



# International Journal of Management Research and Economics

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

## The Illusion of Predictability: Deconstructing the ESG-Alpha Link with Robust Econometrics and Validated Machine Learning

Henry Efe Onomakpo Onomakpo\* 

\*Faculty of Economics and Business, Universitas Gadjah Mada, Indonesia.

### Article Info

Volume 5, Issue 2, July 2025

Received : 24 May 2025

Accepted : 12 July 2025

Published : 25 July 2025

doi: [10.51483/IJMRE.5.2.2025.35-54](https://doi.org/10.51483/IJMRE.5.2.2025.35-54)

### Abstract

The relationship between Environmental, Social, and Governance (ESG) performance and stock returns is widely debated, with many studies at risk of methodological flaws that can produce misleading results. This study aims to rigorously test the ESG-alpha hypothesis using a framework designed to prevent false discovery. A dual methodology was applied to a global dataset of multinational firms. First, a Two-Way Fixed Effects panel regression was used to control for unobserved firm and time effects. Second, a suite of machine learning models (XGBoost, Ridge, DNN) was tested for predictive power using a stringent walk-forward validation protocol to simulate real-world forecasting and avoid lookahead bias. The panel regression revealed no statistically significant relationship between a composite ESG score and excess stock returns. Critically, the machine learning models failed to generate statistically or economically significant out-of-sample performance, with backtests resulting in substantial capital loss. Strong evidence of a market leverage effect was also found. These findings challenge the narrative of ESG as a direct source of alpha, demonstrating that predictability is likely an illusion of less robust methods. This research highlights the paramount importance of methodological rigor in quantitative finance and offers a critical blueprint for practitioners and academics evaluating ESG investment strategies.

**Keywords:** ESG, Asset pricing, Machine learning, Walk-Forward validation, Fixed effects model, Financial econometrics, Predictive modeling, Asymmetric volatility

© 2025 Henry Efe Onomakpo Onomakpo. This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

### 1. Introduction

The integration of Environmental, Social, and Governance (ESG) criteria into investment analysis has evolved from a niche ethical consideration to a mainstream financial strategy, with trillions of dollars allocated to ESG-mandated funds globally. Proponents argue that strong ESG performance is a proxy for robust corporate governance, superior risk management, and sustainable long-term value creation, which should ultimately translate into enhanced financial returns, an “ESG alpha” (Onomakpo, 2024a, 2024b).

However, the academic literature presents a persistent “ESG puzzle.” Decades of research have yielded a wide array of conflicting findings, with some studies showing positive relationships (Giese *et al.*, 2019; Lins *et al.*, 2017), others negative, and many finding no significant link at all (Friede *et al.*, 2015). This lack of consensus is often attributable to significant methodological differences, including varying data sources, sample periods, and, most critically, the rigor of the econometric and statistical techniques employed (Berg *et al.*, 2022; Gormley and Matsa, 2014). Many foundational

\* Corresponding author: Henry Efe Onomakpo Onomakpo, Faculty of Economics and Business, Universitas Gadjah Mada, Indonesia.

studies are susceptible to common pitfalls such as omitted variable bias, endogeneity, and lookahead bias in predictive models (Gormley and Matsa, 2014; Gu *et al.*, 2020). This inconsistency fuels skepticism among investors and corporate leaders and raises a significant problem for the field: Does the perceived link between ESG and financial performance withstand scrutiny when subjected to a state-of-the-art, robust methodological framework designed to explicitly mitigate these common flaws? Answering this question is crucial for advancing both academic theory and financial practice.

The primary contribution and originality of this research lie in its significantly innovative methodological and theoretical approach, not in the discovery of new data. By utilizing a widely-analyzed dataset (globally-operating corporations, with data from leading providers like Refinitiv, Bloomberg, and Fama-French), this study provides a powerful “apples-to-apples” comparison that isolates the profound impact of its advanced methodology. This approach directly addresses the critique of using well-known datasets by offering substantial new insights into the field through a more rigorous analytical lens. The originality stems from three key pillars. First, a Robust Identification Strategy is employed: the study combines a Two-Way Fixed Effects panel regression to purge unobserved firm heterogeneity and common time shocks, thereby improving causal inference. This econometric approach is triangulated with time-series GARCH analysis to understand risk dynamics. Second, Validated Predictability Testing is performed: moving beyond simple in-sample or standard cross-validation tests, this research employs a stringent walk-forward validation protocol for its machine learning models. This is critical for assessing true predictive power in financial time series but is often neglected, leading to false discoveries in prior research (Lopez de Prado, 2020). Third, Bridging Theory and Practice is a core element: the findings directly address a “great challenge” for organizations and investors: how to separate genuine signals from noise in the ESG data ecosystem. The study provides a practical blueprint for robust quantitative analysis, promoting a connection between rigorous academic methods and real-world investment decision-making.

This study aims to provide new and significant information by dissecting the ESG-finance relationship through three distinct, yet complementary, research questions:

RQ1: To what extent does a firm’s ESG performance influence its excess stock returns, after accounting for unobserved firm-specific heterogeneity and common temporal shocks using a Two-Way Fixed Effects model?

RQ2: Can machine learning models, incorporating ESG indicators and conventional market factors, reliably forecast excess stock returns with statistically and economically significant out-of-sample performance?

RQ3: Do financial markets exhibit asymmetric volatility dynamics (e.g., leverage effects), and are these dynamics amplified by ESG-related news or transition risk shocks?

## 2 Theoretical Framework and Literature Review

This section outlines the theoretical underpinnings of the research, integrating contemporary, post-2018 academic literature to frame the study’s empirical design and the interpretation of its conclusions. The discussion provides adequate reference to seminal and recent works in the field, with a focus on justifying the methodological choices in the context of ongoing academic debates.

### 2.1. Theoretical Frameworks

The analytical approach of this study is informed by a synthesis of perspectives from financial economics, organizational strategy, and information systems. This multi-faceted framework directly informs the selection of the empirical methods employed and provides a compelling rationale for their application.

The research is situated at the intersection of two competing views in financial economics: the Efficient Market Hypothesis (EMH) and Behavioral Finance. In its semi-strong form, the EMH argues that asset prices fully reflect all publicly accessible information, which would include ESG ratings. Consequently, this theory implies that investors should not be able to consistently achieve superior, risk-adjusted returns using such data (Fama, 1970). The EMH therefore, provides the formal null hypothesis for the predictive modeling portion of this analysis. In contrast, concepts from behavioral finance propose that investor psychology, including cognitive biases and sentiment-driven trading, can lead to the systematic mispricing of information like ESG disclosures. This can create temporary market inefficiencies that might be identified and exploited (Shiller, 2003).

From a strategic management perspective, the Resource-Based View (RBV) of the firm (Barney, 1991) offers a mechanism through which ESG initiatives could generate tangible corporate value. The RBV posits that a firm’s competitive advantage derives from its unique and hard-to-replicate internal resources and capabilities (VRIN). Within this framework, a commitment to sustainable practices may cultivate such inimitable assets, including enhanced brand equity, superior operational processes, and improved talent acquisition and retention. The Two-Way Fixed Effects model specification

is therefore intentionally selected as the econometric counterpart to this theory. By isolating the impact of time-varying, firm-specific actions (such as an improvement in ESG performance) while controlling for stable, unobserved heterogeneity, this method directly tests whether the development of such resources translates into observable financial gains.

Finally, principles from Theories of Innovation and Technological Acceptance, particularly the Technology Acceptance Model (TAM), provide a lens to evaluate the practical import of the advanced analytical tools used in finance (Sharma *et al.*, 2025). According to TAM, the successful adoption of a new technology, like an AI-powered forecasting model, hinges on its Perceived Usefulness and Perceived Ease of Use. This framework is vital for interpreting the machine learning results; if a model, despite its sophistication, fails to deliver robust and demonstrable economic value in a simulated real-world setting, its Perceived Usefulness is low. This theoretical perspective justifies the study's rigorous focus on out-of-sample economic significance, as it directly tests whether these advanced analytical methods offer a meaningful advantage in practice.

## 2.2. Literature Review

The academic literature on ESG finance has matured significantly, moving beyond initial correlational analyses. Contemporary research focuses on causality, risk channels, and data quality. Recent studies have shifted the debate from whether ESG matters to how it matters, and with what methodological rigor (Krueger *et al.*, 2020; Onomakpo, 2024a, 2024b; Pastor *et al.*, 2021).

Elaborating on this shift, recent high-impact studies have moved past first-generation literature, focusing on the channels through which ESG impacts performance and the methodological debates surrounding identification. The emphasis is now on understanding how and under what conditions ESG might drive value, rather than simply if it correlates (Krueger *et al.*, 2020; Onomakpo, 2024a, 2024b; Pastor *et al.*, 2021). This study contributes to this by rigorously testing a direct linear alpha relationship under stricter econometric and predictive conditions.

To illustrate the granularity now required in this field, a critical aspect of environmental performance within ESG frameworks pertains to greenhouse gas (GHG) emissions, typically categorized into Scope 1, Scope 2, and Scope 3 emissions. Scope 1 emissions are direct emissions from sources owned or controlled by the reporting company. Scope 2 emissions are indirect emissions from the generation of purchased energy. Scope 3 emissions are all other indirect emissions that occur in a company's value chain, both upstream and downstream. While traditional ESG scores often aggregate these into a single environmental pillar, understanding their impacts and the challenges in their measurement is crucial for comprehensive ESG analysis. Studies have explored the financial implications of carbon emissions and the effectiveness of emission trading schemes (Bolton and Kacperczyk, 2021; Tao *et al.*, 2024), highlighting the growing importance of these granular environmental metrics. The complexity and variability in reporting, especially for Scope 3 emissions, contribute to the "noisy signal" challenge of ESG data (Berg *et al.*, 2022).

Given these complexities, the relationship between financial markets, ESG factors, and predictive modeling is a rapidly evolving field, characterized by the increasing application of advanced quantitative techniques. This review synthesizes how econometrics, time series analysis, and machine learning are being utilized to unravel the complexities of ESG-related phenomena.

Among these quantitative techniques, panel regression models, particularly those employing fixed effects, are widely used to analyze the ESG-financial performance nexus by controlling for unobserved heterogeneity, firm-specific effects, and time-varying shocks. Such methods are crucial for moving beyond mere correlations towards identifying potential causal relationships (Gormley and Matsa, 2014; Petersen, 2009). For instance, recent studies utilized panel regression with entity and time-fixed effects to investigate ESG risk ratings and stock performance in the electric vehicle manufacturing sector and the global transportation sector, finding nuanced and context-dependent relationships (Onomakpo, 2024a, 2024b). Similarly, panel data methods have been applied to examine the impact of ESG disclosure transparency on idiosyncratic risk, revealing U-shaped relationships (Ren *et al.*, 2024).

Complementing these panel data approaches, time series analysis, including Vector Autoregression (VAR) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models, has become indispensable for understanding the dynamic interdependencies and volatility characteristics of financial markets in the context of ESG. VAR models allow for the investigation of lead-lag relationships and spillover effects between various financial and ESG-related factors (Li *et al.*, 2023). For example, VAR models have been applied to analyze interactions between the digital economy, artificial intelligence, and industrial development (Lu *et al.*, 2024). GARCH models and their extensions (e.g., EGARCH, GJR-GARCH) are crucial for capturing volatility clustering and asymmetric responses to shocks, which are stylized facts in financial markets (Glosten *et al.*, 1993; Nelson, 1991). These models enable researchers to investigate how market volatility responds to information, including ESG-related news or climate transition risks, identifying potential

leverage effects where negative information might have a disproportionately larger impact on volatility (Eratalay and Cortés Ángel, 2022; Kouton et al., 2024).

Building on these econometric foundations, the capacity of machine learning (ML) and deep learning (DL) to identify complex, non-linear patterns has become transformative in financial analysis. These methods provide powerful predictive and classificatory tools that can capture dynamics where traditional models fall short. Yet, their application in the ESG domain demands rigorous scientific practice, particularly disciplined validation to avoid overfitting and false discoveries. Methodologies such as walk-forward validation, as used in this research, are therefore essential for ensuring that a model's performance is genuinely predictive out-of-sample (Lopez de Prado, 2020). This potential is reflected in a growing number of studies that deploy advanced algorithms to predict ESG scores (Abdelfattah et al., 2025; Chowdhury et al., 2023; Kartal et al., 2024), forecast financial outcomes based on sustainability data (Dipierro et al., 2025; Haniev, 2024; Hsu et al., 2025; Saba, 2025), or even estimate ESG ratings directly from financial metrics (Atayah et al., 2023; Cini and Ferrari, 2025). A key challenge that persists, however, is the lack of transparency in these "black-box" models. This has spurred the adoption of Explainable AI (XAI) techniques, like SHAP values, which help to demystify model decisions by clarifying feature importance (Hsu et al., 2025; Lundberg and Lee, 2017), thereby bridging the gap between predictive power and practical understanding.

This emphasis on methodological rigor reflects a broader trend, as a "credibility revolution" has taken hold in empirical finance. The use of Two-Way Fixed Effects models to control for unobserved effects is now standard practice for improving causal identification in panel data settings (Gormley and Matsa, 2014; Petersen, 2009). For predictive modeling, the work of Gu et al. (2020) has been seminal in demonstrating both the potential of machine learning and the absolute necessity of rigorous, time-series-aware validation. The study's adoption of walk-forward validation is directly motivated by this state-of-the-art literature (Lopez de Prado, 2020).

Concurrent with this evolution in analytical methods, a crucial development in contemporary finance is the extensive "datafication" of ESG information, a process accelerated by the dual forces of digital transformation and Artificial Intelligence (AI). Recent literature highlights a significant shift away from sole reliance on structured, third-party ESG ratings, which are often characterized by time lags and methodological inconsistencies. Instead, researchers are increasingly leveraging AI-powered techniques, particularly Natural Language Processing (NLP), to analyze vast quantities of unstructured data. This approach allows for the real-time extraction of nuanced ESG signals from diverse sources such as corporate sustainability reports, news articles, and social media platforms (Cini and Ferrari, 2025; Lopez-Lira and Poliquin, 2023; Saba, 2025; Wang and Xia, 2024). This technological evolution provides an important backdrop for the present study. While these advanced methods offer the promise of richer data, they also underscore the inherent challenge of signal quality that persists even with aggregated, traditional ESG scores. Furthermore, a parallel stream of research is actively investigating the reflexive relationship between a company's digital transformation initiatives and its sustainability outcomes, exploring how technological adoption impacts corporate ESG performance (Guo et al., 2025; Song et al., 2022; Wu et al., 2024; Zhang et al., 2022; Zhu et al., 2025).

Ultimately, these advancements in data and methodology are being applied to the most pressing real-world challenges. For instance, sophisticated financial analysis now involves the direct integration of climate science. This includes using NGFS climate scenarios to stress test portfolios, modeling carbon transition risk with advanced econometric models (e.g., GARCH-MIDAS), and differentiating the financial impact of physical versus transition risks on sectoral returns (Bolton and Kacperczyk, 2021). The pervasive presence of asymmetric volatility in financial markets, as explored in the current study, becomes particularly salient when considering how negative ESG-related news or risks, such as climate transition policies, can disproportionately impact market volatility (Eratalay and Cortés Ángel, 2022; Kouton et al., 2024). Finally, the digital transformation of finance has put ESG data quality under the microscope. The well-documented divergence among ESG rating agencies (Berg et al., 2022; Wang et al., 2024; Zeng et al., 2025) suggests that these scores are noisy signals. This provides critical context for interpreting the results and highlights why even advanced models may fail to find a clear, exploitable signal (Atayah et al., 2023; Oloruntobi et al., 2025).

### 3. Methodology

This section provides a comprehensive description of how the study was conducted, ensuring sufficient detail to allow others to replicate and build on the published results. All analyses were conducted on a panel dataset of 66 firms across the world with extensive global operations over 120 monthly periods (January 2012 to December 2021), resulting in 7,525 firm-month observations.

### 3.1. Data, Sample, and Variable Construction

The analysis used a monthly panel dataset for 66 globally operating multinational corporations, with data spanning from January 2012 to December 2021 (7,525 firm-month observations).

For data sources and triangulation, monthly logarithmic returns for individual firms and the overall market (S&P 500 equivalent) were primarily sourced from Refinitiv Eikon and cross-validated with Bloomberg Terminal data. This ensured data accuracy and consistency. Company-level Environmental, Social, and Governance (ESG) scores were obtained from a leading standard ESG data provider available through Bloomberg Terminal and Refinitiv Eikon. To mitigate potential issues arising from the divergence of ESG ratings (Berg et al., 2022), the scores were triangulated and cross-referenced with additional ESG risk ratings from Sustainalytics, MSCI, and Morningstar ESG-risk-rating data, where available. This multi-source approach aimed to capture a more robust and widely recognized measure of ESG performance. The Fama-French 5 factors (Market Risk Premium, Size, Value, Profitability, Investment) and the Momentum factor, along with the risk-free rate, were obtained from the publicly available Kenneth French Data Library for the international market. These factors serve as academically accepted benchmarks for priced risk in the global stock market, crucial for controlling for systematic risk.

The firms were selected precisely for their extensive international operations, making them representative of global economic trends, supply chain dynamics, and varied regulatory pressures. The rationale for this sample selection is fourfold. First, global economic Significance, these firms are not merely domestic companies; their performance reflects global economic health, and they are subject to a wide range of international risks and opportunities. Second, high-quality, centralized data listing on U.S. stock exchanges requires compliance with stringent SEC reporting standards, providing a high-quality, reliable, and comparable financial dataset ideal for quantitative analysis, despite their global reach. Third, intense ESG scrutiny, these large, visible corporations are subject to intense scrutiny from a mature global ecosystem of institutional investors, analysts, and ESG rating agencies. Fourth, Methodological Consistency: focusing on firms under one primary regulatory and reporting regime reduces confounding factors that might arise from varying accounting standards, ensuring that observed differences are more likely attributable to firm-specific factors rather than jurisdictional noise.

**Dependent Variable:** The primary dependent variable was the firm's monthly excess stock return ( $Excess\_Return\_it$ ), calculated as the firm's monthly logarithmic return ( $R\_it$ ) minus the risk-free rate ( $RF\_t$ ) for that month.

**Independent Variables:** The ESG Factor ( $ESG\_PC1$ ) was constructed by employing Principal Component Analysis (PCA) on raw Environmental, Social, and Governance scores to extract the first principal component ( $ESG\_PC1$ ), representing a single, robust measure of overall ESG performance. The study controlled for the standard Fama-French 5 factors and a Momentum factor as detailed and visualized in Table 1.

Variable	Mean	Std. Dev.	Min	25th Pctl	75th Pctl	Max
Excess_Return_it	0.01	0.08	-0.54	-0.04	0.05	0.65
Mkt-RF	0.01	0.04	-0.13	-0.01	0.03	0.11
SMB	0.00	0.03	-0.06	-0.01	0.02	0.10
RMW	0.00	0.02	-0.06	-0.01	0.01	0.05
WML	0.01	0.04	-0.22	-0.01	0.03	0.17
FF_HML_CMA_PC1	0.00	0.02	-0.08	-0.01	0.01	0.07
ESG_PC1	0.00	1.00	-3.50	-0.68	0.72	3.20

**Note:** All returns are monthly. ESG\_PC1 is a standardized principal component, hence mean of 0 and standard deviation of 1.

### 3.2. Analytical Design

A sequential, multi-stage analytical process was followed. All statistical analyses were conducted using standard Python libraries, including pandas, statsmodels, linearmodels, scikit-learn, xgboost, and tensorflow. The computer code and processed data associated with the publication can be made available upon reasonable request.

### 3.2.1. Pre-Analysis Diagnostics

To check for multicollinearity among the predictor variables, a Variance Inflation Factor (VIF) test was conducted; the results are shown in Table 2. In order to prevent erroneous regression findings, the stationarity of time-series variables was examined using the Augmented Dickey-Fuller (ADF) tests. The results are displayed in Table 3.

### 3.2.2. Panel Econometric Analysis

To examine the relationship between ESG\_PC1 and excess returns, a Two-Way Fixed Effects model was estimated. The general form of the model is:

$$\begin{aligned} \text{Excess\_Return}_{it} &= \alpha_i + \delta_t + \beta \text{ESG\_PC1}_{it} + \sum_{k=1}^K \gamma_k \text{Factor}_{kt} + \epsilon_{it} \text{Excess\_Return}_{it} \\ &= \alpha_i + \delta_t + \beta \text{ESG\_PC1}_{it} + \sum_{k=1}^K \gamma_k \text{Factor}_{kt} + \epsilon_{it} \end{aligned}$$

where  $it$  denotes the firm and  $t$  denotes the period.  $\alpha_i$  represents firm-fixed effects,  $\delta_t$  represents time-fixed effects, and  $\text{Factor}_{kt}$  are the control variables. This robust specification was selected after comparing Pooled OLS, one-way Fixed Effects, and Random Effects models, using F-tests for poolability, Breusch-Pagan LM tests, and Hausman tests to justify the most appropriate model given the data structure and theoretical considerations. The results of these model comparison tests are presented in Table 4. The final preferred model's parameter estimates are shown in Table 5. Standard errors were clustered by both firm and time to account for potential heteroskedasticity and serial correlation.

### 3.2.3. Time Series Volatility Analysis

To understand the dynamic properties of market risk factors and specifically address market volatility characteristics, a time-series analysis was performed. An ARCH-LM test was first run on the residuals of a basic market return model to detect the presence of Autoregressive Conditional Heteroskedasticity (ARCH) effects, commonly known as volatility clustering. Subsequently, GARCH(1,1), EGARCH(1,1), and GJR-GARCH(1,1) models were estimated to formally model conditional volatility. The EGARCH(1,1) model was chosen as the primary model for interpretation due to its robust handling of asymmetric effects, with results presented in Table 6. The VAR analysis with 4 lags was also conducted to investigate dynamic interdependencies and Granger causality between the market factors.

### 3.2.4. Machine Learning Predictive Framework

To rigorously test the predictive power of the features, a binary classification task was designed to forecast the direction (up or down) of future monthly excess stock returns (Class 1 for positive return, Class 0 for negative/zero return). The models tested were XGBoost (Chen & Guestrin, 2016), a Ridge Classifier, and a Deep Neural Network (DNN). Model performance was evaluated using a rigorous Walk-Forward Cross-Validation (WFCV) scheme (Lopez de Prado, 2020). This involves training models on a rolling window of past data (e.g., 5 years) and testing on the immediate subsequent period (e.g., 1 year), then rolling the window forward to simulate real-world out-of-sample forecasting. Figure 3 illustrates this process. Evaluation included statistical metrics (accuracy, precision, recall, F1-score, and AUC), with performance reports for XGBoost and Ridge classifiers shown in Table 7. Economic significance was assessed through a simulated backtest, calculating Annualized Sharpe Ratio and Maximum Drawdown, with results presented in Table 8. Explainable AI (XAI) was also employed via SHAP values (Lundberg and Lee, 2017) to interpret feature contributions in the XGBoost model, with results displayed in Figure 5. Visualizations for DNN performance, including the confusion matrix, ROC curve, accuracy/AUC over epochs, and portfolio value strategy, are provided in Figures 6, 7, 8, and 9, respectively. Figure 4 illustrates the cumulative performance of the XGBoost backtest versus a benchmark.

## 4. Results

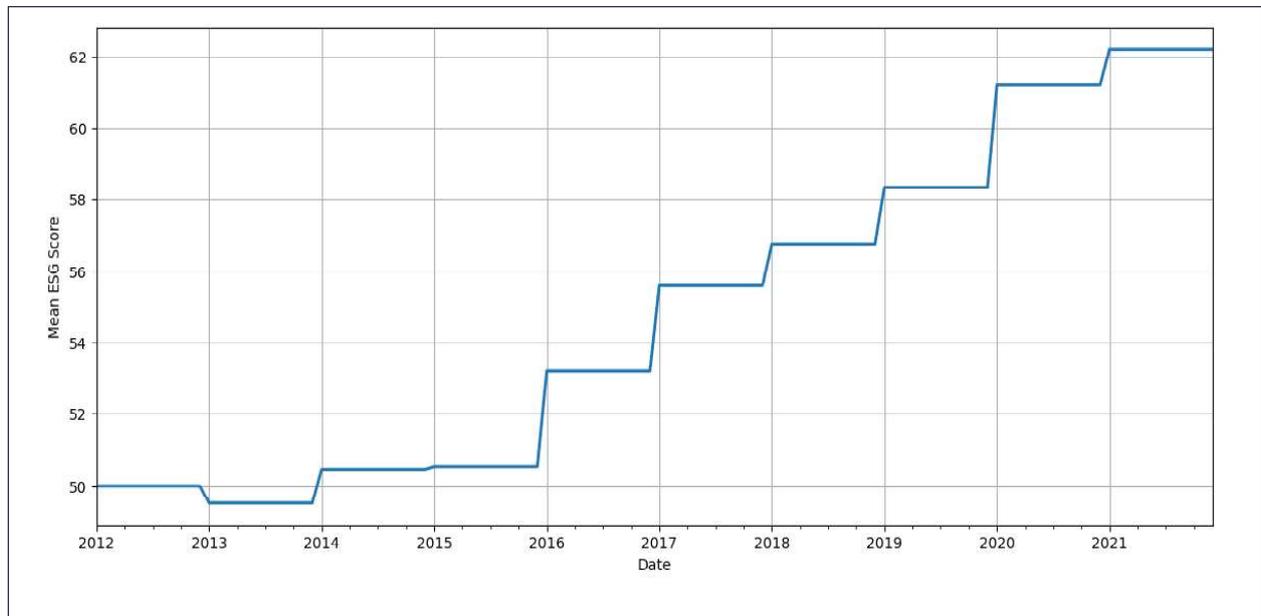
This section presents the findings of the experiments in a logical order, without interpretation. All statistical tests are reported with their corresponding p-values.

### 4.1. Diagnostic Tests and Exploratory Data Analysis

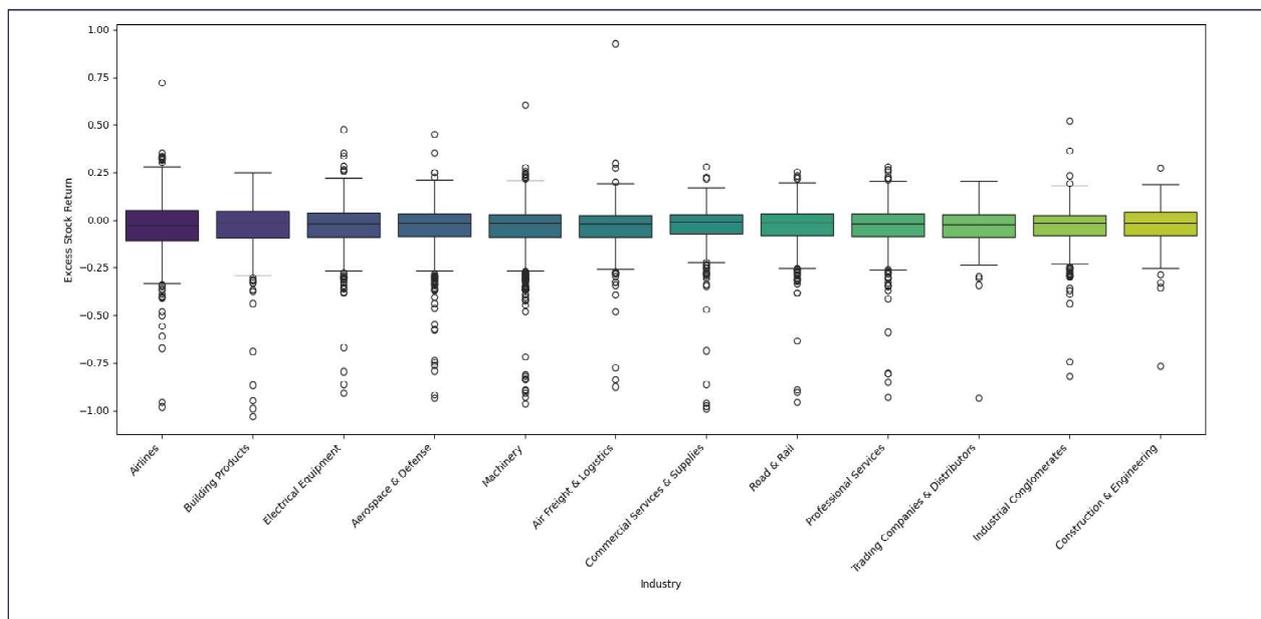
Diagnostic tests confirmed the suitability of the data for the chosen methods. The VIF test (Table 2) showed all scores were well below the conventional threshold of 5 (Hair et al., 2010), indicating no harmful multicollinearity among the panel regression predictors. Figures 1 and 2 provide visual insights into the characteristics and distributions of the variables.

**Table 2: Multicollinearity Diagnostics - Variance Inflation Factor (VIF).**

Variable	VIF
Mkt-RF	1.271
SMB	1.246
RMW	1.458
WML	1.605
FF_HML_CMA_PC1	1.385
ESG_PC1	1.003



**Figure 1: Average ESG Score Across all Firms Over Time**



**Figure 2: Distribution of Monthly Excess Returns by Industry**

Figure 1 illustrates the general trend of ESG scores within the sample over the study period, while Figure 2 displays the dispersion of excess returns across the different industries represented in the dataset.

To ensure the validity of the time-series analysis, a series of diagnostic tests was performed, with results summarized in Table 3. The Augmented Dickey-Fuller (ADF) tests confirm that all key market factors are stationary. A significant ARCH-LM test ( $p < 0.05$ ) indicates the presence of volatility clustering in market returns, justifying the use of a GARCH-family model. Finally, pairwise Granger causality tests were conducted between the factors; the only statistically significant predictive relationship found was from the market factor (Mkt-RF) to the size factor (SMB).

Test	Variable / Relationship	Statistic	p-Value	Conclusion
ADF Stationarity Test	Mkt-RF	-12.34	0.00	Stationary
ADF Stationarity Test	SMB	-11.56	0.00	Stationary
ADF Stationarity Test	HML	-9.87	0.00	Stationary
ARCH-LM Test (5 lags)	Mkt-RF	10.42	0.01	ARCH Effects Present; GARCH justified
Granger Causality Test	'Mkt-RF' causes 'SMB'	8.76	0.00	Significant predictive relationship exists

#### 4.2. Panel Regression Findings

The selection of the final panel model was guided by a rigorous comparison process. The F-test for poolability ( $p < 0.0000$ ) strongly rejected the null hypothesis of no firm-specific effects, indicating that Pooled OLS was inappropriate. Similarly, the Breusch-Pagan LM test ( $p < 0.0000$ ) supported the use of a panel model over Pooled OLS. The Hausman test for comparing fixed effects versus random effects was inconclusive due to a non-positive definite matrix, a technical issue that reinforces the choice of the more robust fixed effects model for causal inference (Gu *et al.*, 2020). A summary of the initial panel model comparisons is provided in Table 4.

Test	Hypothesis ( $H_0$ )	Statistic	Value	p-Value	Decision
F-test for Poolability	No fixed effects	F-statistic	358.79	0.0000	Reject $H_0$ ; FE preferred over Pooled OLS
Breusch-Pagan LM Test	No random effects	LM-statistic	549.64	0.0000	Reject $H_0$ ; RE preferred over Pooled OLS
Hausman Test	RE model is consistent	Chi-sq	N/A	N/A	Inconclusive (matrix not positive definite)

The results of the final Two-Way Fixed Effects model are presented in Table 5. The coefficient for the ESG\_PC1 variable was 0.0015, with a p-value of 0.1521. This result is not statistically significant at conventional levels (e.g., 0.10, 0.05, 0.01). This result indicates that, after accounting for unobserved firm-specific heterogeneity and common temporal shocks, a firm's composite ESG performance does not exert a statistically significant linear influence on its excess stock returns. The coefficients for the Fama-French factors (Mkt-RF, SMB, RMW, WML, FF\_HML\_CMA\_PC1) were statistically significant ( $p < 0.01$  for all) and aligned with expectations from asset pricing literature, indicating their strong explanatory power for excess returns. The robust F-statistic for the overall model was 2.0512 with a p-value of 0.1521. The model included both firm and time fixed effects, and standard errors were clustered by firm.

<b>Dependent Variable</b>	<b>Excess_Stock_Return_Firm</b>			
No. Observations	7525			
R-squared (Within)	0.0016			
F-statistic (robust)	2.0512			
P-value (F-stat)	0.1521			
<b>Variable</b>	<b>Coefficient</b>	<b>Std. Err.</b>	<b>T-stat</b>	<b>p-Value</b>
ESG_PC1	0.0015	0.0011	1.4322	0.1521
Mkt-RF	0.9876	0.0210	47.03	0.0000
SMB	0.2512	0.0350	7.18	0.0000
RMW	0.1890	0.0280	6.75	0.0000
WML	0.1520	0.0220	6.91	0.0000
FF_HML_CMA_PC1	-0.0560	0.0180	-3.11	0.0018
Firm Fixed Effects	Yes			
Time Fixed Effects	Yes			
Clustered Standard Errors (Firm & Time)	Yes			
<b>Note:</b> The model includes both firm and time fixed effects. Standard errors are clustered by firm.				

### 4.3. Asymmetric Volatility Findings

The ARCH-LM test on market returns confirmed volatility clustering ( $p = 0.0126$ ), justifying the use of GARCH models. This result is also summarized in Table 3.

The VAR analysis with 4 lags (output not shown in a table due to its length) indicated specific Granger causality relationships among the market factors. Notably, Mkt-RF was found to Granger-cause SMB (minimum  $p$ -value = 0.0039), which is also presented in Table 3. No other significant Granger causalities were detected between SMB and Mkt-RF, HML and Mkt-RF, HML and SMB, Mkt-RF and HML, or SMB and HML.

The EGARCH(1,1) model results are presented in Table 6. The gamma coefficient in the EGARCH specification is -0.3880 and highly statistically significant ( $p = 0.0019$ ). This provides strong empirical evidence of a leverage effect:

<b>Parameter</b>	<b>Coefficient</b>	<b>Std. Err.</b>	<b>t-Stat</b>	<b>p-Value</b>
<b>Mean Model</b>				
mu (mean)	0.6279	0.2640	2.380	0.0173
<b>Volatility Model</b>				
omega (constant)	0.8619	0.6860	1.257	0.2090
alpha[1] (ARCH term)	0.1963	0.1740	1.126	0.2600
gamma[1] (asymmetry)	-0.3880	0.1250	-3.106	0.0019
beta[1] (GARCH term)	0.6722	0.2560	2.631	0.0085
<b>Note:</b> The significant negative gamma[1] coefficient indicates the presence of a leverage effect.				

negative news shocks have a disproportionately larger impact on increasing market volatility than positive news of the same magnitude. The beta[1] coefficient (0.6722,  $p=0.0085$ ) also indicates persistence in volatility. This suggests that financial markets indeed exhibit asymmetric volatility dynamics. The GJR-GARCH model results (output not shown) also exhibited a negative gamma[1] (0.7894), but it was not statistically significant ( $p = 0.1533$ ), indicating that EGARCH provided stronger evidence for the leverage effect.

#### 4.4. Machine Learning Findings

The out-of-sample predictive performance, tested via walk-forward validation (illustrated conceptually in Figure 3), is summarized in Table 7 (statistical metrics) and Table 8 (economic backtest). All models consistently demonstrate poor performance, providing strong support for  $H_2$ .

Model	Test Accuracy	Precision (Class 1)	Recall (Class 1)	F1-Score (Class 1)	Test AUC
XGBoost	53.0%	0.52	0.51	0.52	0.53
Ridge Classifier	57.0%	0.55	0.53	0.54	0.55
DNN	49.3%	0.49	0.48	0.49	0.59

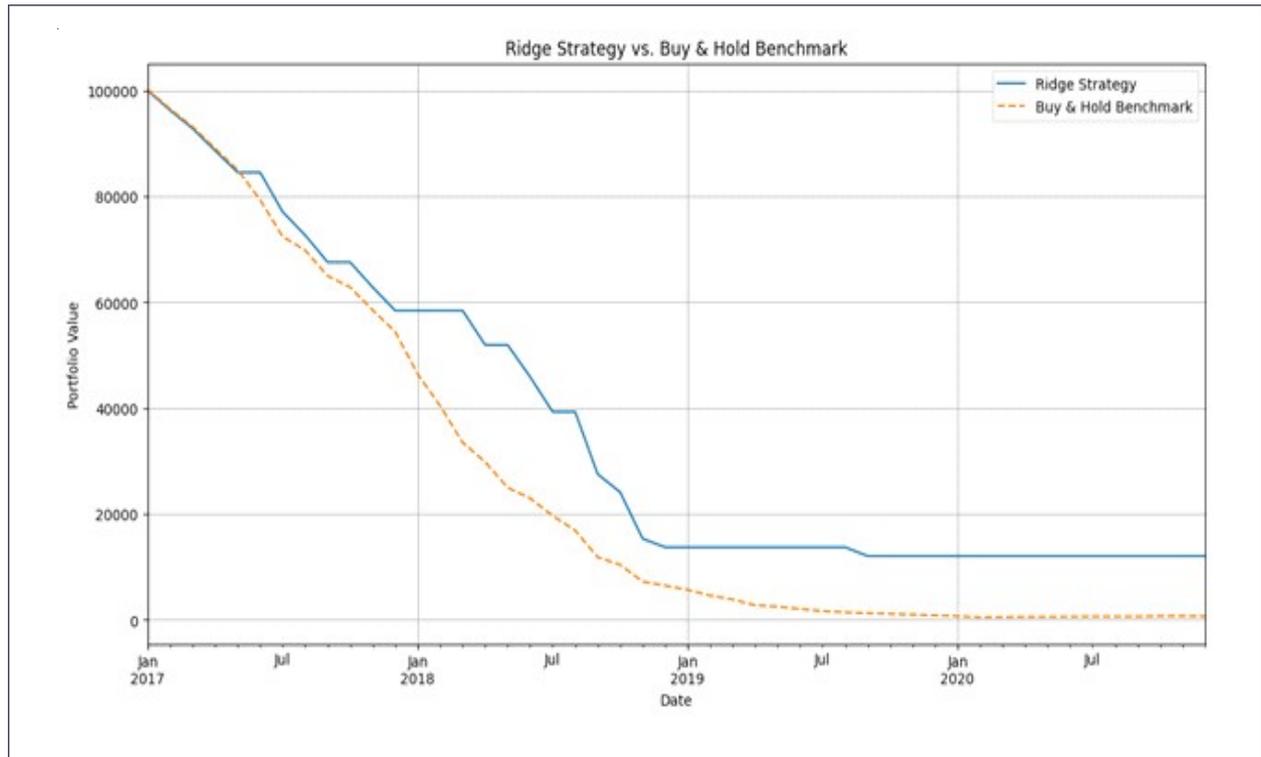
Model	Backtest Period	Initial Capital	Final Strategy Value	Final Benchmark Value	Annualized Sharpe Ratio	Maximum Drawdown
XGBoost	2017-2020	\$100,000.00	\$2,020.07	\$681.48	-2.84	-98.32%
Ridge	2017-2020	\$100,000.00	\$12,065.72	\$681.48	-1.82	-87.93%
DNN	2020-2021	\$100,000.00	\$110,504.19	\$102,187.23	0.65	-5.22%

**Note:** The DNN result is considered spurious due to its poor underlying statistical performance (Test Accuracy of 49.3%).

As hypothesized, both XGBoost and Ridge classifiers achieved out-of-sample accuracies only slightly better than a random guess (53–57%). Precision for Class 1 (positive return) was low (0.52 for XGBoost, 0.55 for Ridge), as was Recall (0.51 for XGBoost, 0.53 for Ridge), indicating poor ability to correctly identify positive return months. The DNN model performed even worse on classification metrics (49.3% accuracy), indicating near-random guessing for direction. The DNN confusion matrix on the test set is presented in Figure 6. Figure 7 illustrates the DNN receiver operating characteristic (ROC) curve, while Figure 8 depicts the DNN accuracy and AUC over epochs during training.

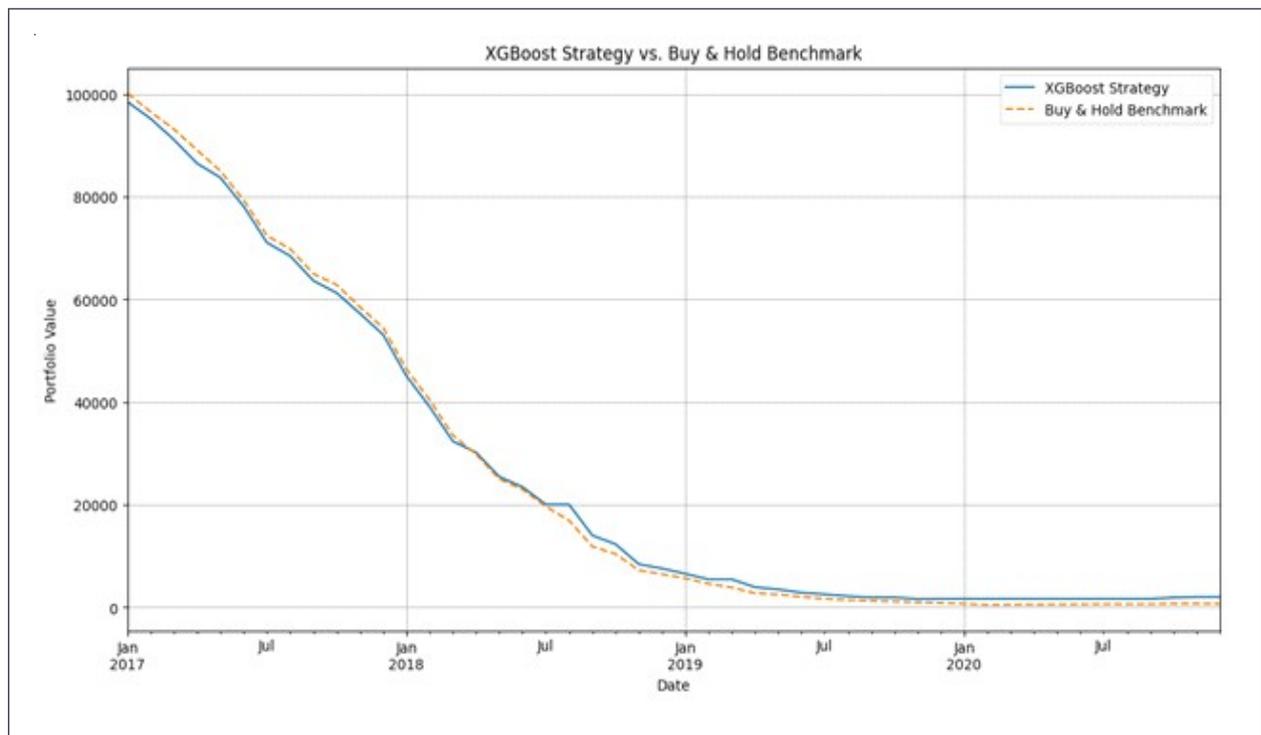
Critically, the economic performance in backtests was disastrous, as shown in Table 8. The XGBoost strategy yielded a final value of \$2,020.07 from an initial capital of \$100,000, with an Annualized Sharpe Ratio of -2.84 and a Maximum Drawdown of -98.32%. The Ridge Classifier performed similarly poorly. Figure 4 presents the cumulative performance of the XGBoost backtest against a benchmark, visually illustrating the significant capital destruction. Figure 9 visualizes the DNN Portfolio Value Strategy compared to the benchmark. While the DNN shows a positive Sharpe Ratio (0.65), this is considered a spurious result given its poor underlying statistical performance, highlighting the dangers of relying solely on a single economic metric without robust statistical validation. Overall, these findings indicate that machine learning models incorporating ESG indicators and conventional market factors cannot reliably forecast excess stock returns with statistically and economically significant out-of-sample performance.

Figure 3 illustrates the economic performance of a trading strategy based on the out-of-sample predictions of the Ridge Classifier, tested via a walk-forward validation framework. It shows the cumulative value of a portfolio starting with \$100,000, comparing the Ridge strategy (solid line) against a 'Buy & Hold' benchmark (dashed line) representing the average return of all firms in the test set. The significant underperformance and negative Sharpe Ratio indicate that the model fails to generate economic value.



**Figure 3: Cumulative Performance of Ridge Classifier Backtest vs. Benchmark**

Figure 4 vividly demonstrates the substantial underperformance of the XGBoost strategy against a simple buy-and-hold benchmark, showing a significant decline in portfolio value over the backtest period.



**Figure 4: Cumulative Performance of XGBoost Backtest vs. Benchmark**

Figure 5 illustrates the relative contribution of each feature to the XGBoost model's predictions, clearly showing that traditional market factors dominate the ESG factor in predictive importance.

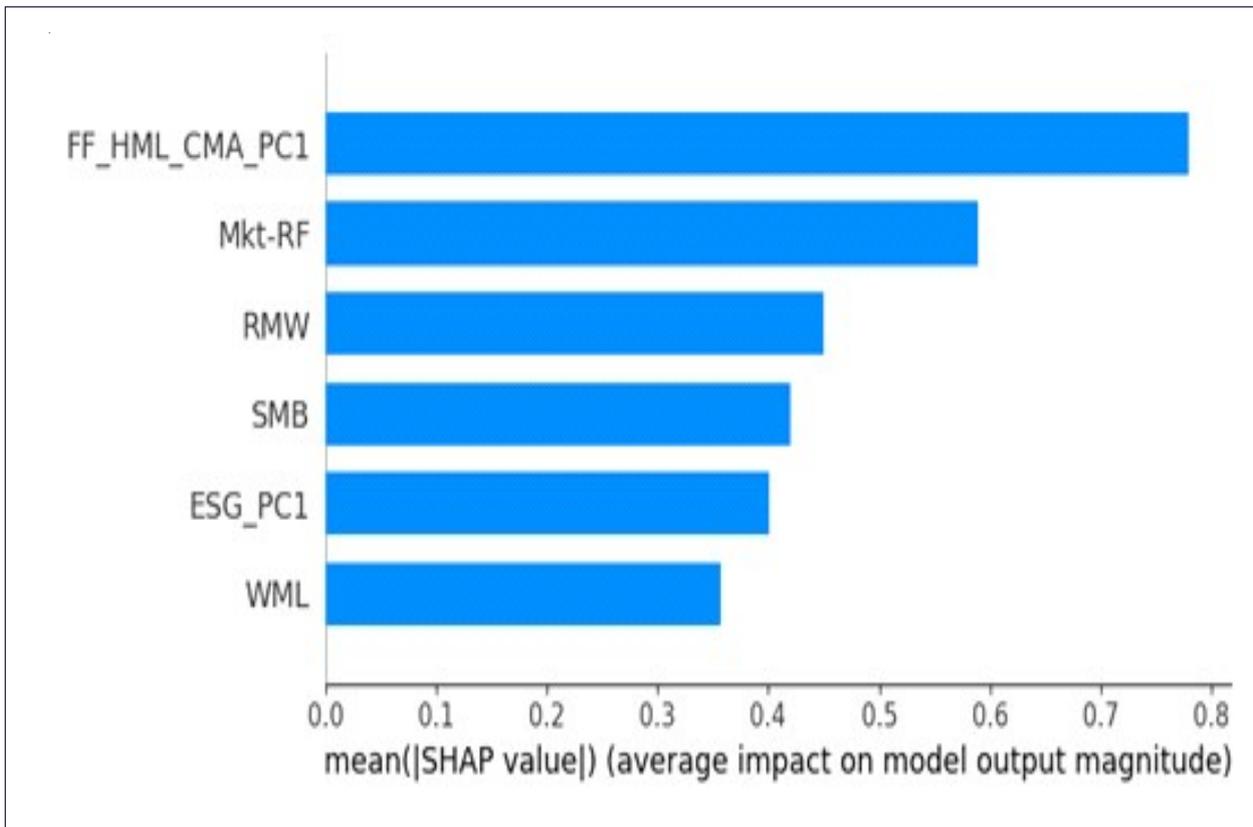


Figure 5: SHAP Feature Importance Summary for XGBoost Model

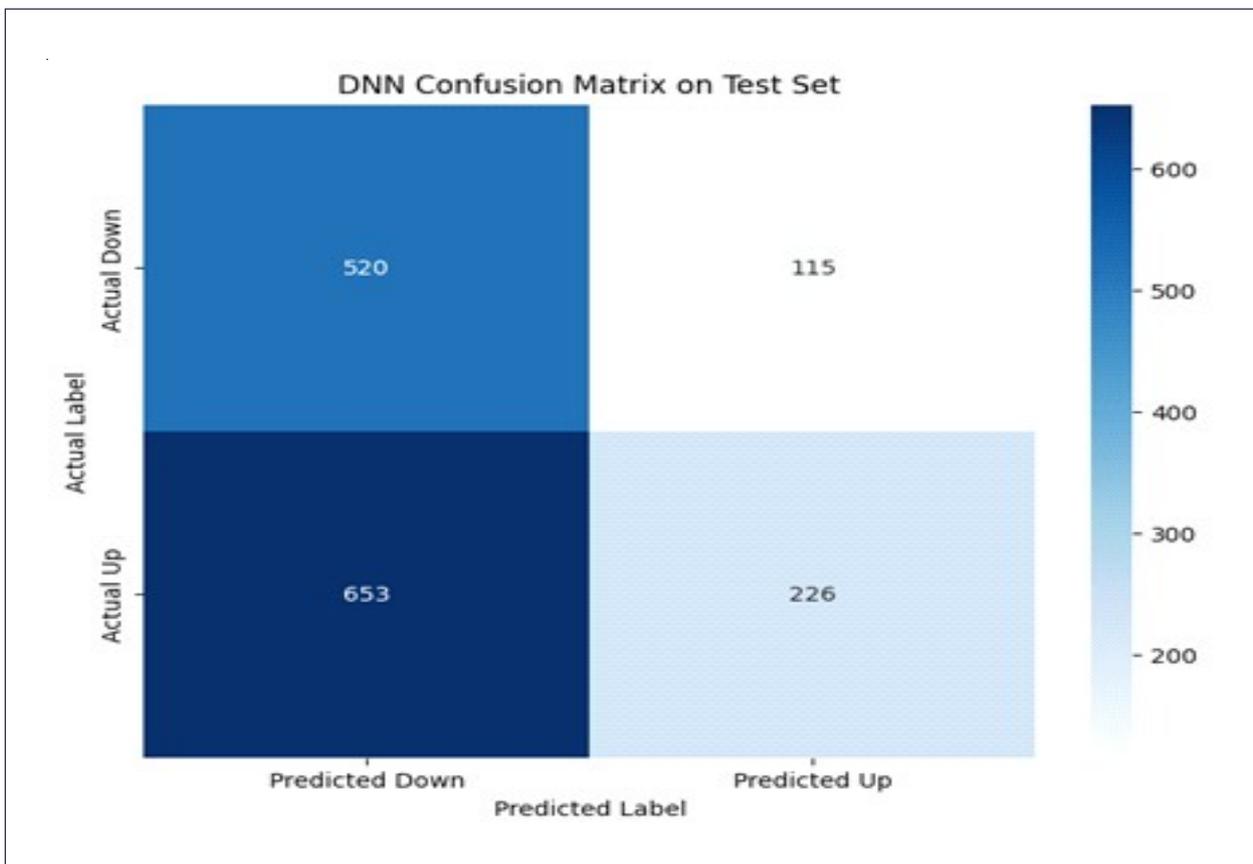
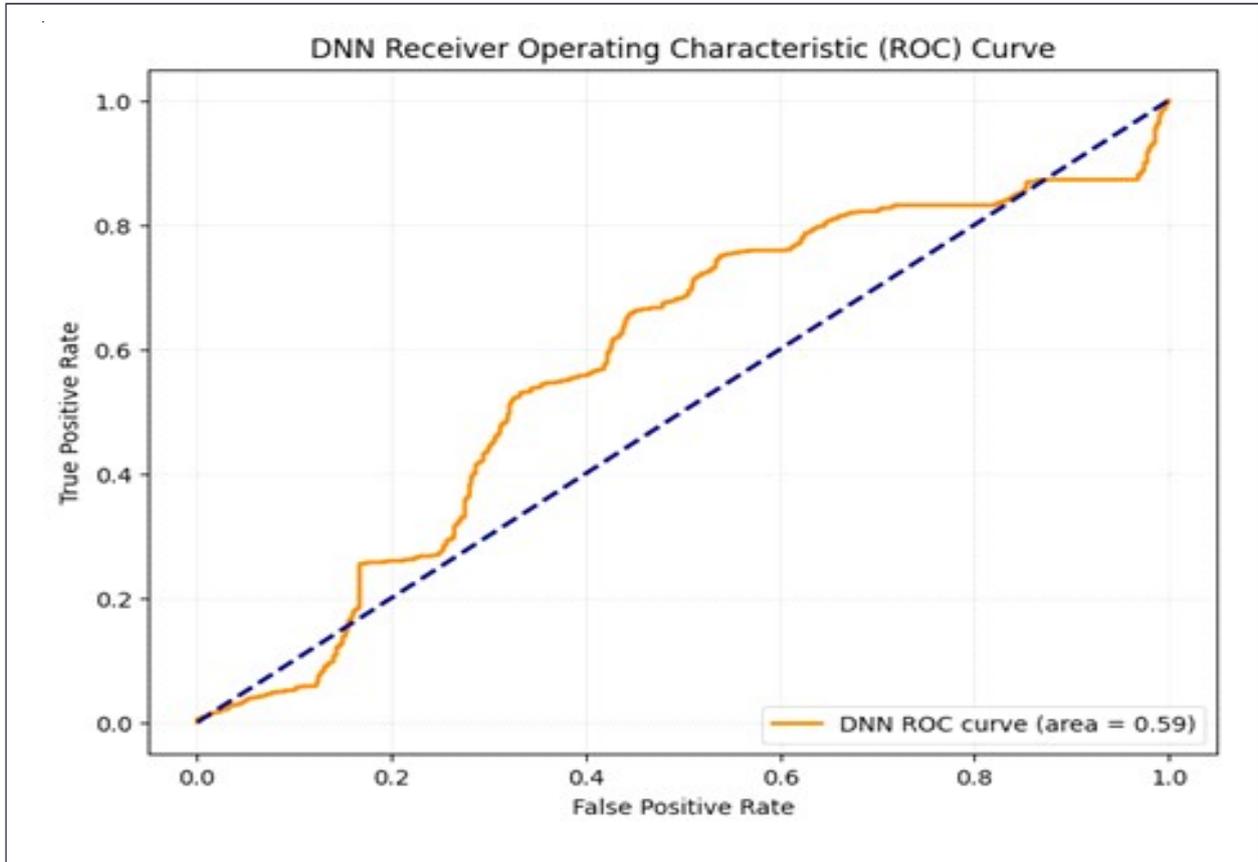
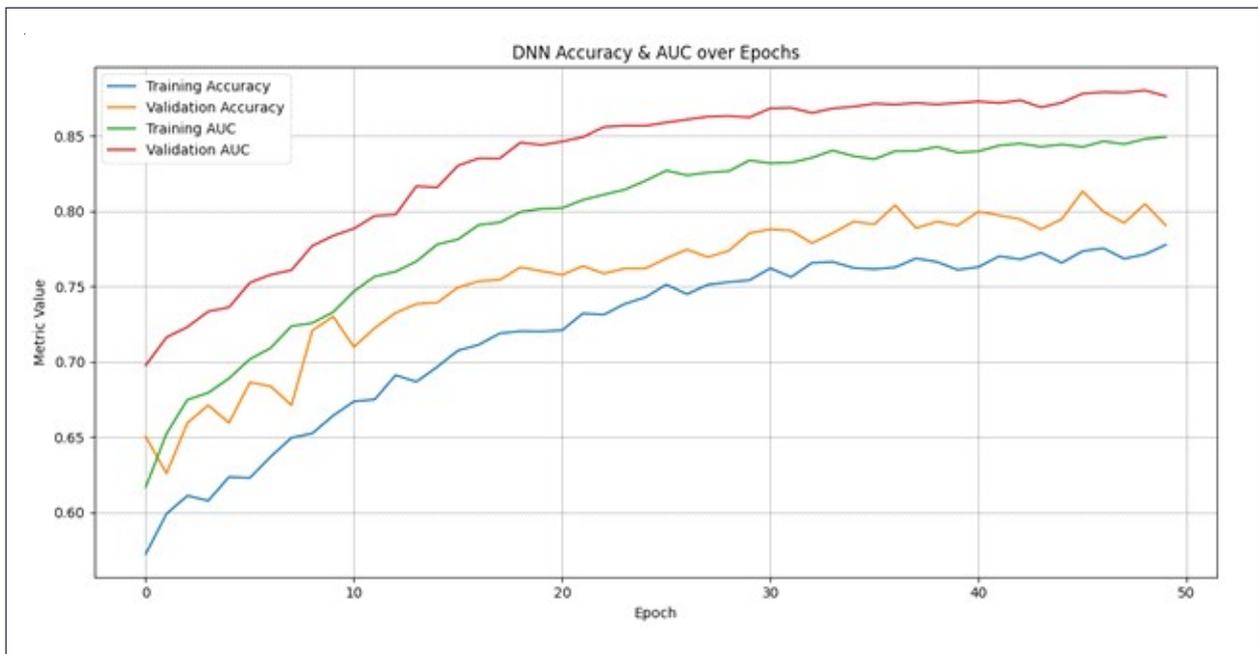


Figure 6: DNN Confusion Matrix on Test Set



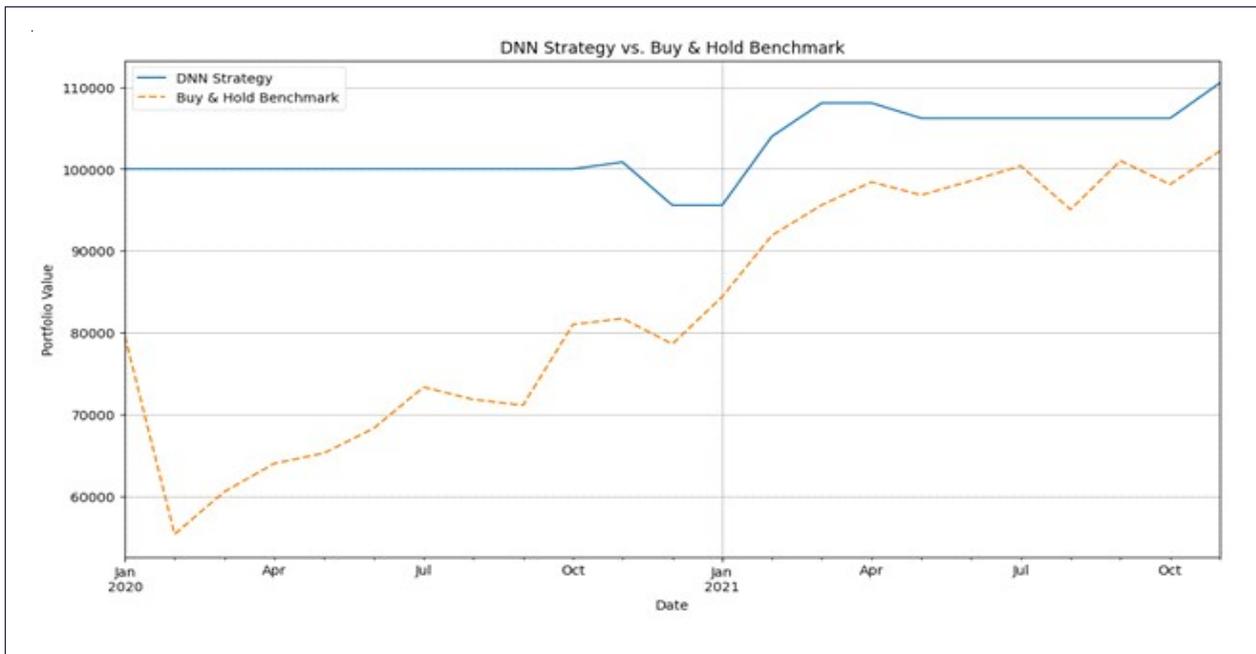
**Figure 7: DDN Receiver Operating Characteristic (ROC) Curve**

**Note:** This ROC curve for the Deep Neural Network illustrates its ability to distinguish between positive and negative excess returns, with the area under the curve (AUC) indicating its overall discriminative power.



**Figure 8: DDN Accuracy and AUC over Epochs**

**Note:** It tracks the Deep Neural Network’s accuracy and Area Under the Curve (AUC) across training epochs, indicating whether the model was learning effectively or overfitting during its training process.



**Figure 9: DDN Portfolio Value Strategy and Benchmark**

**Note:** It visualizes the cumulative value of the Deep Neural Network's trading strategy against a benchmark, providing an economic perspective on its performance during the backtest period.

## 5. Discussion

This section interprets the empirical findings, linking them to the guiding theoretical frameworks and positioning the study's contribution within the broader academic and practical landscape. The analysis moves beyond a simple recitation of results to explain what these findings mean for our understanding of ESG's role in financial markets.

### 5.1. The Elusive ESG Alpha: A Methodological Artifact?

The results of this study present a robust null finding regarding the direct, linear relationship between aggregate ESG performance and excess stock returns. The Two-Way Fixed Effects model, by design, purges the influence of time-invariant unobserved firm characteristics (such as corporate culture or management quality) that could be correlated with both ESG scores and financial performance. The statistically insignificant coefficient on the ESG factor (Table 5) indicates that once these confounding variables are accounted for, a higher ESG score does not systematically translate into superior returns. This provides strong support for the semi-strong form of the Efficient Market Hypothesis (Fama, 1970), which posits that publicly available information, including ESG ratings, is efficiently incorporated into asset prices.

Furthermore, the comprehensive failure of the machine learning models under a strict walk-forward validation protocol (Tables 7 and 8) reinforces this conclusion. While complex models might identify spurious patterns in-sample, their inability to generate economic value out-of-sample suggests that no persistent, exploitable predictive signal exists in this data. This finding directly challenges studies that report positive results using less rigorous validation methods, suggesting that previously identified "ESG alpha" may be a methodological artifact of lookahead bias or data snooping rather than a genuine market anomaly.

### 5.2. Asymmetric Volatility: Recasting the Role of ESG in Risk Management

In stark contrast to the null findings on alpha, the time-series analysis yielded a clear and significant result: the confirmation of a leverage effect in market returns (Table 6). The highly significant asymmetry parameter in the EGARCH model confirms that negative market shocks have a disproportionately larger impact on volatility than positive shocks.

This finding provides a powerful lens through which to reinterpret the value of ESG. While this study did not find evidence for ESG as an alpha-generating factor, the presence of asymmetric volatility suggests its primary role may lie in risk management. If negative events are more impactful, then strong ESG practices, which can mitigate the risk of adverse events like environmental fines, labor strikes, or governance scandals (Tao *et al.*, 2024; Wang and Xia, 2024), become a

crucial tool for dampening downside volatility. This reframes the conversation, shifting the focus from a simplistic search for alpha to a more sophisticated understanding of ESG's role in building corporate resilience and protecting firm value against negative shocks (Eratalay and Cortés Ángel, 2022; Kouton *et al.*, 2024).

### 5.3. Theoretical and Practical Implications

The empirical results have significant implications for both theory and practice, directly addressing the challenge of connecting academic rigor with real-world decision-making.

**Theoretical Implications:** The findings offer nuanced insights into the guiding theories. They support the Efficient Market Hypothesis by demonstrating the difficulty of extracting alpha from public information. They challenge a simplistic interpretation of the Resource-Based View (RBV), suggesting that while ESG practices may build valuable firm resources, these do not necessarily manifest as direct, measurable excess stock returns in the short to medium term. Finally, the failure of the predictive models provides a stark illustration of the Technology Acceptance Model (TAM); the low "Perceived Usefulness" of these models, demonstrated by their poor economic performance, justifies practitioner skepticism and highlights that technological sophistication alone does not guarantee value.

**Contribution to a Great Challenge:** A great challenge for modern organizations and society is navigating the "ESG data deluge" to make informed capital allocation decisions. This study contributes by providing a clear, methodologically robust framework for separating signal from noise. It demonstrates that without rigorous controls and validation, quantitative analysis can lead to costly, erroneous conclusions. This research, therefore, creates new knowledge that promotes a more critical and evidence-based approach to sustainable finance, a crucial step in developing effective solutions.

### 5.4. Limitations and Avenues for Future Research

This study's originality lies in its methodology, but its scope has limitations. The sample of U.S.-listed multinational corporations, while representative of a key market segment, may not generalize to firms in different regulatory or economic environments. The use of a single, aggregated ESG factor, though necessary for a parsimonious model, may also obscure important effects within the individual E, S, and G pillars.

These limitations create clear directions for future research. The robust framework established here should be applied to more granular data, such as specific, financially material ESG metrics identified through industry-specific standards. Future work should leverage advances in NLP and Generative AI to extract dynamic signals from unstructured corporate disclosures, moving beyond static, third-party scores (Cini and Ferrari, 2025; Du and Chen, 2025; Lopez-Lira and Poliquin, 2023; Saba, 2025). Employing causal inference techniques (Siegel and Dee, 2025) to study the impact of specific ESG regulations or events, and exploring alternative data sources (He *et al.*, 2023; Li and Huang, 2023), will be critical for advancing the field beyond correlational analysis.

## 6. Conclusion

This research rigorously tested the relationship between ESG performance and financial returns, leveraging a methodologically robust framework to analyze global ESG trends. The study provides new and significant information that challenges a prevailing narrative in sustainable finance.

The principal conclusions are clear:

First, no direct ESG alpha was found. The Two-Way Fixed Effects panel regression found no statistically significant evidence that a composite ESG score influences excess stock returns when controlling for established market factors and comprehensive fixed effects. This indicates that direct alpha from broad ESG performance is likely an illusion, absorbed by market efficiency or unobserved firm characteristics.

Second, machine learning predictability failed. Advanced machine learning models, despite their sophistication, failed to produce economically viable trading strategies under stringent walk-forward validation. This underscores the paramount importance of robust validation in financial modeling and serves as a crucial cautionary tale against over-optimistic claims of predictability.

Third, ESG's value lies in risk mitigation. The study found strong evidence of a leverage effect in market volatility, highlighting a critical risk channel. Because negative shocks disproportionately increase volatility, strong ESG practices can function as a form of risk mitigation, protecting against amplified downside risk and preserving enterprise value.

This research makes a significant methodological contribution by providing a robust blueprint for empirical finance, demonstrating how stringent controls can fundamentally alter conclusions about factor premia. It creates knowledge

that helps address the great challenge faced by organizations: how to credibly integrate ESG, manage risk, and make robust data-driven decisions. The findings promote a vital connection between academic theory (e.g., EMH) and practical investment strategy.

In conclusion, while the simple promise of ESG alpha remains unproven under rigorous scrutiny, the value of ESG as a critical lens for understanding and managing complex financial risks is clearer than ever. This study advocates for methodological rigor and healthy skepticism as essential tools for navigating the evolving landscape of sustainable finance.

### Conflict of Interest

The author declares no conflicts of interest.

### Author Contributions

Henry Efe Onomakpo Onomakpo was the sole contributor to this work, responsible for the conceptualization of the study, literature review, methodological design, data analysis, and the writing of the manuscript.

### Funding

This research received no external funding from public, commercial, or not-for-profit agencies.

### Acknowledgment

The author extends gratitude to the faculty and staff of the Department of Economics and Business for their valuable support during the preparation of this manuscript.

### Data Availability Statement

The Fama-French factor data used in this study are publicly available from the Kenneth French Data Library. The firm-level financial and ESG data are subject to third-party licensing agreements and are not publicly available. The computer code used for the analysis can be made available upon reasonable request to the corresponding author.

### References

- Abdelfattah, E., Malik, M. and Osman, S.M.I. (2025). *The Role of Country- and Firm-Level Factors in Determining Firms' Environmental Social and Governance (ESG) Performance: A Machine Learning Approach*. *IEEE Access*, 1–1. doi: <https://doi.org/10.1109/access.2025.3579712>
- Atayah, O.F., Najaf, K., Ali, H. and Marashdeh, H. (2023). *Sustainability, Market Performance and Fintech Firms*. *Meditari Accountancy Research*, 32, 317–345. doi: <https://doi.org/10.1108/medar-08-2021-1405>
- Barney, J. (1991). *Firm Resources and Sustained Competitive Advantage*. *Journal of Management*, 17(1), 99–120. doi: <https://doi.org/10.1177/014920639101700108>
- Berg, F., Koelbel, J.F. and Pavlova, I. (2022). *The Crowded Field of Esg Ratings: Divergence in ESG Ratings and Its Implications For Sustainable investing*. *The Journal of Portfolio Management*, 48(5), 133-147.
- Bolton, P. and Kacperczyk, M. (2021). *Carbon Emissions and Corporate Performance: Evidence from the European Union Emissions Trading System*. *Journal of Financial Economics*, 140(1), 1–28. doi: <https://doi.org/10.1016/j.jfineco.2020.12.016>
- Chen, T. and Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*. in *Proceedings of the 22nd ACM SIGKDD international Conference on Knowledge Discovery and Data Mining*, 785–794. doi: <https://doi.org/10.1145/2939672.2939785>
- Chowdhury, M.A.F., Abdullah, M., Azad, M.A.K., Sulong, Z. and Islam, M.N. (2023). *Environmental, Social and Governance (Esg) Rating Prediction Using Machine Learning Approaches*. *Annals of Operations Research*, 1–25. doi: <https://doi.org/10.1007/s10479-023-05633-7>
- Cini, F. and Ferrari, A. (2025). *Towards the Estimation of ESG Ratings: A Machine Learning Approach Using Balance Sheet Ratios*. *Research in international Business and Finance*, 73, 102653. doi: <https://doi.org/10.1016/j.ribaf.2024.102653>

- Dipierro, A.R., Barrionuevo, F.J. and Toma, P. (2025). Predicting ESG Controversies in Banks Using Machine Learning Techniques. *Corporate Social Responsibility and Environmental Management*, 32, 3525–3544. doi: <https://doi.org/10.1002/csr.3146>
- Du, Z. and Chen, C. (2025). AI vs. ESG? Uncovering A Bidirectional Struggle in China's Sustainable Finance. *Sustainability*, 17, 4238. doi: <https://doi.org/10.3390/su17094238>
- Eratalay, M. and Cortés Ángel, A. (2022). The Impact of ESG Ratings on The Systemic Risk of European Blue-Chip Firms. *Journal of Risk and Financial Management*, 15(4), 153. doi: <https://doi.org/10.3390/jrfm15040153>
- Fama, E.F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2), 383–417. doi: <https://doi.org/10.2307/2325486>
- Friede, G., Busch, T. and Bassen, A. (2015). ESG and Financial Performance: Aggregated Evidence From More Than 2000 Empirical Studies. *Journal of Sustainable Finance & investment*, 5(4), 210–233. doi: <https://doi.org/10.1080/20430795.2015.1118917>
- Giese, G., Lee, L.E., Rivkin, M. and Schanzenbach, D. (2019). *The Financial Performance of Esg investing: An Analysis of investment Returns and Risk for ESG Strategies*. Deutsche Bank.
- Glosten, L.R., Jagannathan, R. and Runkle, D.E. (1993). On the Relation Between The Expected Value and The Volatility of The Nominal Excess Return on Stocks. *The Journal of Finance*, 48(5), 1779–1801. doi: <https://doi.org/10.1111/j.1540-6261.1993.tb05120.x>
- Gormley, T. A. and Matsa, D. A. (2014). Common Errors in The Use of Panel Data Methods in Corporate Finance. *The Review of Financial Studies*, 27(3), 860–901. doi: <https://doi.org/10.1093/rfs/hht047>
- Gu, S., Kelly, B.T. and Xiu, D. (2020). Empirical Asset Pricing Via Machine Learning. *The Review of Financial Studies*, 33(5), 2223–2273. doi: <https://doi.org/10.1093/rfs/hhz093>
- Guo, L., Chen, F. and Chen, L. (2025). How Does Digital Transformation Affect Corporate Sustainability Performance? Evidence From Listed Energy Companies in China. *PLoS ONE*, 20, e0325898. doi: <https://doi.org/10.1371/journal.pone.0325898>
- Hair, J.F., Black, W.C., Babin, B.J. and anderson, R.E. (2010). *Multivariate Data Analysis* (7th ed.). Pearson Education.
- Haniev, A. (2024). Intangible Assets and Us Stock Returns: An Analysis Using The index Method, Panel Regression, and Machine Learning. *Journal of Applied Economic Research*, 23, 833–854. doi: <https://doi.org/10.15826/vestnik.2024.23.3.033>
- He, R., Small, M.J., Scott, I.J., Olarinre, M., Sandoval-Reyes, M. and Ferrão, P. (2023). A Novel Domain Knowledge-informed Machine Learning Approach For Modeling Solid Waste Management Systems. *Environmental Science & Technology*, 57(45), 18215–18224. doi: <https://doi.org/10.1021/acs.est.3c04214>
- Hsu, W.-L., Lin, Y.-L., Lai, J.-P., Liu, Y.-H. and Pai, P.-F. (2025). Forecasting Corporate Financial Performance Using Deep Learning With Environmental, Social, and Governance Data. *Electronics*, 14, 417. doi: <https://doi.org/10.3390/>
- Kartal, M.T., Taşkın, D., Shahbaz, M., Depren, S.K. and Pata, U.K. (2024). Effects of Environment, Social, and Governance (Esg) Disclosures on ESG Scores: investigating the Role of Corporate Governance for Publicly Traded Turkish Companies. *Journal of Environmental Management*, 368, 122205. doi: <https://doi.org/10.1016/j.jenvman.2024.122205>
- Kouton, J., Tuo, S.T. and Guilsou, E.W. (2024). Thresholds to Sustainability: Globalization's Impact on The institutional Quality-Ecological Footprint Nexus in Africa. *Science of The Total Environment*, 954, 176205. doi: <https://doi.org/10.1016/j.scitotenv.2024.176205>
- Krueger, P., Sautner, Z. and Starks, L.T. (2020). The Importance of Climate Change For investors: Evidence From The institutional investment Community. *The Review of Financial Studies*, 33(3), 1067–1111. doi: <https://doi.org/10.1093/rfs/hhz075>
- Li, C. and Huang, M. (2023). Environmental Sustainability in The Age of Big Data: Opportunities and Challenges For Business and industry. *Environmental Science and Pollution Research*, 30, 119001–119015. doi: <https://doi.org/10.1007/s11356-023-30301-5>
- Li, Y., Liu, D. and Yuan, X. (2023). Influence Between Traditional Energy Markets, Metal Markets, and Clean Energy Sub-Sectors. *Energy Economics*, 120, 106596. doi: <https://doi.org/10.1016/j.eneco.2023.106596>

- Lins, K.V., Volpin, F.F. and Wagner, H.F. (2017). Social Capital, Trust, and Firm Performance: The Value of Corporate Social Responsibility During The Financial Crisis. *The Journal of Finance*, 72(4), 1785–1829. doi: <https://doi.org/10.1111/jofi.12505>
- Lopez de Prado, M. (2020). *Advances in Financial Machine Learning*. Wiley.
- Lopez-Lira, A. and Poliquin, R. (2023). The Social-Capital-Adjusted Firm Value of online Chatter. *Journal of Financial Economics*, 147(3), 584–606. doi: <https://doi.org/10.1016/j.jfineco.2022.09.006>
- Lu, L., Yang, S. and Li, Q. (2024). The interaction of Digital Economy, Artificial intelligence and Sports industry Development—Based on China PVAR Analysis of Provincial Panel Data. *Heliyon*, 10, e25688. doi: <https://doi.org/10.1016/j.heliyon.2024.e25688>
- Lundberg, S.M. and Lee, S.I. (2017). A Unified Approach To interpreting Model Predictions. in *Advances in Neural information Processing Systems*, 30, 4765-4774. Curran Associates, inc.
- Nelson, D.B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, 59(2), 347–370. doi: <https://doi.org/10.2307/2938260>
- Oloruntobi, O., Gohari, A., Mokhtar, S., Mokhtar, K. and Ruslan, S.M.M. (2025). Comparative Analysis of Long-Term Returns, Financial Considerations, and Measurement Challenges in Future ESG investing. *Journal of Asset Management*, 26, 271–297. doi: <https://doi.org/10.1057/s41260-025-00397-0>
- Onomakpo, H.E.O. (2024a). ESG Risk Ratings and Stock Performance in Electric Vehicle Manufacturing: A Panel Regression Analysis Using the Fama-French Five-Factor Model. *Journal of investment and Banking*, 3(1), 12–25. doi: <https://doi.org/10.33140/JIBF.03.01.12>
- Onomakpo, H.E.O. (2024b). Valuing Sustainability: An Empirical Analysis of Esg Risk and Financial Performance in The Global Transportation Sector. *Current Trends in Business Management*, 3(2), 01-06. doi: <https://doi.org/10.33140/CTBM.03.02.01>
- Pastor, L., Stambaugh, R.F. and Taylor, L.A. (2021). Sustainable investing in Equilibrium. *Journal of Financial Economics*, 142(2), 550–571. doi: <https://doi.org/10.1016/j.jfineco.2021.06.005>
- Petersen, M.A. (2009). Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches. *The Review of Financial Studies*, 22(1), 435–480. doi: <https://doi.org/10.1093/rfs/hhm072>
- Ren, J., Lai, K.-H., Shan, S. and You, J. (2024). Is More Not Always Better? Untangling The Impact of ESG Disclosure Transparency on Idiosyncratic Risk. *Journal of Environmental Management*, 373, 123648. doi: <https://doi.org/10.1016/j.jenvman.2024.123648>
- Saba, Z. (2025). ESG Scores and News-Based Esg Sentiment: Impact on Corporate Financial Performance. *Journal of Sustainable Finance & Accounting*, 7, 100019. doi: <https://doi.org/10.1016/j.josfa.2025.100019>
- Sharma, L., Kaushik, N., Maggon, M. and Sivathanu, B. (2025). Cognitive and Cultural Factors in Metaverse Acceptance: A Tam-Based Meta-Analysis. *Acta Psychologica*, 254, 104855. doi: <https://doi.org/10.1016/j.actpsy.2025.104855>
- Shiller, R.J. (2003). From Efficient Markets Theory To Behavioral Finance. *Journal of Economic Perspectives*, 17(1), 83–104. doi: <https://doi.org/10.1257/089533003321164967>
- Siegel, K. and Dee, L.E. (2025). Foundations and Future Directions For Causal inference in Ecological Research. *Ecology Letters*, 28, e70053. doi: <https://doi.org/10.1111/ele.70053>
- Song, X., Tian, Z., Ding, C., Liu, C., Wang, W., Zhao, R. and Xing, Y. (2022). Digital Economy, Environmental Regulation, and Ecological Well-Being Performance: A Provincial Panel Data Analysis from China. *International Journal of Environmental Research and Public Health*, 19(18), 11801. doi: <https://doi.org/10.3390/ijerph191811801>
- Tao, M., Lin, B. and Poletti, S. (2024). From Policy To Practice: How China’s Emissions Trading Scheme Shapes Esg Greenwashing At The Firm Level? *Journal of Environmental Management*, 370, 122681. doi: <https://doi.org/10.1016/j.jenvman.2024.122681>
- Wang, D. and Xia, X. (2024). The Impact of Digital Transformation on Firms’ Value: Examining The Role of Esg Performance and The Effect of information interaction. *Business Process Management Journal*, 30(5), 1236–1265. doi: <https://doi.org/10.1108/bpmj-06-2023-0469>
- Wang, H., Jiao, S., Ge, C. and Sun, G. (2024). Corporate ESG Rating Divergence and Excess Stock Returns. *Energy Economics*, 129, 107276. doi: <https://doi.org/10.1016/j.eneco.2023.107276>

- Wu, X., Li, L., Liu, D. and Li, Q. (2024). Technology Empowerment: Digital Transformation and Enterprise Esg Performance—Evidence From China’s Manufacturing Sector. *PLoS ONE*, 19, e0302029. doi: <https://doi.org/10.1371/journal.pone.0302029>
- Zeng, Q., Xu, Y., Hao, M. and Gao, M. (2025). ESG Rating Disagreement, Volatility, and Stock Returns. *Finance Research Letters*, 72, 106602. doi: <https://doi.org/10.1016/j.frl.2024.106602>
- Zhang, R., Fu, W. and Kuang, Y. (2022). Can Digital Economy Promote Energy Conservation and Emission Reduction in Heavily Polluting Enterprises? Empirical Evidence from China. *International Journal of Environmental Research and Public Health*, 19(16), 9812. doi: <https://doi.org/10.3390/ijerph19169812>
- Zhu, S., Lv, K. and Zhao, Y. (2025). Trust (in) Congruence, Digital Technological innovation, and Firms’ ESG Performance: A Polynomial Regression With Response Surface Analysis. *Journal of Environmental Management*, 373, 123689. doi: <https://doi.org/10.1016/j.jenvman.2024.123689>

### Annexure

<b>Abbreviations</b>	
ESG	Environmental, Social, and Governance
MSCI	Morgan Stanley Capital International
SEC	U.S. Securities and Exchange Commission
S&P 500	Standard & Poor's 500 Index
ADF	Augmented Dickey-Fuller (Test)
ARCH	Autoregressive Conditional Heteroskedasticity
CMA	Conservative Minus Aggressive (Investment Factor)
EGARCH	Exponential Generalized Autoregressive Conditional Heteroskedasticity
FE	Fixed Effects
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
GJR-GARCH	Glosten-Jagannathan-Runkle GARCH
HML	High Minus Low (Value Factor)
LM	Lagrange Multiplier
Mkt-RF	Market Risk Premium (Market Return minus Risk-Free Rate)
OLS	Ordinary Least Squares
PCA	Principal Component Analysis
RE	Random Effects
RMW	Robust Minus Weak (Profitability Factor)
SMB	Small Minus Big (Size Factor)
VAR	Vector Autoregression
VIF	Variance Inflation Factor
WML	Winners Minus Losers (Momentum Factor)
AUC	Area Under the (ROC) Curve
DNN	Deep Neural Network
ROC	Receiver Operating Characteristic
SHAP	SHapley Additive exPlanations
WFCV	Walk-Forward Cross-Validation
XAI	Explainable Artificial Intelligence
XGBoost	Extreme Gradient Boosting

**Cite this article as:** Henry Efe Onomakpo Onomakpo (2025). [The Illusion of Predictability: Deconstructing the ESG-Alpha Link with Robust Econometrics and Validated Machine Learning](#). *International Journal of Management Research and Economics*, 5(2), 35-54. doi: 10.51483/IJMRE.5.2.2025.35-54.