Final Submission 04.13.2026

Institutional Isomorphism and the Predictive Structure of MSCI's Environmental Pillar

Authors: Gabrielle Pestana, Mukund Bharadwaj

Abstract: ESG ratings increasingly govern the allocation of climate-aligned capital, yet what these scores actually measure remains empirically contested. This study decomposes the predictive structure of MSCI's Environmental pillar scores using a panel of 73,134 firm-year observations spanning 7,342 companies across 42 countries over the period 2014–2025. A gradient-boosted ensemble trained exclusively on financial variables and lagged peer-group means explains 78% of out-of-sample variance in Environmental scores. Adding approximately 25 direct environmental performance metrics, including GHG intensity, renewable energy share, and science-based target indicators, degrades rather than improves predictive accuracy ($\Delta R^2 = -0.054$). Shapley decomposition attributes 68% of predictive importance to peer-group features and 32% to financial features; environmental variables contribute only 7% even when explicitly included. Group permutation analysis confirms that financial features are seven times more important than environmental features in sustaining model accuracy. These results are invariant across six regional blocs, including emerging markets, and strengthen rather than weaken over a decade spanning multiple MSCI methodology revisions. Score persistence analysis reveals that over half of all firm-year transitions are effectively unchanged. The findings indicate that MSCI's Environmental pillar scores are, in their predictive structure, substantially a product of financial scale and institutional peer-group conformity rather than direct environmental performance, with implications for the $1.1 trillion in ESG and climate-themed index products benchmarked against them.

Keywords: ESG ratings, MSCI, institutional isomorphism, single materiality, double materiality, environmental performance, machine learning, SHAP, sustainable finance

JEL Classification: G11, G24, M41, Q56

1. Introduction

MSCI holds 70% of the Equity Sustainability and Climate index market and quadruples its closest competitor, benchmarking $17 trillion in assets and over $1.1 trillion in ESG and climate-themed products (MSCI, 2025, p. 2). Such concentration exemplifies infrastructural power (Fichtner et al., 2024), positioning the firm as a quasi-standard-setter in global capital allocation. ESG Ratings are structured to assess companies' resilience to financially relevant, industry-specific sustainability risks and opportunities, using an industry-relative letter rating from AAA to CCC (MSCI, 2024a). Methodological choices in this framework have broad implications for investment flows and the allocation of capital toward climate-related objectives.

The rapid institutionalisation of ESG ratings has been accompanied by exponential growth in scholarly attention, expanding at an annual rate of 26.81% between 2013 and 2023 (Al Azizah & Haron, 2025). Yet, fundamental questions about what scores capture remain unresolved: ratings from different providers show average pairwise correlations of only 38% to 71% (Berg et al., 2022), and greater disclosure amplifies rather than resolves rating disagreement (Christensen et al., 2022). These findings showcase a field that has documented the symptoms of rating dysfunction without resolving their empirical origins, a gap the present study addresses by decomposing the drivers of MSCI's Environmental pillar scores.

The climate finance gap is particularly acute given that, while global liquidity is adequate to achieve climate goals, actual financial flows for mitigation fall three to six times short of what is needed by 2030, with the deficit most severe in developing and emerging economies (IPCC, 2022). Here, weak risk-return profiles, economic vulnerability, and institutional limitations restrict commercial investment. By adopting an investor-focused materiality framework that prioritises firm-level financial risk over environmental outcomes, MSCI's approach may reinforce existing barriers to capital allocation for high-impact climate solutions (Crona et al., 2025). However, the precise effects of this dynamic remain empirically untested.

Prior research shows ESG ratings are predictable from structured company data (Choi et al., 2024), add limited explanatory power to asset pricing models beyond observable financial characteristics (Lindsey et al., 2025), and do not consistently correlate with stronger environmental regulatory compliance (Raghunandan & Rajgopal, 2022). Legal analysis further documents that proprietary rating methodologies are opaque, favour firms with greater disclosure capacity with high environmental scores correlating positively with greater CO2 emissions, suggesting environmental ratings may not reliably capture actual environmental performance (Boffo et al., 2020, as cited in Mazzacurati, 2021). Finally, bibliometric mapping also shows that governance and sustainable development are well-established in ESG research, while environmental impact remains underexplored (Al Azizah & Haron, 2025).

The opacity of proprietary ESG rating methodologies has prompted research aimed at reconstructing these processes using machine learning. However, existing efforts remain constrained along one or more critical dimensions: geographic scope, temporal depth, sample breadth, industry coverage, and out-of-sample validation, that individually and collectively limit the generalisability of their findings. The results of the present study expose these limitations as shared across the prior literature: no study to date has combined a multi-country panel, a decade-length observation window, broad industry representation, and rigorous out-of-sample evaluation within a single analytical framework applied to MSCI's Environmental pillar.

Belkhiria et al. (2025) are confined to French-listed firms, with a total sample size of 204 firms making up their dataset. Choi et al. (2024) are similarly confined to the Korean market across a three-year window from 2020 to 2022, with data from 2,485 companies. Aue et al. (2025) focus on using news sources for their predictions, looking at 3,343 U.S. companies rated by Refinitiv from 2018 to 2020. Del Vitto et al. (2023) provide a machine learning model to decompose Refinitiv scoring, but limit the dataset to a single year, 2021, focusing only on three industries, selected by the NAICS Sector Classification, and to the economic regions of the USA, EU, and China. Ferro et al. (2025) span 2010–2023 but reconstruct LSEG's architecture rather than MSCI's and critically, do not report out-of-sample accuracy.

By contrast, the present study addresses all five limitations simultaneously. It assembles 73,134 firm-year observations for 7,342 companies domiciled across 42 countries over the period 2014–2025, spanning all RBICS economy-level sectors and covering both developed and emerging markets across six regional blocs. To our knowledge, this constitutes the largest and most geographically diverse panel yet employed to decompose an ESG rating provider's environmental pillar. Crucially, all results are reported on a strictly held-out test set of 1,416 firms whose data never enter the training process, providing the out-of-sample validation that is absent from several prior studies. This breadth permits tests of temporal stability, cross-regional heterogeneity, and industry-level robustness that are structurally unavailable in narrower samples.

Beyond scale, the study introduces a methodological contribution absent from the existing reverse-engineering literature: the explicit operationalisation of institutional isomorphism as a predictive feature. While institutional theory predicts that firms converge on similar rating outcomes through mimetic and normative pressures (DiMaggio & Powell, 1983), prior machine-learning studies have relied exclusively on firm-level financial and disclosure variables, leaving peer-group conformity pressures unmeasured. The present study constructs lagged leave-one-out group means along eight dimensions: industry, sector, economy, subsector, country, region, market-capitalisation tier, and an industry×size interaction, capturing the rating trajectory of a firm's institutional peer group while preventing information leakage. Their inclusion enables the analysis to decompose the variance in environmental ratings into three theoretically distinct sources: firm-specific financial capacity, field-level institutional conformity, and direct environmental performance. The decomposition is designed to test whether environmental performance constitutes an independent predictive source or is subsumed by the other two.

Taken together, these contributions yield the first large-scale, multi-country, decadal-length decomposition that isolates the relative predictive weight of financial structure, isomorphic peer pressure, and ecological footprint in the construction of MSCI's Environmental pillar scores. By integrating institutional theory with interpretable machine learning, specifically, Shapley-based variance attribution and group permutation importance, the study provides not only a descriptive account of what predicts environmental ratings, but a theoretically grounded explanation of why financial variables coupled with peer-group features dominate. In doing so, it moves the literature beyond the documentation of rating opacity toward a structural diagnosis of the mechanisms through which financial materiality is reproduced in the architecture of environmental assessment.

2. Theoretical Framework

2.1 Sustainable Finance and the Persistence of Financial Materiality

Sustainable finance has evolved through three distinct phases. Initially, it relied on ethical exclusion, removing sectors that conflicted with investor values. The early 2000s marked a shift toward integrating ESG factors into investment analysis to enhance risk-adjusted returns. The current phase aims to align capital allocation with measurable environmental and social outcomes (Busch et al., 2021).

Despite advances in ESG theory, investment practice remains anchored in short-term priorities (Cunha, 2021). Institutional investors overwhelmingly use ESG data for its financial relevance, with 82% prioritizing investment performance over ethical considerations (Amel-Zadeh & Serafeim, 2018). Broad ESG funds eclipse Paris-aligned strategies (Fichtner et al., 2024). At the market level, Sankar et al. (2024) identify a significant carbon premium for Scope 1 emissions, suggesting that investors price transition risks rather than penalize environmental harm. Thus, financial materiality frameworks are deeply embedded, meeting the needs of rating agencies, investors, and regulated firms, and shaping disclosure standards and portfolio mandates (Crona et al., 2025).

Stakeholder-oriented materiality frameworks are marginalised due to their limited compatibility with investment decisions (Bose, 2020). One consequence is that structural barriers to decarbonisation, such as fossil fuel subsidies, which the IPCC (2022) identifies as major obstacles to emissions reduction, remain largely outside the scope of financially oriented ESG assessment. In credit markets, environmental information is viewed mainly through the lens of credit risk management, rather than as a tool for mitigating environmental degradation (Hirschbühl et al., 2025). This persists even though projections price the economic consequences of inaction at $38 trillion by 2050, with estimates spanning from $19 trillion to $59 trillion (Potsdam Institute for Climate Impact Research, 2024).

Institutional investors hold high-ESG firms, yet ESG scores are often negatively correlated with portfolio weightings, raising concerns of greenwashing (Lopez-de-Silanes et al., 2024). Governance scores drive investment decisions, while environmental scores are typically linked to reduced capital allocation. This asymmetry underscores the dominance of financially interpretable ESG elements, especially governance, over environmental performance. Consequently, ESG funds prioritise disclosure quality and rating performance over substantive compliance or meaningful emissions reductions (Raghunandan & Rajgopal, 2022).

This institutional orientation persists even though 88% of global individual investors want to invest sustainably, and 80% believe financial returns can align with positive environmental or social outcomes, especially among younger generations (Morgan Stanley, 2025).

Institutional lock-in, as described by Crona et al. (2025), entrenches financially driven measurement frameworks within disclosure rules, portfolio mandates, and regulations. This creates path dependencies that marginalize impact-oriented standards like the Global Reporting Initiative (GRI), deemed less practical for investment decisions (Bose, 2020), and fosters what Avetisyan and Hockerts (2017) call institutional convergence, a narrowing of ESG approaches toward standardized, finance-centric frameworks. The result is a lock-in of unsustainability: unreliable ESG data leads to capital misallocation, weakens incentives for real environmental progress, and perpetuates poor data quality, reinforcing the status quo (Crona, 2025). This dynamic poses considerable macroeconomic risks. While the IPCC (2022, SPM C.12) notes that limiting warming to 2°C offers net economic benefits, the transition necessary to achieve that target entails major stranded asset risks for fossil fuel infrastructure.

Non-impact-oriented investing frameworks, as described by Cunha et al. (2024) and Busch et al. (2021), can price transition risks but are not designed to mitigate them and prioritises financial exposure over decarbonization. Institutional lock-in thus acts as both methodological convergence and a structural barrier to reallocating capital toward climate goals. If single-materiality frameworks are deeply embedded, environmental ratings will remain stable and dominated by financial and market characteristics, rather than respond to improvements in environmental data. This logic is reproduced through the institutional architecture of sustainable finance itself, not through a lack of demand: environmental information primarily enters capital markets via frameworks that translate it into financially relevant risk measures, reinforcing the system's lock into characteristics of Sustainable Finance 2.0.

2.2 ESG Rating Divergence and Informational Properties

The geographic concentration of prior studies limits the generalisability of conclusions about financial materiality's dominance. Most empirical work on ESG rating determinants draws on developed-market samples, predominantly from North America and Western Europe. Emerging-market firms, which face distinct institutional environments, weaker regulatory infrastructure, thinner disclosure ecosystems, and different capital-market structures, are underrepresented (Lozano and Martínez-Ferrero, 2022; Martiny et al., 2024).

This issue of geographic sampling bias foregrounds a critical methodological challenge: determining whether the observed dominance of financial materiality in ESG ratings reflects institutional differences among firms or is a consequence of how ESG ratings are constructed. If financial predictors dominate only in developed-market samples, this would imply an institutional explanation; however, if such dominance persists across both developed and emerging markets, it would suggest that the phenomenon is intrinsic to rating methodologies. The complexity of this question is heightened by notable methodological inconsistencies among major ESG rating providers, which further obscure the origins and reliability of observed patterns.

ESG ratings from major providers (including MSCI, Sustainalytics, and LSEG) frequently assign divergent ratings to the same entity because they define materiality, select indicators, and apply weights differently (Martiny et al., 2024). Berg et al. (2022) decompose this divergence across six agencies into measurement differences (56%), scope differences (38%), and weighting differences (6%), revealing the absence of a shared framework for interpreting sustainability data. Christensen et al. (2022) further demonstrate that greater ESG disclosure is associated with greater rating disagreement. The opposite pattern is observed in credit ratings, suggesting that raters disagree most fundamentally about how to interpret disclosed information.

The informational properties of ESG ratings have been examined from multiple angles. Lindsey et al. (2025), using data from four major providers, find that ESG metrics do not improve risk prediction or portfolio performance once conventional financial variables are controlled for. This finding raises the possibility that ESG ratings may be overlapping with the financial characteristics they are intended to supplement. Consistent with these findings, Mazzacurati (2021) finds on the regulatory front that ESG rating methodologies systematically advantage firms with substantial non-financial reporting capacity, placing smaller firms at a disadvantage, a bias rooted in the design of these systems.

From a portfolio perspective, Raghunandan and Rajgopal (2022) observe that ESG funds typically favor companies with robust disclosure and high ESG ratings, rather than those demonstrating regulatory compliance or lower emissions. Furthermore, Kathan et al. (2025) show that ESG scores are positively correlated with greenwashing cases, and that apparent rather than real environmental performance is measured, with analyst coverage density mediating this relationship.

Collectively, these studies point to a consistent trend with investment decisions being shaped more by geography, reputational and financial proxies than by actual environmental outcomes. Across organizational, portfolio, and market contexts, ESG ratings and investment flows are closely aligned with firms' internal strengths and external capacities rather than direct environmental measurement (Drempetic et al., 2020; Raghunandan & Rajgopal, 2022; Sankar et al., 2024).

2.3 Institutional Isomorphism and Means-Ends Decoupling in ESG Ratings

Institutional theory, as articulated by Meyer and Rowan (1977), posits that organisations adopt formal structures and practices primarily to secure legitimacy by conforming to institutionalised rules, rules that function as myths conferring legitimacy, stability, and access to resources, regardless of their technical efficacy, matching the findings of our literature review. This process fosters organisational isomorphism, where firms converge on similar practices to meet external expectations, often resulting in activities decoupled from substantive outcomes.

These dynamics are directly observable in the ESG ratings landscape. ESG ratings function as instruments of social legitimacy, valued as much for their symbolic conformity to prevailing norms as for their accuracy in measuring environmental performance. DiMaggio and Powell (1983) distinguish between coercive, normative, and mimetic isomorphism, with the latter particularly salient under uncertainty, prompting firms to imitate sector and peer-group norms. Such mimetic processes may produce homogeneity in ESG scores, diminishing differentiation based on actual firm performance. This study empirically tests whether peer-group means are the dominant predictors of firm environmental scores, thus assessing the embeddedness of mimetic mechanisms in ESG rating systems such as MSCI's.

Extending this analysis, Bromley and Powell (2012) introduce the concept of means-ends decoupling: organisations may invest heavily in compliance infrastructures (audits, reporting, and specialist roles) without achieving commensurate improvements in actual outcomes, reflecting a persistent misalignment between measurement activity and meaningful impact. Empirical studies across domains reinforce this: Jabbouri et al. (2019) identify micro-mechanisms that perpetuate the gap between implemented practices and outcomes; Kaufmann, Krlev, & Brown (2024) show that, in sustainable finance, impact measurement often serves relational rather than performance-management purposes, with decoupling intensifying as systemic opacity increases. Stål and Corvellec (2021) demonstrate how structural compartmentalisation sustains, over time, a decoupling that blurs efficiency boundaries between compliance activities and actual outcomes.

Applied to ESG ratings, means-ends decoupling suggests that the proliferation of reporting frameworks and compliance activities may be only weakly linked to genuine environmental improvements. This study, therefore, tests whether direct environmental performance metrics offer explanatory power beyond financial and institutional variables; if not, it would indicate that ESG ratings primarily reflect the construction and maintenance of compliance infrastructure rather than substantive environmental achievement.

2.4 Financial Capacity, Ceremonial Compliance, and Structural Persistence

The compliance infrastructure described above is not equally accessible. Because rating methodologies rely heavily on self-reported disclosures, the strong relationship between firm size and ESG scores documented by Drempetic et al. (2020) implies that financial resources operate as a gatekeeper: larger firms can build the bureaucratic apparatus needed to produce favourable ratings, independent of their actual environmental performance.

Selective disclosure within this infrastructure is increasingly well-documented. For example, Roszkowska-Menkes et al. (2024), analysing MSCI's controversies database, found that 69% of negative sustainability events were selectively reported in corporate sustainability reports, and that neither the GRI guidelines nor external assurance mechanisms effectively prevent such omissions. As a result, if MSCI's environmental pillar scores are largely based on self-reported disclosures that systematically conceal negative performance, these scores will reflect firms' capacity for disclosure management rather than their true environmental standing.

Moreover, this dynamic contributes to the persistence and structural lock-in of ESG scores. Once a firm has invested in the compliance infrastructure needed to achieve a certain rating, that status tends to persist, with scores exhibiting high year-over-year autocorrelation and resisting revision in response to actual environmental changes. This inertia is characteristic of the audit society logic, where systems built around ceremonial compliance reward consistent institutional positioning rather than substantive improvement (Power, 1997). Crona et al. (2025) identify similar self-reinforcing feedback loops at the financial system level, where reliance on ESG ratings that do not adequately capture environmentally material information impedes the reallocation of capital toward genuinely sustainable outcomes.

Cepêda et al. (2025), through a systematic bibliometric review, confirm that ESG decoupling is widespread and persistent across institutional contexts, with institutional theory remaining the dominant explanatory framework. The convergence of these findings points to a predictive structure of environmental ratings that is shaped substantially by financial capacity and institutional conformity, and that naturally tends toward structural persistence over time.

The present study examines these propositions empirically through panel analysis of 7,342 firms (2014–2024), year-over-year score persistence analysis, and a structural break test around MSCI's version 4.0 methodology revision in 2021. If the financial-isomorphic predictive structure strengthens, rather than weakens, after a major methodology revision, this will indicate that methodological refinements have deepened, rather than disrupted, the ceremonial compliance logic embedded in the rating system.

Synthesising the theoretical and empirical evidence leads to a central expectation: if ESG ratings function primarily as instruments of social legitimacy, anchored in institutional isomorphism and underpinned by financial capacity, then these ratings should be driven chiefly by financial and isomorphic variables, with direct measures of environmental performance adding little explanatory power. In other words, the combination of institutional theory, means-ends decoupling, and the observed persistence of ceremonial compliance together predicts that ESG scores are better explained by disclosure management and organisational resources than by actual environmental outcomes.

To empirically evaluate this proposition, this study examines whether MSCI Environmental Pillar scores are best predicted by financial capacity and isomorphic positioning, or whether the inclusion of direct environmental metrics (such as GHG intensity, renewable energy share, science-based targets, and regulatory context) significantly enhances explanatory power. This approach directly tests the institutional perspective against the possibility that financial resources might, in fact, enable real environmental improvements, making any observed correlation between financial variables and ESG scores substantive rather than merely ceremonial.

This core hypothesis is interrogated through four complementary empirical strategies: (1) establishing the baseline explanatory power of financial and isomorphic features; (2) assessing the incremental contribution of approximately 25 direct environmental performance variables; (3) decomposing predictive importance across feature groups using SHAP values; and (4) quantifying the marginal impact of each group through group permutation importance analysis. The robustness of these findings is further evaluated across diverse geographic and institutional contexts, as well as through a structural break analysis surrounding the 2021 MSCI methodology revision.

If the results show that direct environmental metrics provide negligible improvement over the financial-isomorphic baseline, this will constitute strong evidence for the ceremonial compliance account. Alternatively, if substantive environmental variables emerge as significant predictors, it would suggest that financial resources are indeed facilitating genuine improvements, warranting a re-evaluation of the institutional perspective. In this way, the empirical analysis is designed to distinguish between ratings driven by institutional conformity and financial capacity, versus those reflecting authentic environmental achievement.

3. Methodology

3.1 Data and Sample Construction

The analysis draws on a panel compiled from FactSet's screening universe, encompassing all firms for which MSCI assigns an Environmental pillar score over the period 2014–2025. The resulting dataset comprises 73,134 firm-year observations across 7,342 unique companies domiciled in 42 countries. Table 1 summarises the panel's composition by region, market-capitalisation tier, and year.

Table 1: *Global Dataset Description*

Descriptor	
Total firm-year observations	73,134
Evaluation observations (2015–2024)	68,636
Unique firms	7,342
Countries	42
Train companies	5,926
Test companies	1,416

Panel A — Observations by Region

Region	Firms	Obs	Mean ENV	SD ENV
DM_AMERICAS	2,386	21,575	5.001	2.291
DM_APAC	1,758	19,264	5.093	2.249
DM_EMEA	1,265	19,315	5.865	2.304
EM_AMERICAS	191	1,551	5.295	2.218
EM_APAC	1,605	10,150	4.427	2.070
EM_EMEA	137	1,279	5.095	2.399
Total	**7,342**	**73,134**	**5.182**	**2.301**

Panel B — Observations by Market-Capitalisation Tier

Tier	Firms	Obs	Mean ENV	SD ENV	Median MktVal ($M)
Micro	125	203	4.061	1.968	259
Small	4,146	25,570	4.608	2.143	1,042
Mid	4,562	29,955	5.148	2.248	4,164
Large	2,202	16,347	6.113	2.340	21,888
Mega	83	662	6.532	1.714	261,924

Panel C — Observations by Year

Year	Firms	Obs	Mean ENV	SD ENV	p25	p75	Role
2014	3,508	4,498	5.097	2.233	3.500	6.400	Lag init.
2015	3,842	4,875	5.021	2.148	3.500	6.500	
2016	4,128	5,219	5.090	2.140	3.500	6.600	
2017	4,813	5,963	5.058	2.187	3.400	6.500	
2018	5,212	6,479	4.994	2.226	3.300	6.500	
2019	5,584	6,927	5.135	2.259	3.500	6.700	
2020	6,231	7,634	5.105	2.315	3.400	6.700	
2021	6,655	8,135	5.218	2.357	3.400	6.700	v4.0
2022	7,342	8,887	5.237	2.390	3.400	6.700	
2023	6,768	8,221	5.417	2.422	3.600	7.000	
2024	6,296	6,296	5.464	2.385	3.700	7.000	

Panel D — ENV Score Distribution by Rating Band

Grade	Range	Obs	% of Sample	Mean ENV
CCC	[0.000, 1.429)	2,467	3.4%	0.926
B	[1.429, 2.857)	8,990	12.3%	2.262
BB	[2.857, 4.286)	16,318	22.3%	3.575
BBB	[4.286, 5.714)	17,790	24.3%	4.970
A	[5.714, 7.143)	13,832	18.9%	6.436
AA	[7.143, 8.571)	6,222	8.5%	7.792
AAA	[8.571, 10.001)	7,515	10.3%	9.593

Note: The target variable is the MSCI Environmental pillar score, a continuous measure on a 0–10 scale. The letter-grade boundaries shown in Panel D are applied to the continuous score for descriptive purposes only; the Environmental pillar itself is not published as a letter grade. 2014 serves exclusively as a lag-initialisation period for peer-group means. The ‘v4.0’ annotation in Panel C marks the year of MSCI’s methodology revision. Market-capitalisation tiers follow FactSet classifications.

The target variable is the MSCI ESG Environmental pillar score, a continuous measure on a 0–10 scale derived from MSCI's proprietary weighting of industry-specific environmental Key Issues. Unlike the overall ESG rating, which is condensed into a letter grade (CCC–AAA), the Environmental pillar score retains its continuous form, permitting regression-based analysis without the information loss inherent in ordinal classification. The sample mean is 5.18 (SD = 2.30); the distribution by rating band and its evolution over the panel period are reported in Table 1, Panels C and D.

Three features of the panel's structure are methodologically consequential. First, the sample is unbalanced: MSCI's coverage universe expands from 3,508 firms in 2014 to a peak of 7,342 in 2022 before modest attrition in 2023–2024 (Table 1, Panel C). The expanding panel introduces a survivorship consideration: firms entering the sample later may differ systematically from early entrants. The company-stratified train–test split mitigates this concern by ensuring that all observations for a given firm fall exclusively within one partition, regardless of entry year.

Second, the panel exhibits meaningful cross-regional variation in both the level and dispersion of Environmental scores. DM_EMEA firms carry the highest mean score (5.87), consistent with the EU's comparatively stringent non-financial disclosure regime, while EM_APAC firms score lowest (4.43). This 1.44-point gap — equivalent to roughly one full rating band — provides the variation necessary to test whether the financial and peer-group predictive structure operates uniformly across institutional contexts or is attenuated in environments with weaker disclosure infrastructure.

Third, financial data from FactSet are reported with a one-year lag relative to the Environmental pillar score: the financial features for year *t* correspond to fiscal-year reporting for *t*, while MSCI assigns Environmental scores prospectively. To accommodate this lag structure, Environmental pillar scores for 2025 are incorporated into the panel, enabling the model to pair the most recent financial data (fiscal year 2024) with a contemporaneous rating outcome. The year 2014 serves the converse function: it provides the initialisation period for lagged leave-one-out peer-group means but is excluded from all model estimation and evaluation. The effective estimation window is therefore 2015–2025, comprising 68,636 evaluation observations.

Six external country-year datasets are merged onto the panel to characterise the regulatory and institutional environment at the country level. These include the OECD's Economic Complexity Index, proxying for the productive sophistication of the national economy (Hidalgo & Hausmann, 2009); the World Bank's Worldwide Governance Indicators for rule of law and regulatory quality (Kaufmann, Kraay & Mastruzzi, 2010); the World Bank's Carbon Pricing Dashboard, providing the maximum carbon price in USD alongside binary indicators for the presence of an emissions trading system and a carbon tax; the OECD's Environmental Policy Stringency index; and the ND-GAIN Country Index, which captures climate adaptation capacity through composite readiness and vulnerability scores. These variables enter the model exclusively as environmental features in M2, allowing the analysis to disentangle country-level regulatory pressure from firm-level environmental performance and financial capacity.

3.2 Feature Architecture

Features are organised into three theoretically motivated groups designed to operationalise distinct predictive channels through which environmental ratings may be determined.

Financial features (16 variables) capture firm-level economic capacity and market positioning: return on average invested capital, EBITDA margin, debt-to-asset ratio, asset-equity ratio, debt-equity ratio, return on average total equity, price-to-earnings ratio, interest coverage, capital expenditure intensity, R&D intensity, net sales, employee revenue ratio, market capitalisation, market-cap tier, analyst consensus score, and the Economic Complexity Index of the firm's country of domicile. These variables proxy for the financial resources and market standing that institutional theory associates with the capacity for ceremonial compliance (Drempetic et al., 2020).

Peer-group features (8 variables) capture the central tendency of Environmental pillar scores among a firm's reference group in the prior period. For each firm-year observation, we compute the lagged leave-one-out (LOO) mean Environmental pillar score of the firm's peers along eight classification dimensions: RBICS economy, sector, industry, and subsector; country and region of domicile; market-capitalisation tier; and an industry×market-cap interaction. The LOO construction excludes the focal firm's own score from its group mean, and the one-year lag ensures strict temporal precedence, preventing information leakage. When a firm enters the panel for the first time, the group mean is forward-filled from the earliest available period. These features test a specific empirical proposition: that a firm's Environmental score is substantially predictable from the scores that its institutional neighbours received in the prior year, operationalising the field-level convergence mechanisms theorised in Section 2.3.

Environmental features (included only in the robustness model M2; approximately 25 variables after zero-variance pruning) capture direct environmental performance and the regulatory context. Firm-level measures include greenhouse gas emission intensity (Scope 1+2, forward-filled to address sparsity), renewable energy share, and a suite of Science Based Targets initiative (SBTi) indicators covering target validation status, temperature alignment, scope coverage, and commitment horizon. Transition Pathway Initiative (TPI) management quality scores provide an independent, third-party assessment of corporate climate governance. Country-level regulatory variables include World Bank Governance Indicators (rule of law, regulatory quality), carbon pricing instruments (maximum carbon price, ETS and carbon tax indicators), OECD Environmental Policy Stringency scores, and ND-GAIN climate adaptation indices.

A notable feature of the environmental data is its sparsity. GHG emission intensity and renewable energy share are available for only a subset of firm-year observations. SBTi indicators are binary and heavily skewed, reflecting the fact that validated targets remain the exception across the global corporate population, and TPI management scores are concentrated among large-cap firms in carbon-intensive sectors. To address this, we apply forward-filling within firms for the GHG and energy variables and retain explicit availability flags as separate features, allowing the model to distinguish between firms that disclose zero emissions and those that do not report. Features with zero variance or more than 60% zero-inflation across the panel are dropped prior to estimation. This sparsity is not randomly distributed; it correlates with the firm-size and disclosure-capacity gradients discussed in Section 2.4, making the availability flags themselves informative features rather than mere data-quality artefacts.

All continuous features are scaled using a robust scaler, which centres each variable on the median and scales by the interquartile range rather than the mean and standard deviation. This choice attenuates the influence of extreme outliers, which are prevalent in financial variables such as price-to-earnings ratios and interest coverage, without distorting the distributional information available to the ensemble. Missing values are retained in their native form rather than imputed. LightGBM handles missing data natively by learning optimal split directions for absent values during training (Ke et al., 2017), a property that preserves the informational distinction between a genuinely low value and an unreported one. Imputation strategies, whether mean-filling, zero-filling, or model-based, would collapse this distinction and risk introducing systematic bias in precisely the environmental features whose sparsity is most theoretically consequential.

3.3 Modelling Strategy

The primary model (M1) uses only financial and peer-group features. The robustness model (M2) augments M1 with the full set of environmental features. This sequential design operationalises H1 and simultaneously adjudicates between two theoretically distinct accounts of any observed correlation between financial variables and Environmental pillar scores.

Under the ceremonial compliance account, financial resources enable firms to construct disclosure infrastructure and maintain peer-group positioning, producing high scores without commensurate environmental improvement. Direct environmental performance metrics should therefore add negligible explanatory power beyond financial and peer-group variables, because scores are determined by disclosure capacity rather than by what is disclosed. Under the genuine improvement account, financial resources fund real environmental investments, and the correlation between financial variables and scores reflects substantive environmental achievement. Direct metrics such as GHG intensity, renewable energy share, and SBTi target status should therefore substantially improve predictive accuracy, because they capture the environmental outcomes that financial investment produces. The sequential M1/M2 comparison, combined with SHAP decomposition and group permutation analysis, is designed to distinguish between these predictions empirically. A third outcome is also possible: environmental features could degrade out-of-sample accuracy, indicating that environmental data as currently disclosed introduces measurement noise rather than signal, a result that would carry direct implications for data providers and the informational value of current firm-level environmental reporting.

It should be noted that means-ends decoupling is inferred from the predictive structure rather than directly measured. The methodology tests what Environmental scores reflect; the absence of environmental metric predictive power is interpreted as evidence consistent with decoupling between compliance infrastructure and environmental outcomes. A direct measurement of decoupling would require paired data on both reported commitments and verified environmental outcomes at the firm level, which is beyond the scope of the current panel.

Both models employ gradient-boosted regression trees via LightGBM (Ke et al., 2017), chosen for its capacity to capture non-linear relationships and interaction effects without imposing distributional assumptions on the feature space. Hyperparameters are set at 100 leaves, a learning rate of 0.01, L1 and L2 regularisation of 0.1, and a minimum of 10 child samples per leaf (20 for M2). Training proceeds for up to 3,000 boosting rounds with early stopping triggered after 50 rounds of no improvement on the validation set (minimum delta = 0.01). All features are scaled using a robust scaler to mitigate the influence of outliers.

The train–test split is stratified at the company level: 5,926 firms (80.7%) are assigned to the training set and 1,416 firms (19.3%) to the held-out test set, ensuring that all observations for a given firm appear exclusively in one partition. This design prevents temporal information leakage that would arise from within-firm data appearing in both sets. After initial estimation, features are pruned to retain only those contributing to 95% of cumulative SHAP importance, yielding 18 features in M1 and a comparable reduction in M2.

3.4 Evaluation Framework

Model performance is assessed using out-of-sample R^2, mean absolute error (MAE), and root mean squared error (RMSE) on the held-out test set. Statistical inference is conducted via percentile bootstrap (500 iterations, seed = 42) to construct 95% confidence intervals for all metrics. A naïve mean baseline (predicting the training-set mean for all test observations) provides a lower bound. Year-by-year evaluation tracks the stability of predictions across the decade.

The four tests operationalising H1 are implemented as follows. *Test 1* (Financial Explanatory Power) is assessed by the out-of-sample R^2 of M1. *Test 2* (Marginal Impact Gains) compares the incremental R^2 from M1 to M2. *Test 3* (SHAP Variance Decomposition) employs TreeSHAP (Lundberg & Lee, 2017) computed on a random subsample of 1,000 test observations, with mean absolute SHAP values aggregated by feature group. *Test 4* (Group Permutation Importance) permutes all features within each group simultaneously across 50 repetitions, recording the resulting decline in R^2.

Regional robustness is evaluated by stratifying the test set into six regional blocs and computing within-region R^2, MAE, bias, and independent SHAP decompositions. Temporal stability is assessed through year-by-year R^2 and MAE, with a structural break test around MSCI’s methodology version 4.0 revision in 2021.

Score persistence is additionally assessed through year-over-year autocorrelation of Environmental pillar scores and OLS slope estimation of consecutive-year transitions, operationalising the theoretical prediction that ceremonial compliance produces structural inertia rather than dynamic environmental responsiveness.

4. Results

4.1 Test 1: Financial Explanatory Power

The primary model (M1), using only financial and peer-group features, achieves an out-of-sample R^2 of 0.780 (95% CI: 0.772–0.788), a MAE of 0.777 (95% CI: 0.764–0.789), and an RMSE of 1.109 (95% CI: 1.086–1.131), as reported in Table 2 and Figure 1. The naïve mean baseline yields an R^2 of −0.002, confirming that the model captures substantive structure beyond unconditional central tendency. These results indicate that nearly 78% of the out-of-sample variance in MSCI Environmental pillar scores is explained by financial and peer-group characteristics alone, without any direct environmental performance metrics.

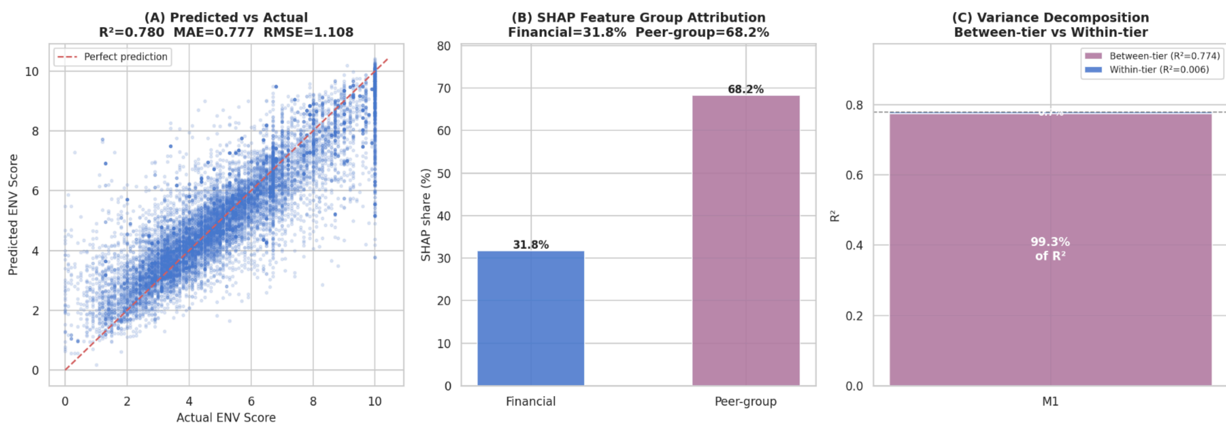

Figure 1: Financial + Peer-group Model Overview (M1)

A tier decomposition reveals that the between-tier R^2 (grouping firms into low, mid, and high terciles of the score distribution) accounts for 0.774, or 82.2% of total variance, while within-tier R^2 contributes only 0.006. This indicates that M1's explanatory power is concentrated in correctly sorting firms into broad rating categories—a function that is overwhelmingly determined by financial and institutional positioning rather than fine-grained environmental differentiation.

4.2 Test 2: Marginal Environmental Gains

The robustness model (M2) augments the M1 feature set with approximately 25 environmental variables, encompassing GHG intensity, renewable energy share, SBTi target indicators, TPI management scores, and country-level regulatory indices. M2 yields an out-of-sample R^2 of 0.726 (95% CI: 0.717–0.736), representing an incremental ΔR^2 of −0.054 relative to M1. All three performance metrics deteriorate: MAE rises from 0.777 to 0.894 and RMSE from 1.109 to 1.238. Figure 2 presents the bootstrap distributions for both models, confirming that the M1 and M2 R^2 intervals are entirely non-overlapping across all 500 iterations.

***Table 2**: Model Performance with 95% Bootstrap Confidence Intervals*

Model	R^2 [95% CI]	MAE [95% CI]	RMSE [95% CI]
Naïve mean baseline	−0.002 [−0.003, −0.001]	1.921 [1.899, 1.945]	2.365 [2.341, 2.387]
M1: Financial + Peer-group	**0.780 [0.772, 0.788]**	**0.777 [0.764, 0.789]**	**1.109 [1.086, 1.131]**
M2: M1 + Environmental	*0.726 [0.717, 0.736]*	*0.894 [0.880, 0.908]*	*1.238 [1.215, 1.258]*

Note: Percentile bootstrap, 500 iterations. M2 reported for robustness; primary analysis uses M1. Incremental ΔR^2 from M1 to M2 = −0.054.

This negative increment is robust to the SHAP-based feature pruning applied to M2, which retains only variables within the 95% cumulative importance threshold, and holds across the full bootstrap distribution (Figure 2C). The pattern is consistent with a structural property of the environmental data. Gradient-boosted ensembles are sensitive to sparse, zero-inflated features: when a large proportion of observations carry missing or zero values, as is the case for GHG intensity, renewable energy share, and SBTi indicators across the global panel, the algorithm identifies spurious splits on availability patterns in the training data that fail to generalise to held-out firms.

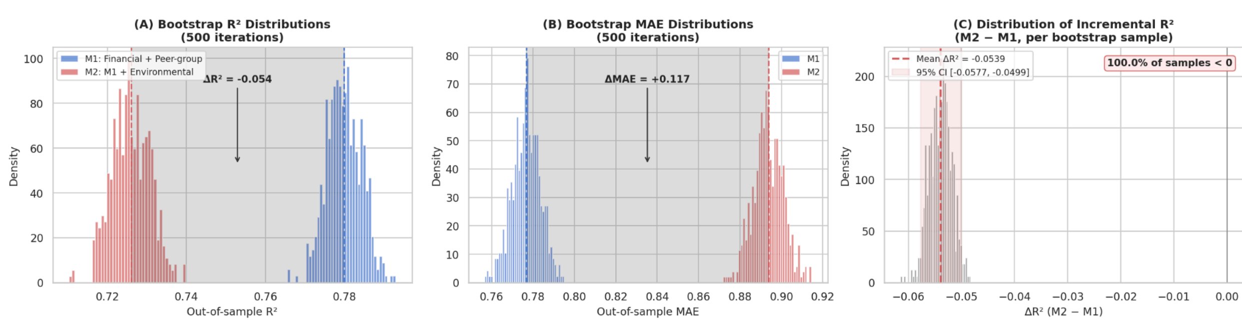

Figure 2: Environmental Features Degrade Predictive Accuracy

This sparsity is itself unevenly distributed, consistent with the disclosure asymmetry discussed in Section 2.4: the model conflates disclosure capacity with environmental performance. Environmental data as currently disclosed therefore introduces measurement noise rather than genuine performance signal, and its addition to the feature set degrades out-of-sample accuracy. The implications of this finding for the informational value of current firm-level environmental reporting are examined in the Discussion.

4.3 Test 3: SHAP Variance Decomposition

SHAP decomposition of M1 assigns 68.2% of mean absolute contribution to peer-group features and 31.8% to financial features. The single most important predictor is the lagged LOO industry-year mean Environmental score (mean |SHAP| = 0.747), followed by the market-cap tier mean (0.314) and firm-level market capitalisation (0.312). The top five features are all peer-group or size-related, and the top three peer-group variables alone account for 47% of total importance (Figure 3(A)). The full feature ranking is reported in Table 3.

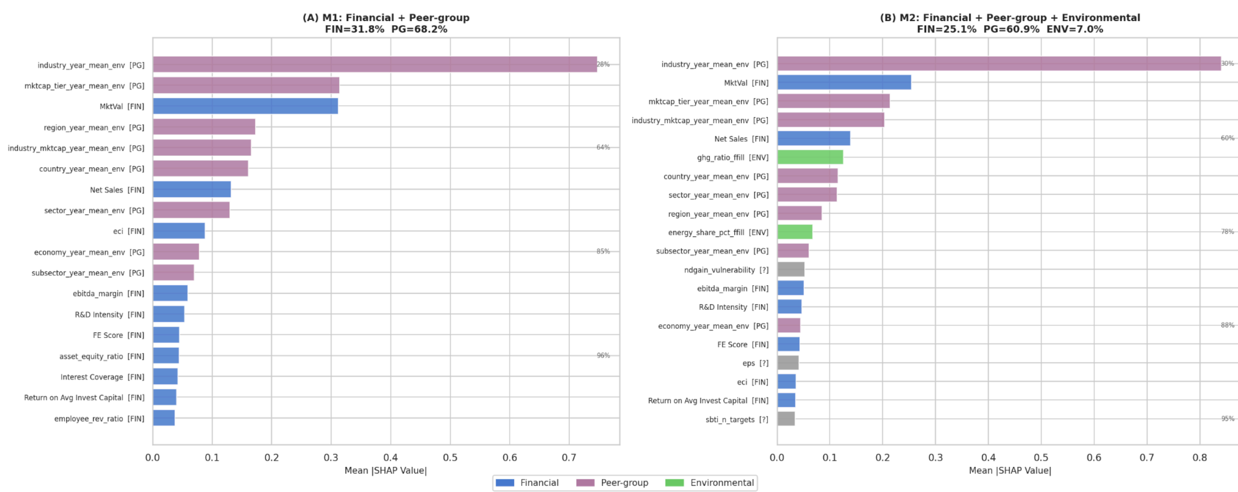

Figure 3: SHAP Feature Rankings: M1 vs M2

When M2 is decomposed by the same procedure, the environmental group accounts for only 7.0% of total SHAP importance, while financial and peer-group features retain a combined share exceeding 86% (Figure 3(B)). The environmental contribution concentrates in the GHG disclosure flag and country-level regulatory indices rather than in direct measures of emissions or resource use, indicating that what the environmental features capture in M2 is a further dimension of institutional context rather than firm-level environmental performance. The theoretical implications of this concentration, and of the peer-group dominance observed in M1, are examined in the Discussion.

Table 3: *Top 10 Features by Mean |SHAP| Value (M1)*

Rank	Feature	Group	Mean |SHAP|	Cum. %
1	Industry-year mean ENV (LOO, t−1)	PG	0.747	28.5%
2	MktCap tier-year mean ENV	PG	0.314	40.5%
3	Market capitalisation	FIN	0.312	52.4%
4	Region-year mean ENV	PG	0.173	59.0%
5	Industry×MktCap-year mean ENV	PG	0.166	65.3%
6	Country-year mean ENV	PG	0.161	71.4%
7	Net sales	FIN	0.132	76.5%
8	Sector-year mean ENV	PG	0.130	81.4%
9	Economic Complexity Index	FIN	0.088	84.8%
10	Economy-year mean ENV	PG	0.078	87.8%

Note: PG = Peer-group; FIN = Financial. SHAP computed on 1,000 random test-set observations (M1, two-group decomposition).

Figure 4 disaggregates the SHAP results into the directional structure of individual feature effects (Panels A and B). Among financial features (Panel A), market capitalisation and net sales exhibit a clear monotonic relationship with Environmental scores: higher values consistently push predictions upward, confirming the size-to-score channel documented in the aggregate decomposition. The Economic Complexity Index displays a similar positive directionality, with firms domiciled in more economically sophisticated countries receiving systematically higher predicted scores. By contrast, profitability and leverage ratios (EBITDA margin, asset-equity ratio, return on invested capital) show substantially greater dispersion in their SHAP contributions, with high feature values generating both positive and negative effects depending on the observation. This pattern is consistent with non-linear relationships or interaction effects, though it may also partly reflect heterogeneity within the robust-scaled feature distribution. In either case, the dispersion indicates that financial capacity operates through multiple channels rather than through a single monotonic pathway.

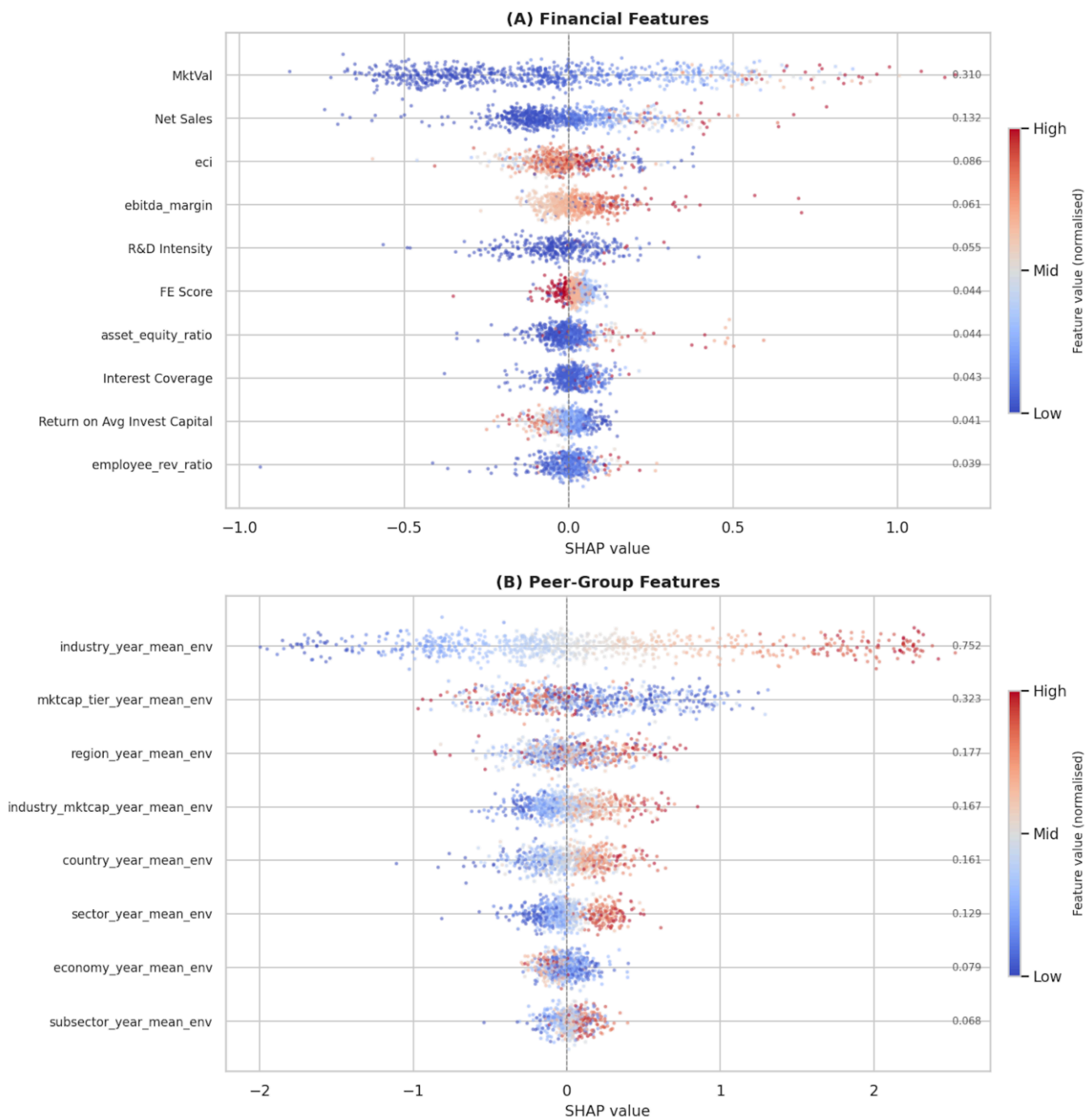

Figure 4: SHAP Attribution: Directional Feature Effects (M1)

Among peer-group features (Panel B), the directional structure is unambiguous. For every peer-group dimension, higher lagged group means push the predicted score upward and lower group means push it downward, with minimal dispersion around this central tendency. The industry-year mean exhibits the widest SHAP range (approximately −2.5 to +2.5), reflecting the substantial between-industry variation in Environmental scores. The tightness of the colour gradient along the SHAP axis confirms that the model is learning a convergence function: a firm's predicted score is pulled toward the mean of its peer group in the prior year, with the strength of the pull proportional to the distance between the firm and its group. This is the empirical signature of the peer-conformity mechanism examined in the Discussion.

4.4 Test 4: Group Permutation Importance

Group permutation importance, computed on M2 to enable direct comparison across all three feature groups, confirms the hierarchy observed in SHAP decomposition. Permuting all peer-group features yields the largest performance decline ($\Delta R^2 = -0.980 \pm 0.011$), followed by financial features ($\Delta R^2 = -0.174 \pm 0.003$). Environmental features produce a decline of only $\Delta R^2 = -0.025 \pm 0.001$. The financial-to-environmental permutation ratio is 7.08×, indicating that financial variables are approximately seven times more important than environmental variables in sustaining the model’s predictive accuracy. These results provide unambiguous support for H1 under Test 4. The full results are shown in Table 4 below.

Table 4: *Group Permutation Importance (M2, 50 Repeats)*

Feature Group	**ΔR² (mean ± SD)**	**95% CI**	**Relative to ENV**
Peer-Group	−0.980 ± 0.011	[0.977, 0.983]	39.84×
Financial	−0.174 ± 0.003	[0.173, 0.175]	7.08×
Environmental	−0.025 ± 0.001	[0.024, 0.025]	1.00×

Note: Permutation conducted on M2 to allow three-group comparison. ΔR² represents mean decline from base R² when group features are shuffled.

4.5 Regional Heterogeneity

Table 5: *Regional Model Performance and SHAP Decomposition (M1)*

Region	n	Mean ENV	R^2	MAE	Bias	FIN%	PG%
DM_AMERICAS	3,994	5.01	0.785	0.757	−0.015	32%	68%
DM_APAC	3,570	5.18	0.783	0.765	+0.013	33%	67%
DM_EMEA	3,717	6.05	0.768	0.775	−0.001	31%	69%
EM_AMERICAS	273	5.62	0.671	0.931	−0.114	31%	69%
EM_APAC	1,842	4.39	0.692	0.807	+0.041	31%	69%
EM_EMEA	173	4.61	0.764	0.933	−0.098	29%	71%

Note: FIN% and PG% from independently computed within-region SHAP decompositions (500 observations per region where available).

Regional stratification reveals consistent performance across developed markets (DM_AMERICAS: R^2 = 0.785; DM_APAC: 0.783; DM_EMEA: 0.768) and moderately lower but still substantial explanatory power in emerging markets (EM_AMERICAS: 0.671; EM_APAC: 0.692; EM_EMEA: 0.764), as reported in Figure 5(A) and Table 5. The DM–EM R^2 gap averages 0.086. Systematic bias (Figure 5(B)) is near zero across all three DM blocs, while EM_AMERICAS and EM_EMEA exhibit small negative biases (−0.114 and −0.098 respectively), partly attributable to smaller sample sizes (n = 273 and n = 173).

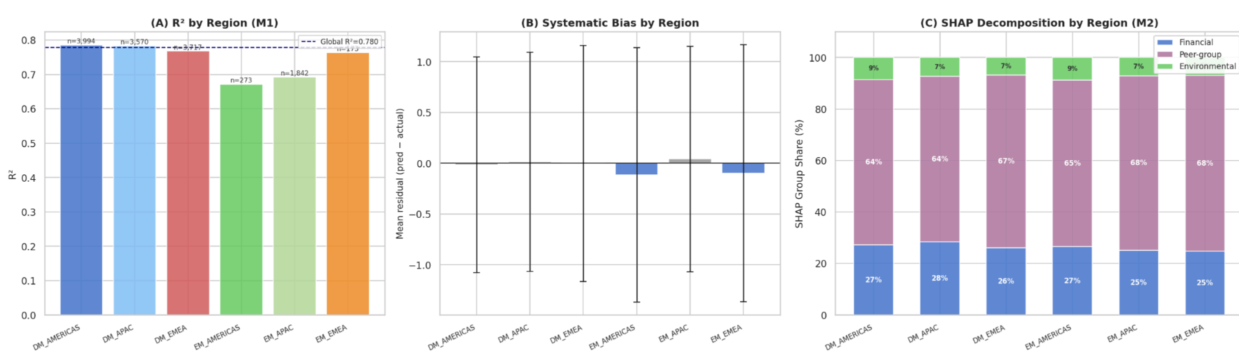

Figure 5: Regional Performance and SHAP Decomposition

Figure 5(C) reports the within-region SHAP decomposition computed on M2 to enable three-way comparison across financial, peer-group, and environmental feature groups. The decomposition is remarkably stable across all six blocs. Peer-group features account for 58–63% of mean absolute SHAP values in every region, financial features contribute 24–29%, and environmental features remain between 6–9%. No region exhibits a qualitatively different feature hierarchy. The environmental share does not expand in DM_EMEA, where mandatory disclosure regimes and carbon pricing mechanisms provide the richest environmental signal in the panel, nor does it compress further in EM blocs where environmental data is sparsest.

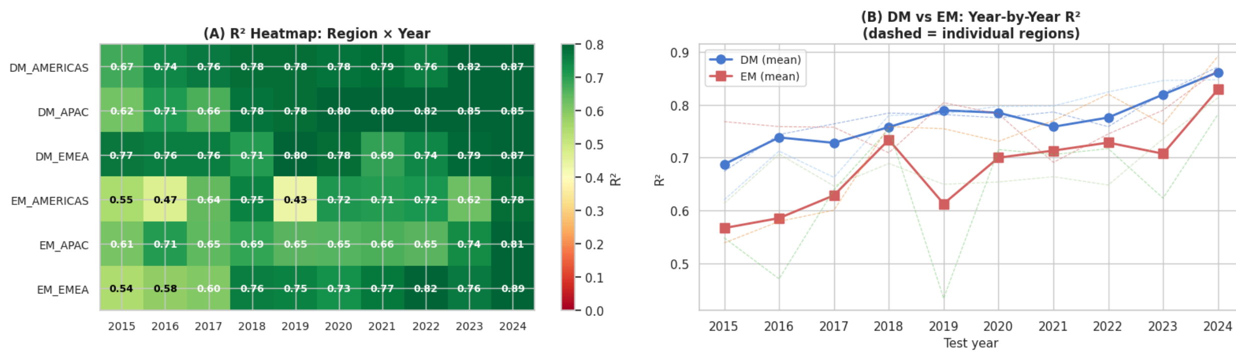

Figure 6: Regional x Temporal R² Structure

Figure 6 extends this analysis along the temporal dimension. The region-by-year R^2 heatmap (Panel A) shows that developed markets maintain consistently high accuracy ($R^2 \geq 0.69$) throughout the decade, while emerging markets exhibit greater year-to-year volatility, partly attributable to smaller sample sizes. The DM–EM gap narrows over time: Panel B shows that the mean EM R^2 rises from approximately 0.59 in 2016 to 0.83 in 2024, converging toward the DM trajectory. This convergence indicates that the financial and peer-group predictive structure has strengthened in emerging markets over the decade as these economies have become more integrated into the global ESG rating infrastructure.

4.6 Industrial Heterogeneity

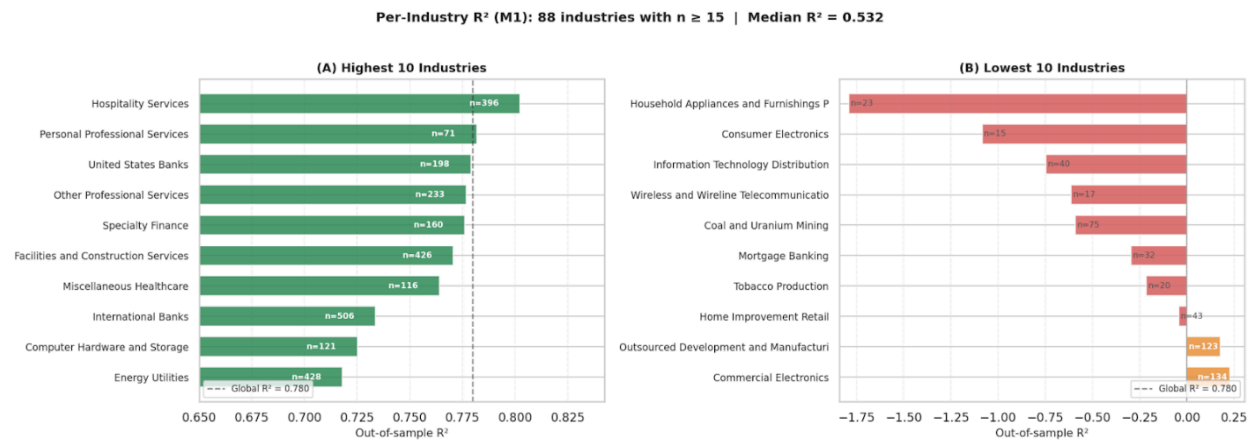

Figure 7: Per-Industry R²

Figure 7 reports the per-industry R^2 for M1, restricted to industries with a minimum of 15 test-set observations (88 qualifying industries). Panel A displays the 10 highest-performing industries, all of which approach or exceed the global R^2 of 0.780, with Hospitality Services (R^2 = 0.83, n = 396) and Personal Professional Services (R^2 = 0.82, n = 71) at the top of the distribution. The median industry-level R^2 is 0.532, and the top 10 industries span a range of sample sizes from n = 71 to n = 506, indicating that high accuracy is not confined to the largest sectors. Panel B displays the 10 lowest-performing industries, several of which exhibit negative R^2 values, indicating performance below the naïve mean baseline. These industries are uniformly characterised by small test-set samples (n = 15–134), and the negative R^2 values are consistent with estimation instability in sparse subsamples rather than a qualitative failure of the predictive structure. The concentration of negative R^2 among niche, thinly represented sectors, combined with broadly positive and stable accuracy across the remaining 78 industries, confirms that the financial and peer-group predictive structure generalises across the large majority of the industry distribution, with degradation confined to sectors where test-set size is insufficient for reliable evaluation.

4.7 Temporal Stability and Score Persistence

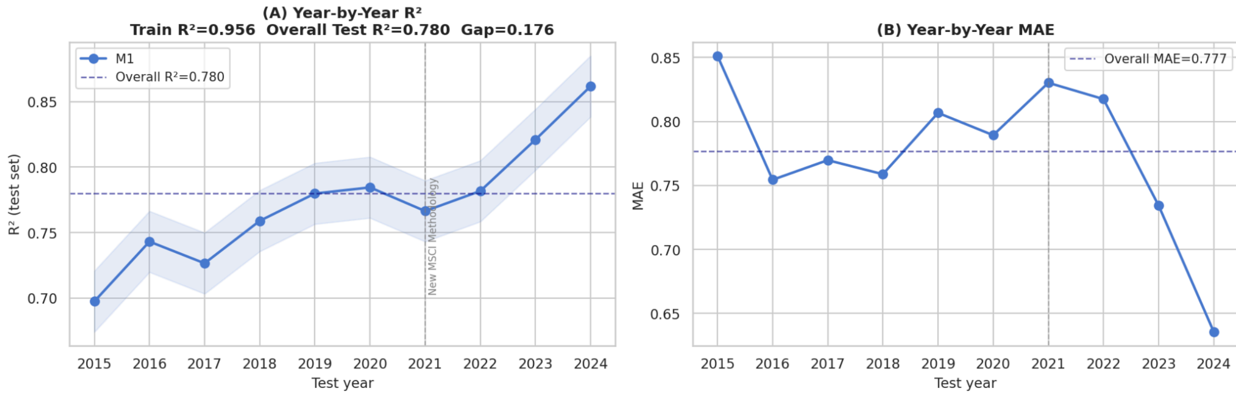

Figure 8: Temporal Stability of Explanatory Power

Year-by-year evaluation demonstrates that M1's predictive accuracy is stable and improving over the decade (Figure 8). The R^2 rises from 0.743 in 2016 to 0.862 in 2024, with a positive linear trend of +0.013 R^2 units per year. The mean R^2 for the period preceding MSCI's version 4.0 methodology revision (2016–2020) is 0.759, compared with 0.808 for the post-revision period (2021–2024), a difference of +0.049. The MAE declines correspondingly from 0.754 in 2016 to 0.636 in 2024. This trajectory is the opposite of what one would expect if MSCI were progressively incorporating environmental performance information orthogonal to financial characteristics: in that scenario, a model built exclusively on financial and peer-group features should lose accuracy over time as environmental signals become more salient in the rating. Instead, the increasing R^2 indicates that successive methodology revisions have deepened rather than disrupted the financial and peer-group predictive structure.

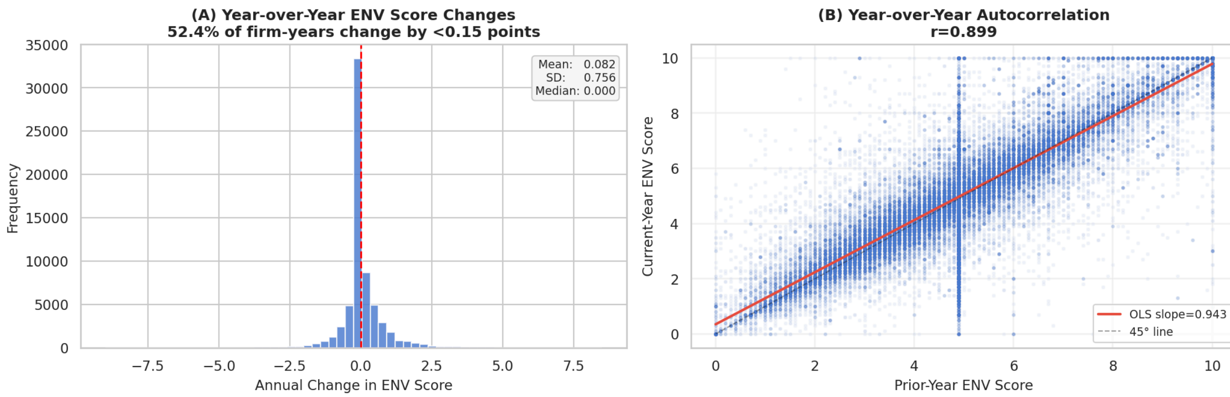

Figure 9: ENV Score Persistence Analysis

Score persistence analysis reveals high autocorrelation between consecutive years (r = 0.899), with an OLS slope of 0.943, a median annual change of exactly zero, and 52.4% of firm-year transitions exhibiting a change of less than 0.15 score points (Figure 9). The annual change standard deviation is 0.756 points. Environmental scores appear to reflect stable structural characteristics, principally sector membership, firm size, and geographic location, rather than dynamic year-on-year changes in firm-level environmental performance. The theoretical and capital-market implications of this persistence are examined in the Discussion.

4.8 Residual Diagnostics

Residual analysis (Figure 10) reveals approximate normality (mean = +0.001, SD = 1.109, skewness = +0.018, kurtosis = 3.570), with the Shapiro–Wilk test rejecting the null of exact normality ($p < 0.001$) owing to sample size. The per-score bias structure is systematic and monotonic: the model over-predicts at the lower end of the Environmental pillar scale (mean residual of +1.74 for firms scoring below 1.43, declining to +0.34 for firms in the 2.86–4.29 range) and under-predicts at the upper end (mean residual of −0.51 for firms scoring between 7.14 and 8.57, reaching −1.02 for firms above 8.57), with scores in the centre of the distribution (4.29–5.71) essentially unbiased (+0.03).

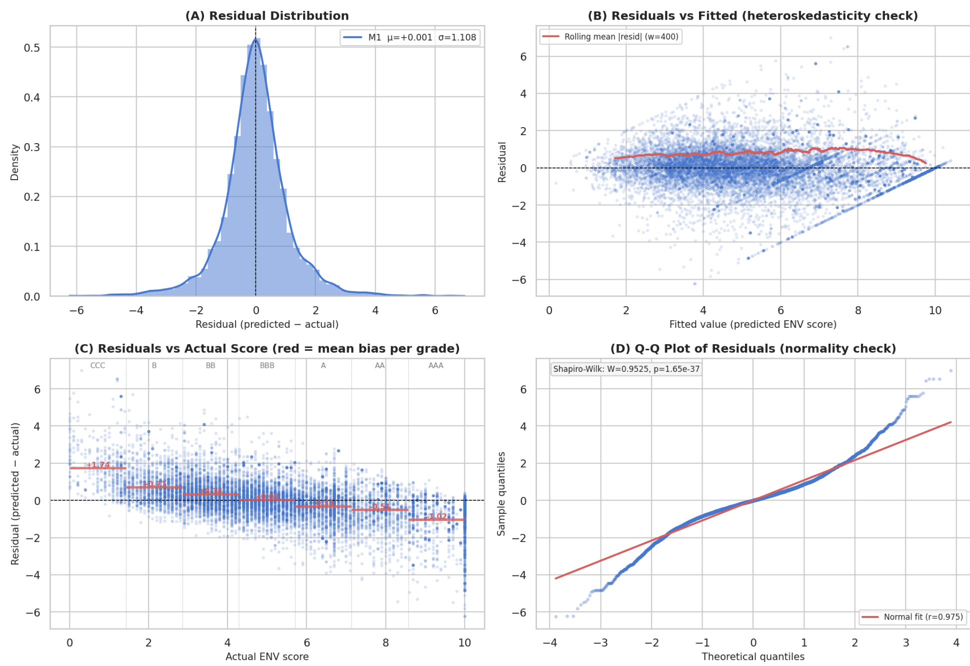

Figure 10: Residuals Analysis (M1)

This regression-to-the-mean pattern is characteristic of ensemble predictions on bounded, heavy-tailed distributions and does not invalidate the variance decomposition, though it indicates that the extremes of the scoring scale are governed by factors that the financial and peer-group feature set does not fully capture. The sources of this tail residual, whether arising from scale-boundary compression, reduced sample density, unobserved qualitative inputs in MSCI's rating process, or some combination thereof, are examined further in the Discussion.

4.9 Summary of Hypothesis Tests

Across all four tests, the evidence uniformly supports H1. Test 1 establishes that financial and peer-group features alone explain 78% of out-of-sample variance (Table 3, Figure 1). Test 2 demonstrates that environmental features add no incremental explanatory power; their inclusion degrades out-of-sample accuracy by 5.4 percentage points (Table 2, Figure 2). Test 3 shows that peer-group features account for 68% and financial features for 32% of SHAP importance in M1, with environmental features contributing only 7% even when explicitly included in M2 (Figure 3). Test 4 confirms that financial features are 7.08

times more important than environmental features in sustaining model accuracy under group permutation (Table 4). The regional analysis establishes that these patterns hold across six geographically and institutionally distinct blocs, with the three-way SHAP decomposition effectively invariant across regions (Figure 5). The temporal analysis demonstrates improving rather than declining accuracy across a decade spanning multiple MSCI methodology revisions (Figure 6), and score persistence analysis confirms that over half of all firm-year transitions are effectively unchanged (Figure 9). The theoretical interpretation of these findings is developed in the Discussion.

5. Discussion

The predictive dominance of peer-group features in the SHAP decomposition constitutes direct empirical evidence for the field-level convergence mechanism that DiMaggio and Powell (1983) theorise under the label of mimetic isomorphism. The beeswarm analysis (Figure 4B) reveals the operative mechanism: for every peer-group dimension, a firm's predicted Environmental score is pulled toward the lagged mean of its reference group, with the strength of the pull proportional to the distance between the firm and its peers. This is a convergence function, not a control variable. The model learns that a firm scoring well below its industry-year mean in period $t{-}1$ will receive a score closer to that mean in period t, and vice versa. The industry-year mean alone accounts for 28.5% of cumulative SHAP importance, and the top three peer-group variables jointly account for 47%. That the financial features contributing most strongly (market capitalisation, net sales) are themselves measures of organisational scale rather than environmental investment reinforces the institutional interpretation: what the rating architecture appears to reward, on the basis of the predictive decomposition, is not environmental achievement but the structural characteristics that position firms within their institutional field.

The cross-regional invariance of this decomposition sharpens the argument. If the peer-group dominance were an artefact of developed-market institutional characteristics, the SHAP structure should shift across the six regional blocs in response to variation in disclosure infrastructure, regulatory stringency, and environmental data quality. The M2 three-way decomposition (Figure 5C) shows that it does not: peer-group features account for 58–63% of predictive importance in every region, financial features contribute 24–29%, and environmental features remain between 6–9%, including in DM_EMEA where mandatory non-financial reporting directives, carbon pricing mechanisms, and the most complete firm-level environmental data in the panel should give environmental variables their strongest predictive advantage. That the environmental share does not expand even under these favourable conditions indicates that the financial and peer-group structure is embedded in MSCI's rating methodology itself rather than arising from data poverty in less regulated markets.

A further identification consideration concerns the collinearity between the peer-group features and the country-level environmental variables. Because the lagged leave-one-out group means are constructed along country, region, and industry dimensions, they absorb regulatory signals, including carbon pricing regimes, environmental policy stringency, and mandatory disclosure mandates, that also enter M2 as explicit environmental features. The near-zero incremental contribution of these variables in M2 may therefore reflect collinearity with the peer-group structure rather than genuine regulatory irrelevance. This interpretation reinforces rather than undermines the institutional argument. If national environmental policies influence ratings primarily by shifting the peer-group norm upward rather than by differentiating firms within a jurisdiction according to their individual regulatory compliance, this constitutes coercive isomorphism operating through the field-level channel that DiMaggio and Powell (1983) describe: the regulatory signal is transmitted institutionally, through group-level convergence, rather than through firm-specific assessment. What the decomposition cannot adjudicate is the relative weight of genuine regulatory effect versus pure mimetic conformity within the peer-group share. Future work employing instrumental-variable designs that exploit discrete regulatory shocks, such as the phased adoption of carbon pricing across jurisdictions, could isolate the coercive component from the mimetic residual and thereby refine the theoretical attribution that the present analysis establishes in aggregate.

The score persistence findings provide a temporal dimension to this institutional reading. The autocorrelation between consecutive years ($r = 0.899$), the OLS slope of 0.943, and the finding that 52.4% of firm-year transitions exhibit a change of less than 0.15 score points collectively indicate that Environmental pillar scores reflect stable structural positioning rather than dynamic environmental responsiveness. This inertia is consistent with the institutional lock-in mechanism that Crona et al. (2025) identify at the financial-system level and with Kaufmann et al.'s (2024) qualitative finding that means-ends decoupling is prevalent in sustainable finance: the compliance infrastructure that generates high scores persists independently of the environmental outcomes it nominally represents. Ruan et al. (2025), analysing Chinese A-share firms under a different rating provider, find that larger year-over-year ESG rating adjustments increase asset mispricing and reduce capital market pricing efficiency through information asymmetry. Although the institutional context differs, their findings suggest that the high score persistence documented here may have capital-market consequences beyond misallocation: by structurally limiting the information content of year-on-year changes, the rating architecture may suppress the price-discovery function that ESG scores are intended to serve.

The residual diagnostics reveal a systematic regression-to-the-mean pattern at the extremes of the scoring scale. The model over-predicts for firms at the lower end of the distribution (mean residual of +1.74 for scores below 1.43) and under-predicts at the upper end (mean residual of −1.02 for scores above 8.57), while the centre of the distribution is essentially unbiased. Several mechanisms may account for this pattern. Scale-boundary compression on the bounded 0–10 interval mechanically limits the range of predictions. Reduced sample density at the tails (CCC and AAA together comprise 13.7% of observations) provides fewer training examples from which to learn extreme-score relationships. Unobserved qualitative inputs in MSCI's rating process, such as analyst judgements applied disproportionately to firms whose environmental profiles deviate most sharply from their peer-group norm, may introduce variation at the extremes that the financial and peer-group feature set cannot capture. These explanations are not mutually exclusive. What the residual pattern confirms is that the financial and peer-group model accounts for the broad structure of Environmental scores with high fidelity but does not fully explain the rating process at the tails, where additional, potentially discretionary factors appear to operate.

The finding that environmental data as currently disclosed degrades rather than improves predictive accuracy ($\Delta R^2 = -0.054$) carries implications for the informational value of current firm-level environmental reporting. The sparsity of GHG intensity, renewable energy share, and SBTi indicators across the global panel, combined with the uneven distribution of disclosure capacity documented by Drempetic et al. (2020) and Mazzacurati (2021), means that environmental variables as they currently exist conflate reporting infrastructure with environmental performance. This corroborates Roszkowska-Menkes et al.'s (2024) observation that negative sustainability events are frequently excluded from corporate reports. The implication is that the barrier to environmentally meaningful ratings is not methodological opacity alone but the quality and integrity of the underlying data. Disclosure mandates that increase the volume of reported information without addressing verification, standardisation, and the systematic exclusion of adverse outcomes will, on the evidence presented here, be absorbed into the peer-group structure rather than producing firm-level differentiation based on environmental outcome.

Recent and forthcoming regulatory initiatives acknowledge elements of this diagnosis. The FCA's 2026 consultation on ESG rating regulation, SEBI India's working group on ESG disclosure, the European Commission's SFDR reforms informed by the SF 3.0 advisory platform, and mandatory standardised disclosure frameworks advancing in Korea and China each represent natural experiments through which the findings of this study can be prospectively tested (Financial Conduct Authority, 2026; Securities and

Exchange Board of India, 2026; European Commission, 2026; Korea Corporate Governance Service, 2026; Ministry of Finance of the People's Republic of China, 2026). The present analysis provides a diagnostic baseline: if these reforms are effective in shifting rating architecture toward substantive environmental assessment, future decompositions should show an increase in the predictive contribution of direct environmental metrics and a corresponding reduction in the peer-group share. If the decomposition remains stable, the persistence of current patterns will indicate that the underlying data infrastructure, rather than methodological transparency, remains the primary barrier to environmentally meaningful ratings.

Several avenues for future research emerge from the limitations of the present study. The feature architecture primarily operationalises mimetic and, through the peer-group structure, coercive isomorphism; normative isomorphism, understood as the diffusion of shared cognitive frameworks through professional training, consultancy networks, and the circulation of ESG personnel across rating agencies and rated firms, has no direct analogue in the feature set and remains unmeasured. Whether analogous patterns characterise other pillars, or the rating architectures of rival providers such as Sustainalytics and LSEG, is an open empirical question that would determine whether the findings are provider-specific or structural. Complementary qualitative investigation of MSCI's internal rating process could further illuminate how analyst judgements, peer-group benchmarks, and disclosure inputs interact in practice.

6. Conclusion

This study assesses whether MSCI's Environmental pillar scores capture substantive environmental performance or are predominantly determined by financial resources and peer-group conformity.

Drawing on a panel of 7,342 firms across 42 countries over a decade, and applying the four-test evaluation framework reported in Section 4, the findings unequivocally support the latter interpretation.

The dominant predictor of a firm's Environmental score is the prior-year score of its institutional peer group. Direct environmental variables, including GHG intensity, renewable energy share, and validated science-based targets, add no incremental explanatory power to a model built on financial and peer-group features alone; their inclusion degrades out-of-sample accuracy rather than improving it. Even where environmental features register in the augmented model, their contribution concentrates in the GHG disclosure flag and country-level regulatory indices rather than in measures of actual emissions or resource use, indicating that scores reward the presence of disclosure infrastructure rather than the environmental outcomes that infrastructure purports to measure.

The predictive structure is not an artefact of developed-market institutional characteristics. The three-way SHAP decomposition is effectively invariant across all six regional blocs, with the environmental share remaining marginal even in DM_EMEA where data quality and regulatory infrastructure are most favourable. The temporal trajectory reinforces this: accuracy has improved over the decade, and the post-2021 methodology revision has deepened rather than disrupted the financial and peer-group predictive structure. Score persistence analysis confirms that over half of all firm-year transitions are effectively unchanged (median annual change of zero, $r = 0.899$), indicating that Environmental scores reflect stable structural positioning rather than dynamic environmental responsiveness.

For the $1.1 trillion in ESG and climate-themed index products benchmarked against these scores, the implication is direct. Environmental tilts in passive strategies function as tilts toward large-cap, financially stable, sector-average firms rather than toward environmental outperformers. Capital intended for climate-aligned investment flows along paths shaped by institutional positioning rather than environmental achievement. The reallocation of capital toward genuine environmental achievement requires a rating architecture grounded in verified environmental outcomes, an architecture that this study demonstrates does not yet exist.

These conclusions are subject to important caveats. The study identifies the predictive structure of MSCI's Environmental pillar scores; it does not directly observe the rating process, and the dominance of financial predictors may reflect causal pathways, such as financial resources genuinely enabling environmental improvements, that the predictive design cannot fully adjudicate. The absence of any incremental explanatory contribution from direct environmental performance metrics, even when modelled alongside the financial variables that could theoretically mediate such improvements, shifts the evidential weight toward the institutional interpretation, though it cannot rule out the substantive channel entirely. Finally, the analysis is confined to MSCI's Environmental pillar; the generalisability to other pillars and providers remains an open empirical question.

Bibliography

Al Azizah, U. S., & Haron, R. (2025). A trend analysis of research on environmental, social, and governance (ESG) performance: A bibliometric analysis. *Social Sciences and Humanities Open*, *11*, 101670. https://doi.org/10.1016/j.ssaho.2025.101670

Amel-Zadeh, A., & Serafeim, G. (2018). Why and how investors use ESG information: Evidence from a global survey. *Financial Analysts Journal, 74*(3), 87–103. https://doi-org.mu.idm.oclc.org/10.2469/faj.v74.n3.2

Aue, T., Jatowt, A., & Färber, M. (2025). Predicting Company ESG Ratings from News Articles Using Multivariate Timeseries Analysis. *Companion Proceedings of the ACM on Web Conference 2025, WWW '25*, 1774–1780. https://doi.org/10.1145/3701716.3717509

Avetisyan, E., & Hockerts, K. (2017). The consolidation of the ESG rating industry as an enactment of institutional retrogression. *Business Strategy and the Environment*, *26*(3), 316–330. https://doi.org/10.1002/bse.1919

Belkhiria, S., Lajmi, A., & Sayed, S. (2025). Predicting Environmental Social and Governance Scores: Applying Machine Learning Models to French Companies. Journal of Risk and Financial Management, 18(8), 413. https://doi.org/10.3390/jrfm18080413

Berg, F., Kölbel, J. F., & Rigobon, R. (2022). Aggregate confusion: The divergence of ESG ratings. *Review of Finance*, *26*(6), 1315–1344. https://doi.org/10.1093/rof/rfac033

Bose, S. (2020). Evolution of ESG reporting frameworks. In S. Bose, G. Dong, & A. Simpson (Eds.), *The financial ecosystem: The role of finance in achieving sustainability* (pp. 13–33). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-32197-0_2

Bromley, P., & Powell, W. W. (2012). From smoke and mirrors to walking the talk: Decoupling in the contemporary world. *Academy of Management Annals*, *6*(1), 483–530. https://doi.org/10.1080/19416520.2012.684462

Busch, T., Bruce-Clark, P., Derwall, J., Eccles, R., Hebb, T., Hoepner, A., Klein, C., Krueger, P., Paetzold, F., Scholtens, B., & Weber, O. (2021). Impact investments: A call for (re)orientation. *SN Business & Economics*, *1*(2), Article 33. https://doi.org/10.1007/s43546-020-00033-6

Choi, J.-C., Chen, Q., & Lee, S.-J. (2024). Predicting ESG Ratings by Machine Learning and Analyzing Influencing Factors by XAI. Proceedings of the 2024 7th International Conference on Software Engineering and Information Management, ICSIM ’24, 120–125. https://doi.org/10.1145/3647722.3647742

Christensen, D. M., Serafeim, G., & Sikochi, A. (2022). Why is corporate virtue in the eye of the beholder? The case of ESG ratings. *The Accounting Review*, *97*(1), 147–175. https://doi.org/10.2308/TAR-2019-0506

Crona, B., Peterson, G., Meacham, M., Parlato, G., Lade, S. J., Rocha, J. C., & Galaz, V. (2025). A systems approach to sustainable finance: Actors, influence mechanisms, and potentially virtuous cycles of sustainability. *iScience,* *28*(7), Article 112785. https://doi.org/10.1016/j.isci.2025.112785

Cunha, F. A. F. D. S., Meira, E., & Orsato, R. J. (2021). Sustainable finance and investment: Review and research agenda. *Business Strategy and the Environment*, *30*(8), 3821–3838. https://doi.org/10.1002/bse.2842

Del Vitto, A., Marazzina, D., & Stocchi, M. (2023). ESG ratings explainability through machine learning techniques. *Annals of Operations Research*. https://doi.org/10.1007/s10479-023-05514-z

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*(2), 147–160. https://doi.org/10.2307/2095101

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

European Commission. (2026). *Platform on sustainable finance*. https://finance.ec.europa.eu/sustainable-finance/overview-sustainable-finance/platform-sustainable-finance_en

European Commission. (2026). *European climate resilience and risk management integrated framework*. https://climate.ec.europa.eu/eu-action/adaptation-and-resilience-climate-change/european-climate-resilience-and-risk-management-integrated-framework_en

Financial Conduct Authority. (2026, February). *Broadening the regulatory framework for ESG ratings providers* (Consultation Paper CP25/34). https://www.fca.org.uk/publication/consultation/cp25-34.pdf

Ferro, A., D. Marazzina, and D. Stocco. 2025. " Uncovering ESG Ratings: The (Im)Balance of Aspirational and Performance Features." Corporate Social Responsibility and Environmental Management 32, no. 5: 5895–5917. https://doi-org.mu.idm.oclc.org/10.1002/csr.70007

Fichtner, J., Jaspert, R., & Petry, J. (2024). Mind the ESG capital allocation gap: The role of index providers, standard-setting, and "green" indices for the creation of sustainability impact. *Regulation & Governance*, *18*(2), 479–498. https://doi.org/10.1111/rego.12530

Hirschbühl, D., Ceglar, A., Cojoianu, T., Emambakhsh, T., Qi, Y., Rho, C., Hu, E., Petracco Giudici, M., Biganzoli, F., Jager, A. de, Herrero, L. G., Mandrici, A., & Pasqua, C. (2025, December). *The climate-biodiversity-pollution nexus: The pricing of environmental credit risks for European industrial polluters* (ECB Working Paper No. 2025/3164). https://doi.org/10.2139/ssrn.5897106

Intergovernmental Panel on Climate Change. (2022). *Approved summary for policymakers: Working Group III contribution to the IPCC Sixth Assessment Report – Climate Change 2022: Mitigation of climate change*. https://www.ipcc.ch/site/assets/uploads/2022/04/P56_Doc_3_Approved_Summary_for_Policymakers.pdf

Jabbouri, R., Truong, Y., Schneckenberg, D., & Palmer, M. (2019). Institutional means-ends decoupling work in industrial R&D project implementation. *Industrial Marketing Management, 80*, 296–311. https://doi.org/10.1016/j.indmarman.2019.01.012

Kathan, M. C., Utz, S., Dorfleitner, G., Eckberg, J., & Chmel, L. (2025). What you see is not what you get: ESG scores and greenwashing risk. *Finance Research Letters, 74*, 106710. https://doi.org/10.1016/j.frl.2024.106710

Kaufmann, L., Krlev, G., & Brown, M. (2024). *In plain sight: Mechanisms of means-ends decoupling in impact investing*. **Organization Studies, 46**(5), 667–692. https://doi.org/10.2139/ssrn.5197024

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In *Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS 2017), Long Beach, CA, USA* (pp. 3146–3154). https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf

Korea Corporate Governance Service. (2026). *Press release*. https://www.cgs.or.kr/news/press_view.jsp?no=234

Lindsey, L. A., Pruitt, S., & Schiller, C. (2025). ESG and the conditional pricing of risk. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3975077

Lopez-de-Silanes, F., McCahery, J. A., & Pudschedl, P. C. (2024). Institutional investors and ESG preferences. *Corporate Governance: An International Review*, *32*(6), 1060–1086. https://doi.org/10.1111/corg.12583

Lozano, M. B., & Martínez-Ferrero, J. (2022). Do emerging and developed countries differ in terms of sustainable performance? Analysis of board, ownership and country-level factors. *Research in International Business and Finance, 62*, 101688. https://doi.org/10.1016/j.ribaf.2022.101688

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.

Marquis, C., Toffel, M. W., & Zhou, Y. (2016). Scrutiny, norms, and selective disclosure: A global study of greenwashing. *Organization Science*, *27*(2), 483–504. https://doi.org/10.1287/orsc.2015.1039

Martiny, A., Taglialatela, J., Testa, F., & Iraldo, F. (2024). Determinants of environmental social and governance (ESG) performance: A systematic literature review. *Journal of Cleaner Production*, *456*, Article 142213. https://doi.org/10.1016/j.jclepro.2024.142213

Mazzacurati, J. (2021). ESG ratings: Status and key issues ahead. *ESMA Report on Trends, Risks and Vulnerabilities*, No. 1, 2021, 105–115. https://www.esma.europa.eu/sites/default/files/trv_2021_1-esg_ratings_status_and_key_issues_ahead.pdf

Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, *83*(2), 340–363. https://doi.org/10.1086/226550

Ministry of Finance of the People's Republic of China. (2026, January 27). *[Policy release]*. https://kjs.mof.gov.cn/zhengcefabu/202601/t20260127_3982643.htm

Morgan Stanley. (2025, April 30). *Sustainable investing: Interest remains strong in 2025*. Morgan Stanley. Retrieved February 27, 2026, from https://www.morganstanley.com/insights/articles/sustainable-investing-interest-2025

MSCI (2025). *MSCI index composition viewer* [Brochure]. https://www.msci.com/downloads/documents/access/brochures/MSCI%20Index%20Composition%20Viewer.pdf

MSCI. (2024a). *ESG Ratings Methodology*. MSCI ESG Research LLC. https://www.msci.com/documents/1296102/34424357/MSCI+ESG+Ratings+Methodology.pdf

MSCI. (2024b). *ESG Ratings Methodology – Process*. MSCI ESG Research LLC. https://www.msci.com/documents/1296102/34424357/MSCI+ESG+Ratings+Methodology+-+Process.pdf

MSCI. (2021). *MSCI response to the European Commission's consultation on the establishment of a European Single Access Point (ESAP)*. https://www.msci.com/documents/1296102/23400696/MSCI+Response+%281%29.pdf/be0fdb29-d82c-b8cb-060d-9e21542c397d

Potsdam Institute for Climate Impact Research. (2024, April 17). *38 trillion dollars in damages each year: World economy already committed to income reduction of 19 % due to climate change*. https://www.pik-potsdam.de/en/news/latest-news/38-trillion-dollars-in-damages-each-year-world-economy-already-committed-to-income-reduction-of-19-due-to-climate-change (Retrieved February 28, 2026).

Power, M. (1997). *The audit society: Rituals of verification* (1st ed.). Oxford University Press.

Raghunandan, A., & Rajgopal, S. (2022). Do ESG funds make stakeholder-friendly investments? *Review of Accounting Studies*, *27*(3), 822–863. https://doi.org/10.1007/s11142-022-09693-1

Roszkowska-Menkes, M., Aluchna, M., & Kamiński, B. (2024). True transparency or mere decoupling? The study of selective disclosure in sustainability reporting. *Critical Perspectives on Accounting, 98*, Article 102700. https://doi.org/10.1016/j.cpa.2023.102700

Ruan, L., Li, J., & Huang, S. (2025). ESG rating adjustment and capital market pricing efficiency: Evidence from China. *International Review of Economics & Finance*, *98*, 103845. https://doi.org/10.1016/j.iref.2025.103845

Sankar, N. G., Nag, S., Chakrabarty, S. P., & Basu, S. (2024). The carbon premium: Correlation or causality? Evidence from S&P 500 companies. *Energy Economics*, *134*, Article 107635. https://doi.org/10.1016/j.eneco.2024.107635

Stål, H. I., & Corvellec, H. (2021). Organizing means–ends decoupling: Core–compartment separations in fast fashion. *Organization Studies*, 61(4). https://doi.org/10.1177/00076503211001856

Securities and Exchange Board of India. (2026, February). *SEBI constitutes working group to review ESG rating providers (ERPs) regulatory framework*. https://www.sebi.gov.in/media-and-notifications/press-releases/feb-2026/sebi-constitutes-working-group-to-review-esg-rating-providers-erps-regulatory-framework_99864.html

UK Government. (2026). *UK sustainability reporting standards*. https://www.gov.uk/guidance/uk-sustainability-reporting-standards