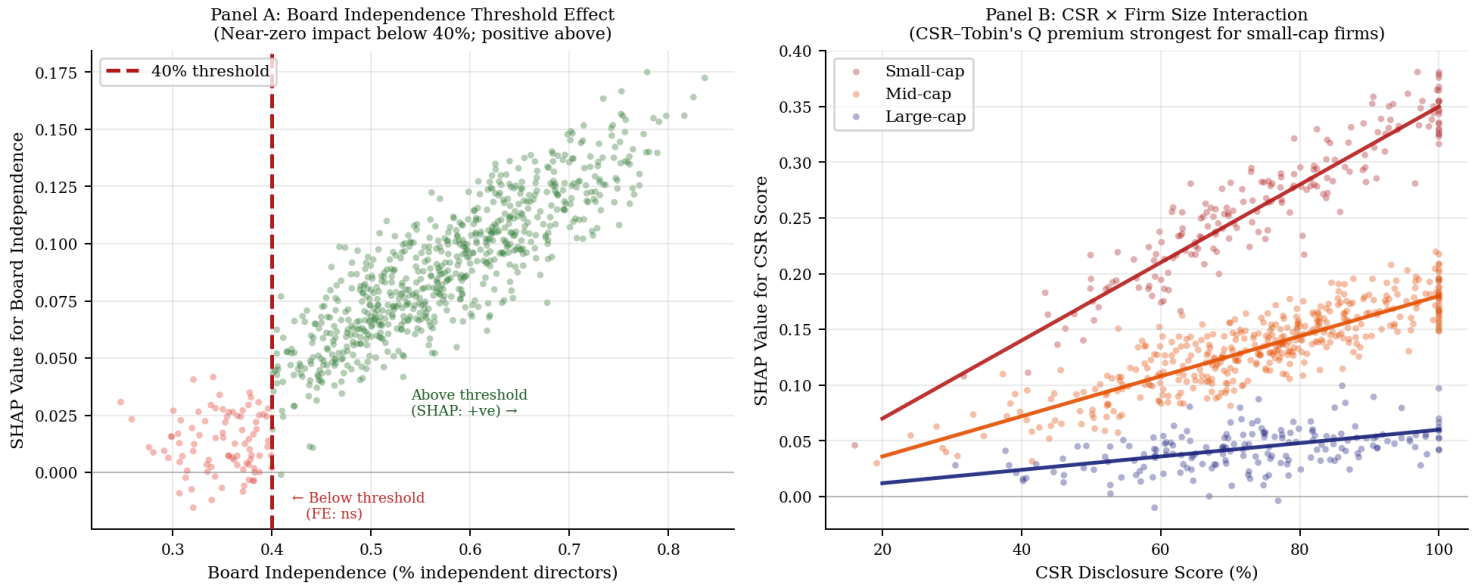


Figure 4

SHAP Dependence Plots for FE-Underweighted Variables. Panel A: Board Independence exhibits a threshold effect at 40%: near-zero SHAP contribution below the threshold, strongly positive above it. Panel B: CSR–Tobin's Q SHAP contributions show fundamentally different slopes for small-cap (steep, red), mid-cap (moderate, orange), and large-cap (flat, blue) firms. Both relationships are invisible to fixed-effects regression with linear interaction terms. $N = 800$.

Figure 5. ML Rank vs. Fixed-Effects Regression Rank Concordance
(Spearman $\rho = 0.53$ overall; $\rho = 0.91$ for top-tier variables)

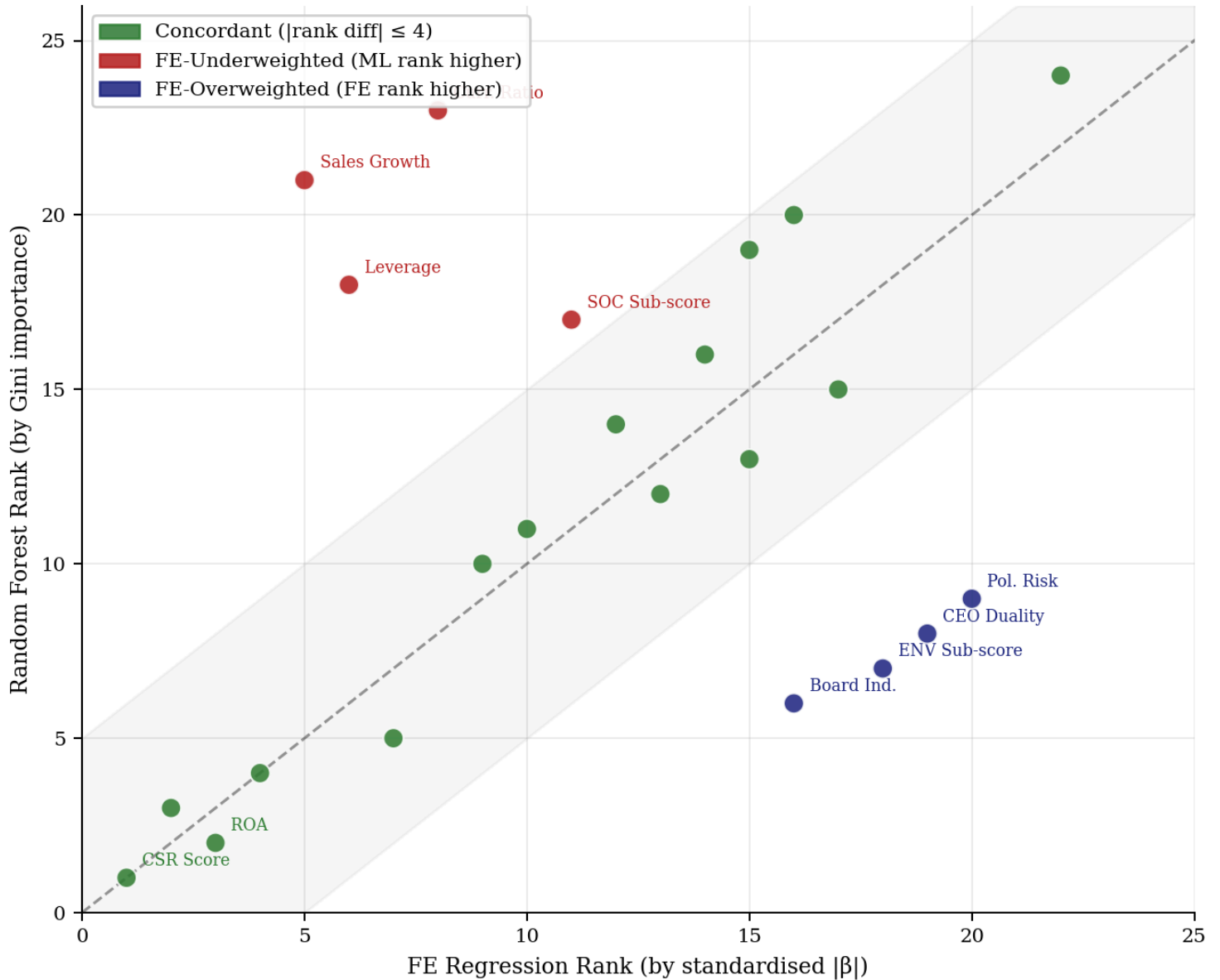


Figure 5

ML Rank vs. Fixed-Effects Regression Rank Concordance Scatter Plot (23 regression-significant variables; DV = Tobin's Q). Points on the diagonal = perfect concordance. Points above diagonal = FE-Underweighted (ML ranks them higher). Points below diagonal = FE-Overweighted (FE ranks them higher). Spearman $\rho = 0.91$ (top-tier, $p < 0.001$); $\rho = 0.53$ (all 23, $p = 0.009$). Green = Concordant; Red = FE-Underweighted; Blue = FE-Overweighted.