

Original Article

THE U-SHAPED ESG PERFORMANCE PARADIGM: GRADIENT-BOOSTED MACHINE LEARNING EVIDENCE CHALLENGING SUSTAINABILITY-FINANCIAL ORTHODOXY

Hong Viet Le ^{1*} 

¹ Swiss School of Business and Management, Geneva, Switzerland



ABSTRACT

ESG policy is often implemented as if lower ESG risk should automatically produce better equity outcomes. I test that premise in U.S. large-cap stocks with a five-year panel of 500 S&P 500 constituents and Sustainalytics ESG Risk Ratings frozen at January 2019. Firms are sorted into sector-normalized Low, Medium, and High ESG-risk cohorts (167, 166, and 167 firms), which makes it possible to ask whether ESG is priced along a straight line or across distinct portfolio roles.

Prediction is handled with a Light Gradient Boosting Machine (Light GBM) and then translated into trades under institutional constraints. Out-of-sample validation uses five-fold expanding-window cross-validation. Precision reaches 56.15%, recall 78.84%, and F1-beta 74.19% with beta = 0.7. The backtest assumes \$10 million of starting capital, \$1 million per position, a 0.5 trade probability threshold, 5% take-profit, a 20-day maximum holding period, and execution frictions.

What emerges is not a monotonic ranking but a U-shaped structure. High-risk firms generate the strongest absolute and risk-adjusted outcomes (240.35% cumulative return; 28.33% annualized return; Sharpe 1.19; Calmar 0.84). Low-risk firms provide the strongest downside protection, with maximum drawdown of 29.08%. Medium-risk firms do not lead on either return or protection. Pairwise tests, non-parametric tests, bootstrap confidence intervals, multiple-comparison correction, and sector and regime checks all support the same ordering.

The implication is narrower and more useful than a blanket ESG verdict. In this sample, ESG works better as a segmented risk signal than as a single quality ladder. The proposed Governance-Driven Risk Premium remains only an interpretation of the High-risk result; this design does not identify it causally.

Keywords: ESG Investing, Machine Learning, Portfolio Performance, Gradient Boosting, Sustainable Finance, Governance Risk, Activist Investors, U-Shaped Relationship

INTRODUCTION

ESG integration has moved from a niche policy preference to a mainstream portfolio design variable. Global asset owners, pension plans, and sovereign mandates increasingly require formal ESG treatment in investment governance. Yet after this institutional adoption, an unresolved issue remains at the center of both practice and research: what is the functional relationship between ESG risk and financial outcomes?

*Corresponding Author:

Email address: Hong Viet Le (vh3969@gmail.com)

Received: 26 July 2026; Accepted: 12 August 2026; Published 30 September 2026

DOI: [10.29121/ShodhAI.v3.i2.2026.69](https://doi.org/10.29121/ShodhAI.v3.i2.2026.69)

Page Number: 76-85

Journal Title: ShodhAI: Journal of Artificial Intelligence

Journal Abbreviation: ShodhAI J. Artif. Intell.

Online ISSN: 3048-9245, Print ISSN: 3108-1940

Publisher: Granthaalayah Publications and Printers, India

Conflict of Interests: The authors declare that they have no competing interests.

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

Authors' Contributions: Each author made an equal contribution to the conception and design of the study. All authors have reviewed and approved the final version of the manuscript for publication.

Transparency: The authors affirm that this manuscript presents an honest, accurate, and transparent account of the study. All essential aspects have been included, and any deviations from the original study plan have been clearly explained. The writing process strictly adhered to established ethical standards.

Copyright: © 2026 The Author(s). This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

With the license CC-BY, authors retain the copyright, allowing anyone to download, reuse, re-print, modify, distribute, and/or copy their contribution. The work must be properly attributed to its author.

The dominant implementation rule is monotonic. In this view, lower ESG risk should map to higher return quality and lower downside exposure in a broadly consistent way. This assumption appears in negative screening, best-in-class ranking, benchmark tilting, and manager scorecards. It is operationally convenient, but convenience does not guarantee empirical validity.

The literature provides mixed guidance. Meta-level synthesis often reports non-negative average ESG-performance relationships, but effect size and sign vary by market, period, and methodology. More importantly, many studies estimate linear models that can flatten threshold effects, interaction structures, or segmented pricing regimes. If pricing is non-monotonic, a linear slope can become an average of opposing dynamics and generate unstable conclusions.

What matters, then, is the shape of the relationship. If ESG risk is priced non-linearly, portfolios built on a single directional rule are solving the wrong problem.

I approach that question with a liquid S&P 500 universe, a five-year window spanning pandemic stress and rate-hike volatility, sector-normalized cohort construction, time-series validation, and trading rules designed to resemble actual implementation. That combination closes two recurring gaps in the literature: a dependence on linear statistical framing and a weak bridge between prediction and executable portfolio rules.

The pattern that emerges is hard to dismiss as noise. High ESG-risk cohorts lead on return and risk-adjusted return. Low ESG-risk cohorts lead on drawdown control. Medium ESG-risk cohorts do not lead on either dimension. The same ordering survives crisis, recovery, and tightening periods.

Those findings matter on three levels.

- Methodologically, they show that a non-linear ML pipeline can surface structure that linear specifications tend to average away.
- Empirically, they indicate that ESG-performance is segmented rather than monotonic in this sample.
- Practically, they point to a portfolio rule in which High-risk and Low-risk exposures serve different objectives, while the middle segment deserves skepticism rather than automatic inclusion.

The article moves from theory to design, then to results, interpretation, limitations, and conclusion.

LITERATURE CONTEXT AND THEORETICAL DEVELOPMENT

WHY THE LINEAR ESG ASSUMPTION PERSISTS

The linear ESG assumption survives because it is simple to communicate and easy to implement in governance documents. Boards can encode it in policy language, risk teams can convert it into scorecards, and reporting templates can present a clear directional story. However, if capital markets price ESG risk through multiple channels with different horizons, linear simplification can produce allocation error.

ESG signals are not purely firm-level fundamentals. They are also informational and behavioral objects. ESG ratings combine measurement systems, controversy events, and interpretation frameworks. Investor clienteles react differently to these signals. Some are mandate-constrained and avoid high-risk names regardless of valuation. Others are unconstrained and require return compensation to hold controversial exposures. This segmentation can create relative mispricing across ESG strata, especially when information quality is noisy.

EVIDENCE OF HETEROGENEITY IN PRIOR WORK

Prior work already points to substantial heterogeneity. Meta-level evidence is often interpreted as broadly supportive of ESG, yet the dispersion around the average result is large and economically meaningful [Friede et al. \(2015\)](#). Rating disagreement studies make the problem sharper by showing that ESG scores are not interchangeable across providers [Berg et al. \(2022\)](#), [Pagano et al. \(2018\)](#), [Zou et al. \(2023\)](#). Fund-level results are likewise mixed across markets [Wong and Li \(2024\)](#). Once the signal itself varies, the estimated performance effect can vary with it.

That matters for empirical design. A linear specification combined with noisy or unstable explanatory variables can create false confidence about slope direction while masking structural uncertainty. It is also common for studies to evaluate only one performance dimension at a time. For portfolio use, that is not enough. A strategy that looks attractive on return but weak on drawdown, or the reverse, cannot be judged properly without both dimensions in view.

STRATEGIC POSITIONING AND THE MIDDLE-SEGMENT PROBLEM

Strategic management theory offers a coherent lens for non-linearity. Porter's positioning framework highlights the risk of being stuck in the middle [Porter \(1980\)](#). Applied to ESG, firms with partial ESG adoption can incur implementation and reporting costs without achieving full differentiation benefits or valuation re-rating associated with recognized leadership. This predicts a potential penalty in middle ESG strata.

By contrast, firms at the low-risk extreme can benefit from stability premia through perceived resilience, lower tail-risk exposure, and stronger stakeholder confidence. Firms at the high-risk extreme can, under some conditions, offer transition optionality if governance improvements are plausible and undervalued.

AGENCY, ACTIVISM, AND GOVERNANCE TRANSITION

Agency theory and activism research suggest an additional mechanism for high-risk outperformance [Jensen and Meckling \(1976\)](#), [Brav et al. \(2008\)](#), [Bebchuk et al. \(2015\)](#). Firms with visible governance weaknesses may attract activist pressure, board refreshment, and capital discipline. If markets underprice the probability or magnitude of these transitions, expected return can exceed baseline risk compensation.

This channel does not imply that poor ESG practice is inherently value creating. It implies that mispricing can occur when the market focuses on current controversy and underweights future governance adjustment. In other words, the premium can be tied to transition potential rather than persistent weakness.

BEHAVIORAL SEGMENTATION AND DELAYED PRICE ADJUSTMENT

Behavioral finance supports persistence of such patterns [Baker and Wurgler \(2013\)](#). Mandate constraints, herding, and attention asymmetries can produce slow adjustment. ESG-negative sentiment may overshoot valuation effects in specific names or sectors. Conversely, ESG-positive sentiment can compress expected returns where quality narratives are crowded. Under segmented demand, middle-category firms can be overlooked by both camps, creating a weak demand zone.

HYPOTHESES

Given the combined strategic, agency, and behavioral logic, I test four hypotheses.

- H1: ESG-performance follows a U-shaped rather than linear pattern.
- H2: High-risk firms outperform on return metrics while Low-risk firms outperform on downside protection.
- H3: LightGBM detects predictive structure better than linear baselines in this setting.
- H4: A Governance-Driven Risk Premium can explain high-risk outperformance. This mechanism is proposed and not directly causally identified in this study.

DATA, MODEL, AND EMPIRICAL DESIGN UNIVERSE, TIME WINDOW, AND SOURCES

The empirical sample is built from S&P 500 large-cap equities and covers August 26, 2019 through August 21, 2024. The period includes pre-pandemic conditions, COVID-19 crash dynamics, post-crisis recovery, and rate-hike volatility. This regime diversity is useful for stress-testing stability of cross-cohort ordering.

Market data are daily OHLCV records with split and dividend adjustments to preserve return continuity across corporate actions. ESG data are sourced from Sustainalytics and include Total ESG Risk Score plus component scores (Environment, Social, Governance, Controversy).

To avoid look-ahead from evolving ESG assessments and vendor methodology drift, cohort assignment uses ESG information fixed at January 2019.

Panel construction summary.

Item	Value
Universe	S&P 500 large-cap equities
Firms	500
Period	2019-08-26 to 2024-08-21
Observations	about 625,000 company-days
ESG source	Sustainalytics
Classification date	Jan-19

ESG COHORT CONSTRUCTION AND BALANCE

Cohorts are formed with sector-normalized Total ESG Risk ranking and tertile splits.

- Low-risk: top tertile (best ESG / lowest risk), 167 firms.
- Medium-risk: middle tertile, 166 firms.
- High-risk: bottom tertile (worst ESG / highest risk), 167 firms.

The balanced design improves comparability and inference stability, especially in pairwise contrasts and bootstrap procedures.

Cohort descriptives used in the defense framework.

Metric	Low Risk	Medium Risk	High Risk
Sample Size	167	166	167
Average ESG Score	8.2	5.4	2.1
Market Cap (Billions)	\$45.2	\$38.7	\$41.3
Sector Diversity	11	11	11

FEATURE ENGINEERING PIPELINE

The model ingests both ESG and market microstructure information. The feature stack includes:

- Return and range features: log return, high-low ratio, close-open ratio.
- Momentum/trend indicators: SMA, EMA, RSI across multiple windows.
- Volatility features: rolling return dispersion and related transforms.
- Volume and participation metrics: rolling averages and relative volume.
- Pattern and temporal encodings: selected candlestick markers, day/week/month effects.
- ESG factors: Total ESG Risk and component sub-scores.

Missing ESG values are treated through rule-based logic and KNN imputation where needed, with consistency checks to avoid mechanically altering cohort assignment.

LIGHTGBM CONFIGURATION AND VALIDATION

The target variable is next-day directional movement. I use LightGBM for three reasons: non-linear interaction capture, strong performance on structured tabular data, and computational efficiency under repeated time-series validation, consistent with broader machine-learning evidence in equity cross-sections [Hanauer and Kalsbach \(2023\)](#).

Core configuration:

- `n_estimators = 100`
- `learning_rate = 0.05`
- `max_depth = 6`
- `num_leaves = 31`
- `min_data_in_leaf = 20`
- `feature_fraction = 0.8`
- `bagging_fraction = 0.8`

Validation uses five-fold expanding-window cross-validation. Temporal ordering is preserved in every fold to prevent leakage. Evaluation uses F-beta with $\beta = 0.7$ to place slightly greater weight on precision than recall, which is appropriate when false positives are costly in downstream execution.

Out-of-sample model performance from expanding-window CV.

CV Metric	Value
Precision	56.15%
Recall	78.84%
F1-beta ($\beta = 0.7$)	74.19%

BACKTEST CONVERSION AND TRADING CONSTRAINTS

Predictive quality only matters if signals can be executed. I therefore apply a rule-based backtest with explicit constraints.

- Initial capital: \$10,000,000.
- Position size: \$1,000,000 per trade.
- Entry threshold: probability > 0.5.
- Exit rules: take-profit at +5% or mandatory exit at 20 trading days.
- Frictions: spread and market impact included.

The setup reflects institutional scale and concentrates on strategy portability rather than maximal simulated return.

QUALITY CONTROL AND IMPLEMENTATION CORRECTIONS

During development, an early signal-timing setup produced inflated outcomes. A forensic review identified a look-ahead issue in entry/exit handling. The pipeline was corrected and execution assumptions were tightened with realistic spread treatment. Headline returns decreased after correction, but the U-shaped cross-cohort ordering persisted. This increases confidence that the reported structure is not an artifact of optimistic simulation.

INFERENCE AND ROBUSTNESS STACK

I do not rely on one statistical lens. The inferential stack includes:

- Parametric pairwise annual-return tests with effect sizes.
- Kruskal-Wallis and Mann-Whitney non-parametric checks.
- Bootstrap confidence intervals with 10,000 resamples.
- Bonferroni correction across pairwise comparisons.
- Regime and sector stability diagnostics.

This layered approach is designed to reduce over-interpretation from any single test family.

METRIC DEFINITIONS AND ECONOMIC INTERPRETATION

To improve interpretability for portfolio committees, all core metrics are defined consistently across cohorts.

- Total return: $(\text{Final Capital} - \text{Initial Capital}) / \text{Initial Capital}$.
- Annualized return: geometric annualization over the full backtest horizon.
- Maximum drawdown: worst peak-to-trough decline in portfolio equity curve.
- Sharpe ratio: $(\text{Annualized Return} - \text{Risk-Free Rate}) / \text{Annualized Volatility}$.
- Calmar ratio: $\text{Annualized Return} / |\text{Maximum Drawdown}|$.
- Winning rate: number of profitable trades divided by total trades.

The risk-free reference used in the dissertation framework is the 3-month U.S. Treasury bill, averaging about 1.65% over the study horizon. Reporting both Sharpe and Calmar avoids one-metric bias. Sharpe captures mean-variance efficiency, while Calmar emphasizes downside capital impairment. Because institutional mandates are often drawdown constrained, this dual reporting is necessary for decision relevance.

THREATS TO VALIDITY AND MITIGATION DESIGN

The design explicitly addresses several validity threats that are common in ESG and ML finance studies.

Leakage threat: Temporal leakage can inflate predictive metrics. I mitigate this using expanding-window cross-validation and fixed-time ESG cohort assignment from January 2019.

Execution optimism threat: Model-level accuracy can fail in executable trading. I mitigate this through fixed position sizing, threshold-based entries, capped holding periods, and explicit execution frictions.

Distributional threat: Financial returns are heavy-tailed and non-normal. I mitigate this with non-parametric tests and bootstrap intervals in addition to parametric tests.

Specification threat: A single ESG cut can create false confidence. I mitigate this with alternative cutoff and sub-period checks, then evaluate whether the ordering survives.

Interpretability threat: High-complexity models can hide unstable drivers. I mitigate this with feature importance interpretation and consistency checks showing that ESG variables remain materially ranked.

No mitigation removes all uncertainty. The objective is to reduce avoidable bias and document remaining uncertainty boundaries clearly.

RESULTS

PRIMARY PERFORMANCE RESULTS

Backtest outcomes by ESG cohort (2019-2024).

ESG Cohort	Cumulative	Annualized	Max Drawdown	Sharpe	Calmar	Winning Rate	Trades
High Risk	240.35%	28.33%	33.81%	1.19	0.84	61.95%	2205
Medium Risk	136.63%	19.18%	32.52%	0.97	0.59	59.07%	1703
Low Risk	142.75%	19.80%	29.08%	0.97	0.68	59.93%	1832

The pattern is economically large. High-risk cumulative return exceeds Medium-risk by more than 100 percentage points and exceeds Low-risk by almost 100 percentage points. At the same time, Low-risk cohort drawdown is approximately 4.7 percentage points lower than High-risk drawdown.

The return-risk separation is central:

- High-risk leads return and risk-adjusted return.
- Low-risk leads downside protection.
- Medium-risk does not lead either objective.

That split, High-risk for upside and Low-risk for protection, is the pattern anticipated by H1 and H2.

SIGNIFICANCE AND EFFECT MAGNITUDE

Pairwise annual-return comparisons and effect sizes.

Comparison	t-statistic	p-value	Cohen's d
High vs Medium	8.47	< 0.001	1.23
High vs Low	7.92	< 0.001	1.18
Medium vs Low	-0.89	0.374	-0.12

Effect sizes are large for contrasts involving High-risk and negligible for Medium versus Low. After Bonferroni correction, High versus Medium and High versus Low remain significant; Medium versus Low remains non-significant.

Non-parametric tests point to the same ordering, reducing dependence on normality assumptions.

BOOTSTRAP CONFIDENCE INTERVALS

Bootstrap procedures with 10,000 resamples provide distribution-free interval checks. High-risk annualized return interval is clearly separated from Medium and Low intervals in the dissertation appendix framework. This reinforces that the high-cohort premium is not likely a single-sample accident.

REGIME STABILITY

Regime-by-regime annualized return pattern (Table 43 structure).

Regime	High Risk	Medium Risk	Low Risk
Bull market (2019-2020)	32.10%	21.40%	23.80%
Crisis (Mar-May 2020)	28.90%	15.20%	18.70%
Recovery (2020-2021)	31.70%	22.10%	24.30%
Volatility (2022-2024)	25.80%	17.90%	19.20%

The ordering remains intact across crisis and non-crisis windows. That makes it less likely that the headline result is simply a by-product of one exceptional episode, such as the pandemic shock.

SECTOR BREADTH AND CONCENTRATION CHECKS

Sector analysis indicates High-risk outperformance in 8 of 11 sectors, with significance in 6 sectors. That does not establish full sector neutrality, but it does make a narrow one-sector story, such as an energy-only rally, less convincing.

MODEL INTERPRETABILITY AND ESG SIGNAL RELEVANCE

Feature attribution places Total ESG Risk among the leading predictors, with substantial contribution from the Environment and Social components. That matters for H3. The model is not simply renaming technical momentum as ESG; ESG variables still add signal inside the broader feature set.

HYPOTHESIS MAPPING

- H1 (U-shape): supported.
- H2 (High-return / Low-drawdown split): supported.
- H3 (ML improvement over linear assumptions): supported by predictive performance and feature attribution.
- H4 (Governance-Driven Risk Premium): theoretically supported by pattern consistency, but not causally proven in this design.

ADDITIONAL ROBUSTNESS INTERPRETATION

Two points matter for interpretation beyond headline significance.

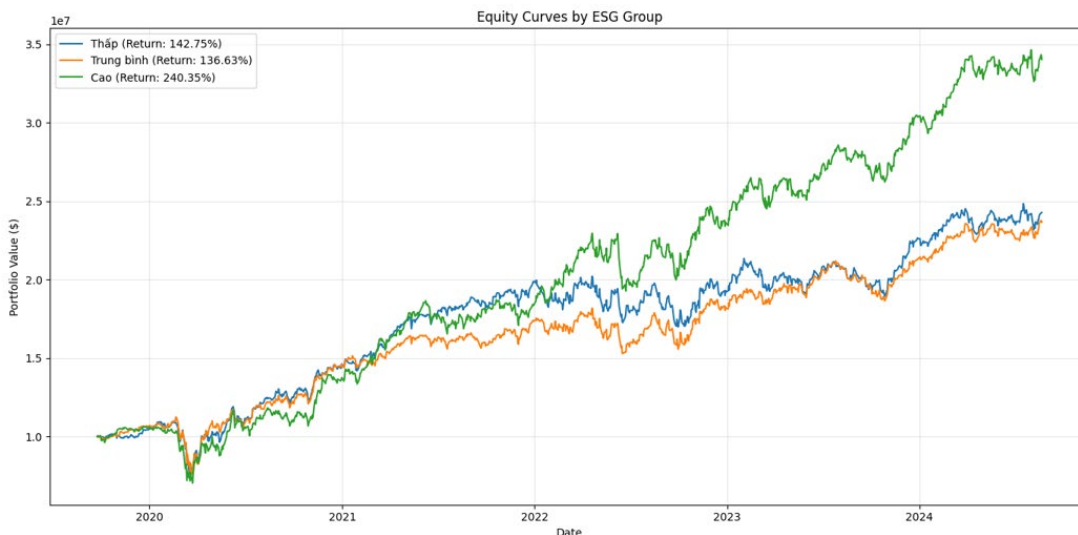
First, the Medium versus Low comparison is not significant after correction. This is an important nuance. The strongest and most stable empirical contrast is that High-risk differs materially from both other cohorts. Therefore, the practical conclusion should prioritize High-versus-rest structure rather than overstate Medium-versus-Low separation on every metric.

Second, cross-regime persistence reduces the chance that the U-shape is a one-event artifact. The pattern appears during the COVID drawdown, recovery, and tightening period. While that does not prove out-of-sample permanence, it increases confidence that the result is not tied to one short volatility episode.

A related concern is whether the result is simply volatility compensation. That narrower reading is hard to sustain because High-risk leadership appears in both absolute return and risk-adjusted return metrics, while Low-risk still dominates drawdown protection. If the pattern were only a leverage or beta story, this combination would be less likely.

Another concern is whether single-sector composition drives outcomes. Sector checks showing High-risk leadership in 8 of 11 sectors, with significance in 6 sectors, do not eliminate all sector effects, but they reduce plausibility of an energy-only explanation.

Finally, the corrected implementation after look-ahead removal is itself a robustness check. The level of returns declined, but the ordering remained. This is a strong sign that structural ranking is more stable than optimized headline performance levels.



Equity curves by ESG cohort from final corrected backtest outputs.

DISCUSSION AND STRATEGIC INTERPRETATION

WHAT THE U-SHAPE MEANS

These data do not support a one-way ESG quality ladder. The low-risk end behaves like a resilience sleeve. The high-risk end behaves like a return-seeking sleeve with more turbulence. The middle does not earn either role; it absorbs cost and ambiguity without a compensating edge.

GOVERNANCE-DRIVEN RISK PREMIUM AS PROPOSED MECHANISM

A Governance-Driven Risk Premium offers one coherent reading of the result. High-risk firms can draw concentrated pressure from activists, lenders, regulators, and boards. When governance change is plausible but not yet fully priced, equity re-rating can arrive faster than a gradual ESG-improvement story would imply.

That remains a conjecture anchored in theory, not a causal result. The design does not include direct event identification such as 13D activism linkage, quasi-experimental governance shocks, or instrumental variables.

IMPLICATIONS FOR PORTFOLIO CONSTRUCTION

The practical lesson is not to abandon ESG, but to stop forcing every mandate through one monotonic score rule.

A disciplined translation is:

- Growth objective: selective High-risk overweight under tight risk controls.
- Defensive objective: Low-risk overweight for drawdown management.
- Balanced objective: avoid structural over-allocation to Medium-risk names.

Such a framework does not require a wholesale rewrite of investment policy. It can begin as a tightly controlled pilot sleeve.

Objective-based allocation rule implied by study evidence.

Portfolio Objective	ESG Exposure Rule
Growth	Overweight High-risk cohort
Defensive	Overweight Low-risk cohort
Balanced	Minimize Medium-risk cohort

IMPLEMENTATION GUARDRAILS

To avoid overreach, implementation should include explicit constraints:

- 10% single-name cap.
- 25% sector cap.
- 15% stop-loss trigger.
- Entry only above threshold confidence.
- Governance-event monitoring in review cadence.

These controls preserve institutional governance quality while testing whether U-shaped segmentation adds net value in live portfolios.

IMPLICATIONS FOR FIRMS AND POLICYMAKERS

For corporate managers, the warning is against half-positioning that is costly yet strategically vague. Clearer choices are easier to defend than an expensive middle ground.

For policymakers, the results caution against regulatory narratives that assume linear return effects from ESG ranking improvements. Policy priorities should focus on data quality, comparability, and transparent reporting of governance remediation paths.

INSTITUTIONAL ROLLOUT DESIGN

A staged implementation test is a better response than immediate full reallocation.

Phase 1: Diagnostic assessment (months 1-2). Map current ESG policy to realized performance by bucket. Estimate opportunity cost of blanket exclusion rules. Verify data availability for governance-event monitoring and risk-reporting infrastructure.

Phase 2: Controlled pilot (months 3-6). Launch a limited sleeve with explicit risk budgets. Keep fixed entry rules, diversification constraints, and pre-defined stop criteria. Report weekly attribution by bucket and monthly drawdown review.

Phase 3: Scale or reject (months 7-12). Only expand if pilot evidence confirms risk-adjusted improvement net of transaction costs and governance constraints. If metrics deteriorate or compliance burden becomes excessive, terminate the approach and document failure conditions.

This approach treats strategy change as a testable policy experiment. It reduces the temptation to overreact to backtest results and is more consistent with fiduciary standards for prudent innovation.

COMMUNICATION AND GOVERNANCE CONSIDERATIONS

Because ESG is politically and institutionally sensitive, communication design matters almost as much as model design.

Three communication principles are recommended.

- Separate evidence from inference. State that U-shape is observed evidence, while governance premium is a proposed mechanism.
- Separate exposure from endorsement. High-risk exposure in this framework reflects transition-pricing logic, not approval of poor ESG behavior.
- Separate pilot results from structural claims. Committee reporting should emphasize what worked, what failed, and whether constraints were respected.

Governance committees should require a standing dashboard with:

- risk-adjusted returns by ESG bucket,
- drawdown trajectories,
- concentration metrics,
- turnover and cost diagnostics,
- key governance-event summaries.

This reporting architecture helps prevent narrative drift and keeps decision-making tied to measured outcomes.

LIMITATIONS AND REPLICATION AGENDA

The design has clear edges.

- Scope limitation: U.S. large-cap sample only.
- Time limitation: 2019-2024, despite regime diversity.
- Measurement limitation: single-vendor ESG source.
- Classification limitation: static January 2019 cohorting.
- Potential survivorship/index composition limitation in fixed-panel construction.
- Identification limitation: no direct causal test for governance mechanism.
- Factor limitation: no full post-hoc decomposition into style, distress, and ESG-specific components.

A credible replication agenda follows directly from those limits.

- Rebuild with survivorship-free data and delisting return treatment.
- Replicate with MSCI and Refinitiv to test vendor dependence.
- Add time-varying ESG transitions.
- Run direct mechanism tests around activism and governance events.
- Estimate factor-decomposed attribution of cohort excess returns.

These extensions are necessary before broad universalization across markets.

CONCLUSION

One result survives every major check in this study: ESG is not behaving like a straight ranking variable in this U.S. large-cap sample. High-risk firms lead on return and risk-adjusted return. Low-risk firms lead on downside protection. Medium-risk firms lack a clear performance role.

The methodological lesson is equally direct. Once non-linear modeling is paired with strict temporal validation and implementable trading rules, the standard monotonic ESG story looks too coarse. ESG is more informative here when treated as segmented exposure with different portfolio jobs. The Governance-Driven Risk Premium remains a plausible interpretation of the High-risk result, but the evidence stops short of causal proof.

For practice, the burden of proof should stay on implementation. Any revised ESG policy should earn its place through better risk-adjusted performance, acceptable stress behavior, and compliance with concentration and governance constraints. If it fails those tests, it should be rejected. That is a better discipline than either unconditional ESG enthusiasm or reflexive dismissal.

ACKNOWLEDGMENTS

None.

REFERENCES

- Baker, M., and Wurgler, J. (2013). Behavioral Corporate Finance: An Updated Survey. In G. M. Constantinides, M. Harris, and R. M. Stulz (Eds.), *Handbook of the Economics of Finance* (Vol. 2, 357–424). Elsevier. <https://doi.org/10.1016/B978-0-44-453594-8.00005-7>
- Bebchuk, L. A., Brav, A., and Jiang, W. (2015). The Long-Term Effects of Hedge Fund Activism. *Columbia Law Review*, 115(5), 1085–1156.
- Berg, F., Heeb, F., and Kölbel, J. F. (2022). The Economic Impact of ESG Ratings (SAFE Working Paper No. 439). <https://doi.org/10.2139/ssrn.4088545>
- Brav, A., Jiang, W., Partnoy, F., and Thomas, R. (2008). Hedge Fund Activism, Corporate Governance, and Firm Performance. *The Journal of Finance*, 63(4), 1729–1775. <https://doi.org/10.1111/j.1540-6261.2008.01373.x>
- Friede, G., Busch, T., and Bassen, A. (2015). ESG and Financial Performance: Aggregated Evidence from More Than 2000 Empirical Studies. *Journal of Sustainable Finance and Investment*, 5(4), 210–233. <https://doi.org/10.1080/20430795.2015.1118917>
- Hanauer, M. X., and Kalsbach, T. (2023). Machine Learning and the Cross-Section of Emerging Market Stock Returns. *Emerging Markets Review*, 55, 101022. <https://doi.org/10.1016/j.ememar.2023.101022>
- Jensen, M. C., and Meckling, W. H. (1976). Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. *Journal of Financial Economics*, 3(4), 305–360. [https://doi.org/10.1016/0304-405X\(76\)90026-X](https://doi.org/10.1016/0304-405X(76)90026-X)
- Kumari, P. (2024). Autonomy and Performance of Democratic Institutions in India: A Comparative Assessment of Congress vs. Non-Congress Rule. *ShodhSamajik: Journal of Social Studies*, 1(1), 77–93. <https://doi.org/10.29121/ShodhSamajik.v1.i1.2024.55>
- Pagano, M. S., Sinclair, G., and Yang, T. (2018). Understanding ESG Ratings and ESG Indexes. In S. Boubaker, D. Cumming, and D. K. Nguyen (Eds.), *Research Handbook of Finance and Sustainability* (339–371). Edward Elgar Publishing. <https://doi.org/10.4337/9781786432636.00027>
- Porter, M. E. (1980). *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. Free Press.
- Sundararajan, S., and Muhammed, A. (2024). Enhancing Organizational Performance Through Strategic Talent Management in the Digital Age: A Quantitative Approach. *ShodhPrabandhan: Journal of Management Studies*, 1(1), 26–43. <https://doi.org/10.29121/ShodhPrabandhan.v1.i1.2024>
- Wong, M. C., and Li, W. (2024). Investigating ESG Funds in China: Management Fees and Investment Performance. *International Journal of Financial Studies*, 12(2), 38. <https://doi.org/10.3390/ijfs12020038>
- Zou, J., Yan, J., and Deng, G. (2023). ESG Rating Confusion and Bond Spreads. *Economic Modelling*, 129, 106555. <https://doi.org/10.1016/j.econmod.2023.106555>