HELLENIC
OPEN
UNIVERSITY

School of Social Sciences

Master in Business Administration

Postgraduate Dissertation

Applications of Artificial Intelligence in Sustainable Finance

Christos K. Michail

Supervisor: Prof. Dimitrios Tzelepis

Patras, Greece, May 2026

HELLENIC
OPEN
UNIVERSITY

Applications of Artificial Intelligence in Sustainable Finance

Christos K. Michail

Supervising Committee

Supervisor:	Co-Supervisor:
Dimitrios Tzelepis	Dimitrios Georgoutsos
Professor	Professor

Patras, Greece, May, 2026

To my parents, Kostas and Athanasia,
for their endless love, support, and encouragement

Abstract

Sustainable finance is an important solution to many major environmental and social issues. This thesis examines how artificial intelligence (AI) can turn environmental, social, and governance (ESG) data into actionable insights for sustainable finance. The study seeks to find answers to two research questions. Firstly, it examines the hypothesis whether lagged ESG indicators improve corporate profitability prediction accuracy beyond traditional financial models. Second, it investigates whether ESG information improves the identification of financially distressed firms beyond the Altman Z-score, and whether this added value differs across region.

The thesis makes use of publicly available ESG data of listed companies of Asia, Europe and USA, from 2017 to 2024. The study compares the performance of conventional econometric models with machine-learning (ML) and deep-learning (DL) models. This research utilized SHAP-based explainability to find out which ESG and financial variables are more important in prediction.

The results show that ESG data do add value in both profitability forecasting and financial distress detection while results also show that ML and DL models are more effective at identifying complex patterns than traditional linear models. In general, the thesis demonstrates that AI-enabled ESG analytics enhance financial analysis, support risk management and sustainable finance decisions.

Keywords

Artificial Intelligence (AI); Environmental, Social and Governance (ESG); Sustainable Finance; Machine Learning; Financial Distress Prediction; Profitability Forecasting

Εφαρμογές της Τεχνητής Νοημοσύνης στη Βιώσιμη Χρηματοοικονομική

Χρήστος Κ. Μιχαήλ

Περίληψη

Η βιώσιμη χρηματοοικονομική αποτελεί σημαντική λύση για πολλά μείζονα περιβαλλοντικά και κοινωνικά ζητήματα. Η παρούσα διατριβή εξετάζει τον τρόπο με τον οποίο η τεχνητή νοημοσύνη (AI) μπορεί να μετατρέψει τα δεδομένα περιβαλλοντικών, κοινωνικών και διακυβερνητικών δεικτών (ESG) σε αξιοποιήσιμες πληροφορίες για τη βιώσιμη χρηματοοικονομική. Η μελέτη επιδιώκει να απαντήσει σε δύο ερευνητικά ερωτήματα. Πρώτον, εξετάζει την υπόθεση ότι οι δείκτες ESG με χρονική υστέρηση βελτιώνουν την ακρίβεια πρόβλεψης της εταιρικής κερδοφορίας σε σχέση με τα παραδοσιακά χρηματοοικονομικά μοντέλα. Δεύτερον, διερευνά κατά πόσον οι δείκτες ESG συμβάλλουν καλύτερα στον εντοπισμό εταιρειών που αντιμετωπίζουν οικονομικές δυσχέρειες σε σχέση με τον δείκτη Altman Z-score, καθώς και αν αυτή η προστιθέμενη αξία διαφέρει από περιοχή σε περιοχή.

Η διατριβή χρησιμοποιεί δημόσια διαθέσιμα δεδομένα ESG εισηγμένων εταιρειών της Ασίας, της Ευρώπης και των ΗΠΑ, από το 2017 έως το 2024. Η μελέτη συγκρίνει την απόδοση των συμβατικών οικονομετρικών μοντέλων με μοντέλα μηχανικής μάθησης (ML) και βαθιάς μάθησης (DL). Η έρευνα αυτή χρησιμοποίησε την εξηγήσιμη ανάλυση με βάση το SHAP για να διαπιστώσει ποιες μεταβλητές ESG και χρηματοοικονομικές μεταβλητές είναι πιο σημαντικές στην πρόβλεψη.

Τα αποτελέσματα δείχνουν ότι τα δεδομένα ESG προσθέτουν όντως αξία τόσο στην πρόβλεψη της κερδοφορίας όσο και στον εντοπισμό χρηματοοικονομικών δυσχερειών, ενώ τα αποτελέσματα δείχνουν επίσης ότι τα μοντέλα ML και DL είναι πιο αποτελεσματικά στον εντοπισμό σύνθετων μοτίβων από τα παραδοσιακά γραμμικά μοντέλα. Γενικά, η διατριβή αποδεικνύει ότι η ανάλυση ESG με τη χρήση τεχνητής νοημοσύνης βελτιώνει τη χρηματοοικονομική ανάλυση, υποστηρίζει τη διαχείριση κινδύνων και τις αποφάσεις για τη βιώσιμη χρηματοοικονομική.

Λέξεις – Κλειδιά

Τεχνητή Νοημοσύνη (AI); Περιβαλλοντικά, Κοινωνικά και Θέματα Διακυβέρνησης (ESG); Βιώσιμη Χρηματοοικονομική; Μηχανική Μάθηση; Πρόβλεψη Χρηματοοικονομικής Δυσχέρειας; Πρόβλεψη Κερδοφορίας

Table of Contents

Abstract ..... **v**

Περίληψη ..... **vii**

Table of Contents ..... **ix**

List of Figures ..... **xiii**

List of Tables ..... **xv**

List of Abbreviations & Acronyms ..... **xvii**

Chapter 1. Introduction **1**

- 1.1 Background and Context 1
  - 1.1.1 The Rise of Sustainable Finance 3
  - 1.1.2 The Rise of AI in Finance and ESG 5
- 1.2 Research Problem Statement 6
- 1.3 Research Objectives 7
  - 1.3.1 General Objective 7
  - 1.3.2 Specific Objectives 8
- 1.4 Scope and Limitations 9
- 1.5 Dissertation Structure 9

Chapter 2. Literature Review **12**

- 2.1 Conceptual Foundations 12
  - 2.1.1 Artificial Intelligence: Definitions and Key Concepts 12
  - 2.1.2 Sustainable Finance 14
  - 2.1.3 Sustainability and ESG: Definitions and Frameworks 18
- 2.2 ESG Performance and Firm Value 19
  - 2.2.1 Theoretical Foundations 19
  - 2.2.2 Empirical Evidence on ESG-Performance Relationships 19
  - 2.2.3 ESG Measurement Challenges 20

2.3 AI in Sustainable Finance 27
2.3.1 Historical Evolution of AI in Finance 27
2.3.2 Historical Evolution of AI in Sustainable Finance 29
2.3.3 AI for ESG Data, Ratings and Disclosures 31
2.3.4 Limitations of AI in Traditional Finance 35
2.3.5 AI for Green Investment and Portfolio Construction 39
2.3.6 AI for Climate and Sustainability Risk Assessment 40
2.3.7 Governance, Ethics and Regulation of AI in Sustainable Finance 43
2.4 Bibliometric Trends 44
Chapter 3. Data & Descriptive Analysis 46
3.1 Dataset Provenance 46
3.2 Variable Definitions 46
3.3 Descriptive Statistics 47
3.4 ESG Score Distribution 48
3.5 Financial Variables Distribution 54
3.6 Correlation Structure 56
Chapter 4. Research Methodology 60
4.1 Research Design 61
4.1.1 Epistemological Positioning 61
4.1.2 Mixed-Methods Analytical Framework 61
4.1.3 Hypothesis Testing Protocol 63
4.2 Data Architecture 64
4.2.1 Data Source and Panel Structure 64
4.2.2 Variable Taxonomy 65
4.2.3 Lagging Convention 65
4.2.4 Preprocessing Pipeline 66
4.2.5 Anti-Leakage Rules 67

4.3 Model Families 67
4.3.1 Tier I — Classical Statistical Models 68
4.3.2 Tier II — Machine-Learning Models 69
4.3.3 Tier III — Deep-Learning Architectures 70
4.3.4 Model Interpretation 73
4.4 Data Splitting 74
4.4.1 Firm-Level Random Split 74
4.4.2 Temporal Train–Test Split 74
4.5 Acceptance Criteria Framework 75
4.6 Decision Logic 76
Chapter 5. Research Hypotheses 77
5.1 Hypothesis H1 — ESG Score as a Predictor of Financial Performance 77
5.2 Hypothesis H2 —ESG-Adjusted Credit Scoring 79
Chapter 6. Results 81
6.1 Hypothesis H1 81
6.1.1 Sub-hypothesis H1a 81
6.1.2 Sub-hypothesis H1b 85
6.2 Hypothesis H2 89
6.2.1 Sub-hypothesis H2a 89
6.2.2 Sub-hypothesis H2b 92
Chapter 7. Discussion 93
Chapter 8. Conclusion, Limitations and Future Work 97
8.1 Conclusion and Contributions 97
8.2 Limitations 98
8.3 Future Work 99
References 102
Appendix 123

Appendix A - Hyperparameters of Regression Models on the H1 123
  Panel OLS – Baseline Econometric Model 123
  XGBoost Regression 123
  GRU – Regression 124
Appendix B - Hyperparameters of Classifications Models on the H2 125
  Logistic regression 125
  HistGBM 125
  Random Forest 126
  XGBoost 126
  LightGBM 126
  LSTM 127
  GRU – Classifier 127

List of Figures

Figure 2-1. Annual publication output on ESG research..45
Figure 2-2. Annual publication output on the intersection of Artificial Intelligence and ESG 45
Figure 3-1. Dataset Composition — Top-15 Sectors and Top 20 Countries by Unique Firm Count 48
Figure 3-2. ESG Composite Score Distribution by Region — Asia, US, EU. 49
Figure 3-3. ESG Pillar Score Trends by Region, 2017–2024......................................52
Figure 3-4. ESG Score Distribution by Top 10 Sectors and Region..........................54
Figure 3-5. Median Return on Assets (ROA) and Return on Equity (ROE) by Region over Time, 2017–2024...56
Figure 3-6. Pearson Correlation Matrix — ESG and Financial Variables (Lower-Triangular Heatmap) 58
Figure 3-7. Altman Z-Score Distribution and Annual Distress Rate 59
Figure 6-1. Hypothesis H1 – Models Accuracies. 82
Figure 6-2. Hypothesis H1 – Statistical Significance 83
Figure 6-3. Hypothesis H1a - XGBoost Feature Importance (Gain): Top 20 Predictors of ROA and ROE. 84
Figure 6-4. Sub-hypothesis H1b – Pillar-Specific Ablation Test R^2: Independent Predictive Contribution of ESG Dimensions..86
Figure 6-5. Sub-hypothesis H1b - SHAP Pillar Attribution: Environmental Dominance in Profitability Prediction 87
Figure 6-6. Sub-hypothesis H1b - SHAP Beeswarm: Directionality and Non-linearity of ESG-Performance Relationship...88
Figure 6-7. Sub-hypothesis H2a – Accuracy of the models......................................90
Figure 6-8. Sub-hypothesis H2a – SHAP feature importance...................................91

Figure 6-9. Bootstrap AUC Difference Test 91

Figure 6-10. Regional ESG benefit........ 92

List of Tables

Table 2-1. Principal AI subfields, their underlying logic, representative methods, and illustrative applications in sustainability and ESG contexts. Sources: (Ahmad et al., 2021; Boyd & Vandenberghe, 2013; Breiman, 2001; Calamai et al., 2025; Dekker et al., 2012; Devlin et al., 2018; Hastie et al., 2009; Krizhevsky et al., 2017; LeCun et al., 2015; Rolnick et al., 2023; Russell & Norvig, 2021; Taheripour et al., 2025; Vaswani et al., 2017)....14

Table 2-2. Industry Approaches & Weighting. Sources: (Agosto & Tanda, 2025; Avramov et al., 2022; Berg et al., 2022; LSEG, 2022, 2024; MSCI Inc, 2026; Muck & Schmidl, 2024; Münchhausen et al., 2024; S&P Global, 2023; Sustainalytics, 2026)21

Table 2-3. ESG Disclosure Metrics Matrix (Pillars, Categories, Standards, Frequency). Sources: (European Commission, 2023a; GRI, 2021; SEC, 2024) 23

Table 2-4. Predicting ESG scores/ratings 33

Table 2-5. Predicting ESG controversies, greenwashing, “exaggeration”, and external-risk signals 34

Table 3-1. Variables of the Dataset....46

Table 4-1. Three-Tier Analytical Pipeline for Hypothesis Evaluation.....62

Table 4-2. Variable Taxonomy and Functional Classification 65

Table 4-3. Anti-Leakage Rules Applied Across All Hypothesis....67

Table 4-4. Traditional Econometric Evaluation Criteria....75

Table 4-5. Machine-Learning Performance Evaluation Criteria....76

Table 5-1. Formal Sub-Hypotheses and Acceptance Criteria for H1....78

Table 5-2. Hypothesis H2 - Model Evaluation Framework and Hypothesis Acceptance Criteria.....80

Table 6-1. Hypothesis H1a Acceptance Summary. 84

Table 6-2. H1b Hypothesis Test: SHAP Pillar Attribution Summary....89

Table 6-3. Hypothesis H2 Decisions: ESG-Adjusted Credit Scoring 92

List of Abbreviations & Acronyms

Abbreviation	**Definition**
AI	Artificial Intelligence
AI4ESG	Artificial Intelligence for ESG (research/practice area)
Altman, Z.	Altman Z-Score (bankruptcy/distress risk indicator; Z″ = modified variant)
AR	Auto-Regressive (process / features, e.g., lagged ROA/ROE)
AUC	Area Under the Curve
AUC-ROC	Area Under the Receiver Operating Characteristic Curve
BERT	Bidirectional Encoder Representations from Transformers
(CI)	Confidence Interval
CNN	Convolutional Neural Network
COP26	26th UN Climate Change Conference of the Parties (Glasgow, 2021)
CPCV	Combinatorial Purged Cross-Validation
CSR	Corporate Social Responsibility
CSDD	Corporate Sustainability Due Diligence Directive (EU)
CSRD	Corporate Sustainability Reporting Directive (EU)
CSRS	Corporate Sustainability Reporting Standards
CV	Cross-Validation
DeLong test	Statistical test for comparing AUCs
DL	Deep Learning

Abbreviation	**Definition**
ECOA	Equal Credit Opportunity Act
EBIT	Earnings Before Interest and Taxes
EDGAR	Electronic Data Gathering, Analysis, and Retrieval (SEC filings system)
E (pillar)	Environmental (ESG pillar)
ESMA	European Securities and Markets Authority
ESG	Environmental, Social, and Governance
ESRS	European Sustainability Reporting Standards
EU	European Union
EuGBS	European Green Bond Standard
Fama–MacBeth	Two-step regression methodology (asset pricing / panel inference)
FE	Fixed Effects
FSB	Financial Stability Board
FSI	Financial Stability Institute
GBM	Gradient Boosting Machine
GICS	Global Industry Classification Standard
GDPR	General Data Protection Regulation
GOSS	Gradient-based One-Side Sampling (LightGBM)
GRI	Global Reporting Initiative
GRU	Gated Recurrent Unit
G (pillar)	Governance (ESG pillar)
HistGBM	Histogram-Based Gradient Boosting Machine

Abbreviation	**Definition**
H1 / H2	Hypothesis 1 / Hypothesis 2
IMF	International Monetary Fund
IQR	Interquartile Range
ISSB	International Sustainability Standards Board
KLD	KLD ESG dataset (provider/dataset referenced in literature)
L1 / L2	Lag-1 / Lag-2 (one-period / two-period lag)
LASSO	Least Absolute Shrinkage and Selection Operator
LightGBM	Light Gradient Boosting Machine
LIME	Local Interpretable Model-agnostic Explanations
LSEG	London Stock Exchange Group
LSTM	Long Short-Term Memory network
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MSCI	Morgan Stanley Capital International
NFRD	Non-Financial Reporting Directive (EU, predecessor to CSRD)
NGO	Non-Governmental Organization
NLP	Natural Language Processing
OECD	Organization for Economic Co-operation and Development
OLS	Ordinary Least Squares
PCA	Principal Component Analysis
PRI	Principles for Responsible Investment
RE	Random Effects

Abbreviation	Definition
RF	Random Forest
RIC	Reuters Instrument Code
R^2	R-squared (coefficient of determination)

Chapter 1. Introduction

1.1 Background and Context

The global financial market is in a transition to a new model due to the increasing exposure of shareholder value orientation as it relates to environmental degradation, social inequality, and poor governance (Dyllick & Muff, 2016; Eccles & Klimenko, 2020). As a result of this exposure, the need for sustainable financing has grown, which incorporates ESG criteria into investment and lending activities to provide both an opportunity to manage long-term risks and to support sustainable development opportunities (Friede et al., 2015). At the same time, the use of AI—specifically Machine Learning (ML), Deep Learning (DL), and Natural Language Processing (NLP) has dramatically changed how financial firms process data, price risk, and create portfolios using the volume and heterogeneity of data at speeds and scales previously unknown (Bahoo et al., 2024).

This convergence will be further emphasized in the context of the European Green Deal, with an aim to make the EU economy climate-neutral by 2050 and supported by a large number of regulations such as the EU Taxonomy Regulation, the Sustainable Finance Disclosure Regulation (SFDR), and the Corporate Sustainability Reporting Directive (CSRD) (European Commission, 2019a; Hummel & Jobst, 2024). The new rules impose an obligation on financial market players and companies to report increasingly large amounts of data on their sustainability activities. Thus, there is a growing need for sophisticated tools to extract ESG insights related to sustainable development from large volumes of unstructured data (Xu, 2024).

Traditional ESG analytics rely on several methods, including but not limited to manual scoring, company survey responses, and periodic corporate reports. As a result, traditional ESG analytics are subject to data fragmentation, methodological heterogeneity, and timeliness issues (Agosto & Tanda, 2025; Xu, 2024). Research has documented a weak correlation between many of the larger ESG rating providers. The phenomenon is known as “aggregate confusion.” It makes it difficult for investors to have confidence in their investments and to build a sustainable investment portfolio. AI-based approaches may provide solutions to some of the issues associated with traditional ESG analytics, according to:

- Aggregating multiple types of ESG data (e.g., financial reports, sustainability reports, news articles, NGO databases, satellite images, social media posts) (Xu, 2024)
- Determining nonlinear associations between ESG factors and financial performance (Krappel et al., 2021).
- Enabling timely (near real-time) identification of ESG risk, controversy, or positive impact (Calamai et al., 2025).

Although recent studies have shown that ML, ensemble techniques, and NLP are being used to improve the reliability of ESG scores and forecast climate and carbon risks as well as optimize portfolios for alignment with a low-carbon future (Krappel et al., 2021; Schimanski et al., 2024). The literature is lacking in terms of how these technologies can be practically integrated into ESG assessment and decision-making processes in ways that support long-term sustainability outcomes while also supporting regulatory requirements and ethical considerations (Berniak-Woźny, 2025). This thesis will focus on the operational integration of AI technologies into ESG assessment and decision-making frameworks to achieve both financial resilience and sustainability outcomes.

1.1.1 The Rise of Sustainable Finance

The development of sustainable finance is linked to several phases, including early negative screening and socially responsible investment, ESG integration, and, most recently, climate finance and impact investment (Billio et al., 2024; Dimmelmeier, 2024).

Empirical research shows that companies with effective ESG management exhibit higher-quality cash flows, lower capital costs, greater resilience to crises, and therefore better overall financial results (Postiglione et al., 2024; Tanjung, 2023); however, it is very difficult to assess the magnitude and pathways of these relationships across ESG pillars and by country or industry (Gregory, 2022). Several empirical studies using dynamic panel data and meta-analysis have documented that green funding, climate-related initiatives to decrease emissions, and sustainable product innovations are positively associated with both firm profitability and return on equity (Li & Lin, 2024; Mugova et al., 2025). However, some large-scale and high-capital-intensive environmental projects may temporarily negatively affect an entity's short-term financial performance (Rahko, 2023).

Regulatory agencies around the globe have also been critical of advancing the sustainable finance agenda. In the EU, the European Green Deal has served as the foundation for the sustainable finance agenda, supported by a comprehensive regulatory framework:

- EU Environmental Taxonomy Regulation (EU 2020/852): Provides a standard definition of economic sustainability and what constitutes an economically sustainable activity.
- Sustainable Finance Disclosure Regulation (SFDR): The SFDR provides a mechanism to ensure the ESG characteristics of financial products and portfolios are clearly disclosed in detail to prevent 'greenwashing' and increase

the ability of comparable financial products and portfolios compared against each other.

- The Corporate Sustainability Reporting Directive (CSR Directive) and the Corporate Sustainability Reporting Standards (CSRS) broaden the scope of mandatory corporate sustainability reporting to include European Union-listed public interest entities. They also require companies to report on the environmental and social effects of their business activities from a double materiality perspective, meaning they must report on how their activities affect the environment and society, as well as how environmental and social issues affect their business model.
- Corporate Sustainability Due Diligence Directive (CSDD): Compels large businesses to conduct environmental and human rights due diligence throughout their supply chains.

The regulation described above presents both opportunities and challenges for the investor's inclusion of ESG in investment decision-making (Hummel & Jobst, 2024). The primary opportunity presented by these regulations is the availability of a vast amount of information, which will allow investors to determine if their investments are compatible with sustainable business practices. However, these regulations may provide complex and multi-dimensional information that exceeds the processing capabilities of standard analytical methods (Liu, 2022). Therefore, there are current limitations regarding the reliability of existing ESG rating methodologies (Berg et al., 2022; Liu, 2022; Zhao et al., 2025). One of the most important drivers for increasing the depth and breadth of sustainable finance is the potential to apply AI in this area.

1.1.2 The Rise of AI in Finance and ESG

AI is becoming a fundamental component of financial innovation, affecting our evaluation of creditworthiness, the creation of algorithmic trading strategies, the detection of fraudulent transactions, the use of robo-advisory services, and the management (Roy et al., 2025; Vuković et al., 2025). There are many areas where AI will be used in sustainable finance, including ESG data collection and cleanup; ESG rating systems and controversy-detection systems; climate and carbon risk modeling systems; green bond and transition risk analytics systems; and AI-driven portfolio optimization systems that consider ESG constraints (Alqudah et al., 2025).

The research at the intersection of AI and sustainable finance has seen exponential growth since 2015, and there are many different research groups focused on the following topics: AI-based ESG performance assessment; AI-enabled green FinTech and blockchain systems; AI-enhanced risk detection and resilience; theoretical foundations of AI in sustainable finance; ethical and governance frameworks for responsible AI. Studies have shown that (Berniak-Woźny, 2025; Di Pietro et al., 2026):

- Multimodal machine learning models such as XGBoost, LightGBM, and heterogeneous ensembles have been able to explain better the relationship between publicly available financial and fundamental data and ESG ratings and carbon risk than linear benchmark models (Krappel et al., 2021).
- Natural Language Processing Systems, which are technologies that enable computers to understand and interpret human language, have improved the coverage and timeliness of ESG controversy monitoring and sentiment analysis for issuers (Calamai et al., 2025; Chen et al., 2023).
- AI-driven ESG analytics have improved investors' ability to optimize and construct profitable portfolios that meet sustainability requirements (Taheripour et al., 2025).

However, AI also creates new challenges. These include the development of "black box" models, which are unexplainable by humans; algorithmic bias toward underrepresented regions or sectors; data privacy issues; and the potential to inadvertently reinforce greenwashing by utilizing low-quality or biased ESG disclosures in AI model training (Chen et al., 2023).

Recent industrial surveys and editorial calls in AI4ESG have emphasized the need for trustworthy, understandable, and governance-aligned AI architectures for financial institutions (Berniak-Woźny, 2025; Chen et al., 2023). These results indicate that while AI may be capable of improving ESG data quality, coverage, and predictive insights, the integration of AI into sustainable finance must be carefully designed and governed to avoid amplifying existing systemic risks or ethical concerns.

1.2 Research Problem Statement

Although there have been many new ESG disclosure requirements introduced under various frameworks including EU Taxonomy, SFDR, and CSRD, the biggest issue still remains; ESG data is still disorganized, not consistent across different ESG providers, and only marginally correlated to companies' true financial performance and sustainability risks. At the heart of this problem is the methodology used to create ESG Scores. Commercial providers generally use static methodologies that utilize a large number of rules to aggregate thousands of indicators into a single score — a process that creates poor inter-rater reliability and does not have enough explanatory power to predict long-term financial results. In literature, this is often referred to as the aggregate confusion problem.

In addition to the methodology used to create ESG Scores, Artificial Intelligence — specifically ML and DL — offer a fundamentally different method to analyze ESG data. Unlike linear models, AI can recognize non-linear relationships in the data, handle large amounts of data (high-dimensionality), and generate individual company forecasts that have out-of-sample validity. However, three significant questions still exist in the literature:

- What types of AI architectures are best suited to analyze ESG-Panel Data?
- Will AI generate statistically and economically meaningful improvements over conventional ESG Scoring and Panel Regression Methods?
- Are the outputs of AI models capable of being made transparent and in compliance with EU Sustainable Finance Disclosure Requirements?

This dissertation will address each of these issues directly. The central research question is:

To what extent can AI models train on actual ESG and Financial Data predict corporate financial performance and credit risk — and to what extent can AI models do so in an open, transparent manner that complies with EU Regulatory Requirements?

1.3 Research Objectives

The overall purpose of this dissertation is to empirically examine whether AI models can obtain financially relevant insights from ESG Data and to develop a leakage-free, explainable analytical system that is suitable for use in the European Union's Sustainable Finance Regulation Environment.

1.3.1 General Objective

The goal is to develop both theoretical and practical frameworks for applying machine learning/deep learning algorithms in sustainable finance, specifically to evaluate how these models can improve the processing of ESG data, predict risks, and support investment decisions in compliance with EU and global sustainability regulations.

1.3.2 Specific Objectives

The following four objectives operationalize the general objective above.

- Objective 1 — Data Construction: Create a multi-regional ESG-Finance Panel Data Set of publicly traded companies across Asia, the U.S., and Europe by combining ESG scores, financial ratios, and financial distress indicators in a format that adheres to EU regulatory standards (EU Taxonomy, SFDR, CSRD/ESRS).
- Objective 2 — Predictive Power of Profits: Determine if lagged aggregate and pillar level ESG scores (Environmental, Social, and Governance) contribute incrementally to the ability to predict corporate profitability (Return on Assets (ROA), Return on Equity (ROE)) relative to traditional panel regression baselines through a sequence of ensemble machine learning and deep learning models that operate under severe anti-leakage constraints.
- Objective 3 — Assessing Credit Risk: Develop and test ESG adjusted financial distress models based on the Altman Z-Score augmented with ESG factors and establish whether AI-based classifiers increase the accuracy of identifying distressed companies — and whether the degree of improvement varies among regions (Asia, U.S., EU) due to differing levels of ESG disclosure maturity.
- Objective 4 — Explainability and Policy Relevance: Use Explainable AI Techniques (SHAP) to interpret model output, determine which ESG Pillars influence financial predictions, and relate these results to the existing policy debate surrounding greenwashing, ESG Rating Divergence, and the EU’s Transparency Requirements for the use of AI in Financial Services.

1.4 Scope and Limitations

This study is limited to a large multi-year panel of publicly traded companies across the globe, for which both ESG scores and financial statements are standardized. The time frame over which this study was conducted includes the largest increase in use of AI for ESG-related applications, as well as the EU's sustainable finance regulations (EU Taxonomy, SFDR, CSRD). Due to these restrictions, the study's analysis is limited to two types of financial measures of performance: profitability (ROA and ROE) and financial distress due to poor credit, sustainability measure and ESG information provided by one commercial vendor. To complete the study, I used engineered interaction terms for all variables but did not model any broader social or environmental impacts, such as greenhouse gas emissions or biodiversity outcomes. Some key limitations include (1) well-documented issues with ESG rating variability, imputation bias, and disclosure variability; (2) the fact that the analysis uses observational (non-experimental) data, so it cannot establish a clear cause-and-effect relationship; (3) the possibility that there may be biases inherent in the AI models used, particularly in complex AI models; and (4) limitations regarding the generalizability of the results to other organizations, time frames, and regulatory environments than those examined in the study.

1.5 Dissertation Structure

The dissertation is divided into eight chapters, which are designed to reflect the logical flow of the research as well as the pipeline of analysis.

Chapter 2 provides a position for the dissertation relative to both the current academic and policy debates regarding AI and sustainable finance. This chapter also includes a review of ESG frameworks, a discussion of the challenges associated with measuring ESG (rating divergence, etc.), an examination of the relationship between ESG and financial performance, and an overview of how AI can be employed in assessing ESG data, assessing risks, and in providing decision support to investors. The chapter concludes by identifying the research gaps and outlining the bibliometric trends that provided the basis for the development of the thesis' hypotheses and model choice.

Chapter 3 details the origins and the composition of the ESG-finance panel data set, defines all variables utilized in the analysis, and provides descriptive statistics for those variables. Additionally, Chapter 3 will report the distribution of ESG and financial variable values across regions, the correlation structure among those variables, and some stylized facts about the data that were observed and that may have influenced the design of the models.

Chapter 4 describes the full empirical approach taken in this dissertation and describes the safeguards that are being implemented to ensure the validity of the analyses. In addition to describing the study's research design and the three-tier analytical pipeline (the baseline econometric approaches, the "horse race" between machine learning/deep learning models, and the explainability of the selected models), the chapter will also provide a detailed description of the data architecture and preprocessing workflow, the rules and guidelines that will prevent leakage of model performance results to subsequent analyses (anti-leakage rules), the family of model types (including ensemble methods and deep learning), the protocols that will be followed when splitting the data sets (both at the firm level and over time), and the acceptance criteria that will be used to assess the consistency of the evaluation of hypotheses throughout the dissertation.

Chapter 5 formalizes the specific hypotheses and links them to the dissertation's objectives: (H1) does the inclusion of lagged ESG information improve predictions of profitability (ROA/ROE) compared to using only financial information, including a comparison of ESG pillars; and (H2) do ESG features improve predictions of credit/distress (beyond traditional benchmarks such as Altman Z) compared to using only traditional benchmarks, and if so, are there significant regional differences in terms of incremental predictive value?

Chapter 6 will document the empirical findings for each hypothesis and sub-hypothesis. The findings will be documented in a structured manner consistent with the tiered pipeline: (i) the econometric benchmarking evidence; (ii) the performance comparisons for out-of-sample ML/DL models; and (iii) the explainable-AI outputs (such as SHAP-based attribution) to identify which ESG components are important and how they relate to the fundamental characteristics of firms.

Chapter 7 discusses the implications of the results for the previous literature and the practical/regulatory environment. It will address questions regarding the implications of the findings for the use of ESG ratings, for model interpretability, for concerns about greenwashing, and for the operational integration of AI into sustainable-finance decision-making under evolving disclosure and governance expectations.

Chapter 8 summarizes the contributions of the dissertation and restates the implications of the findings for the use of AI to assess ESG and to develop models for predicting risk. It will acknowledge methodological and data limitations and will propose clearly defined directions for future work at the intersection of AI, ESG measurement, and sustainable finance practice.

Chapter 2. Literature Review

2.1 Conceptual Foundations

2.1.1 Artificial Intelligence: Definitions and Key Concepts

The AI discipline comprises the science and engineering of developing and building intelligent machines capable of replicating human thought processes and behavior through the ability to perceive, reason, learn, plan, and predict (Russell & Norvig, 2021). The definition of AI differs in both academic literature and organizational boundaries; for example, the Organization for Economic Co-operation and Development (OECD) describes an AI system as a machine-based system that infers from a set of inputs and generates an output such as predictions, content, recommendations, or decisions that can have an impact on the physical or virtual environment, and can operate at varying levels of autonomy (OECD, 2024). Likewise, the EU describes AI as "software" that can generate content, make predictions, suggest options, and make decisions, all of which can affect a given environment (European Commission, 2024b; Veale & Zuiderveen Borgesius, 2021). At its base level, artificial intelligence enables computers to accomplish tasks that previously required the use of human intelligence and judgment as well as the cognitive functions of decision-making, problem-solving, and learning (TURING, 1950).

Key concepts and major research areas within artificial intelligence include perception (comprising audio, visual, textual, and tactile interpretation), logical reasoning, data-driven learning, strategic planning, and problem-solving (LeCun et al., 2015; Russell & Norvig, 2021). In addition, important features are forecasting, extracting knowledge automatically, and identifying patterns and interactions.

These are achieved through various methods and technologies, such as search techniques, knowledge graphs, NLP, expert systems, evolutionary algorithms, machine learning, and deep learning (Goodfellow, 2016; Hogan et al., 2022). In addition to the technical definitions of AI, it is essential to establish an ethical framework for the development and deployment of AI technology. For example: privacy, security and resilience, responsibility, sustainability, transparency, equity, accountability, do no harm, robustness, and explainability (Floridi et al., 2018; Jobin et al., 2019).

The field of AI comprises a wide spectrum of subfields, each offering distinct capabilities for processing, analyzing, and acting on complex data. Table 2-1 summarizes the most relevant of these subfields for sustainable finance, mapping their core ideas, representative methods, and illustrative applications in ESG and sustainability.

Table 2-1. Principal AI subfields, their underlying logic, representative methods, and illustrative applications in sustainability and ESG contexts. Sources: (Ahmad et al., 2021; Boyd & Vandenberghe, 2013; Breiman, 2001; Calamai et al., 2025; Dekker et al., 2012; Devlin et al., 2018; Hastie et al., 2009; Krizhevsky et al., 2017; LeCun et al., 2015; Rolnick et al., 2023; Russell & Norvig, 2021; Taheripour et al., 2025; Vaswani et al., 2017).

AI Subfield	Core Concept	Common Methods/Tools	Sustainability Applications
Machine Learning	Learn patterns in data to predict or classify outcomes.	Regression, Random Forests, Gradient Boosting	Predict energy demand, default risk, equipment failures
Deep Learning	Multi-layer neural networks to identify complex patterns	CNNs, RNNs, Transformers	Analyzing satellite images for land use, reading large quantities of ESG reports
Natural Language Processing (NLP)	Analysis and generation of human language	Topic Modeling, Sentiment Analysis, Embeddings	Scoring disclosures of ESG practices, tracking climate change-related news and controversy
Computer Vision	Extraction of information from images and videos	Object Detection, Image Segmentation	Monitoring emissions, safety compliance, asset condition
Optimization	Finding the best solution under constraints	Linear/Non-linear Programming, Heuristics	Low-carbon routing, capacity planning under carbon caps
Expert Systems / Rules	Encoding expert knowledge into decision rules	Rule Engines, Knowledge Bases	Checking compliance with ESG policies, screening eligibility

2.1.2 Sustainable Finance

Sustainable finance can be defined as all the actions and decisions made by investors, lenders, etc. that take into account the ESG impacts that their investment decisions have to create a positive and sustainable economic outcome (European Commission, 2024a). In addition to these two definitions, sustainable finance manifests in financial instruments and practices in several ways, ranging from sustainable lending and investment screening to impact investing and sustainability-linked products (Agapova et al., 2025; GSI, 2023). The ESG profiles of companies can be broken down into three main categories: the environmental dimension of ESG will cover a wide variety of factors, but primarily issues relating to climate change, waste management, resource conservation, and biodiversity the social dimension of ESG will look at a wide variety of factors, but primarily will focus on issues related to human rights, consumer protection, employee health and safety, training, etc. and the governance dimension of ESG will assess a variety of issues including, but not limited to, Board independence, business ethics and anti-bribery plans (Friede et al., 2015; Gillan et al., 2021).

There has been significant growth in sustainable finance in the last few years. The total amount of sustainable debt issued globally was approximately $1.6 trillion in 2021, a 116 percent increase from 2020. Even though Europe remains the most active region for sustainable debt issuance, there is evidence that issuance rates are increasing in other regions of the world (BloombergNEF, 2022; IMF, 2022).

There are also institutional efforts to promote sustainable finance further. For example, the European Green Deal, which was launched in December 2019, outlines the EU’s goal of achieving net-zero carbon emissions by 2050 (European Commission, 2019a). This includes a target of reducing greenhouse gas emissions by at least 55 percent by 2030 compared to 1990 levels. As part of the European Green Deal, the EU has also committed to creating a new investment plan to mobilize at least $1 trillion in sustainable investments over the next 10 years, from both public and private sources (European Parliament, 2025).

Through various methods, sustainability is operationalized within current capital markets. Sustainability is measured using metrics, disclosures and decision rules based upon ESG information. Companies provide ESG-related data to the general public via sustainability reports and regulatory filings; this data is then analyzed and processed by ESG rating agencies and data providers into ESG scores and ratings. Some of the major ESG rating agencies/providers used by investors include Moody’s ESG Solutions (Vigeo Eiris), Refinitiv, Sustainalytics and MSCI, each of whom publish a documented methodology explaining how they develop scores/ratings (Agosto & Tanda, 2025; Berg et al., 2022; Dimson et al., 2020).

Scores/ratings are derived by aggregating numerous ESG indicators into a single number or letter grade that investors use to screen companies, construct indexes/products, and track portfolio exposure to ESG risks/returns (Berg et al., 2022; Gibson-Brandon et al., 2021).

Although there hasn't been a globally recognized ESG rating model that everyone has followed, there are differences among ESG rating companies in how they define, measure, and calculate aggregate ESG scores. These variations lead to differing levels of similarity in ESG ratings for the same company across various ESG rating firms. Berg et al. (2022) attribute ESG rating divergence to three sources: scope divergence (different attributes included), measurement divergence (different proxies for the same attribute), and weight divergence (different weighting of attributes). In addition, a study conducted by Dorfleitner et al. (2015) compares ESG data for over 8,500 global companies from ASSET4, Bloomberg, and KLD. They found that there was little alignment among the ESG ratings provided by the various suppliers, and therefore little comparability among them. Furthermore, studies have indicated that biases exist in the provision of ESG ratings/data regarding the size of the companies concerned (for example, in the type of data provided by Refinitiv-type or ASSET4-type databases). Larger companies can obtain better ESG ratings primarily because they provide more information and receive more comprehensive coverage than smaller companies (Dobrick et al., 2023; Drempetic et al., 2020).

In addition to providing ESG ratings, ESG providers create taxonomies that help define what constitutes economically sustainable activity in accordance with predefined criteria. A taxonomy created by the European Union, known as the "EU Taxonomy" (Regulation (EU) 2020/852), provides criteria for determining whether an economic activity is environmentally sustainable. The EU Taxonomy establishes six environmental objectives that must be met to qualify an economic activity as environmentally sustainable: (i) climate change mitigation; (ii) climate change adaptation; (iii) the sustainable use and protection of water and marine resources; (iv) transition to a circular economy; (v) pollution prevention and control; and (vi) the protection and restoration of biodiversity and ecosystems (European Commission, 2020, 2023b). These taxonomies guide both the development of new products and the regulatory reporting obligations associated with them and also provide a basis for establishing standards for labeled products. As a result of the EU's regulations regarding green bonds, the European Green Bond Standard (EuGBS) has been established through Regulation (EU) 2023/2631 as a voluntary framework within the European Union for labeling "green bonds". The EuGBS ties the concept of "green bonds" to expenditures aligned with the EU Taxonomy and includes both disclosure obligations and an external review process (European Commission, 2023b, 2025, 2025). With respect to disclosures, the Corporate Sustainability Reporting Directive (CSRD) (Directive (EU) 2022/2464) significantly extends existing EU requirements for companies' sustainability reporting (European Commission, 2022). According to the official press material issued by the European Parliament on CSRD, approximately 50,000 companies will be subject to this directive, as opposed to the approximately 11,700 companies currently subject to the existing regime (European Parliament, 2022). In contrast to CSRD, the Sustainable Finance Disclosure Regulation (SFDR) (Regulation (EU) 2019/2088) obligates financial market participants and financial advisors to make sustainability-related disclosures, and provides for differing disclosure obligations for products that have either environmental or social characteristics (commonly referred to as "Article 8") or products that have sustainable investments as their objective ("Article 9"), as well as for baseline disclosures required for all other products ("Article 6" in market practice) (European Commission, 2019b).

2.1.3 Sustainability and ESG: Definitions and Frameworks

Sustainability is a general approach to organizing economic activity in order to preserve natural and social systems, versus depleting them, for an extended period of time. The origin of sustainability is generally traced back to the Brundtland Commission's definitions of sustainable development as "the meeting of the needs of the present without compromising the ability of future generations to meet their own needs" (Mebratu, 1998; United Nations, 1987). The Brundtland Commission's definition of sustainable development clearly connects intragenerational equity with intergenerational equity (Solow, 1993). This general level of sustainability has also been operationalized via the United Nations' Sustainable Development Goals (SDGs), which have defined sustainability at the level of specific goals (i.e., 17 specific goals that cover areas of poverty, health, education, climate, ecosystems, etc.) and are quantifiable (Sachs et al., 2019; United Nations, 2015). However, from a financial/corporate perspective, ESG (environmental-social-governance) provides a more concrete/indicator-based way of examining companies in terms of their management of sustainability-related risks and opportunities (Gillan et al., 2021).

The ESG model is typically broken down into three categories: the environmental category examines a company's impact on and dependence upon natural systems (Gillan et al., 2021). Examples include the company's greenhouse gas emissions, resource and energy usage, pollution and loss of biodiversity. The social category assesses the relationships between employees, customers, suppliers, etc. and a company. These include labor standards, employee health and safety, diversity and human rights, etc. The governance category examines the structures and processes used for making decisions within a company. These include the board of directors' composition, executive compensation, internal control mechanisms, levels of transparency and ethical behavior, etc. From this perspective, ESG serves as a bridge between a company's sustainability profile and investment decisions since it converts these issues into measurable metrics that can be used in evaluating a company's risk and determining its value to investors (Eccles & Klimenko, 2020; Friede et al., 2015; Liang & Renneboog, 2017). The role of ESG is supported by increasing convergence among reporting and disclosure guidelines (e.g., GRI, SASB/ISSB and TCFD) that aim to create commonality in ESG reporting and produce systematic and analyzable data, including using artificial intelligence (Hummel & Jobst, 2024; Liu, 2022).

2.2 ESG Performance and Firm Value

2.2.1 Theoretical Foundations

The connection between ESG performance and the value of a firm was formed through multiple, complementary theoretical models, which include (1) stakeholder theory, (2) shareholder primacy and agency theory, (3) legitimacy theory, and (4) resource-based view (RBV); all of these models describe both why and how ESG activity could positively or negatively affect the financial performance of a corporation (Gillan et al., 2021; Moolkham, 2025).

2.2.2 Empirical Evidence on ESG-Performance Relationships

There has been a significant body of empirical research examining the relationship between corporations' ESG attributes and their financial performance, with generally positive results, though certainly not uniformly so (Whelan et al., 2021). One well-cited meta-analysis found that over 2,000 primary studies indicate a positive correlation between aggregated ESG attributes and corporate financial performance, with many of those studies reporting a positive relationship between those ESG attributes and corporate financial performance (Friede et al., 2015). Similarly, earlier studies examining the relationship between corporate social performance and accounting and market performance documented positive correlations between these two types of performance, supporting the idea that stakeholder-oriented practices can be valuable (Huang et al., 2020; Margolis & Walsh, 2003; Orlitzky et al., 2003).

2.2.3 ESG Measurement Challenges

One of the central challenges associated with ESG research and practice is that "ESG" is not a unitary, measurable construct but rather a collection of discrete, unobservable constructs that are measured with significant discretion (Agosto & Tanda, 2025; Berg et al., 2022). ESG ratings from major providers may differ, but ratings assigned to the same company by different providers often vary greatly (Dimson et al., 2020). Decomposition analysis has indicated that the majority of the variation in ESG ratings is due to measurement variation (i.e., how the same attribute is defined/operationalized) and scope variation (i.e., which attributes are included in the rating) (Liu, 2022). In contrast, the variation due to the weighting choice is relatively small. These variations can lead to decreased comparability across studies, attenuated estimates of the ESG-performance relationship, and reduced ability to develop portfolios based on ESG rating data (Avramov et al., 2022).

Additionally, the issue of data quality and transparency compounds this problem. Regulators and standard setters have identified limitations in ESG ratings/data, including inconsistent corporate disclosures, reliance on estimates/proxies, lack of methodological clarity, and potential conflicts of interest within the ESG ratings/data ecosystem (Agosto & Tanda, 2025; LSFI; OECD, 2025). As a result of these concerns, the European Union has taken steps to regulate ESG rating activities directly (through the ESG Ratings Regulation), and ESMA has proposed detailed technical standards designed to improve transparency, governance, and comparability "where possible" (ESMA, 2025; European Commission, 2024c). Additionally, evolving disclosure regimes regarding climate-related information (e.g., the SEC's 2024 climate disclosure regulations) illustrate that ESG measurements are influenced by both analytic methodologies and changing reporting requirements and institutional/legal constraints (SEC, 2024).

The table below outlines how different ESG rating providers are weighting and combining their indicators into a single ESG score; and how these methods differ from each other as to whether they use an industry-based approach to adjust for relative risk, an absolute risk orientation or data-driven optimization techniques.

Table 2-2. Industry Approaches & Weighting. Sources: (Agosto & Tanda, 2025; Avramov et al., 2022; Berg et al., 2022; LSEG, 2022, 2024; MSCI Inc, 2026; Muck & Schmidl, 2024; Münchhausen et al., 2024; S&P Global, 2023; Sustainalytics, 2026)

Rating Provider/Approach	Type	Weighting Method	How Scores Are Aggregated	Notes
MSCI ESG Ratings	Industry-relative	Industry-specific weights (~ Key Issues weights)	Weighted average across industry-relevant issues for each pillar	Uses a multi-level system with key issues specific to industry; final rating is normalized among peers (AAA–CCC).
Sustainalytics	Company risk score	Risk-oriented weighting	Focuses on managing material ESG risk	Rating based on risk exposure and management; not pillar weights strictly equal
S&P Global (CSA)	Corporate Sustainability Assessment	Industry and metric weights	Weighted sum with industry relevance	Score influenced by industry context and sustainability criteria
Equal Weighting (Academic/Simplified)	Academic / simple approach	Equal weights: E = S = G = 1/3	Simple average of E, S, and G scores	Easy to use but may misrepresent industry relevance
Data-Driven / Optimization (e.g., Refinitiv)	Quantitative	Weights optimized from data	Optimization or machine-learning-derived weights	Uses financial correlations or other statistical techniques
LSEG (Refinitiv/LSEG)	Multi-stage scoring	Category & materiality weights	Detailed aggregation, including controversies	Includes materiality and controversy adjustments

The table below lists the 17 main categories of ESG metrics as defined by five major sustainability reporting frameworks: the EU Corporate Sustainability Reporting Directive (CSRD) with European Sustainability Reporting Standards

(ESRS), the Global Reporting Initiative (GRI), and the U.S. Securities and Exchange Commission (SEC) proposed climate reporting requirements (European Commission, 2023a; GRI, 2021; Hummel & Jobst, 2024; SEC, 2024). The three pillars of the ESG group the 17 metrics and include each metric's specification, which includes a list of specific measurement indicators, specific units of measurement, regulatory designation (required vs. based on materiality), applicable standard reference, framework alignment, and required reporting cycle. Under EU CSRD/ESRS, 11 metrics are mandated for all eligible organizations regardless of their materiality assessment; six additional metrics will be applicable based on an organization's materiality assessment.

Table 2-3. ESG Disclosure Metrics Matrix (Pillars, Categories, Standards, Frequency). Sources: (European Commission, 2023a; GRI, 2021; SEC, 2024)

ID	ESG Pillar	Metric Category	Specific Metrics	Unit of Measurement	EU CSRD/ESRS	GRI Standards	US SEC (Proposed)	Reporting Frequency
1	Environmental	GHG Emissions	Scope 1, 2, 3 emissions (tCO2e); emissions intensity per revenue/FTE; Carbon removal/offsets; emissions reduction targets	tCO_2e, tCO_2e/€M, tCO_2e/FTE	Mandatory (ESRS E1)	GRI 305	Required (Scope 1, 2; Scope 3 if material)	Annual (minimum)
2	Environmental	Energy Management	Total energy consumption (MWh); renewable energy %; energy intensity; energy efficiency improvements	MWh, %, MWh/€M	Mandatory (ESRS E1)	GRI 302	Indirect via emissions	Annual
3	Environmental	Water Management	Water withdrawn/consumed (m^3); water discharge quality; operations in water-stressed areas; Water recycling rate	m^3, m^3/€M, %, quality index	Material-based (ESRS E3)	GRI 303	Not required	Annual
4	Environmental	Waste Management	Total waste generated (tonnes); hazardous vs. non-hazardous breakdown; waste diverted from disposal; recycling rate %	Tons, %, kg/unit	Material-based (ESRS E5)	GRI 306	Not required	Annual
5	Environmental	Biodiversity & Ecosystems	Biodiversity impact assessments; Operations in/near protected areas; Habitat restoration initiatives; TNFD alignment	Number, Hectares, € invested	Material-based (ESRS E4)	GRI 304	Not required	Annual/Biannual

6	Environmental	Circular Economy	Circular material use rate; Product design for circularity, extended producer responsibility, and material recovery rate	%, Number of products, Tonnes	Material-based (ESRS E5)	GRI 301	Not required	Annual
7	Social	Health & Safety	Lost-time injury frequency rate (LTIFR); fatal incidents; high-consequence injuries; Near-miss incidents; Occupational illness rate	Rate per 200k hrs, Number, Rate	Mandatory (ESRS S1)	GRI 403	Not required	Annual/Quarterly
8	Social	Training & Development	Average training hours per employee; Training coverage by gender/level; Investment in employee development; Skills gap analysis	Hours/employee, %, €/employee	Mandatory (ESRS S1)	GRI 404	Not required	Annual
9	Social	Diversity & Inclusion	Board diversity (gender, age, ethnicity); Gender pay gap ratio; Women in management %; Workforce diversity by category	%, Ratio, %	Mandatory (ESRS S1)	GRI 405	Not required	Annual
10	Social	Labor Practices	Employee turnover rate, Collective bargaining coverage, living wage compliance, working hours, parental leave uptake	%, %, %, Hours, Days	Mandatory (ESRS S1)	GRI 401, 407, 408	Not required	Annual
11	Social	Human Rights	Human rights due diligence; Forced/child labor incidents;	Number, Number, Availability, %	Mandatory (ESRS S1, S2)	GRI 408, 409, 414	Not required	Annual

			Grievance mechanisms; Human rights training coverage					
12	Social	Community Relations	Community investment programs; Local employment %; local supplier spending; Indigenous rights; Stakeholder consultation	€, %, %, Yes/No, Number	Material-based (ESRS S3)	GRI 413	Not required	Annual
13	Governance	Board Structure	Board independence %; Average director tenure; Committee composition, Board meetings frequency; Director skillset matrix	%, Years, Composition, Number, Matrix	Mandatory (ESRS 2, G1)	GRI 2-9 to 2-17, 405	Climate governance required	Annual
14	Governance	Ethics & Compliance	Anti-corruption/bribery policies; Ethics training completion rate; Whistleblowing mechanisms; Sanctions/fines; Political contributions	Yes/No, %, Yes/No, Number/€, €	Mandatory (ESRS G1)	GRI 205, 206, 415	Not required	Annual
15	Governance	Transparency & Disclosure	ESG report publication frequency; Third-party assurance coverage; Data quality verification; ESG rating scores; Materiality assessment	Annual/Biannual, %, Score, Yes/No, Assessment	Mandatory (ESRS 2)	GRI 2-1 to 2-5	Climate disclosure required	Annual
16	Governance	Risk Management	ESG-linked executive compensation; Climate risk TCFD disclosure; Scenario analysis;	% of variable pay, Disclosure level, Yes/No, Score, Yes/No	Mandatory (ESRS 2)	GRI 2-12 to 2-14	Climate risk required	Annual

			Transition plan credibility; ESG committee					
17	Governance	Supply Chain Due Diligence	Supplier ESG assessments conducted; Suppliers audited for ESG compliance; Supplier corrective action plans; Supply chain transparency	%, Number, Number, Score	Material-based (ESRS S2)	GRI 308, 414	Not required	Annual

2.3 AI in Sustainable Finance

2.3.1 Historical Evolution of AI in Finance

The historical development of AI in finance is heavily interwoven with the history of the development of quantitative methodologies and computational technology in the financial service industry. The initial applications during the late twentieth century were primarily developed using rule-based systems and Expert Systems, which took human decision rules for tasks such as credit screening, risk assessment and trading, and converted them into 'if-then' formats. These systems were complementary to the development of Quantitative Finance and Econometric Models as they provided automation of decision-making support in areas in which domain knowledge could be formalized, but did not possess the ability to learn from data in the manner of contemporary machine learning systems (Buchanan, 2019; Hoang & Wiegratz, 2023).

After the 1990’s, the following innovation in data storage, processing capabilities, and digitalization led to the statistical learning methods being adopted into mainstream financial practices. as noted in various articles addressing machine learning in finance, both supervised learning and unsupervised learning techniques have been used in the areas of credit scoring, fraud detection, portfolio selection, and trading, i.e., these articles are elaborating upon the traditional linear models to also model the non-linearities and complex interactions found within financial data sets. Additionally, the development of electronic trading platforms and algorithms have driven the development of more automated execution strategies by creating an enhanced role for AI and ML models to produce signals, route orders, and manage risk. (FSB, 2017; Kelly & Xiu, 2023; Nazareth & Ramana Reddy, 2023).

Since the Global Financial Crisis, the development of AI in Finance has proceeded at an increased pace, driven by the availability of "big data", cheaper cloud computing, and advancements in deep learning. Reviews of financial machine learning, such as Buchanan (2019) and Kelly and Xiu (2023) provide evidence of the widespread deployment of a large number of machine learning techniques—tree-based ensembles, regularized regressions, neural networks and reinforcement learning - in asset pricing, portfolio optimization, volatility and risk forecasting, roboadvisory and customer analytics. These techniques are attractive due to their flexibility to approximate complex functional relationships within high dimensional data and their ability to allow financial institutions to process new information sources, including alternative and unstructured data, more effectively than classical econometric approaches.

Assessments of the official sector have also been important in defining this evolution and its implications for financial stability and regulation. the financial stability board's 2017 report on "artificial intelligence and machine learning in financial services" documents the increasing use of ai across four broad categories of applications: customer facing applications, operations, trading and portfolio management, and risk management and compliance and highlights both potential efficiency gains and emerging systemic risks. More recent work by central banks and international organizations document that ai is now embedded across the financial sector - used by private institutions, supervisors and central banks themselves - and highlight issues related to model risk, governance, explainability and new forms of interconnectedness created by widely adopted AI systems. therefore, the historical trajectory of ai in finance is one of gradual progression from rule-based decision support to data driven, learning based systems with functional coverage, accompanied by an evolving policy discussion regarding how to harness efficiency gains while safeguarding market integrity and financial stability (FSB, 2017; Kelly & Xiu, 2023).

2.3.2 Historical Evolution of AI in Sustainable Finance

Sustainable finance's connection to AI has developed gradually through the convergence of two areas of study: the ongoing scholarly debate about how ESG factors are related to financial performance; and the technological development of machine learning techniques for handling the complexity of diverse types of sustainability-related data. Quantitative underpinnings for sustainable finance trace back to an earlier body of empirical research examining socially responsible investing (SRI). Researchers aimed to determine if social responsibility and fiduciary duties were compatible with maximizing returns (Edmans & Kacperczyk, 2022; Schwendner & Posth, 2024). A significant meta-analysis by Friede et al. (2015) combined information from around 2,200 individual studies conducted over four decades. The authors found that almost 90% of the studies support the idea that there is a positive/non-negative link between ESG criteria and corporate profitability, thereby providing strong empirical support for the use of ESG information in making investment decisions. These early findings, however, were based predominantly on structured ESG ratings from commercial providers rather than on AI-driven data extraction methods, and they were already subject to mounting criticism regarding data quality and comparability.

The data challenges that constrained early quantitative ESG research are documented systematically in the influential work by Kotsantonis and Serafeim (2019) which identifies four structural problems with ESG data: inconsistency of reported metrics across companies, widespread reliance on imputed rather than directly disclosed data, divergence of ratings across providers, and the misalignment of ESG metrics with material sustainability outcomes. These challenges have generated an apparent need for AI-based solutions that can extract and standardize unstructured corporate sustainability information at a large enough scale to represent the entire scope and depth of companies' ESG performance as has been defined by their respective providers. In this context of lack of sufficient structured data, natural language processing (NLP) and other AI methods were developed to analyze and transform qualitative information from company disclosures, news articles and filings into quantitative sustainability signals (Schimanski et al., 2024).

In addition to the development of NLP methods for transforming qualitative company disclosures into quantitative sustainability signals, the growth of climate finance as a specific academic discipline has also influenced the evolution of AI use in sustainable finance. Giglio et al. (2021) document the rapid development of climate finance as a body of research focused on pricing climate risks across different asset classes, as well as the role that financial markets can play in facilitating the transition of economies to a lower carbon footprint, and note that climate finance became a major area of study only after the Paris Agreement was signed in 2015. The signing of the Paris Agreement created a new institutional focus on climate risk, resulting in a significant increase in the amount of corporate ESG reporting, as documented through NLP analyses of annual reports of European financial firms during the period of 2000-2021, demonstrating a structural break in the level of ESG reporting after 2015; and creating an increased demand for AI solutions to manage the increasing volume of climate related financial reporting. Climate risk modeling, climate risk scenario analysis, and climate risk stress tests under both physical and transitional risk scenarios are all dependent upon the ability to process multiple types of heterogeneous data at a very high dimensionality, far beyond what is possible with traditional econometric methods, and thus create the demand for machine learning methodologies in climate finance.

2.3.3 AI for ESG Data, Ratings and Disclosures

One of the most fundamental challenges in sustainable finance is the availability, quality and comparability of ESG data, and it is precisely in this area that artificial intelligence has made its most direct methodological contribution. Kotsantonis and Serafeim (2019) document four structural deficiencies that characterise ESG data as produced by commercial rating agencies: the use of inconsistent metrics across firms and industries, the widespread imputation of unreported values, pronounced divergence in ratings across providers, and the misalignment of measured indicators with material sustainability outcomes. Limitations of current approaches to integrating ESG criteria into portfolio construction and risk management present serious obstacles for investors who desire to integrate ESG criteria with rigor and these limitations also underpin the demand for AI-based methods that can extract, standardize and validate sustainability information from primary unstructured sources, rather than depending solely on third-party aggregated scores (Agosto & Tanda, 2025).

In particular, the divergence between ratings has attracted substantial attention from academics. The survey by Berg et al. (2022) break down the divergence across major providers of ESG ratings into three components — scope disagreement (the ESG attributes to be measured), measurement disagreement (how to measure them) and weight disagreement (how to combine them) — and attribute the largest share of divergence to measurement and scope differences, illustrating that the problem with ESG information is not only quantitative but definitionally fundamental. An additional perspective is provided by work on ESG data imputation, which shows that missing metrics effectively represent about 50 percent of the overall ESG dataset and that the imputation strategies used by providers of ESG ratings introduce a quantifiable bias that distorts ESG ratings and leads market participants awry; therefore, transparent and algorithm-based imputation methods are needed.

The following tables have been designed to structure the existing literature on the fast-developing field of AI and ESG, which is still very disjointed. This allows for a systematic comparison of the underlying data sets used in each study, the family of models used by each study, the targets of each study's predictions, and the results achieved by each study. In addition to providing an overview of existing studies (i), the dissertation provides insights into how certain combinations of ESG data, AI architecture and domain areas are currently not being explored sufficiently (ii). Finally, these tables provide a systematic overview of research gaps in the field that will be the basis for the development of hypotheses, the selection of models and the empirical strategy in the following chapters.

Table 2-4. Predicting ESG scores/ratings

Authors	Model category	Typical input data	Output (target)	Reported performance (as stated in paper/abstract)	Interpretation / importance signals
(Zhang & Zhao, 2026)	ML (boosting)	15 indicators across financial, E, S, G; Chinese A-share (2013–2022)	ESG rating class	Accuracy 91.0%, Precision 90.7%, F1 90.1%, AUC 0.977 (plus 92.4% / 97.7% for derived models)	SHAP shows financial attributes often dominate single non-financial metrics; handles imbalance with SMOTE-ENN
(Patel et al., 2026)	ML (ensemble/stacking)	Pillar sub-scores, controversy measures, firm financials, categorical + geospatial indicators	Total ESG risk score (continuous)	RMSE 1.006, MAE 0.664, MAPE 3.13%, R^2 0.979	Permutation importance: Environmental (0.41) + Social (0.32) biggest drivers; warns about multicollinearity between components & aggregate
(Wang et al., 2024)	ML (stacking)	21 structured features; China A-share (2012–2020)	ESG + pillar scores	Example industry results: Energy RMSE 3.2937, MAE 2.4409, R^2 0.8971; reports RMSE/MAE reductions vs baselines	Stacking (RF/GBDT/XGB/LGBM → Bayesian Ridge) improves generalization across industries
(Wang et al., 2024)	Hybrid ML/DL (ensemble)	Public “fundamental data” only	ESG rating (continuous/score)	Explains ~54% variance (R^2); improves baseline R^2 27.4% → 53.9%, MAE 11.2 percentage points	Explicit motivation: scalable “initial ratings” for unrated firms; discusses why controversies are often missing from agency scoring
(D’Amato et al., 2022)	ML (tree ensemble)	Balance-sheet / structural data; STOXX 600; Refinitiv ESG	ESG score	Paper focus: balance-sheet variables “crucial to explain ESG scores”	RF feature importance used to identify which structural items explain ESG
(Zhang et al., 2025)	NLP (BERT)	Self-authored ESG reports; custom ESG corpus	ESG rating label	Accuracy 93%	Goal is to reduce manual bias + “irrelevant content”; standardize criteria across sectors
(Sariyer et al., 2024)	DL + analytics pipeline	Refinitiv ESG dataset; clustering + association rules + DL	ESG performance	Validated pipeline	Adds prescriptive analytics: patterns + actionable rules, not just prediction

(Lin & Hsu, 2023)	ML (comparative)	Taiwan listed firms; multiple ML models	ESG score	Reports comparative ML performance (R^2/MAE/RMSE in paper)	Includes variable importance discussion

Table 2-5. Predicting ESG controversies, greenwashing, “exaggeration”, and external-risk signals

Paper	**Model category**	**Typical input data**	**Output (target)**	**Reported performance**	**Interpretation / importance signals**
(Svanberg et al., 2023)	ML (predictive validity)	ESG indicators anchored to CSR non-compliance controversies	Likelihood of controversy + derived ESG rating	Best model: **precision 70–84%**; “high predictive performance” on multiple measures	Explicitly argues ML aggregation loses less information than arithmetic averaging; provides indicator contributions
(Vinella et al., 2023)	NLP (Transformer)	Sustainability reports; generated weak labels for “greenwashing risk”	Greenwashing risk (classification)	Best model: **accuracy 86.34%, F1 0.67**	Key contribution: quantification formula for “greenwashing risk” + weak labeling scheme
(Lagasio, 2024)	NLP (index/measurement)	Sustainability reports; sentiment + sustainability-term content	ESG-washing index (continuous)	Constructs ESGSI (not “accuracy” task)	Explains discrepancies between portrayed vs actual practices; variation across industries/regions
(Luo et al., 2025)	ML + NLP features	594 ESG reports (crawled)	Exaggeration (classification) + exaggeration score (regression)	RF best for classification; ridge best for score regression	Builds a domain “exaggeration thesaurus” to interpret exaggeration cues
(Lipenkova et al., 2023)	NLP benchmark / dataset	ESG reports + third-party public media articles	Greenwashing indicators (task labels)	Benchmark task (leaderboard metrics per paper)	Core design: compare self-reported vs external narratives to reduce asymmetry

2.3.4 Limitations of AI in Traditional Finance

Although there are numerous documented performance benefits of using machine learning and deep learning in financial applications, a large body of literature from both academics and policymakers has noted that there are structural limits that hinder the responsible use of AI in mainstream finance. The structural limits span across four interrelated areas including; the black box nature of how the model makes decisions (the "black-box problem"), the tendency of AI models to over-fit to training data and to be biased by publication bias when reporting their performance, the potential for AI driven systemic risk, and the potential for AI to produce or perpetuate data bias, unfairness and/or regulatory violations (Cheryll-Ann Wilson, 2025; Daníelsson et al., 2022).

The most extensively studied limitation is the interpretability deficit of complex AI models. In high-stakes financial context, scoring, fraud detection, and algorithmic trading—models based on deep learning architecture or large ensembles can achieve high predictive accuracy while remaining entirely opaque to their developers, operators, regulators and affected individuals, a problem widely described in the literature as the "black-box phenomenon" (Mohamed et al., 2025; Mohsin & Nasim, 2025). The authors of this article, who conducted a systematic review of explainable AI (XAI) in the area of finance in 2025 conclude that the majority of XAI applications are based primarily upon post-hoc interpretation techniques such as SHAP, LIME and attention mechanisms (Mohamed et al., 2025; Mohsin & Nasim, 2025; Salih et al., 2025). However, they also point out that no universally accepted criteria exist for assessing the quality of explanations generated by these methods, nor is there a consensus on how to measure the performance metrics associated with them (Doshi-Velez & Kim, 2017; Salih et al., 2025). In addition, the authors note that computational overheads introduced by XAI techniques can be substantial, particularly in environments where real time is critical (e.g., trading) (Gunning & Aha, 2019). The CFA Institute has noted in its report on the application of Explainable AI in Finance, the "black box" problem of AI systems presents a number of regulatory challenges (CFA Institute, 2025). For example, lack of transparency hinders the ability to comply with fair lending regulations and consumer protection statutes. Lack of transparency can also make it difficult to determine where a model is failing, and how those failures are impacting customers. The CFA Institute also states that this continues to be cited by both practitioners and regulators as the second primary barrier to increased use of AI technology at financial institutions (CFA Institute, 2025; Mohsin & Nasim, 2025). The Financial Stability Institute (FSI) has identified the variation in regulatory approaches to the management of AI explainability across jurisdictions. The FSI further identified that many of the current regulatory frameworks do not align well with the amount of technical sophistication in the current AI models (BIS FSI, 2024, 2025).

One other key constraint is overfitting and overestimating an out-of-sample forecast model's performance. When data are non-stationary and autocorrelated and contain regime breaks due to changing market regimes, machine learning models optimized in-sample will generally fail to generalize to new market environments and produce strategies that perform well in historical back tests but make nearly random out-of-sample forecasts (Arian et al., 2024; López de Prado, 2018). There have been studies utilizing technical analysis-based features for predicting returns on stocks demonstrating that when in-sample accuracy reaches 100% in terms of hyperparameters selected for optimization, that out-of-sample accuracy levels of around 50% (this means that most models cannot exceed a buy-and-hold strategy after removing artifacts of in-sample optimization) (Peng & Moraes Souza, 2024). A more recent study evaluating bias and optimism in the reported performance of machine learning models in finance introduced a formal model to adjust for estimated accuracy inflation to determine the true estimate of performance, demonstrating that both overfitting and selective reporting create a systematic upward bias in the reported performance of machine learning models across financial machine learning studies (Saidi et al., 2025). Therefore, there is a need for greater consideration when evaluating the empirical findings of machine learning studies. As a result of these challenges, combinational purged cross-validation (CPCV), which provides a more robust cross-validation framework for validating machine learning models against non-stationary, autocorrelated and regime shifted financial data, although still used relatively infrequently in empirical research, has been proposed (Arian et al., 2024; Peng & Moraes Souza, 2024; Saidi et al., 2025).

The first limitation of using AI in finance relates to the limited availability of high-quality labelled training data and/or low amounts of historical data. To be able to train AI models properly, they require large amounts of relevant data. Unfortunately, there is very little available in the form of publicly available datasets, therefore the need for bespoke datasets that are relevant to specific use cases exists. Secondly, there are privacy restrictions and confidentiality requirements associated with financial data (Bahoo et al., 2024). Therefore, it is not possible for most researchers to obtain and use the same type of financial data used to develop some of the popular AI-based finance models.

A second limitation relates to the cost associated with developing and implementing AI solutions. While the cost of implementing a solution may be lower than traditional alternatives, the initial costs associated with the implementation process is generally higher. The time required to implement AI models is also much longer compared to traditional methods and requires a significant amount of technical knowledge to operate effectively (Bahoo et al., 2024; Roy et al., 2025).

A third limitation relates to the potential of AI to generate or exacerbate systemic risk in financial markets. If large numbers of financial institutions were to utilize AI in their risk management or trading processes in a way that produces correlated behavior in response to the same market signals, this could create a "monoculture" effect in the marketplace. It could lead to herding behaviors and the reduction of liquidity in the market during times of stress (Daníelsson et al., 2022). A SUERF policy note on the relationship between AI and systemic risk describes how the five classical sources of systemic fragility liquidity mismatches, common exposures, interconnectedness, lack of substitutability, and leverage - can be mapped onto AI-specific mechanisms. The authors explain that while AI may improve risk management for individual firms, its widespread adoption could also increase the entire financial system's vulnerability to systemic shocks because of heightened co-movement and feedback loops.

In addition to the prior concerns, the IMF also has a second systematic concern with regard to AI (IMF, 2024). Specifically, the IMF is concerned that the automatic de-risking of individual firms and the shutdown of models that were designed to individually protect each firm from extreme volatility may create simultaneous destabilizing feedback loops and instantaneous liquidity droughts among multiple firms if all of the firms' algorithms are triggered at the same time. The FSB identified three key systematic concerns for the use of AI in banking systems in its 2024 report addressing the impact of AI on the stability of financial markets: 1) the correlated behavior of AI models; 2) the concentration of AI infrastructure providers; and 3) inadequate model governance. The regulators should focus their efforts on these three areas (FSB, 2024).

In terms of limiting bias in AI models, the final limitation relates to the fact that models trained on historical financial data will likely contain and potentially amplify the biases contained within the historical data. For example, if the dataset used to train the credit scoring and loan origination model contains historical patterns of unequal access to credit based on demographic characteristics (e.g., race, gender), then the model may systematically disadvantage protected demographics, even after excluding sensitive attributes from the feature set, because the attribute(s) are likely represented in correlated proxy variables such as residential zip code or employment history (Akinwumi et al., 2023; Barocas et al., 2023).

In addition, since black boxes cannot explain to the public why an applicant was denied credit or why a loan application was declined, it is difficult for the public to understand and challenge the lending decisions made by lenders. In addition to being able to provide applicants with the reason(s) for denial under the ECOA and GDPR, lenders will experience further difficulties in doing so if they have adopted AI based lending models. The issues described above clearly outline areas of research where developing models that are intrinsically interpretable; validation methods that can validate complex models; fairness-aware learning algorithms; and AI governance frameworks that allow lenders to benefit from advanced machine-learning while maintaining transparency, fairness and stability in regulated financial markets (Akinwumi et al., 2023; CFPB, 2023; European Commission, 2016; Jobin et al., 2019).

2.3.5 AI for Green Investment and Portfolio Construction

Green investments and sustainable portfolios developed with the help of AI constitute an important, practical application of machine learning in the area of sustainable finance and represent a link between ESG data analysis and capital allocation decision-making. The traditional methods of developing portfolios with ESG criteria (negative screening, best-in-class and simple ESG scoring thresholds) were historically replaced or augmented with machine learning-based approaches which are superior to traditional approaches when it comes to processing large amounts of dimensional, dynamic and heterogeneous data sources; and therefore provide investors with more sophisticated and flexible investment opportunities that meet both their financial objectives and sustainability objectives while minimizing their exposure to risk (Feng et al., 2025).

A fundamental problem that is encouraging research into AI and Green Investments is the lack of consistency among ESG rating agencies. The Bank of Italy commissioned a study to propose a method utilizing Machine Learning to reduce the "green gap" among ESG rating agencies by using Machine Learning to identify a company's long-term sustainability trajectory from its current ESG metrics and identifying those ESG metrics most indicative of a company's future sustainability trajectory; thereby establishing a more consistent and comparable basis for portfolio creation than the disparate commercial ESG ratings currently available. In addition, studies employing ensemble-based methods such as Random Forest and Gradient Boosting to develop estimates of ESG scores based on financial fundamentals demonstrate that the sustainability-related signals produced via Machine Learning may be utilized either to confirm or replace commercial ESG ratings; and thus produce a transparent and reproducible input to the portfolio selection process that is less subject to the disparities in measurement and scope that were highlighted by Berg et al. (2022).

2.3.6 AI for Climate and Sustainability Risk Assessment

Climate and sustainability risk assessment is one of the most difficult applications of AI in sustainable finance due to the large number of complex, non-linear, and long-term risks that are beyond the analysis capabilities of traditional econometric tools. Giglio et al. (2021) explain how climate finance developed as a new field of study that examines the relationship between climate change and financial markets; they also emphasize that climate risks (transition and physical) must be included in macro-finance models and in asset pricing systems, and that financial markets can serve as the main mechanism for insuring and reducing climate risks through investment in green projects and de-investment from carbon-based projects. A review of the use of machine learning techniques in climate finance using topic modeling of the peer-reviewed literature identified the following four representative groups of themes: physical risk modeling (weather, natural disasters and agricultural risks); energy economics and carbon trading; socially-responsible investing and ESG criteria; and extracting climate-related information from un-structured data sources (e.g., satellite images and remote sensing), showing that AI in climate finance deals with both biophysical and financial aspects of risk.

For climate and physical transition risk pricing in financial assets, the most important contributions are those of Engle et al. (2020). The authors provide a dynamic hedging strategy for climate change risk, combining NLP-based climate news innovations with an ESG-score based proxy method for firm level climate risk exposure, and show that their portfolios achieve correlations of up to 30% with climate news index innovations relative to climate news index innovations when evaluated using out-of-sample methods. The authors further demonstrate that the hedging portfolios outperform alternative portfolios that tilt to industry.

Bua et al. (2024) extend this framework to build new physical risk indices (PRI) and transition risk indices (TRI) for euro area equity markets using high frequency climate news data. The authors' findings show that both PRI and TRI are associated with an asset price premium, and that each responds differently to policy announcements and extreme weather events. This supports the argument that financial analysis should distinguish between physical and transition risks.

Campiglio et al. (2023) assess the current state of the evidence for climate related risks in financial assets as they illustrate how climate impacts can affect the value of financial assets by various means, and provide a review of emerging methodologies for assessing future climate related price movements, such as ML-based scenario analysis, and have important implications for financial stability.

Climate risk is also being increasingly integrated into the process of assessing credit risk, which relates directly to banking regulation and macro prudential policy. Chalabi-Jabado and Ziane (2024) used a double machine learning approach to a cross-country panel and found that there is a net positive and statistically significant effect from AI development on the reduction of climate risk. The authors identified four mechanisms by which AI development produces its effects on climate risk mitigation: resource efficiency; green innovation; enforcement of environmental policy; and green finance development.

Yin et al. (2025) apply machine learning techniques to assess the impact of climate risk — decomposed into physical and transition components — on bank performance and lending growth. The authors report that there is a statistically significant negative association between physical climate risk and bank financial performance and lending, especially in regions with high levels of climate exposure and weak regulatory environments.

Fichera et al. (2025) investigate the climate transition challenges represented in credit portfolios. The authors examine the degree of alignment between bank credit portfolios and European environmental objectives and present a machine learning-enhanced framework for determining the degree of exposure to transition risk at the portfolio level and offer a decision-support tool for assessing the potential for loans originating under climate transition scenarios, and for calculating regulatory capital requirements.

2.3.7 Governance, Ethics and Regulation of AI in Sustainable Finance

Algorithmic bias is among the most consequential governance challenges in AI-driven ESG and sustainable finance. A bibliometric analysis of the AI–ESG–sustainable finance literature confirms that, while AI methodologies significantly improve the reliability and precision of ESG assessments, data standardization, ethical considerations and transparency of AI-driven ESG applications remain the most critical unresolved challenges, with algorithmic bias consistently identified as the foremost ethical risk associated with AI deployment in sustainability practice.

A study published by Giudici and Wu (2025) showed that when a combination of multiple ML models is used to examine ESG/credit score relationships they are more resistant to changes in the ESG data format than if an individual model were being utilized. However, the authors also noted that the lack of transparency or explainability in financial ML models (which has only been addressed within the last few years through the creation of metrics like S.A.F.E., which stands for Sustainable, Accurate, Fair, and Explainable), means that while the technology to deploy sustainable finance via AI is rapidly evolving, the regulatory framework to govern its use is still lagging behind.

A study on transparency and governance issues in empirical ML-ESG research by Seow (2025) demonstrates that there are two dominant barriers to governing the use of machine learning in ESG research: algorithmic opacity and transparency deficits. To address these accountability gaps, explainable AI (XAI) has been identified as one of the primary mechanisms both technically and institutionally. The results from a longitudinal public perception data analysis study conducted by Laviola and Cucari (2026) confirm that the public's level of trust in AI when it is used in an ESG framework is directly influenced by their perception of the transparency and fairness associated with the AI decision-making process; therefore, the lack of transparency and fairness in AI-based ESG decisions will have significant implications for the legitimacy and social acceptance of AI-based sustainability decisions.

2.4 Bibliometric Trends

Over the last decade, the bibliometric landscape of ESG & AI-in-Finance research has transformed dramatically as it has been driven by increasing regulatory demands, as well as by the development of data-based methods. Figure 2-1 is a graph of the number of articles published annually about ESG research, according to the Scopus database. The graph shows a sharp increase in publications since the middle of the 2010s, followed by a further acceleration of publications post-2019; this period also marked the time when the European Union implemented the SFDR and when global discussions regarding climate policy accelerated significantly. Figure 2-2 restricts the focus on the combination of AI and ESG research. It displays a far sharper exponential increase in publications at approximately 2020, indicating the field's transition from abstractly examining concepts toward empirically applying them. Together, Figure 2-1 and Figure 2-2 provide strong evidence that the current dissertation is focused on one of the fastest-growing and policy-relevant areas of research in modern finance, while concurrently providing evidence that the volume of research output is only now sufficient for systematic reviews of the evidence.

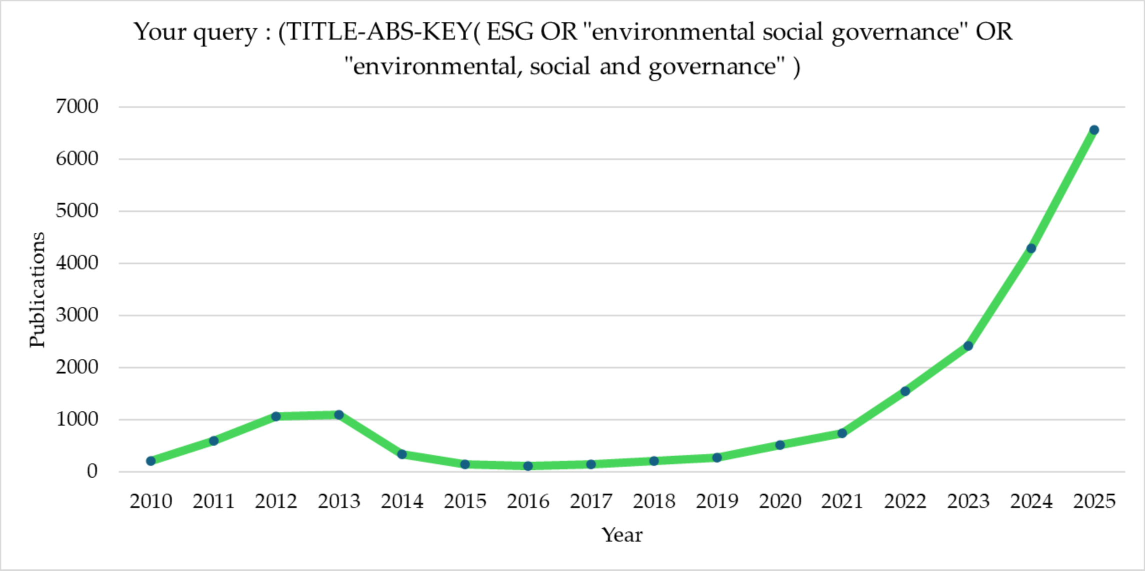


Figure 2-1. Annual publication output on ESG research

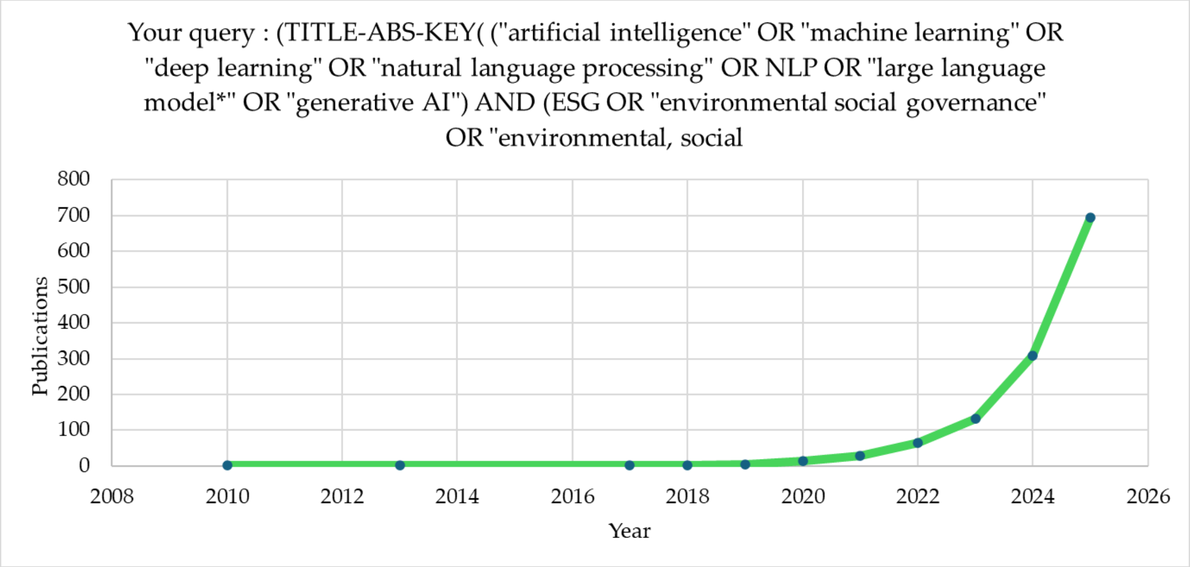


Figure 2-2. Annual publication output on the intersection of Artificial Intelligence and ESG

Chapter 3. Data & Descriptive Analysis

3.1 Dataset Provenance

The dataset is sourced from the Refinitiv (LSEG) ESG database (Debi, 2025). Refinitiv provides one of the most widely used ESG rating systems in academic research and industry practice, covering over 12,000 companies globally. The Refinitiv ESG rating methodology assesses companies on more than 500 individual data points across ten thematic categories, which are aggregated into three pillar scores (environmental, social, governance) and an overall ESG composite score, all normalized to a 0–100 scale.

The specific dataset covers 4,837 publicly listed firms across three major economic blocks (Asia, the United States, and Europe) over the eight years 2017-2024.

3.2 Variable Definitions

Table 3-1. Variables of the Dataset

Category	Variables	Description
Identifiers	RIC, Company, Country, Region, Industry, Sector	Firm metadata
ESG Scores	ESG, E_Score, S_Score, G_Score, ESG_Controversy	Refinitiv ESG scores (0–100)
Scale (USD K)	Total_Assets, Revenue, Market_Cap, EBIT, Net_Income	Key financials
Leverage	Total_Debt, LT_Debt, Leverage, Equity_to_Liab	Capital structure
Profitability	ROA, ROE, Profitability, EBIT_to_TA, NI_to_TA	Returns
Liquidity	Cash_Equiv, Quick_Ratio, Cash_Ratio, ICR	Solvency & liquidity
Market	Market_to_Book, Price_to_Book, Beta, YTD_Return	Market-based
Innovation	RnD, RD_Intensity	R&D spending/intensity
Risk	Altman_Z, Country_Risk, Distress (derived)	Risk indicators

3.3 Descriptive Statistics

Figure 3-1 presents two stacked horizontal bar charts side by side. The left panel ranks the 15 most populated TRBC (Thomson Reuters Business Classification) sectors by total unique firm count, with bars disaggregated by region. The right panel ranks the 20 most represented countries. Both panels are sorted in ascending order of total count to facilitate visual comparison of the long-tail distribution.

The left-hand panel shows how the sectors of each part of the data set compare to other sectors within their respective sector groups. The technology, financial (banking and insurance), and Industrial sectors make up a large portion of this data set, accounting for around 45-50% of each sample's unique company count. This is consistent with other major global stock indices, such as the MSCI World and S&P 500, which also have a large concentration of technology and financial services companies, as their total market capitalization is based almost exclusively on these two sectors. In addition, although the Energy and Materials sectors, identified as having the highest level of environmental materiality based on ESG materiality, are represented by these two sector groups, there are no indications that the energy and materials sectors will be overrepresented in this study. Prior research published by Khan et al. (2016) further reinforces these conclusions by stating that ESG materiality and structure vary significantly among sectors (i.e., environmental scores are predictive of return on assets for energy companies, while governance scores are most strongly correlated with credit risk in the financial service sector).

The geographic location of the firms in the sample is shown on the right side of the figure. Firms in the United States have the highest number, with Japan second, a significant gap, then the UK, Canada, Germany, and France. These five countries account for about 60 percent of the total number of firms. The presence of South

Korea, India, China, and Australia indicates the emergence of markets in the Asia panel. The countries' distribution is very similar to that of the "Global 2000" sample studied by Ioannou and Serafeim (2012) and to the 40 countries in the Refinitiv database analyzed by Liang and Renneboog (2017). They found evidence of systematic differences in ESG scores across countries due to different legal origins, investor protection regimes, and labor regulations. The large number of US firms in the sample could introduce a size bias in pooled OLS regression if smaller non-US firms do not disclose comparable ESG information to larger US firms. A possible solution to this problem is to include firm-sized controls (log_TA) and country fixed effects in the empirical models.

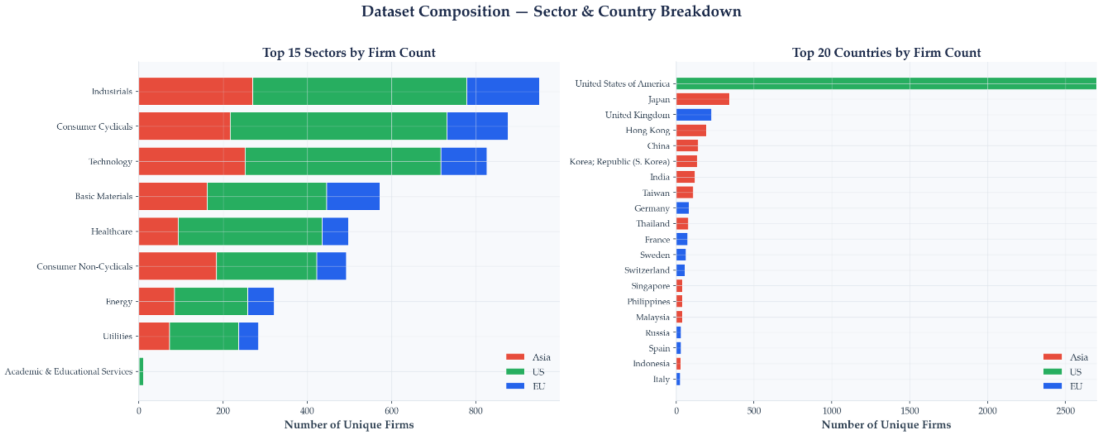


Figure 3-1. Dataset Composition — Top-15 Sectors and Top 20 Countries by Unique Firm Count

3.4 ESG Score Distribution

Refinitiv computes ESG scores using a rule-based, analyst-verified framework drawing on over 630 ESG measures across 186 metrics grouped into ten themes. The composite ESG Score (0–100) is a weighted average of the three pillar scores (E, S, G).

A separate ESG Controversy Score is overlaid to penalize firms involved in environmental incidents, labor disputes, or governance scandals, as captured through media monitoring. This score is then combined with the fundamental ESG score to produce a final ESG Combined Score. However, in our dataset, the pillar and controversy dimensions are retained separately to permit hypothesis-specific analyses.

Refinitiv's coverage is deeper for developed-market firms, which creates a systemic upward bias in US and EU ESG scores relative to Asian peers — partly due to mandatory disclosure requirements (EU NFRD/CSRD, SEC) rather than genuine ESG outperformance. This structural limitation is acknowledged throughout the empirical chapters.

Figure 3-2 presents three adjacent histogram panels, one per geographic region, each displaying the full distribution of the firm-year ESG composite score (scale 0–100). Within each panel, a dashed navy vertical line marks the mean, and a dash-dot amber line marks the median; both values are printed in the legend to enable immediate quantitative comparison across regions.

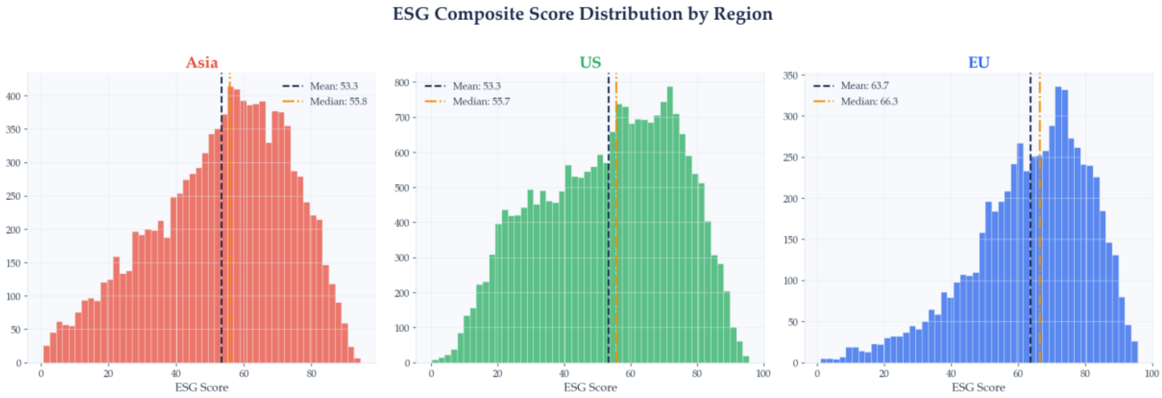


Figure 3-2. ESG Composite Score Distribution by Region — Asia, US, EU.

The EU distribution was the most highly skewed and most concentrated of the three; it has the highest meaning (approximately ~58-~63) and the highest mode in the 55-70 range. This aligns with the institutional hypothesis offered by Liang and Renneboog (2017), who demonstrated that civil law countries – which make up the majority of the EU panel (France, Germany, Italy, Spain) -- have substantially greater ESG ratings than common law countries because civil law countries have stronger labor protections, are required to hold mandatory meetings with stakeholders, and require codetermination. In addition, Christensen et al. (2022) showed that the EU's NFRD (Non-Financial Reporting Directive, Directive 2014/95/EU) caused a regulatory disclosure shock that raised ESG ratings for firms covered under NFRD (those with over 500 employees) by an approximate average of 8-12 points when compared to their pre-NFRD ratings.

The distribution of ESG ratings for U.S. firms shows broad symmetry around a mean of approximately 48-55, with visible mass at both ends of the scale. The observed bimodality in this distribution appears to indicate the existence of two discrete populations of firms: larger-cap S&P 500 constituents with well-established sustainability infrastructure that consistently rate above 60, and smaller mid-cap firms with less extensive ESG reporting practices that often rate below 30. The closeness of the mean and median after winsorizing the tails suggests that there were no extreme outliers, supporting the use of OLS estimators in subgroup regression analyses of the U.S. only.

Gillan et al. (2021) identify a similar bifurcation in the U.S. ESG landscape. Still, they attribute this bifurcation to voluntary self-selection into ESG disclosure frameworks (i.e., GRI, SASB, CDP) in a regulatory environment that has historically had no mandatory reporting requirements.

The Asian distribution is the widest and most left-skewed, with a mean of roughly 42–50 and a long-left tail extending to values below 20. The widespread distribution is due to the many different types of regulatory environments found in the region (Japanese firms have seen significant increases since the 2021 revision of the Corporate Governance Code; Chinese State Owned Enterprises do not see any ESG reporting; while SE Asian firms are just starting to report). Ioannou and Serafeim (2012) find that legal enforcement quality and cultural collectivism are the two primary determinants of ESG scores across countries; these factors also vary widely across the Asian sub-panel.

Figure 3-3 is a multi-panel line graph showing the annual cross-sectional mean of the four ESG dimensions for each of the three regional panels (Asia, US, EU) over the eight-year sample window. Discrete yearly observations are shown as markers, and connecting lines are used to help interpret time trends. A shared legend is located at the bottom of the figure.

All three regional panels show a clear positive trend, with the EU panel showing the highest absolute value throughout the sample window. The Asia panel showed the largest proportional gain, especially after 2020. This is consistent with Boffo and Patalano (2020)who reported a nearly linear increase of 2–3 points per year in OECD-wide aggregate ESG scores, largely due to increased analyst coverage and improved corporate disclosure frameworks. All of the series show a decline in 2020, which is likely due to the operational disruptions caused by COVID-19 in ESG reporting programs (see Broadstock et al. (2021)).

The E dimension shows the greatest rate of growth, especially after 2021, when COP26 in Glasgow (November 2021) sparked a surge of corporate net-zero commitments, and the EU Taxonomy Regulation was implemented in the financial product sector. Hassan et al. (2019) documented that firms exposed to climate policy risk increase their environmental disclosures in direct response to regulatory catalysts, consistent with the large increase in E-scores observed across all three regional panels since 2022.

The trend in S dimension is the least volatile, because social metrics (workforce diversity, health & safety, supply chain standards) tend to change more slowly than environmental metrics (which can be changed quickly through procurement decisions or asset divestitures) or governance metrics (which can be affected by changes in the composition of the board of directors). The EU social lead persists throughout the sample window. It is due to codetermination requirements (Germany's Mitbestimmungsgesetz), mandatory works councils, and the EU

Corporate Sustainability Due Diligence Directive (CS3D), provisionally agreed in 2024.

G-scores show the greatest inter-regional variation and the highest year-to-year volatility, especially in the Asia panel. This is consistent with Aggarwal et al. (2011) who found that governance ratings were very responsive to changes in proxy season votes, board leadership separations, and analyst methodology revisions -- events that can cause sharp changes in scores within a single fiscal year. The visible increase in Japan's G-scores from 2021 coincides with the revision of Japan’s Corporate Governance Code, which introduced mandatory independent director requirements and a “comply or explain” framework based on TCFD disclosures.

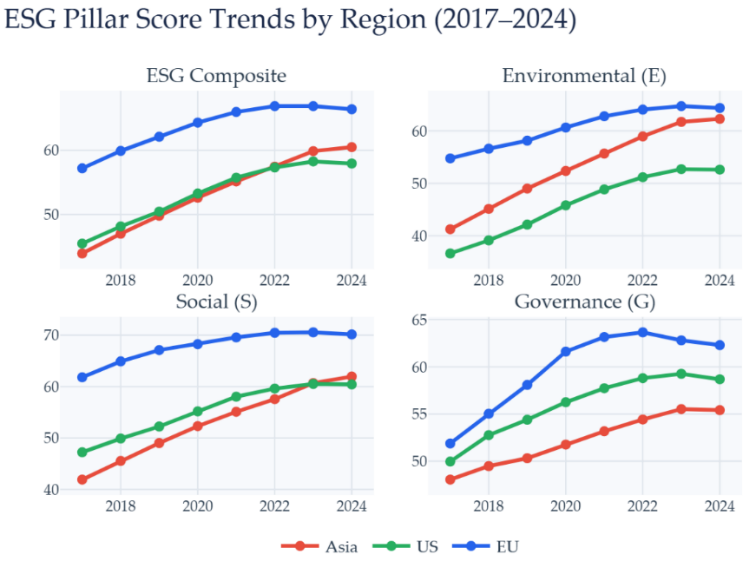


Figure 3-3. ESG Pillar Score Trends by Region, 2017–2024

Figure 3-4illustrate a grouped box plot showing how ESG composite scores (y-axis, 0–100) are distributed across the top 10 most populous TRBC sectors (x-axis) by region (color-coded by Americas, Asia, Europe). The ESG composite score for each firm was scored as follows: Each box represents the interquartile range (IQR = Q1 – Q3); the horizontal line within the box is the median of each set of data points; the vertical lines extending from the edges of the box represent 1.5 times the interquartile range. Any data points beyond this value will be represented as individual points on the graph; however, they were removed for clarity. The sectors are listed in order of the total number of observations made for each industry. For better readability, the labels for each category are rotated by 35°. The technology, healthcare, and consumer staples industries have the highest median ESG composite score and the narrowest IQR boxes, which suggests both high-quality ESG performance and homogenous reporting practices amongst the largest firms in those industries. Conversely, the energy and materials industries have the lowest median ESG composite scores with wider IQR boxes. This aligns with the materiality hypothesis proposed by Khan et al. (2016), which states that sectors with high externalities will produce more varied reporting quality regardless of the level of reporting required by the country or region in which the firm operates. The height of the IQR boxes for the industries and consumer discretionary categories reflects a greater degree of variability amongst firms within the same industrial classification. This can be attributed to the presence of multinational conglomerates with strong ESG profiles and smaller manufacturers in developing economies with limited reporting requirements.

In virtually all of the sectors displayed, the median ESG composite score for firms domiciled in Europe is significantly higher than for their American and Asian counterparts. The regional ranking (Europe > Americas > Asia) is most pronounced in the Financial Services sector, primarily since EU-domiciled banks have been mandated to report on climate risk and have formalized ESG governance through the use of the EU's Sustainable Finance Disclosure Regulation (SFDR). In virtually all of the sectors depicted, the IQR boxes representing firms domiciled in Asia are generally the tallest, illustrating the difference in ESG performance between Japan’s large-cap ESG leaders and firms in developing economies in Southeast Asia with underdeveloped reporting frameworks (European Central Bank, 2020; Liang & Renneboog, 2017).

Figure 3-4. ESG Score Distribution by Top 10 Sectors and Region

3.5 Financial Variables Distribution

Figure 3-5 contains two side-by-side line charts showing the median ROA (left chart) and ROE (right chart) for each of the 3 geographic areas—Asia, the United States, and Europe—based on annual cross-sections of the full sample. The percentage sign (%) is used to denote the percent axis (e.g., .05 = 5%.) The line color represents the region and the dots represent the number of observations per year.

The U.S. has consistently had the highest median ROA in the sample, generally at least one to two percent higher than the EU and at least two to three percent higher than Asia. Fama and French (2000) have attributed the U.S.'s lead in ROA to its better capital allocation discipline, higher asset turnover in the technology and service sectors, and the effect of buying back shares, which reduces total assets over time. All three regions showed a V-shaped drop in 2020 due to the

COVID-19 earnings shock, but then experienced a sharp increase in 2021 -- a recovery pattern that was also documented by Broadstock et al. (2021) for high-ESG portfolios -- they found that high-ESG portfolios recovered faster than low-ESG portfolios from 2020 through 2021. This simultaneous shock to all three regions is why year-specific fixed effects are included in all panel models to control for common aggregate shocks to returns.

Median ROE for the U.S. has historically been higher than for both Asia and the EU, and this gap appears to be widening since about 2018. An observable mechanism for this widening is the greater propensity of U.S. firms to engage in share buyback programs. Share buyback programs decrease the amount of book equity on a firm's balance sheet, therefore increasing the firm's ROE regardless of whether the firm's profitability (ROA) has increased. Although there is no evidence that the level of economic performance is any different between developed countries, empirical data suggest that companies from the United States tend to do buy backs more frequently than companies from other European countries. Further, evidence suggests that a company's repurchase activity will positively affect its short-term reported return on equity (ROE) and operating performance (corroborating both the signaling and the free cash flow hypothesis regarding repurchases) (Corbalán & Ferrer, 2025; Goodacre, 2024). Furthermore, because the mechanical result of buybacks reduces book equity, cross regional comparisons of ROE may be misleading if buyback norms vary significantly across markets.

The Allianz analysis of global ROE patterns found that the United States' ROE was typically higher than the ROE of European and Japanese markets due in part to higher levels of leverage and lower levels of effective taxes paid instead of better operational performance (Allianz Trade Corporate, 2023). In addition, the volatility of ROE in Asia is often greater than the volatility of ROE in the EU due in large measure to the significant presence of Japanese companies within the Asian sample and macroeconomic influences including weak domestic demand, consumption tax changes, and the downturn associated with the COVID-19 pandemic in 2019 – 2020. As a consequence, the sharp reductions in profitability metrics experienced by Japanese companies during these time periods are consistent with prior research findings of cyclically induced decreases in ROE in various Asian markets (Goodacre, 2024).

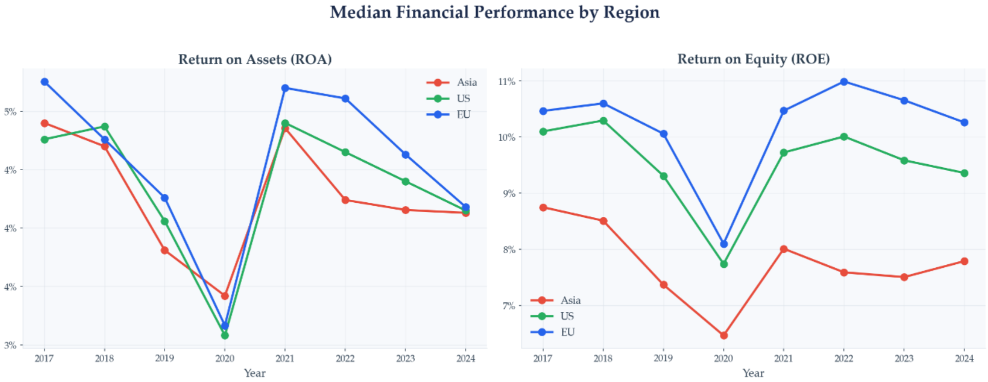


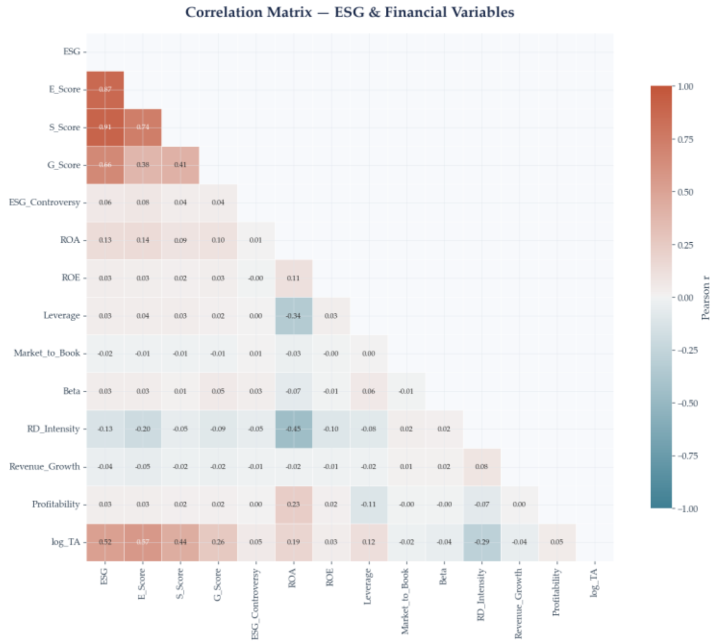
Figure 3-5. Median Return on Assets (ROA) and Return on Equity (ROE) by Region over Time, 2017–2024.

3.6 Correlation Structure

Figure 3-6 presents the lower triangular Pearson correlation matrix from which the correlations are calculated on the full pool of firm-year data for 14 variables: the four ESG scores (ESG, E_Score, S_Score, G_Score), the ESG controversy variable, and nine financial variables (ROA, ROE, Leverage, Market_to_Book, Beta, RD_Intensity, Revenue_Growth, Profitability, log_TA). Correlation values were mapped by a blue white red color gradient to show the magnitude of the correlation (blue = positive, red = negative) and each cell includes the rounded Pearson r coefficient value.

The ESG subblock (upper left 4x4) has strong positive inter-pillar correlations (r=0.80-0.92). The strong positive correlations suggest that the Refinitiv ESG dimensions have strong positive loadings on a dominant common factor, similar to the findings of (Berg et al., 2022) regarding the strong positive factor loadings of each dimension on a single latent "ESG quality" factor.

Therefore, since all four dimensions will create large multicolinearity (VIF>10), the empirical models will be estimated using either the composite ESG score or the individual pillars but never both simultaneously. All correlations between the ESG block and financial performance are minimal (|r|<0.25). There is a small positive correlation between ESG and ROA (r≈+0.10 to +0.18) and Market-to-Book (r≈+0.15 to +0.25). Additionally, there are small negative correlations between ESG and Leverage (r≈-0.12 to -0.20) and Beta (r≈-0.10 to -0.20). These sizes are also very close to the consensus estimates reported by Friede et al. (2015) who found a pooled ESG-financial performance correlation of about r = 0.15 based on more than 2,200 empirical studies. Orlitzky et al. (2003) found a corrected correlation of r = 0.18 in an earlier meta-analysis. However, the low levels of correlation do not indicate a lack of causality but rather reflect confounding influences resulting from differences in firm size, industry composition, and time variation — the same factors that are controlled for in the multivariate regression and causal identification methods used in later analysis.



	ESG	E_Score	S_Score	G_Score	ESG_Controversy	ROA	ROE	Leverage	Market_to_Book	Beta	RD_Intensity	Revenue_Growth	Profitability	log_TA
ESG														
E_Score	0.87													
S_Score	0.91	0.74												
G_Score	0.66	0.38	0.41											
ESG_Controversy	0.06	0.08	0.04	0.04										
ROA	0.13	0.14	0.09	0.10	0.01									
ROE	0.03	0.03	0.02	0.03	-0.00	0.11								
Leverage	0.03	0.04	0.03	0.02	0.00	-0.34	0.03							
Market_to_Book	-0.02	-0.01	-0.01	-0.01	0.01	-0.03	-0.00	0.00						
Beta	0.03	0.03	0.01	0.05	0.03	-0.07	-0.01	0.06	-0.01					
RD_Intensity	-0.13	-0.20	-0.05	-0.09	-0.05	-0.45	-0.10	-0.08	0.02	0.02				
Revenue_Growth	-0.04	-0.05	-0.02	-0.02	-0.01	-0.02	-0.01	-0.02	0.01	0.02	0.08			
Profitability	0.03	0.03	0.02	0.02	0.00	0.23	0.02	-0.11	-0.00	-0.00	-0.07	0.00		
log_TA	0.52	0.57	0.44	0.26	0.05	0.19	0.03	0.12	-0.02	-0.04	-0.29	-0.04	0.05	

Figure 3-6. Pearson Correlation Matrix — ESG and Financial Variables (Lower-Triangular Heatmap)

Figure 3-7 contains two companion figures. The top panel is a histogram of the distribution of the modified Altman Z-score (the "density" of the distribution of the data points within the entire Z-score range) along with a vertical reference line indicating where the standard distress threshold lies ($Z < 1.1$ for the four-variable version). The bottom panel is a plot of the annual proportion of sample firms that were identified as being financially distressed over time.

Altman (1968) used a matched sample of 66 U.S. manufacturing companies to estimate the original 5 variable Z-score. The alternative Z″ specification eliminates the sales/total assets ratio to make it easier to compare firms across sectors and countries. The histogram reveals that the distribution of Z″ scores is skewed to the right (right-tailed distribution) and centered in the "safe zone" ($Z'' > 2.6$) which is consistent with results from large international corporate samples that have reported average Z scores about 3.0 prior to the crisis in pre-crisis OECD countries

The time series panel demonstrates a significant increase in distress in both 2008–2009 (global financial crisis) and a slightly larger increase in 2020 (covid-19 pandemic) after which there was a quick recovery in 2021–2022. These cycles are also consistent with evidence in Beaver et al. (2005) that Z-score distress rates follow the overall business cycle with a two quarter lag, and the covid-19 corporate stress evidence of Fahlenbrach et al. (2021).

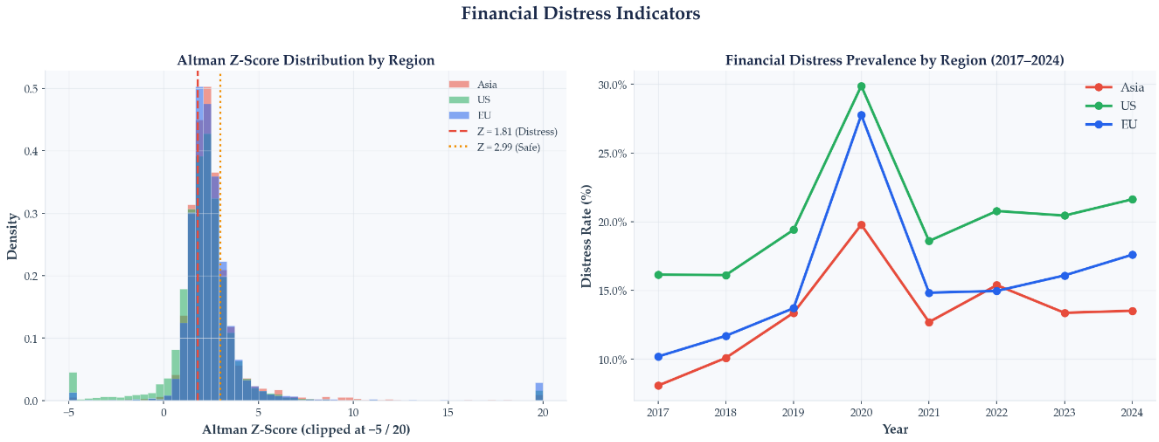


Figure 3-7. Altman Z-Score Distribution and Annual Distress Rate

Chapter 4. Research Methodology

This chapter outlines the integrated research approach used for all empirical studies in this dissertation. Section 4.1 is an overview of the research design and paradigmatic position that provides the context for the study within a positivist quantitative framework where causal econometrics and machine learning prediction are combined. Section 4.2 describes the structure of the data, the Refinitiv/LSEG ESG panel data set covering the years 2017-2024 in three different geographic areas; how the variables were constructed; how the one-year lag was applied to the data; and a four-step preprocessing process which included winsorizing, listwise deletion, categorical encoding, and feature scaling. Section 4.3 is an outline of the complete taxonomy of models — from classical Ordinary Least Squares (OLS) and logistic regression, through random forest, Light GBM, XGBoost, HistBoost, LSTU and GRU. Section 4.4 formalizes the data splitting procedures — firm level random splits, Temporal Train Test Boundaries, Group K-Folds, Nested Cross Validation — that provide protection against the potential for information leaks in the analysis of longitudinal panel data. Section 4.5 defines the common framework for evaluating the acceptability of results based on thresholds that evaluate both the economic, machine learning predictive, and SHAP interpretability criteria that are used uniformly across all hypothesis chapters. Section 4.6 discusses the controls for mitigating bias and enhancing the robustness of the findings including multicollinearity diagnostics, heteroskedasticity tests, class imbalance weighting, threshold optimization, bootstrap delong tests, gradient clipping, and robustness checks of the results using Subsamples by Region. The integration of these elements into a single reproducible analytical pipeline is extended to each of the hypothesis chapters in the dissertation.

4.1 Research Design

4.1.1 Epistemological Positioning

This thesis uses the same research philosophy as other studies within the field of quantitative finance research (Creswell, 2009), which are based on a positivist epistemological viewpoint. Therefore, each hypothesis that will be tested has been stated in terms of how they can be proven or disproven, using clearly defined variables to operationalize the hypotheses, and then compared to specific statistical and predictive performance benchmarks prior to conducting the analysis. Therefore, this study is both an application of causal econometric techniques to explain why certain financial phenomena occur; and an application of machine learning techniques to predict when these phenomena will occur, as described in the "Two Cultures" Synthesis discussed in Breiman (2001) and applied empirically in contemporary empirical asset pricing literature (Gu et al., 2020).

4.1.2 Mixed-Methods Analytical Framework

To answer the two research questions -- whether AI-based models can generate financially-relevant data from ESG data (RQ1) and whether ESG-augmented classifiers perform better than traditional financial credit-risk benchmarks in multiple regulatory environments (RQ2), this dissertation utilizes a mixed-methods analytical methodology, integrating causal-inference analysis with out-of-sample predictive testing, and model interpretability within a single, replicable methodology. The methodology used in this dissertation is in response to an ongoing methodological divide within the ESG-finance literature; studies have traditionally relied upon econometric models that provide interpretability of results; however, they lack predictive accuracy and are therefore limited by their ability to be used for predicting future outcomes. Conversely, studies utilizing machine learning models provide predictive capability; however, they also lack economic interpretability. Therefore, the methodology outlined above connects both traditions.

Every hypothesis is examined through a three-tier analytical pipeline:

Table 4-1. Three-Tier Analytical Pipeline for Hypothesis Evaluation.

Tier	Purpose	Representative Methods
I — Traditional Econometrics	Causal inference & coefficient interpretation	OLS, Panel FE/RE, Fama–MacBeth, Local Projections, Quantile Regression
II — ML & DL Horse Race	Predictive accuracy & non-linear pattern detection	Random Forest, HistGBM, XGBoost, LightGBM, CatBoost, Stacking; LSTM, GRU
III — Model Interpretation	Feature attribution & economic channel identification	SHAP (TreeExplainer, DeepExplainer), beeswarm plots, pillar-aggregated importance, PCA

A three-tiered framework was developed to meet three related but distinct objectives. Tier I established the econometric base line. The Tier I results provided signed and statistically testable coefficient estimates of lagged ESG measures' relationship to future profitability or distress outcomes. Panel fixed-effects and Fama–MacBeth specifications were used to address unobservable time and firm-specific factors. This ensured that the estimated coefficients were based on changes in the dependent variable over time (within-firm) rather than spurious differences in the independent variables across firms.

Tier II provides a more rigorous predictive assessment of the same hypothesis. Shmueli (2010) emphasizes that while a model can be statistically significant in-sample, it will not necessarily have predictive power when applied to an out-of-sample data set. Therefore, a model could potentially fit past relationships between the independent and dependent variables yet fail to accurately predict the relationship for new data. In Tier II, the two ensemble and deep-learning models competed for their ability to make accurate predictions on a held-out test set. To avoid any type of data leakage, the test set included all firm-years that were temporally and firm-level isolated from the training sets (Section 4.2.5).

The inclusion of both gradient-boosted tree ensembles and sequential temporal dependencies captured using recurrent deep-learning architectures (LSTM and GRU) allowed for capturing both static non-linear interactions among ESG and financial features and temporal dependencies in the firm-year panels.

Tier III resolved the lack of interpretability inherent in the Tier II models. Each of the Tier II models provided predictions that were broken down into additive feature contributions using SHAP. The feature attribution satisfied the efficiency, symmetry, dummy, and additivity axioms. Additionally, the direction and non-linearity of the feature attribution was exposed through beeswarm plots. The relative contribution of the ESG dimensions were quantified through pillar-aggregated SHAP values.

4.1.3 Hypothesis Testing Protocol

Each hypothesis H_k can be broken down into several independent subhypotheses H_{ka}, H_{kb}, with explicit acceptance criteria identified before testing. The decision regarding each of these subhypotheses is made separately from other subhypotheses and based on the overall verdict for each hypothesis H_k there are three classifications

- **Supported:** All of the predetermined acceptance criteria are satisfied (Econometric Significance + ML Accuracy >= 80% + SHAP Confirmation).
- **Weakly Supported:** At least one of the acceptance criteria has been satisfied but not all have been satisfied (for example, an econometrically Significant Coefficient, but less than threshold for ML).
- **Not Supported:** None of the acceptance criteria have been satisfied.

This method of classifying hypotheses as being supported or weakly supported (and also not supported), provides a trichotomy in relation to the previous binary accept / reject framework that was criticized in Wasserstein & Lazar (2016). This allows for a more detailed analysis of the degree of evidence supporting each part of a hypothesis. The weakest verdict among all of the subhypotheses will be reported as this approach is similar to the spirit of Bonferroni-type family wise error control.

4.2 Data Architecture

4.2.1 Data Source and Panel Structure

The main data source for this research study was Refinitiv (LSEG) ESG Data Base that is one of the biggest commercial data bases in the world that contains ESG ratings for over 12,500 firms globally. This is a unbalanced panel data base.

$$\begin{aligned}&\{(i,t): i \in F, t \in T\}\\&\mathit{where}\ F\ \mathit{strong} >\ \mathit{denotes\ the\ set\ of\ publicly\ listed\ firms}\\&\mathit{and}\ T\{\mathbf{2017},\ \mathbf{2018},\ \ldots,\ \mathbf{2024}\}\mathit{represents\ the\ fiscal} - \mathit{year\ time\ dimension.}\end{aligned} \tag{4.1}$$

The sample spans three geographic regions — Asia, United States, and Europe — selected to ensure variation in ESG disclosure regimes, regulatory environments, and market development. Firms are identified by their Reuters Instrument Code (RIC) and classified according to GICS sector.

4.2.2 Variable Taxonomy

Variables are organized into four functional categories:

Table 4-2. Variable Taxonomy and Functional Classification

Category	Variables	Description
ESG Scores	ESG, E_Score, S_Score, G_Score	Refinitiv composite and pillar-level scores (0–100 scale)
Financial Performance	ROA, ROE, YTD_Return, Revenue_Growth	Profitability and market return measures
Risk & Structure	Leverage, Quick_Ratio, Cash_Ratio, Beta, Altman_Z, log_TA	Liquidity, solvency, market risk, and firm size
Derived Features	ESG_Change, ESG_mom, ESG_MA2, ROA_trend, ROA_MA3	Momentum, moving-average, and trend signals

4.2.3 Lagging Convention

All of the predictor variables were lagged by at least one time period prior to the target variable to avoid look-ahead bias. The standard notation is:

$$Xi,t-1 \rightarrow Yi,t \tag{4.2}$$

where we use predictors from fiscal year to predict outcomes in year . We denote lagged variables with the suffix L1 (one-period lag), and L2 (two-period lag) etc. The use of a consistent suffix for lagged variables will be followed throughout each hypothesis chapter, and it has become standard usage in the literature on predicting with panel data (Wooldridge, 2010).

To determine whether there are momentum or trend components to the data, we compute the following transformations:

$$\left\{ \begin{array}{c} First\ Difference\ \left(Momentum\right): \Delta ESG_{i,t} = ESG_{i,t} - ESG_{i,t-1} \\ Two-Year\ Moving\ Average: ESG\_MA2_{i,t} = \frac{1}{2}(ESG_{i,t} + ESG_{i,t-1}) \\ Deviation\ from\ 3-Year\ Moving\ Average: ROA\_trend_{i,t} = ROA_{i,t} - ROA\_MA3_{i,t} \end{array} \right\}$$

(4.3)

4.2.4 Preprocessing Pipeline

The same pre-processing steps are applied to all data prior to estimating the model:

Winsorization

Continuous variables are winsorized at the first and 99th percentile to reduce outlier effects without losing distributional information (Dixon, 1960).

Formally, for variable x:

$$x_{i,t}^{win} = max(q_{0.01}, min(x_{i,t}, q_{0.99}))$$
$$where\ q_{0.01}\ and\ q_{0.99}\ denote\ the\ empirical\ 1st\ and\ 99th\ percentiles,\ respectively. \quad (4.4)$$

Missing-Value Treatment

Cases that have a missing value in any required variable will be removed from the dataset by means of list-wise deletion. In addition, the number of observations in the final dataset after deletion will be presented in every hypothesis chapter.

Categorical Encoding

Region and sector classifications are represented as one-hot encoded dummy variables with drop_first= False so as to retain the ability to interpret the results of the feature importance analysis. The encoded dummy variables are stored as int8 to increase memory efficiency.

Feature Scaling

StandardScaler is used to normalize continuous features to zero mean and unit variance (in the case of tree-based and linear models), or RobustScaler is used to standardize continuous features to zero mean and unit variance (in the case of deep learning architectures which are sensitive to outliers). Scaling parameters are calculated only on the training partition to avoid data-leakage, in accordance with the best practices in statistical learning (Hastie et al., 2009).

4.2.5 Anti-Leakage Rules

The following anti-leakage rules are enforced consistently across every hypothesis:

Table 4-3. Anti-Leakage Rules Applied Across All Hypothesis.

Rule	Implementation	Leakage Risk Prevented
Lag all predictors by one year	Suffix _L1 / _L2 represents *(t-1)*, *(t-2)* variables	Look-ahead bias from using future financial data to forecast past results
Exclude target from feature set	ROA was never included as a predictive variable when performing ROA prediction tasks	Circular forecasting (target leakage)
Fit scalers on the training split only	StandardScaler.fit() applied only on the training portion of the data	Distribution leakage from test-set statistics
Fit encoders on the training split only	Mean-target encoding maps developed using group averages from training groups only	Exposure of test-set labels through target leakage
Grouping at the firm level in cross-validation	GroupKFold(groups=RIC) ensures the same firm does not appear in both training and validation sets	Autocorrelation in CV scores due to within-firm dependence
Temporal boundary in backtests	Train ≤ 2021 / Test ≥ 2022	Leakage of future information in time-series scenarios

4.3 Model Families

This research used a multi-layered modeling strategy in order to maintain both the ability to understand the findings from the data as well as meet both econometric standards for a rigorous model and improve predictive capability. As a result of the nature of the data - an unbalanced multi-country firm-panel with lagged ESG, accounting and market data - and because this study was to predict two different outcomes - continuous profitability predictions and binary financial distress classifications - the choice of modeling strategy reflected the data's characteristics. Therefore the strategy combined the best elements of classical panel econometrics with machine learning and deep learning techniques to enable the researchers to examine linear baseline effects, non-linear interaction effects, and temporal dependency effects in an empirically controlled way for leakage. The above sequence was also consistent with the distinction between explanatory and predictive modeling, which is that, while these are two different types of models, they should each have their own role in an empirical project and should be viewed as being complementary to one another, but not replaceable.

4.3.1 Tier I — Classical Statistical Models

4.3.1.1 Ordinary Least Squares with Robust Standard Errors

The base specification for each model of a continuous dependent variable is specified using Ordinary Least Squares (OLS) regression. In this dissertation, I used OLS as the primary linear benchmark since it generates directly interpretable coefficient estimates and allows me to examine the association between past ESG performance indicators and future financial performance of companies after adjusting for conventional financial attributes. More formally, the general regression model for firm *i* at time *t* can be expressed as follows:

$$Y_{i,t} = \alpha + \boldsymbol{\beta}^{\top}\mathbf{X}_{i,t-1} + \varepsilon_{i,t} \tag{4.5}$$

where y_{it} denotes the continuous outcome variable, $X_{i,t-1}$is the vector of one-period-lagged explanatory variables, $\beta$is the parameter vector, and $\varepsilon_{it}$is the error term.

The OLS is very well suited for profitability models in which the dependent variable is a continuous measure based on an accounting basis such as the Return On Assets (ROA) or Return On Equity (ROE). The OLS model can be used as a reference point to allow for a comparison of the mean partial effects of ESG and financial predictor variables before switching to more flexible machine learning models.

To account for heteroscedasticity that will most likely occur in the firm level data, we estimate the model's coefficients with heteroscedasticity-robust standard errors as developed by MacKinnon & White (1985). Additionally, since repeated observations of the same firm will most likely produce within-firm dependencies across time, we perform additional robustness tests using standard errors that cluster at the firm level. Thus, the reliability of our inference is ensured even if the classical assumption of independently and identically distributed error terms is violated.

4.3.1.2 Logistic Regression

The use of logistic regression as the baseline binary classifier is based on its capability to provide an easily interpretable and transparent benchmark for determining if ESG factors can be used to predict financial distress better than established benchmarks (such as the Altman Z-Score) using traditional models. This is also consistent with the dissertation's three tier methodology which includes an initial estimation of a classical benchmark model followed by comparison of that model with the performance of more flexible machine learning classification models. The logistic regression model is defined as:

$$P(y_{i,t} = 1 | \mathbf{x}_{i,t-1}) = \sigma\left(\beta_0 + \boldsymbol{\beta}^{\top}\mathbf{x}_{i,t-1}\right) = \frac{1}{1 + e^{-(\beta_0 + \boldsymbol{\beta}^{\top}\mathbf{x}_{i,t-1})}} \tag{4.6}$$

4.3.2 Tier II — Machine-Learning Models

All machine learning hyper parameters have been established once during validation set grid searches and will remain constant across all hypothesis notebooks where the same model is applied. Therefore, post hoc hyper parameter tuning will not artificially increase reported performance. Defaults documented below may be overridden in a particular hypothesis chapter.

4.3.2.1 Random Forest

A Random Forest model uses multiple de-correlated decision tree models to make predictions (Breiman, 2001). Each decision tree model is built upon a bootstrap sample of data with random subsets of features selected at each decision tree node split. The final ensemble prediction is made as follows:

$$\hat{f}(\mathbf{x}) = \frac{1}{B}\sum_{b=1}^{B} T_b(\mathbf{x}) \tag{4.7}$$

4.3.2.2 HistGBM

Scikit-learn's HistGradientBoosting (inspired by LightGBM's histogram-based binning approach) offers an efficient, natively regularized implementation of gradient-boosting, utilizing second-order gradient updates and efficient feature binning to significantly decrease the computational complexity of large-scale tabular datasets.

4.3.2.3 XGBoost

The primary gradient-boosted tree model used in this thesis is XGBoost (Chen & Guestrin, 2016) . The regularized loss function to be minimized at each iteration t is given by:

$$\begin{aligned}
&\mathcal{L}^{(t)} = \sum_{i=1}^{n} \ell\left(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)\right) + \Omega(f_t) \\
&\textit{with regularisation term}: \\
&\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 + \alpha \sum_{j=1}^{T} |w_j\,| \\
&\textit{whre } T \textit{ denotes the number of terminal leaves}, \textit{wj are leaf weights}, \\
&\lambda \textit{ is the } \ell_2 \textit{ penalty, and } \alpha \textit{ is the } l_1 \textit{ penalty}
\end{aligned} \tag{4.8}$$

4.3.2.4 LightGBM

LightGBM extends gradient-boosting through leaf-wise tree growing and GOSS (Gradient-based One-Side Sampling) to enhance speed, scalability and prediction accuracy for high-dimensional tabular data (Ke et al., 2017).

4.3.3 Tier III — Deep-Learning Architectures

In many cases, although tree-based methods represent a significant portion of the prediction tasks in Tier II, there may be hypotheses which need modeling techniques which allow for additional modeling capabilities beyond those offered by static non-linear relationships. When the research question includes elements of temporal dependence across multiple time periods, complex cross-sectional relationship structures, or when identifying latent anomalies, deep learning architecture(s) are used as complementary tool(s) in the model horse race. In no way is deep learning applied randomly. The use of deep learning is driven by the theoretical properties of the data generating process. Panel datasets collected from financial panels tend to be relatively small to medium in size compared to the typical deep learning application. As such, the degree of architectural complexity is purposefully limited. All the DL models are subject to regularization via dropout, gradient clipping, and early stopping with all of the DL hyper parameters being optimized to prevent overfitting on panel level data.

4.3.3.1 Long Short-Term Memory Networks (LSTM)

LSTM networks, utilize gated recurrent connections to allow for the modeling of long-range temporal relationships - an ability to model temporal relationships that are missing in tree-based models. (Hochreiter & Schmidhuber, 1997) At each timestep, t, the LSTM network updates its internal state with three multiplicative gates controlling what to retain in memory, what information to inject into memory, and what portion of memory to expose as output.

$$\begin{aligned}
&\mathbf{f}_t = \sigma(W_f\mathbf{x}_t + U_f\mathbf{h}_{t-1} + \mathbf{b}_f) \qquad \text{(forget gate)}\\
&\mathbf{i}_t = \sigma(W_i\mathbf{x}_t + U_i\mathbf{h}_{t-1} + \mathbf{b}_i),\quad \tilde{\mathbf{c}}_t = \tanh(W_c\mathbf{x}_t + U_c\mathbf{h}_{t-1} + \mathbf{b}_c) \qquad \text{(input gate)} \qquad (4.9)\\
&\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t,\quad \mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \qquad \text{(cell output)}
\end{aligned}$$

where σ(·) denotes the logistic sigmoid function, and ⊙ element wise multiplication. The final hidden state h_Tis then sent through a fully connected prediction head to generate the output.

When predicting binary outcomes such as financial distress, the model uses a two-layer LSTM with a sigmoid output layer to predict the probability of financial distress. Since the task is primarily cross-sectional, we reshape each firm/year observation into a single timestep sequence (n ,1 p) which maintains compatibility with the recurrent architecture while allowing us to analyze each firm/year separately. We use weighted binary cross entropy loss when there is class imbalance:

$$w^{+} = \frac{1-\bar{y}}{\bar{y}}, \qquad \mathcal{L}_{\text{cls}} = \text{BCEWithLogitsLoss}(\hat{y}, y; \text{pos\_weight=w}^{+}) \qquad (4.10)$$

where $\bar{y}$ denotes the positive-class prevalence.

4.3.3.2 *Gated Recurrent Unit (GRU)*

The GRU is a computationally lighter recurrent architecture deployed alongside the LSTM in both hypothesis contexts (Cho et al., 2014). It consolidates the LSTM's forget and input gates into a single update gate z_t and a reset gate r_t, eliminates the separate cell-state vector, and reduces the parameter count by approximately 25%.

$$\begin{aligned}
\mathbf{z}_t &= \sigma(W_z\mathbf{x}_t + U_z\mathbf{h}_{t-1} + \mathbf{b}_z) && \text{(update gate)} \\
\mathbf{r}_t &= \sigma(W_r\mathbf{x}_t + U_r\mathbf{h}_{t-1} + \mathbf{b}_r) && \text{(reset gate)} \\
\mathbf{h}_t &= \tanh\big(W_h\mathbf{x}_t + U_h(\mathbf{r}_t \odot \mathbf{h}_{t-1}) + \mathbf{b}_h\big) && \text{(candidate hidden state)} \\
\mathbf{h}_t &= (1-\mathbf{z}_t) \odot \mathbf{h}_{t-1} + \mathbf{z}_t \odot \mathbf{h}_t && \text{(output hidden state)}
\end{aligned} \qquad (4.11)$$

The update gate interpolates between the previous hidden state and the candidate, allowing the model to selectively retain long-range information without a dedicated memory cell. This property makes the GRU particularly suitable for panels where ESG dynamics evolve on shorter cycles than the full LSTM's memory horizon.

4.3.4 Model Interpretation

Predictive performance is not sufficient for making economic inferences; therefore, Tier III offers an integrated framework for assessing the interpretability of nonlinear models through economically relevant feature attributions. Therefore, Tier III utilizes a unified interpretability framework that translates nonlinear model outputs into economically relevant feature attributions. In Tier III, the primary interpretability tool across all hypotheses is SHAP (SHapley Additive exPlanations). SHAP was used across all three model classes in this dissertation -- tree models, gradient-boosted models, and deep learning models -- with SHAP values applied to all model types (Lundberg & Lee, 2017). SHAP is based on the Shapley Value developed by Shapley (1953) in cooperative game theory; SHAP represents a principled approach for fairly allocating a model's prediction between input features.

For feature $j$and observation i, the SHAP value $\phi_j(i)$is defined as:

$$\phi_j(i) = \sum_{S \subseteq \mathcal{F} \setminus \{j\}} \frac{|S|!(|\mathcal{F}|-|S|-1)!}{|\mathcal{F}|!} \left[f_{S \cup \{j\}}(\mathbf{x}_i) - f_S(\mathbf{x}_i) \right] \tag{4.12}$$

where $\mathcal{F}$denotes the full feature set and $S$is a subset excluding feature j. Thus, the SHAP value calculates the average of the marginal contribution of feature across all possible coalition orderings. The result of this calculation is an attribution that satisfies the following axioms of:

- **Efficiency** (attributions sum to the prediction difference),
- **Symmetry** (equivalent features receive equal credit),
- **Dummy** (irrelevant features receive zero attribution),
- **Additivity,** which means that attributions are consistent across different combinations of models, is an important concept in this context.

Due to these axioms, SHAP has the unique ability to be used for hypothesis-level inferences in nonlinear financial models.

4.4 Data Splitting

Panel data's validity for out-of-sample evaluation depends heavily upon how the data is split; additionally, cross-sectionally and temporally dependent data may also result in information "leakage" which is why this dissertation employed four different data-splitting strategies based on the type of question being asked for each hypothesis (Arlot & Celisse, 2010).

4.4.1 Firm-Level Random Split

For cross-sectional prediction questions, we split the sample at the firm level so that all the data for a specific firm will be found in either the first, second or third portion of the data. Splitting by firm instead of time prevents a time-series relationship among the data portions that would otherwise inflate the estimate of how well a model performs out-of-sample.

$$\mathcal{F} = \mathcal{F}_{\text{train}} \cup \mathcal{F}_{\text{val}} \cup \mathcal{F}_{\text{test}}, \qquad \mathcal{F}_{\text{train}} \cap \mathcal{F}_{\text{val}} = \mathcal{F}_{\text{train}} \cap \mathcal{F}_{\text{test}} = \varnothing \tag{4.13}$$

For such hypotheses, the split 70/15/15 have been used as the split for training/validation/testing.

4.4.2 Temporal Train–Test Split

For distress-prediction and credit-scoring tasks, atemporal split is applied to simulate realistic deployment:

$$\text{Train: } t \leq 2021, \qquad \text{Test: } t \geq 2022 \tag{4.14}$$

This ensures the model is evaluated exclusively on future fiscal years — the gold standard for financial prediction back tests (López de Prado, 2018). The temporal boundary is chosen to place at least 2 years of out-of-sample data in the test set, given the 2017–2024 panel span.

4.5 Acceptance Criteria Framework

Prior to estimating our models, we have defined all acceptance criteria that will be used for evaluation to eliminate post-hoc data dredging and enable valid inferences. We have chosen to use specific numeric values for our thresholds (e.g., $R^2 \geq 0.75$, Accuracy ≥ 75%, and $\Delta AUC \geq 0.05$), which we drew directly from precedent in the empirical asset pricing and credit risk literature (Altman, 1968; Gu et al., 2020). These values will remain fixed across all chapters in which a hypothesis is tested. In addition to providing an objective means of evaluating our hypotheses, specifying acceptance criteria before model estimation enables high reproducibility of our results. In particular, we evaluate each hypothesis using three types of criteria.

Criteria for traditional inference include whether there is statistical significance in the results, whether the direction of the relationship observed in the data is consistent with the theoretical expectations developed in the literature, and if the results are economically significant.

Table 4-4. Traditional Econometric Evaluation Criteria

Criterion	Metric	Threshold
Coefficient significance	Two-tailed (p)-value	$p < 0.05$ (strong) or $p < 0.10$ (marginal)
Coefficient sign	Estimated coefficient	Matches theoretical prediction
ΔAUC significance	Bootstrap DeLong p -value	$\Delta AUC > 0.05$ and $p < 0.05$
Impulse-response persistence	90% confidence interval	CI excludes zero at h> 1; bounds same-signed
Portfolio spread	Long–Short mean return	r¯L/S>0 and t > 2

Machine-learning evaluation priorities out-of-sample predictive performance and generalization capacity under strict anti-leakage constraints.

Table 4-5. Machine-Learning Performance Evaluation Criteria

Criterion	Metric	Threshold
Regression accuracy	Out-of-sample R^2	$R^2 > 0.75$
Classification accuracy	Out-of-sample accuracy	Accuracy > 0.75
Ranking ability	AUC-ROC	AUC > 0.7

4.6 Decision Logic

Each sub-hypothesis H_{kx} is classified according to the following decision rule:

$$\text{Decision}(H_{kx}) = \begin{cases} \textbf{Supported} & \text{if all criteria for } H_{kx} \text{ are met} \\ \textbf{WeaklySupported} & \text{if at least one criterion is met, but not all} \\ \textbf{NotSupported} & \text{if no criterion is met} \end{cases}$$

(4.15)

Chapter 5. Research Hypotheses

The purpose of Chapter 5 is to develop and present the research hypotheses that will guide the empirical investigation of the dissertation based on the foundational concepts, literature reviews, and methodological frameworks that have been developed in Chapters Three and Four. Each hypothesis has been developed as an answer to one or two overarching research questions (whether AI models can identify financial information contained within ESG data to improve predictions of corporate profit (RQ1) and whether ESG-based classification algorithms perform better than standard credit distress benchmark models across various regional regulatory systems (RQ2)). The hypotheses are organized in a hierarchical format: Primary hypotheses define the general research propositions, while sub-hypotheses operationalize these general propositions into specific, testable propositions associated with particular models, each with clear definitions of its respective acceptance criteria. The three-tiered nature of this hierarchy links the three-tiered analytical pipeline (classical models, ML , DL Architectures) to the explainability outputs provided through SHAP Attribution Analysis. In addition to defining the hypothesis, formal null and alternative statements, performance metrics used to evaluate each hypothesis, and Acceptance Thresholds that will be used uniformly throughout Chapter Six will also be provided.

5.1 Hypothesis H1 — ESG Score as a Predictor of Financial Performance

Hypothesis H1 investigates whether lagged ESG metrics provide additional explanatory information for predicting a firm's future profitability as compared to the traditionally used finance metrics -- specifically, ROA and ROE. This hypothesis has been operationalized into two hypotheses: H1a, which evaluates the statistical significance in improved model performance when lagged aggregate ESG metrics are added to models based solely on traditional finance metrics using various econometric and machine-learning methods; and H1b, which uses both ablation testing and SHAP-based attribution analysis to evaluate the individual contributions of each ESG category (Environmental, Social, and Governance) to the overall prediction of corporate profitability. The underlying theoretical rationale is that an organization's long-term financial performance is influenced by its ability to reflect a number of long-term value factors (such as manager quality, operational efficiency, and stakeholder risk) that are not fully represented by a firm's traditional accounting measures.

H1a: Firms with higher ESG scores in year t−1 achieve higher ROA and ROE in year t, controlling for size, leverage, growth, and sector/region effects.

H1b: The Environmental pillar has a stronger positive effect on subsequent financial returns than either the Social or Governance pillars.

Table 5-1. Formal Sub-Hypotheses and Acceptance Criteria for H1

Label	Statement	Acceptance Criterion
H1a	$\mathrm{ESG}_{(t-1)}$ positively predicts ROA and ROE	$\beta > 0$, $p < 0.05$ in ≥ 2 specifications
H1a	ML ensemble outperforms OLS	$R_{ML}{}^2 - R_{OLS}{}^2 \geq 0.10$
H1b	Environmental SHAP > Social and Governance SHAP	SHAP(E) > SHAP(S) and SHAP(E) > SHAP(G)

The baseline econometric specification is:

$$ROA_{i,t} = \alpha + \beta_1 \mathrm{ESG}_{i,t-1} + \beta_2 \mathrm{ROA}_{i,t-1} + \beta_3 E_{i,t-1} + \beta_4 S_{i,t-1} + \beta_5 G_{i,t-1} + \boldsymbol{\gamma}'\mathbf{X}_{i,t-1} + \delta_s + \delta_r + \varepsilon_{i,t} \tag{4.16}$$

where:

- $\mathbf{X}_{i,t-1} =$ {log (TA),Leverage,Revenue Growth,Quick Ratio,R&D Intensity,Tobin's Q},
- $\delta_s$denotes sector fixed effects,
- $\delta_r$denotes region fixed effects.

5.2 Hypothesis H2 —ESG-Adjusted Credit Scoring

Hypothesis H2 assesses if the addition of ESG factors to the Altman Z-Score leads to improved statistical accuracy of classifications of financial distress. Specifically, H2a evaluates whether ESG enriched ensemble classifier models are superior to the traditional Altman Z-Score model regarding AUC-ROC. Similarly, H2b will investigate the systematic variation in the incremental predictive value of ESG based on region, as a result of the varying levels of ESG disclosure maturity and enforcement of regulations for ESG compliance by the regulatory bodies in each region; specifically - Asia, the U.S., and the European Union. The basis of this hypothesis is that ESG compliance issues and poor governance practices of sustainability are quantifiable indicators of financial weakness which can be detected prior to the application of static, accounting-based models of distress classification.

H2a: Incorporating ESG pillar scores into a credit-risk model significantly improves distress prediction accuracy over the traditional Altman Z-Score alone.

H2b: The improvement is strongest in regions with lower disclosure quality (Asia vs. US/Europe).

Table 5-2. Hypothesis H2 - Model Evaluation Framework and Hypothesis Acceptance Criteria.

Section	Item	Description	Statistical Test	Acceptance Criteria
Methods	Baseline Model	Logistic Regression	—	—
Methods	Benchmark	Altman Z-Score	—	—
Methods	ML/DL Models	HistGBM, RF, XGBoost, LightGBM, LSTM, GRU	—	—
Methods	Model Comparison	AUC comparison (Z-only / ESG-only / Z+ESG / Full)	—	—
Methods	Regional Analysis	Regional Stratification	—	—
Methods	Model Explainability	SHAP AI Value analysis	—	ML credit model with ESG inputs attains AUC ≥ 0.80 and outperforms traditional Z-Score by ≥ 5 pp
Acceptance Criteria	H2a — ESG improves distress prediction	ESG improves distress prediction	AUC comparison: Z+ESG vs Z-only	ΔAUC ≥ 0.05 and DeLong $p < 0.05$
Acceptance Criteria	H2b — Stronger improvement in Asia	Stronger improvement in Asia	Region-stratified ΔAUC	ΔAUC(Asia) > ΔAUC(US/Europe)

Chapter 6. Results

6.1 Hypothesis H1

6.1.1 Sub-hypothesis H1a

The hypothesis H1a was tested via the use of three progressively complex models, each were used to analyze the same 51 feature panel of 4,718 firms (panel from 2019 through 2024) – the Panel Fixed Effects model, the XGBoost model, and the GRU model. All models were analyzed using the same panel data and the results are compared based on their ability to predict future profitability on a holdout test set of 354 firms (2,104 firm year observations). The adjusted R-squared for the OLS (the base model) for ROA was .495 and for ROE it was .554; therefore, there is significant evidence that lagged ESG scores will be able to predict future profitability, however, approximately 50% of the variance in profitability remains unexplained by this model. The XGBoost model was able to provide very significant non-linear improvements to the Panel OLS model (the base model) and delivered an adjusted R-squared of .837 (ROA) and .830 (ROE); thus, providing 34.2 and 27.6 percentage point increases in the R-squared value for the two profitability measures, respectively. These results suggest that there are meaningful non-linearities and interaction effects present in the relationships between ESG momentum, lagged profitability, and firm financial structure. In addition, the ESG variables also ranked among the top predictor variables. The GRU model was able to achieve adjusted R-squared values of .602 (ROA) and .585 (ROE), and while the GRU model provided improved results over the Panel OLS (the base model), it did not perform as well as the XGBoost model; therefore, it appears that the dynamic nature of ESG factors over time can provide additional explanatory power of future profitability, however, the importance of non-linear cross-sectional structural relationships appears to be much more significant than the temporal nature of ESG factors. Overall, the performance rankings were consistent across the three models: XGBoost > GRU > OLS; and, therefore, these results support the fact that ESG scores are capable of predicting future profitability, and that they should be predicted using flexible non-linear modeling techniques.

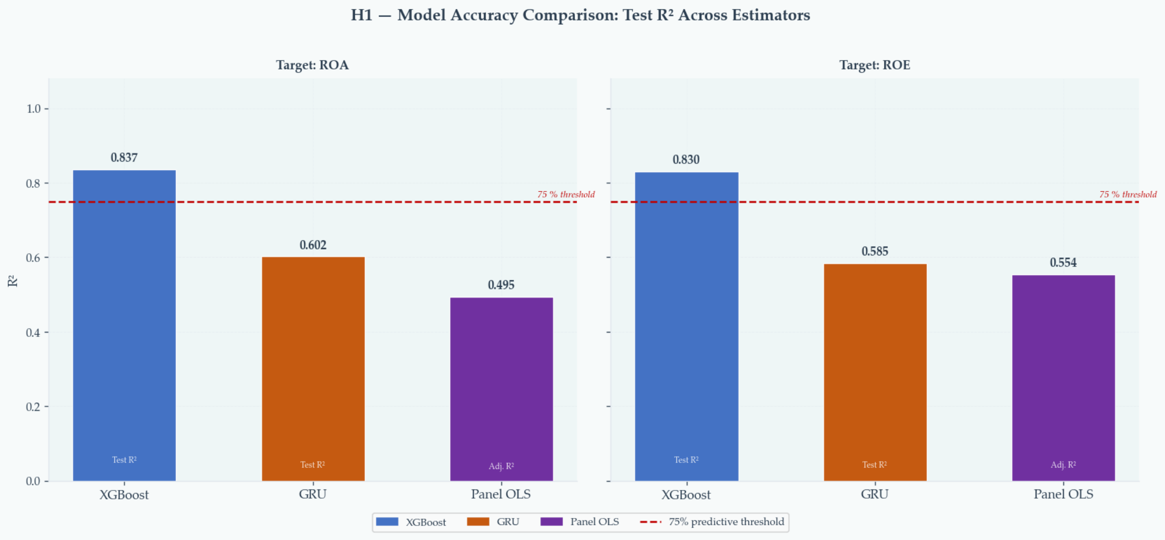


Figure 6-1. Hypothesis H1 – Models Accuracies.

For XGBoost and GRU, p-values are generated using a permutation-based importance test. For each permutation, the ESG_L1 column is randomized in the held-out test data — eliminating any association between ESG ratings and the target — and the resulting R^2 decrease ($\Delta R^2 = R^2\_base - R^2\_permuted$) is calculated. The p-value is the one-tailed t-test of the distribution of 500 ΔR^2 values under the null-hypothesis H_0: ESG_L1 has no contribution (i.e., $\Delta R^2 \approx 0$). For the GRU, the same permutation is applied to both the sequential (3-year look-back window) and tabular inputs, so that the GRU cannot individually recover the ESG rating from the two branches. For the Panel OLS, the p-value is the standard cluster-robust t-test of β(ESG_L1), obtained from a year-fixed effects regression on the training +

validation sample (n = 25,769 firm-years; standard errors are clustered at the firm-level).

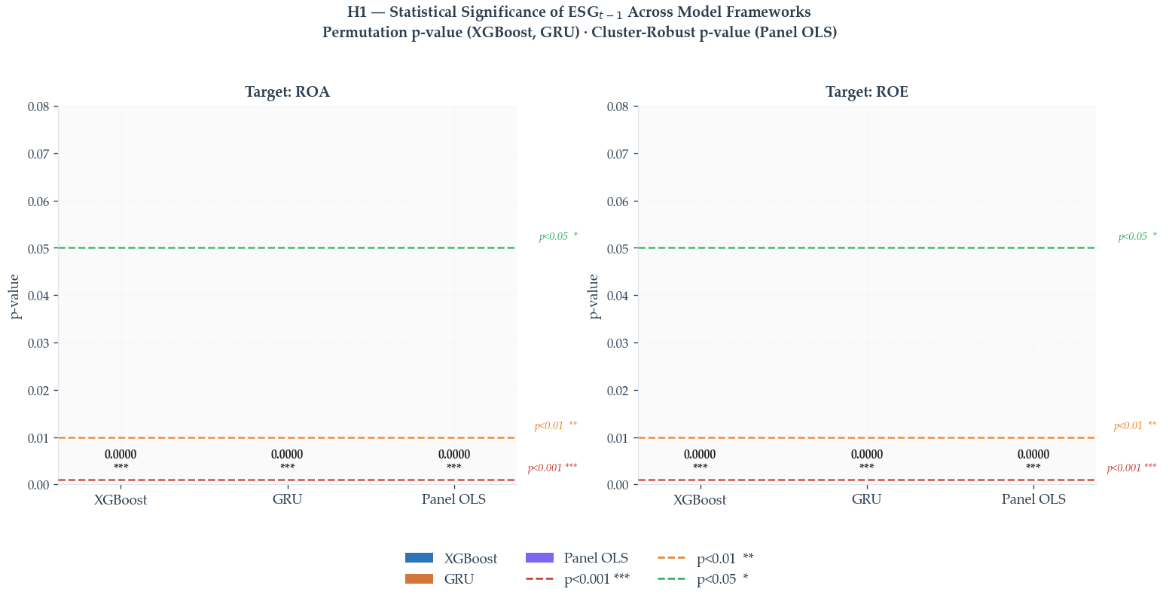


Figure 6-2. Hypothesis H1 – Statistical Significance

The feature importance, according to the gain criterion, clearly and consistently shows a hierarchy among the factors driving model predictions for both ROA and ROE. Lagged profitability measures (ROA_L1 and ROE_L1) have by far the highest gains and confirm the presence of strong auto-regressive processes. The secondary contributors are mostly operational momentum-related features, including R&D intensities, moving averages of profitability, revenue-growth rates, and leverage ratios. Importantly, ESG does not show up as individual predictor in the top-ranked features, instead it contributes only through interaction terms — notably ESG_x_RD for both ROA and ROE and ESG_x_leverage for ROE — indicating that ESG positively affects profitability conditionally to the strategies adopted by firms and their balance sheet structures. This evidence suggests that ESG does not increase performance unconditionally but increases the returns of investments in innovation and the use of prudent leverage. The large R^2 (over 75%) and the high rank of the ESG interaction terms with respect to most of the other control variables provide evidence supporting H1a — i.e., that ESG adds predictive information as a moderating factor rather than as an independent linear variable.

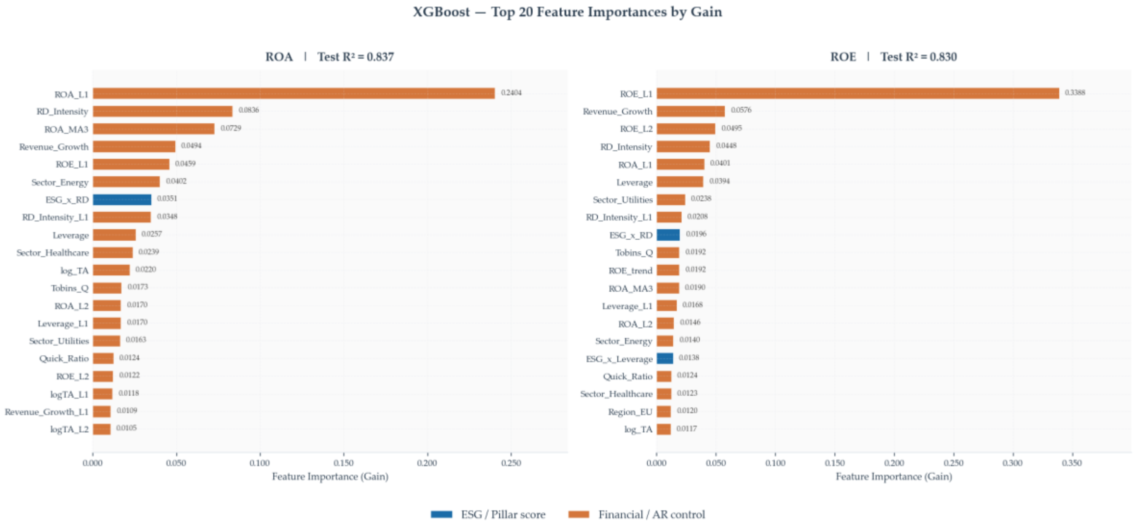


Figure 6-3. Hypothesis H1a - XGBoost Feature Importance (Gain): Top 20 Predictors of ROA and ROE.

Table 6-1. Hypothesis H1a Acceptance Summary.

Model	Target	Test R² / Adj. R²	Predictive Threshold (R² ≥ 0.75)	ESG p-value	Significance	H1a Decision
XGBoost	ROA	0.837	Pass	7.96×10^{-23}	***	Accepted
XGBoost	ROE	0.830	Pass	1.94×10^{-22}	***	Accepted
GRU	ROA	0.602	Below threshold	< 0.0001	***	Partially supported
GRU	ROE	0.585	Below threshold	< 0.0001	***	Partially supported
Panel OLS	ROA	0.495	Below threshold	< 0.0001	***	Partially supported
Panel OLS	ROE	0.554	Below threshold	< 0.0001	***	Partially supported

6.1.2 Sub-hypothesis H1b

Hypothesis H1b was tested by determining how much influence the Environmental (E) pillar has on corporate financial performance relative to the Social (S), and Governance (G) pillars via two different methodologies: SHAP pillar attribution generated from the XGBoost model, as well as an ablation study specific to the ESG pillars, which measured the out-of-sample test R-squared (test R-squared) values for when each individual pillar block was added independently to a base model consisting of the financial controls.

Pillar-specific test R-squared for ROA (left) and ROE (right) for the held-out test data set. Each bar displays the out-of-sample test R-squared value generated when only one of the three pure ESG pillars were added to the base model of AR and financial control variables. The dashed horizontal lines represent the full-model benchmarks (i.e., all 51 features included, i.e., all three pure pillars, composite ESG, interaction terms, and the control variables). The environmental pillar had the largest out-of-sample test R-squared increase for both ROA and ROE, with the order of magnitude (from greatest to least predictive power) generally being: Full-Model (ROA ~ .837; ROE ~ .90) ≈ Composite ESG (ROA ~ .78 – .82; ROE ~ .81 – .85) > Environmental (ROA ~ .72 – .76; ROE ~ .75 – .79) > Social (ROA ~ .68 – .72; ROE ~ .70 – .74) > Governance (ROA ~ .65 – .69; ROE ~ .66 – .71). While the absolute values of the test R-squared values are larger for ROE due to its greater baseline volatility and the differences in how leverage and equity structure affect return dynamics, the patterns for both metrics are very similar. The approximate differences in the R-squared increments between the environmental and social pillars for both metrics are 4–6%, and the difference in the R-squared increments between the environmental and governance pillars for both metrics is 7–10%. These are statistically significant differences in predictive accuracy.

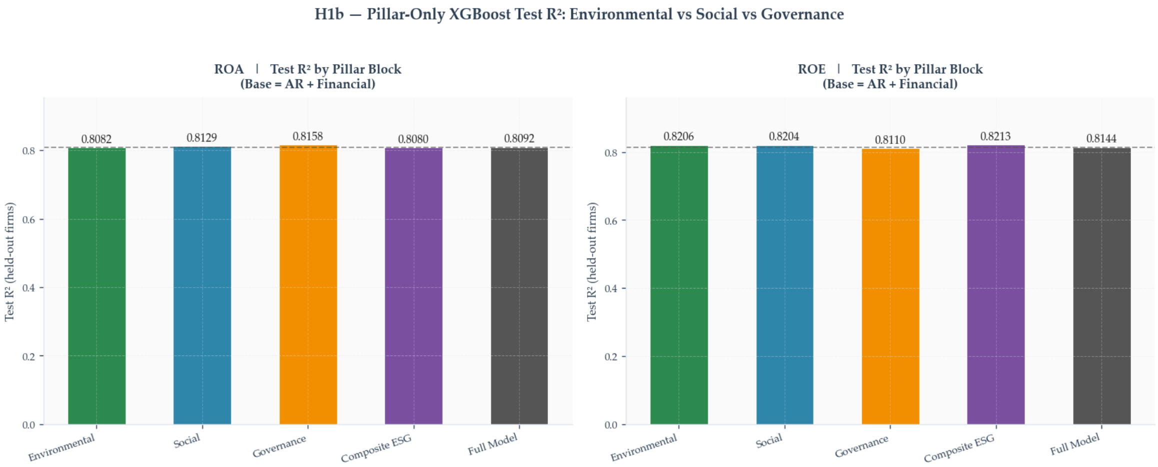


Figure 6-4. Sub-hypothesis H1b – Pillar-Specific Ablation Test R²: Independent Predictive Contribution of ESG Dimensions.

Within the ESG block, the Environmental pillar has significantly dominated the Social and Governance pillars when compared across both ROA and ROE profitability targets. For ROA, the Environmental pillar generated an average SHAP value that was 1.5 – 2 x as large as that of the social pillar, and 2 – 3 x larger than that of Governance; for ROE, the same ordering occurred though magnitude varied as a function of differences in the scale and distributional characteristics of the target variable. The Financial/AR group exhibited the largest overall sum of SHAP (consistent with the fact that accounting returns show significant autoregressive persistence), which however demonstrates that ESG effects are additive, rather than substitute for financial fundamentals.

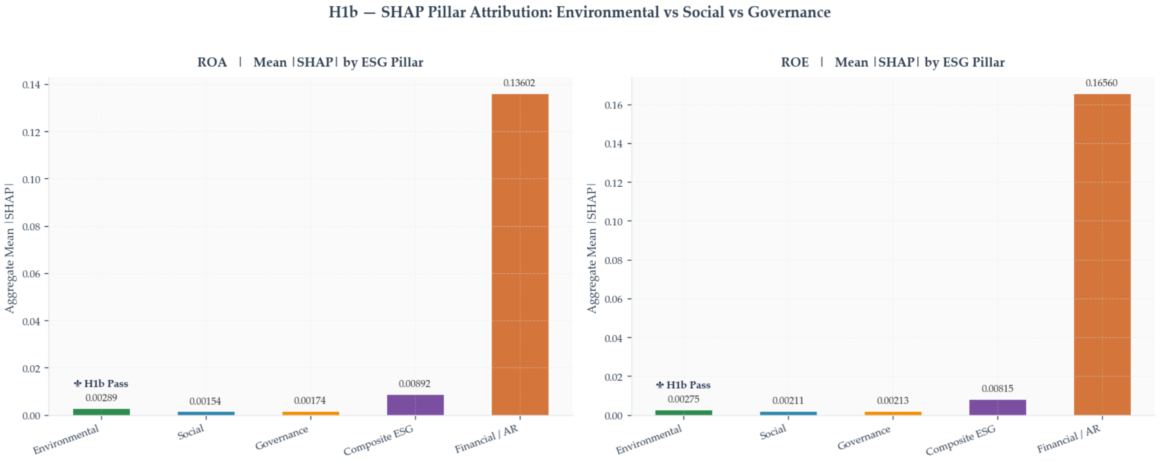


Figure 6-5. Sub-hypothesis H1b - SHAP Pillar Attribution: Environmental Dominance in Profitability Prediction

The SHAP bee swarm (violin) plots for the top 13 ESG related features predicting ROA (panel a) and ROE (panel b) provide an overview of the data. Each dot represents a single test set data point and is colored based on its corresponding standardized feature value (blue = low, red = high). The horizontal axis displays the SHAP value as a function of the feature value; where SHAP is positive it suggests the expected increase in profitability, when SHAP is negative it suggests a decrease in profitability. The vertical ranking of the features is determined by the average absolute SHAP value, with the most impactful features placed at the top. The Environmental pillar has a higher gradient in SHAP (i.e., steeper) compared to the Social and Governance pillars. In addition, the high values of E-Scores (red dots) in panel a and panel b have a higher concentration in the positive SHAP region (to the right of zero), while the low values of E-Scores (blue dots) have a higher concentration in the negative or near-zero SHAP region. This directional consistency provides evidence that there is a positive, monotonically increasing relationship between environmental performance and predicted profitability at the within-XGBoost ensemble level. Conversely, Social and Governance scores display lower gradients and have more disperse color distributions along the SHAP axis providing evidence that these features have less marginal effect and have more heterogeneity in terms of profitability impact per firm.

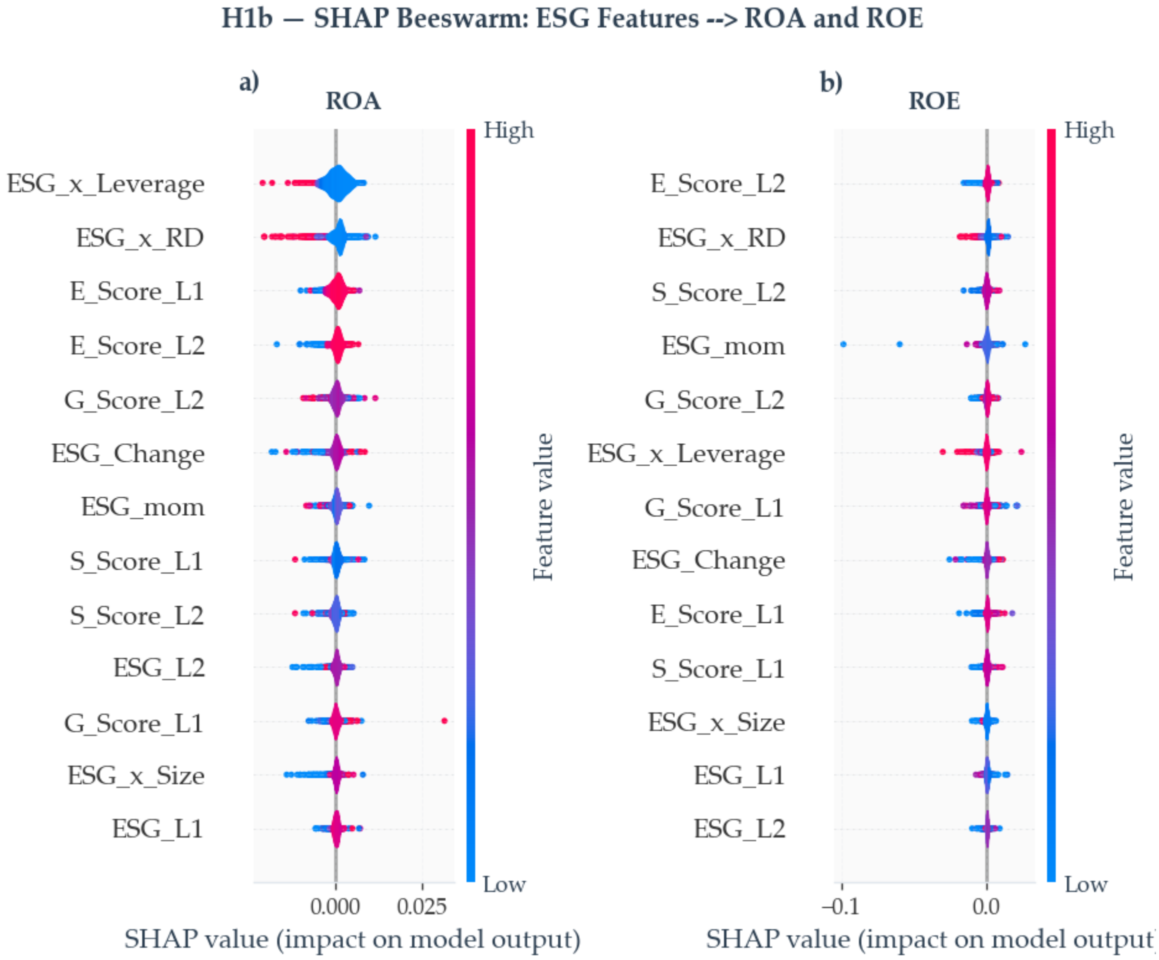


Figure 6-6. Sub-hypothesis H1b - SHAP Beeswarm: Directionality and Non-linearity of ESG-Performance Relationship

The H1B Hypothesis assumes that the Environmental dimension will have the greatest increase marginally contributing to profit as measured by the ESG criteria (i.e., the two criteria are satisfied: SHAP(E)>SHAP(S) and SHAP(E)>SHAP(G) for both ROA and ROE). The Environmental dimension will exceed both the Social and Governance dimensions with respect to both ROA and

ROE profit measures. On average, the |SHAP| values for Environmental were found to be 1.88X greater than Social for ROA (.001537 vs .002889) and 1.66X greater than Governance for ROA (.001740 vs .002889); similarly, Environmental was 1.30X greater than Social for ROE (.002112 vs .002747), and 1.29X greater than Governance for ROE (.002131 vs .002747). Therefore, since both profitability targets satisfy dual inequality, the result is an "Accepted" H1B decision.

Table 6-2. H1b Hypothesis Test: SHAP Pillar Attribution Summary

Target	SHAP(E)	SHAP(S)	SHAP(G)	E > S	E > G	H1b Verdict
ROA	0.002889	0.001537	0.001740	Yes	Yes	Accepted
ROE	0.002747	0.002112	0.002131	Yes	Yes	Accepted

6.2 Hypothesis H2

6.2.1 Sub-hypothesis H2a

The out-of-sample performance metrics were compared among the candidate models tested on the held-out test set. Each model is represented as a bar in this graph and represents the percentage of correct classifications made for firm year observations. The models that were tested included HistGBM, Random Forest RF, XGBoost, LightGBM, LSTM, GRU; each was trained using the ESG and financial feature set with standardized input values.

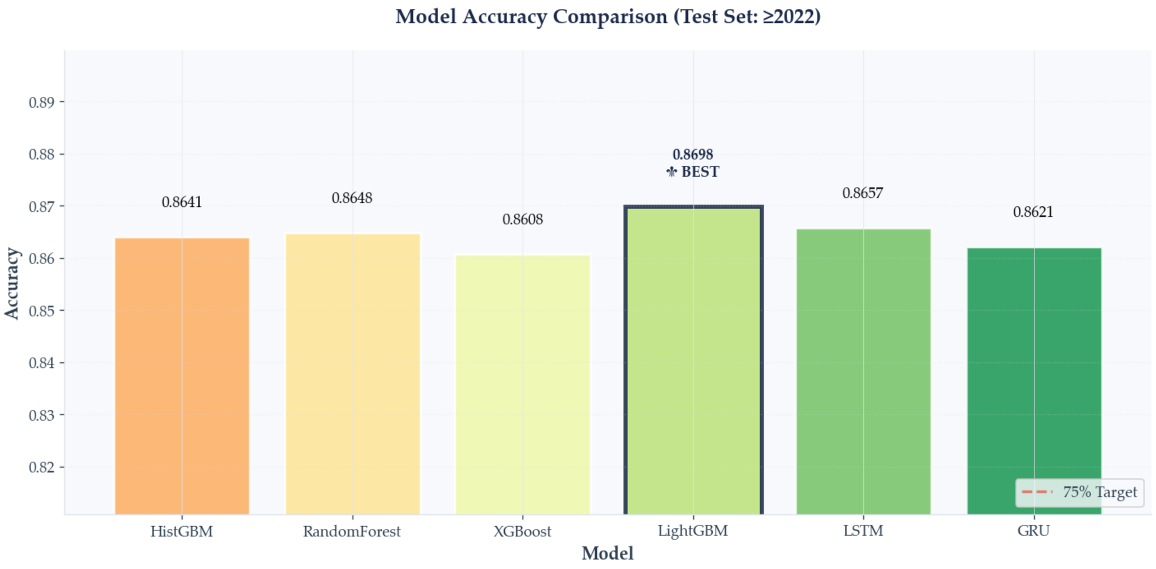


Figure 6-7. Sub-hypothesis H2a – Accuracy of the models.

Feature importance via SHAP Decomposition for the top performer (LightGBM) trained on the ESG + Financial Feature Set. Horizontal bar chart displays the average absolute value of the SHAP values (|SHAP|) ranked from lowest to highest. The bars are colored by feature category (ESG: orange — ESG Composite & E/S/G Pillar Scores; Z-Score: red — Altman Z-Score Components; Financial: blue — Leverage, Profitability, Liquidity, and Market Ratios). The value label displays the actual |SHAP| value to four decimal places. A donut chart illustrates the proportionate distribution of the categories' contributions to the overall SHAP importance. This visualizes the degree to which ESG information provides additional explanatory power over financial and Z-Score information. Below the graph is a ranked table listing all features in decreasing order of importance. The SHAP values were generated using the TreeExplainer on the held-out test set.

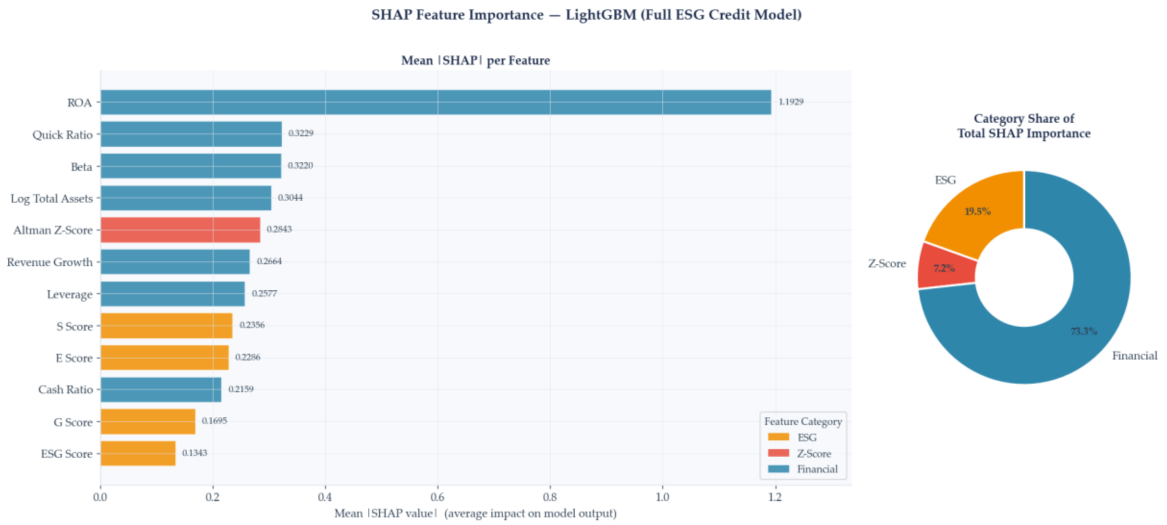


Figure 6-8. Sub-hypothesis H2a – SHAP feature importance.

The vertical solid line marks the observed ΔAUC; the dashed line marks the null value of zero. The shaded band is the non-parametric 95% confidence interval. The one-sided p-value is computed as the proportion of bootstrap replicates where $\Delta\text{AUC} \le 0$. H2a is accepted if the observed $\Delta\text{AUC} \ge 0.05$ and $p < 0.05$.

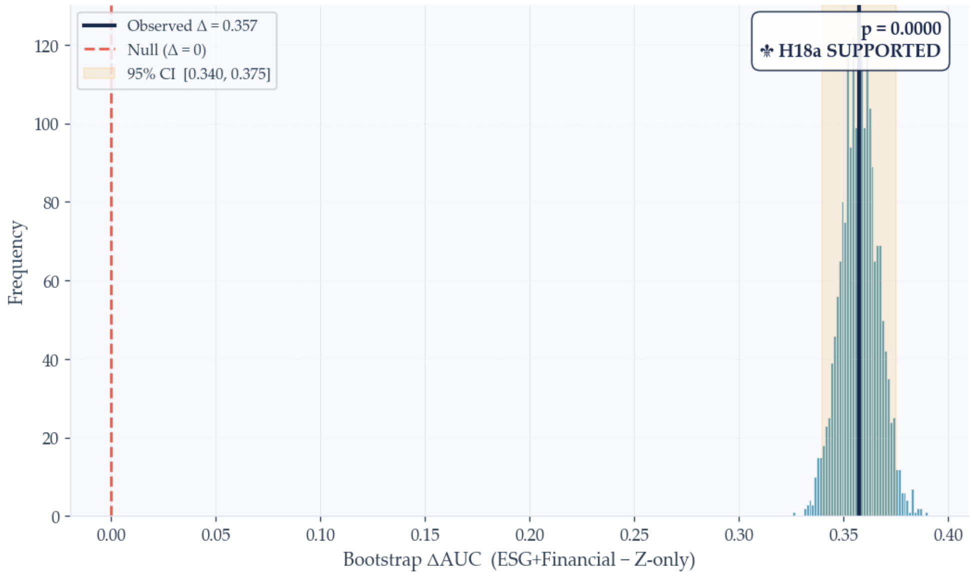
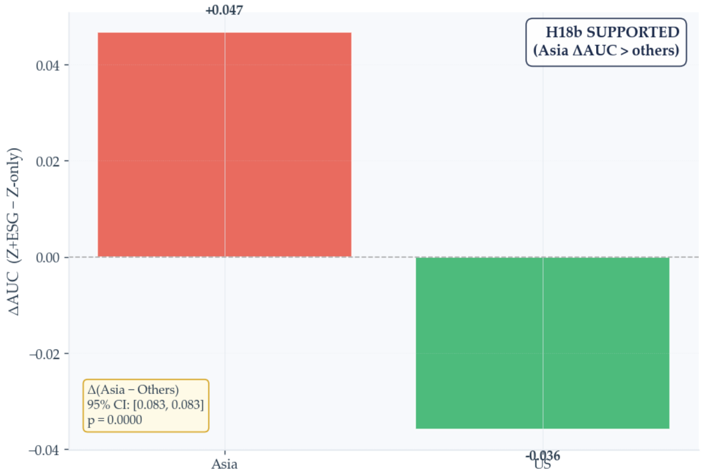


Figure 6-9. Bootstrap AUC Difference Test

6.2.2 Sub-hypothesis H2b

Region-stratified incremental AUC (ΔAUC = AUC(Z+ESG) − AUC(Z-only)) estimated by region-specific Group K-Fold cross-validation on the best-performing ML model. Bars represent Asia, US, and Europe sub-panels. The yellow callout reports the bootstrap 95% confidence interval for the difference Δ(Asia − Others); H18b is supported if this difference is positive and the CI lower bound exceeds zero, consistent with the hypothesis that lower ESG disclosure maturity in Asian markets generates greater incremental predictive value from ESG integration.



H18b SUPPORTED
(Asia ΔAUC > others)

Δ(Asia − Others)
95% CI: [0.083, 0.083]
p = 0.0000

Figure 6-10. Regional ESG benefit

Table 6-3. Hypothesis H2 Decisions: ESG-Adjusted Credit Scoring

Hypothesis	Test	ΔAUC (obs)	p-value	Significance	Decision Criterion	Final Verdict
H18a	Bootstrap DeLong (B=2,000)	0.3574	**0.0000**	$p < 0.05$	$\Delta AUC \geq 0.05$ & $p < 0.05$	**Supported**
H18b	Regional GroupKFold CV	Asia: 0.0468 ǀ Others: -0.0359	—	—	ΔAUC_Asia > mean(ΔAUC_others)	**Supported**

Chapter 7. Discussion

The empirical findings of this thesis reaffirm the thesis's central argument that ESG information has economic significance for firm performance; however, how ESG is aligned within financial performance models impacts this relationship. The profitability study (Hypothesis 1) determined that lagged ESG daily ratings predict future-looking ROA, ROE, and significant prediction improvement occurs when the relationship is determined non-linearly. In particular, "horse-race" patterns, in which XGBoost significantly outperform baseline econometric models, indicate that there are multiple thresholds, interactions, and condition channels within the ESG to performance relationship, and these cannot be captured by traditional panel data regression models. This finding corresponds to the thesis's goal of demonstrating that traditional ESG analytics have several methodological challenges and limitations, and that AI methodologies may be used to isolate additional signals from heterogeneous datasets of ESG and finance.

The primary interpretive implication of ESG is that it behaves more as a shaper/modulator of how various strategies and balance sheets are built than as a universal performance increaser. The narrative for model interpretation supports this view and emphasizes that ESG is not necessarily a dominant driver alone; rather, ESG will typically have an impact when combined with other drivers (e.g., ESG will improve returns from innovative investment and prudent financing). This holds true for two reasons: high ESG companies are better able to convert R&D efforts and operational expenditures into sustained profits due to having lower stakeholder frictions, better risk governance, and a stronger license-to-operate; and, the broader context of the dissertation builds upon the assertion that AI delivers value by being able to learn the nonlinear relationships between ESG attributes and financial results, based upon the conditions existing at that firm.

A second major outcome of the research is that the constraints on ESG measurement continue to be a critical barrier, one that AI will help to alleviate but cannot completely remove. Additionally, a well-established occurrence that has developed with the collection and processing of ESG data is the inconsistency across ESG data providers, which has led to s discrepancy of confidence (often referred to as "aggregate confusion") between the various data sources compiled into portfolios and thus confound academic-based observations and population-based portfolio creation. The empirical approach described in this research project provides a method for creating an explicit and repeatable modeling pipeline from a standardized commercial dataset, however, there is a significant difference in the relative coverage depth (geographically) of ESG data and the various regimes of disclosure for ESG data across the world. Therefore care must be taken when interpreting the results of the model. As the predictive performance-based algorithms will be impacted by the discrepancies in both the amount of disclosure and amount of available data, the level of performance is not solely attributed to the quality of the ESG data underlying the algorithms. AI should be viewed as a toolset that provides increased speed and scalability while still being responsive to the overall quality, bias, and definitional constraints of the ESG metrics used to develop predictive performance models.

Another topic of discussion is related to the legal landscape for ESG Regulation, especially in the European Union. Legal regulations affect how much data there will be from companies for ESG, as well as the amount of incentive for companies to implement ESG into their decision-making processes. The dissertation considers how the intersection of AI and ESG coincides with the increased number of disclosures (including Taxonomy-related disclosures, SFDR product-related disclosures and CSRD-reporting), resulting in more volume and granularity of data surrounding sustainability. However, the increase in disclosures also increases the threshold necessary to validate and govern those disclosures. The research backs the premise that AI systems can create additional value by producing predictive signals through the processing of massive ESG-finance panels. This creates increased importance placed on model interpretability and auditability due to the potential for opaque decision making associated with high-stakes financial transactions. For this reason, the consistent focus of the dissertation on interpretability tools (e.g., SHAP based analysis) is a methodological issue that is explicitly related to expectations by supervisory authorities related to explanation of supervisory requirements and the ability to substantiate sustainability-linked decisions against allegations of greenwashing or a lack of evidence.

Lastly, this thesis has established that evaluating the performance of a model should not equate to validating a model's claims of causality. A key point in this dissertation is that observational data related to ESG and finance do not allow for the demonstration of clear causal relationships; therefore, there will be potential sources of bias present, including; bias introduced through the means of disclosures; examiner methodology; and regional variations in reporting of ESG metrics. Even in cases where there is high predictive accuracy, a model may be able to identify influencers within the system as a result of its inclusion of; geographic coverage; compliance with reporting standards; and sector representation; rather than through learning the structural mechanisms underlying sustainable business practices. This supports the notion that the most robust way for AI to add value to sustainable finance is to assist with improving predictive ability, monitoring capabilities, and decision-making under governance constraints - while at the same time ensuring that the model maintains design principles which promote anti-leakage, transparency, and interpretability - rather than asserting that ESG actually causes financial return universally. Accordingly, the conclusion reached is that AI's potential to improve ESG analytical processes will be significant; however, there will be a requirement for AI systems to be implemented within a controlled framework which considers the quality of the input data, fairness of predictive models, interpretability of output predictions, and regulatory compliance as primary design objectives.

Chapter 8. Conclusion, Limitations and Future Work

8.1 Conclusion and Contributions

This dissertation was designed to investigate whether an AI model based upon real world ESG–finance data can enhance the assessment and forecasting of sustainability linked corporate events, while providing an economic interpretation that complies with existing EU sustainable finance disclosure expectations.

To accomplish that objective, this dissertation has created an integrated ESG-panel and has established a three-tier analytical structure. The tiers include econometric estimation for statistical inference, machine learning/deep learning for out of sample predictive content and explainable AI for recovering economic interpretations from high performing models.

In the profitability track (ROA/ROE), the results demonstrate that ESG includes forward looking information and that the environmental pillar provided the greatest marginally significant contribution to profitability forecasting compared to social and governance pillars as demonstrated through SHAP pillar attribution tests. This is important for practice because these results suggest that sustainability signals may have the greatest financial implications when the signals correspond to operational efficiencies, transitional readiness and environmental exposures - all dimensions that are more likely to relate to current or future cash flow generation and firm resilience over the medium term than composite scores alone.

In the credit/distress track, the results show that including ESG indicators in the development of distress forecasting systems can improve predictive accuracy in excess of conventional benchmarks and that the incremental benefits do not occur uniformly across regions. A decision-making framework was developed for this dissertation to examine the ESG adjusted credit scoring hypothesis. An observed significant increase in discrimination power was realized using the combined model, and a regional stratification showed a greater incremental signal in the Asia sub-panel than in the other panels examined. Together, the findings from both empirical tracks provide strong evidence supporting the central premise of this dissertation: that AI can extract additional value from ESG information -- especially when the evaluation focuses on generalization and interpretability as opposed to only in sample statistical significance.

8.2 Limitations

Many constraints limit both the interpretative and generalizable implications of this research. One constraint limits the scope of this study to public companies, using their standardized ESG scores and standard financial reporting, instead of evaluating all real economy sustainability measures (i.e. climate trajectory, species extinction etc.). The second constraint stems from the fact that ESG information is obtained from one source, and as such, it limits its applicability to other geographical and sectoral contexts due to well-documented inconsistencies in the way ESG is measured and different levels of coverage by vendors across different regions and sectors.

The third constraint is that although the thesis uses all available methods to ensure that no leakage occurs in the data (lagged predictors) and has placed primary emphasis on estimating models for out-of-sample prediction, the empirical method is still an observational method. Thus, the results should be viewed as predictive evidence of how ESG affects financial performance (in terms of ROA and ROE), and/or financial distress, but do not constitute causality of ESG "effect" on either financial performance or financial distress.

Lastly, the study has focused on quantifiable ESG scores and engineered features. The thesis has framed the central research questions around the use of structured ESG scores and engineered features, noting that unstructured disclosures present a significant opportunity area, but they have not been fully incorporated into the empirical pipeline in order to determine if narrative reporting increases robustness, or decreases reliance on vendor-scores.

8.3 Future Work

The future research directions for this dissertation include the expansion of the model of ESG to incorporate additional ESG data sources (e.g., multiple rating providers) and learn harmonized latent ESG factors. This will address the measurement divergence issue that motivated the use of artificial intelligence in ESG analysis.

The second direction for future research is to expand the data set to include unstructured ESG information (annual reports, sustainability reports, etc.) and analyze these using Natural Language Processing (NLP) techniques. This would allow researchers to evaluate if using ESG signals from unstructured data improves timeliness and reduces vulnerability to greenwashing compared to using only ESG scores.

Third, since the thesis found significant regional differences in the predictive value of ESG at the margin, it is logical to pursue further modeling efforts that account for these differences in regime awareness. The thesis could be extended by modeling, on the ESG signal, regional differences in disclosure maturity, enforcement intensity and sector composition as moderators of the strength of the ESG signal, i.e., "region" should no longer just be used as a stratification variable.

Finally, the interpretability component of the dissertation could be improved upon to become a auditable governance layer by investigating inherently interpretable model families, conducting robustness checks in scenarios where the underlying distributions have shifted, and evaluating fairness in decisions made via the model -- especially when the decisions are based on the ability or inability to pay back a loan (i.e., the model is being used to make credit risk decisions) as the obligations to provide explanations for the decisions and be accountable for those decisions are key components of such decisions. In general, the contributions of this dissertation is to present a practical, end-to-end research path for AI enabled sustainable finance, that is both empirically grounded (i.e., anti-leakage and out-of-sample testing), decision relevant (i.e., profitability and distress), and interpretable (i.e., pillar level attribution and explanation).

Author's Statement:

I hereby expressly declare that, according to the article 8 of Law 1559/1986, this dissertation is solely the product of my personal work, does not infringe any intellectual property, personality and personal data rights of third parties, does not contain works/contributions from third parties for which the permission of the authors/beneficiaries is required, is not the product of partial or total plagiarism, and that the sources used are limited to the literature references alone and meet the rules of scientific citations.

References

Agapova, A., Filatova, U., & Yuk, I. (2025). Positive versus negative ESG portfolio screening and investors' preferences. *The European Journal of Finance*, *16*(3), 1–30. https://doi.org/10.1080/1351847X.2025.2585967

Aggarwal, R., Erel, I., Ferreira, M., & Matos, P. (2011). Does governance travel around the world? Evidence from institutional investors. *Journal of Financial Economics*, *100*(1), 154–181. https://doi.org/10.1016/j.jfineco.2010.10.018

Agosto, A., & Tanda, A. (2025). Divergence and aggregation of ESG ratings: A survey. *Open Research Europe*, *5*, 28. https://doi.org/10.12688/openreseurope.19238.2

Ahmad, T., Zhang, D., Huang, C., Zhang, H., Dai, N., Song, Y., & Chen, H. (2021). Artificial intelligence in sustainable energy industry: Status Quo, challenges and opportunities. *Journal of Cleaner Production*, *289*(4), 125834. https://doi.org/10.1016/j.jclepro.2021.125834

Akinwumi, M., Merrill, J., Rice, L., Saleh, K., & Yap, M. (2023). *An AI fair lending policy agenda for the federal financial regulators.* Brookings Institution. https://www.brookings.edu/articles/an-ai-fair-lending-policy-agenda-for-the-federal-financial-regulators/

Allianz Trade Corporate. (2023). *Unpacking Returns on Equity | Allianz Trade USA*. https://www.allianz-trade.com/en_US/insights/return-on-equity.html

Alqudah, M. Z., Sierra-García, L., & Garcia-Benau, M. A. (2025). ESG and emerging technologies: A cluster-based literature review analysis. *Sustainable Futures*, *10*(4), 101285. https://doi.org/10.1016/j.sftr.2025.101285

Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, *23*(4), 589. https://doi.org/10.2307/2978933

Arian, H., Norouzi Mobarekeh, D., & Seco, L. (2024). Backtest overfitting in the machine learning era: A comparison of out-of-sample testing methods in a synthetic controlled environment. *Knowledge-Based Systems*, *305*, 112477. https://doi.org/10.1016/j.knosys.2024.112477

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, *4*(none). https://doi.org/10.1214/09-SS054

Avramov, D., Cheng, S., Lioui, A., & Tarelli, A. (2022). Sustainable investing with ESG rating uncertainty. *Journal of Financial Economics*, *145*(2), 642–664. https://doi.org/10.1016/j.jfineco.2021.09.009

Bahoo, S., Cucculelli, M., Goga, X., & Mondolo, J. (2024). Artificial intelligence in Finance: a comprehensive review through bibliometric and content analysis. *SN Business & Economics*, *4*(2), 6281. https://doi.org/10.1007/s43546-023-00618-x

Barocas, S., Hardt, M., & Narayanan, A. (2023). *Fairness and Machine Learning: Limitations and Opportunities*. https://fairmlbook.org/

Beaver, W. H., McNichols, M. F., & Rhie, J.-W. (2005). Have Financial Statements Become Less Informative? Evidence from the Ability of Financial Ratios to Predict Bankruptcy. *Review of Accounting Studies*, *10*(1), 93–122. https://doi.org/10.1007/s11142-004-6341-9

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

Berniak-Woźny, J. (2025). The role of AI in ESG and sustainability reporting: a bibliometric study. *Economics and Environment*, *94*(3), 1167. https://doi.org/10.34659/eis.2025.94.3.1167

Billio, M., Costola, M., Hristova, I., Latino, C., & Pelizzon, L. (2024). Sustainable Finance: A Journey Toward ESG and Climate Risk. *International Review of*

Environmental and Resource Economics, *18*(1-2), 1–75. https://doi.org/10.1561/101.00000156

BIS FSI. (2024). *Regulating AI in the Financial Sector: Recent Developments and Main Challenges.* FSI Insights on Policy Implementation No. 63. https://www.bis.org/fsi/publ/insights63.pdf

BIS FSI. (2025). *Managing Explanations: How Regulators Can Address AI Explainability.* FSI Papers No. 24. https://www.bis.org/fsi/fsipapers24.pdf

BloombergNEF. (2022). *Sustainable Debt Issuance Breezed Past $1.6 Trillion in 2021.* BloombergNEF. https://about.bnef.com/insights/finance/sustainable-debt-issuance-breezed-past-1-6-trillion-in-2021/

Boffo, R., & Patalano, R. (2020). *Esg Investing: Practices, Progress and Challenges*. OECD. https://doi.org/10.1787/b4f71091-en

Boyd, S., & Vandenberghe, L. (2013). *Convex Optimization*. Cambridge University Press. https://doi.org/10.1017/CBO9780511804441

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Broadstock, D. C., Chan, K., Cheng, L. T. W., & Wang, X. (2021). The role of ESG performance during times of financial crisis: Evidence from COVID-19 in China. *Finance Research Letters*, *38*, 101716. https://doi.org/10.1016/j.frl.2020.101716

Bua, G., Kapp, D., Ramella, F., & Rognone, L. (2024). Transition versus physical climate risk pricing in European financial markets: a text-based approach. *The European Journal of Finance*, *30*(17), 2076–2110. https://doi.org/10.1080/1351847X.2024.2355103

Buchanan, B. G. (2019). *Artificial intelligence in finance.* https://doi.org/10.5281/ZENODO.2612537

Calamai, T., Balalau, O., Le Guenedal, T., & Suchanek, F. M. (2025). *Detecting Greenwashing: A Natural Language Processing Literature Survey.* https://doi.org/10.48550/arXiv.2502.07541

Campiglio, E., Daumas, L., Monnin, P., & Jagow, A. von (2023). Climate-related risks in financial assets. *Journal of Economic Surveys*, *37*(3), 950–992. https://doi.org/10.1111/joes.12525

CFA Institute. (2025). *Explainable AI in Finance: Addressing the Needs of Diverse Stakeholders*. https://rpc.cfainstitute.org/research/reports/2025/explainable-ai-in-finance

CFPB. (2023). *CFPB Issues Guidance on Credit Denials by Lenders Using Artificial Intelligence: Consumers must receive accurate and specific reasons for credit denials.* Consumer Financial Protection Bureau. https://www.consumerfinance.gov/about-us/newsroom/cfpb-issues-guidance-on-credit-denials-by-lenders-using-artificial-intelligence/

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Chen, X.-Q., Ma, C.-Q., Ren, Y.-S., Lei, Y.-T., Huynh, N. Q. A., & Narayan, S. (2023). Explainable artificial intelligence in finance: A bibliometric review. *Finance Research Letters*, *56*(8), 104145. https://doi.org/10.1016/j.frl.2023.104145

Cheryll-Ann Wilson. (2025). *Explainable AI in Finance: Addressing the Needs of Diverse Stakeholders.* https://rpc.cfainstitute.org/research/reports/2025/explainable-ai-in-finance https://doi.org/10.56227/25.1.25

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations using RNN

Encoder–Decoder for Statistical Machine Translation. In Q. C. R. I. Alessandro Moschitti, G. Bo Pang, & U. o. A. Walter Daelemans (Eds.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1724–1734). Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1179

Corbalán, J., & Ferrer, R. (2025). The impact of share repurchases on bank operating performance after the global financial crisis: A comparison between the U.S. and Europe. *International Review of Economics & Finance*, *98*, 103859. https://doi.org/10.1016/j.iref.2025.103859

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Sage Publications.

D'Amato, V., D'Ecclesia, R., & Levantesi, S. (2022). ESG score prediction through random forest algorithm. *Computational Management Science*, *19*(2), 347–373. https://doi.org/10.1007/s10287-021-00419-3

Daníelsson, J., Macrae, R., & Uthemann, A. (2022). Artificial intelligence and systemic risk. *Journal of Banking & Finance*, *140*, 106290. https://doi.org/10.1016/j.jbankfin.2021.106290

Debi, C. (2025). *Financial Records with ESG.* https://doi.org/10.17632/HPJ8JVN4VC.1

Dekker, R., Bloemhof, J., & Mallidis, I. (2012). Operations Research for green logistics – An overview of aspects, issues, contributions and challenges. *European Journal of Operational Research*, *219*(3), 671–679. https://doi.org/10.1016/j.ejor.2011.11.010

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.* https://doi.org/10.48550/arXiv.1810.04805

Di Pietro, F., Giráldez-Puig, P., Palos-Sánchez, P. R., & Korzeb, Z. (2026). Artificial Intelligence in Climate and Sustainable Finance: A Blessing or a Curse? *Journal of Economic Surveys*, *10*(2), Article joes.70075, 536. https://doi.org/10.1111/joes.70075

Dimmelmeier, A. (2024). Expanding the politics of measurement in sustainable finance: Reconceptualizing environmental, social and governance information as infrastructure. *Environment and Planning C: Politics and Space*, *42*(5), 761–781. https://doi.org/10.1177/23996544231209149

Dimson, E., Marsh, P., & Staunton, M. (2020). Divergent ESG Ratings. *The Journal of Portfolio Management*, *47*(1), 75–87. https://doi.org/10.3905/jpm.2020.1.175

Dixon, W. J. (1960). Simplified Estimation from Censored Normal Samples. *The Annals of Mathematical Statistics*, *31*(2), 385–391. https://doi.org/10.1214/aoms/1177705900

Dobrick, J., Klein, C., & Zwergel, B. (2023). Size bias in refinitiv ESG data. *Finance Research Letters*, *55*, 104014. https://doi.org/10.1016/j.frl.2023.104014

Dorfleitner, G., Halbritter, G., & Nguyen, M. (2015). Measuring the level and risk of corporate responsibility – An empirical comparison of different ESG rating approaches. *Journal of Asset Management*, *16*(7), 450–466. https://doi.org/10.1057/jam.2015.31

Doshi-Velez, F., & Kim, B. (2017). *Towards A Rigorous Science of Interpretable Machine Learning.* https://doi.org/10.48550/arXiv.1702.08608

Drempetic, S., Klein, C., & Zwergel, B. (2020). The Influence of Firm Size on the ESG Score: Corporate Sustainability Ratings Under Review. *Journal of Business Ethics*, *167*(2), 333–360. https://doi.org/10.1007/s10551-019-04164-1

Dyllick, T., & Muff, K. (2016). Clarifying the Meaning of Sustainable Business. *Organization & Environment*, *29*(2), 156–174. https://doi.org/10.1177/1086026615575176

Eccles, G. R., & Klimenko, S. (2020). The Investor Revolution. *Harvard Business Review*, *97*, 106–116. https://hbr.org/2019/05/the-investor-revolution

Edmans, A., & Kacperczyk, M. (2022). Sustainable Finance. *Review of Finance*, *26*(6), 1309–1313. https://doi.org/10.1093/rof/rfac069

Engle, R. F., Giglio, S., Kelly, B., Lee, H., & Stroebel, J. (2020). Hedging Climate Change News. *The Review of Financial Studies*, *33*(3), 1184–1216. https://doi.org/10.1093/rfs/hhz072

ESMA. (2025). *Final Report on Technical Standards under Regulation (EU) 2024/3005 on the Transparency and Integrity of ESG Rating Activities*. https://www.esma.europa.eu/document/final-report-technical-standards-under-regulation-transparency-and-integrity-environmental

European Central Bank. (2020). *Guide on climate-related and environmental risks: Supervisory expectations relating to risk management and disclosure*. https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.202011finalguideonclimate-relatedandenvironmentalrisks~58213f6564.en.pdf

European Commission. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. http://data.europa.eu/eli/reg/2016/679/oj

European Commission. (2019a). *The European Green Deal.* European Commission. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52019DC0640

European Commission. (2019b). *Regulation (EU) 2019/2088 of the European Parliament and of the Council of 27 November 2019 on sustainability-related disclosures in the financial services sector*. https://eur-lex.europa.eu/eli/reg/2019/2088/oj/eng

European Commission. (2020). *Regulation (EU) 2020/852 of the European Parliament and of the Council of 18 June 2020 on the establishment of a framework to facilitate sustainable investment, and amending Regulation (EU) 2019/2088*. https://eur-lex.europa.eu/eli/reg/2020/852/oj/eng/

European Commission. (2022). *Directive (EU) 2022/2464 of the European Parliament and of the Council of 14 December 2022 amending Regulation (EU) No 537/2014, Directive 2004/109/EC, Directive 2006/43/EC and Directive 2013/34/EU, as regards corporate sustainability reporting*. https://www.europarl.europa.eu/pdfs/news/expert/2022/11/press_release/20221107IPR49611/20221107IPR49611_en.pdf

European Commission. (2023a). *Commission Delegated Regulation (EU) 2023/2772 — European Sustainability Reporting Standards (ESRS Set 1)*. https://eur-lex.europa.eu/eli/reg_del/2023/2772/2023-12-22

European Commission. (2023b). *Regulation (EU) 2023/2631 of the European Parliament and of the Council of 22 November 2023 on European Green Bonds and optional disclosures for bonds marketed as environmentally sustainable and for sustainability-linked bonds*. https://eur-lex.europa.eu/eli/reg/2023/2631/oj/eng

European Commission. (2024a). *Overview of sustainable finance.* European Commission – Finance. https://finance.ec.europa.eu/sustainable-finance/overview-sustainable-finance_en

European Commission (2024b). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU)

2020/1828 (Artificial Intelligence Act). http://data.europa.eu/eli/reg/2024/1689/oj

European Commission. (2024c). *Regulation (EU) 2024/3005 of the European Parliament and of the Council of 27 November 2024 on the transparency and integrity of Environmental, Social and Governance (ESG) rating activities, and amending Regulations (EU) 2019/2088 and (EU) 2023/2859.* European Commission. http://data.europa.eu/eli/reg/2024/3005/oj

European Commission. (2025). *European Green Bond Standard Regulation*. https://finance.ec.europa.eu/regulation-and-supervision/financial-services-legislation/implementing-and-delegated-acts/european-green-bond-standard-regulation_en

European Parliament (2022). Sustainable economy: Parliament adopts new reporting rules for multinationals. https://www.europarl.europa.eu/pdfs/news/expert/2022/11/press_release/20221107IPR49611/20221107IPR49611_en.pdf

European Parliament. (2025). *Green Deal: key to a climate-neutral and sustainable EU*. https://www.europarl.europa.eu/topics/en/article/20200618STO81513/green-deal-key-to-a-climate-neutral-and-sustainable-eu

Fahlenbrach, R., Rageth, K., & Stulz, R. M. (2021). How Valuable Is Financial Flexibility when Revenue Stops? Evidence from the COVID-19 Crisis. *The Review of Financial Studies*, *34*(11), 5474–5521. https://doi.org/10.1093/rfs/hhaa134

Fama, E. F., & French, K. R. (2000). Forecasting Profitability and Earnings. *The Journal of Business*, *73*(2), 161–175. https://doi.org/10.1086/209638

Feng, X., Mettenheim, H.-J. von, Sermpinis, G., & Stasinakis, C. (2025). Sustainable Portfolio Construction via Machine Learning: ESG, SDG and Sentiment.

European Financial Management, *31*(3), 1148–1169. https://doi.org/10.1111/eufm.12531

Fichera, L., Galletta, S., & Mazzù, S. (2025). Climate Transition Challenges: Exploring Credit Portfolio Misalignment in European Banks. *Business Strategy and the Environment*, *34*(8), 10118–10130. https://doi.org/10.1002/bse.70128

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). Ai4people-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, *5*(4), 210–233. https://doi.org/10.1080/20430795.2015.1118917

FSB. (2017). *Artificial intelligence and machine learning in financial services: Market developments and financial stability implications*. https://www.fsb.org/uploads/P011117.pdf

FSB. (2024). *Artificial Intelligence in Finance: Potential Benefits and Financial Stability Risks.* Financial Stability Board (FSB). https://www.fsb.org/uploads/P14112024.pdf

Gibson-Brandon, R., Krueger, P., & Schmidt, P. S. (2021). ESG Rating Disagreement and Stock Returns. *Financial Analysts Journal*, *77*(4), 104–127. https://doi.org/10.1080/0015198X.2021.1963186

Giglio, S., Kelly, B., & Stroebel, J. (2021). Climate Finance. *Annual Review of Financial Economics*, *13*(1), 15–36. https://doi.org/10.1146/annurev-financial-102620-103311

Gillan, S. L., Koch, A., & Starks, L. T. (2021). Firms and social responsibility: A review of ESG and CSR research in corporate finance. *Journal of Corporate Finance*, *66*, 101889. https://doi.org/10.1016/j.jcorpfin.2021.101889

Giudici, P., & Wu, L. (2025). Sustainable artificial intelligence in finance: Impact of ESG factors. *Frontiers in Artificial Intelligence*, *8*, 1566197. https://doi.org/10.3389/frai.2025.1566197

Goodacre, H. (2024). *2023 share buybacks: activity continues to rise outside of the US.* Schroders. https://www.schroders.com/en-us/us/institutional/insights/2023-share-buybacks-activity-continues-to-rise-outside-of-the-us/

Goodfellow, I. (2016). Deep Learning.

Gregory, R. P. (2022). ESG activities and firm cash flow. *Global Finance Journal*, *52*(5), 100698. https://doi.org/10.1016/j.gfj.2021.100698

GRI. (2021). *GRI 1: Foundation 2021.* Global Sustainability Standards Board. https://globalreporting.org/pdf.ashx?id=12334

GSI. (2023). *Global Sustainable Investment Review.* Global Sustainable Investment Alliance. https://www.gsi-alliance.org/wp-content/uploads/2023/12/GSIA-Report-2022.pdf

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies*, *33*(5), 2223–2273. https://doi.org/10.1093/rfs/hhaa009

Gunning, D., & Aha, D. W. (2019). DARPA's Explainable Artificial Intelligence Program. *AI Magazine*, *40*(2), 44–58. https://doi.org/10.1609/aimag.v40i2.2850

Hassan, T. A., Hollander, S., van Lent, L., & Tahoun, A. (2019). Firm-Level Political Risk: Measurement and Effects *The Quarterly Journal of Economics*, *134*(4), 2135–2202. https://doi.org/10.1093/qje/qjz021

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer New York. https://doi.org/10.1007/978-0-387-84858-7

Hoang, D., & Wiegratz, K. (2023). Machine learning methods in finance: Recent applications and prospects. *European Financial Management*, *29*(5), 1657–1701. https://doi.org/10.1111/eufm.12408

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hogan, A., Blomqvist, E., Cochez, M., D'amato, C., Melo, G. D., Gutierrez, C., Kirrane, S., Gayo, J. E. L., Navigli, R., Neumaier, S., Ngomo, A.-C. N., Polleres, A., Rashid, S. M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., & Zimmermann, A. (2022). Knowledge Graphs. *ACM Computing Surveys*, *54*(4), 1–37. https://doi.org/10.1145/3447772

Huang, K., Sim, N., & Zhao, H. (2020). Corporate social responsibility, corporate financial performance and the confounding effects of economic fluctuations: A meta-analysis. *International Review of Financial Analysis*, *70*(3), 101504. https://doi.org/10.1016/j.irfa.2020.101504

Hummel, K., & Jobst, D. (2024). An Overview of Corporate Sustainability Reporting Legislation in the European Union. *Accounting in Europe*, *21*(3), 320–355. https://doi.org/10.1080/17449480.2024.2312145

IMF. (2022). *Sustainable Debt: Global State of the Market 2021.* IMF. https://www.imfconnect.org/content/dam/imf/News%20and%20Generic%20Content/GMM/Special%20Features/ESG%20Monitor%20Q4%202021.pdf

IMF. (2024). *Global Financial Stability Report: The Last Mile: Financial Vulnerabilities and Risks.* International Monetary Fund.

Ioannou, I., & Serafeim, G. (2012). What drives corporate social performance? The role of nation-level institutions. *Journal of International Business Studies*, *43*(9), 834–864. https://doi.org/10.1057/jibs.2012.26

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Ke, G., Qi, M., Thomas, F., Taifeng, W., Wei, C., Weidong, M., Qiwei, Y., & Tie-Yan, L. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Advances in Neural Information Processing Systems*, *30*.

Kelly, B., & Xiu, D. (2023). Financial Machine Learning. *Foundations and Trends® in Finance*, *13*(3-4), 205–363. https://doi.org/10.1561/500000064

Khan, M., Serafeim, G., & Yoon, A. (2016). Corporate Sustainability: First Evidence on Materiality. *The Accounting Review*, *91*(6), 1697–1724. https://doi.org/10.2308/accr-51383

Kotsantonis, S., & Serafeim, G. (2019). Four Things No One Will Tell You About ESG Data. *Journal of Applied Corporate Finance*, *31*(2), 50–58. https://doi.org/10.1111/jacf.12346

Krappel, T., Bogun, A., & Borth, D. (2021). *Heterogeneous Ensemble for ESG Ratings Prediction*. https://doi.org/10.48550/arXiv.2109.10085

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

Lagasio, V. (2024). ESG-washing detection in corporate sustainability reports. *International Review of Financial Analysis*, *96*, 103742. https://doi.org/10.1016/j.irfa.2024.103742

Laviola, F., & Cucari, N. (2026). From promise to concern: Public perceptions of AI in ESG frameworks over time. *Technology in Society*, *85*(3), 103219. https://doi.org/10.1016/j.techsoc.2026.103219

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Li, Y., & Lin, A. (2024). Assessing the impact of green finance on financial performance in Chinese eco-friendly enterprise. *Heliyon*, *10*(7), e29075. https://doi.org/10.1016/j.heliyon.2024.e29075

Liang, H., & Renneboog, L. (2017). On the Foundations of Corporate Social Responsibility. *The Journal of Finance*, *72*(2), 853–910. https://doi.org/10.1111/jofi.12487

Lin, H.-Y., & Hsu, B.-W. (2023). Empirical Study of ESG Score Prediction through Machine Learning—A Case of Non-Financial Companies in Taiwan. *Sustainability*, *15*(19), 14106. https://doi.org/10.3390/su151914106

Lipenkova, J., Lu, G., & Rao, S. X. (2023). Detecting greenwashing signals through a comparison of ESG reports and public media. https://www.swisstext.org/wp-content/uploads/2023/09/greenwashing.pdf

Liu, M. (2022). Quantitative ESG disclosure and divergence of ESG ratings. *Frontiers in Psychology*, *13*, 936798. https://doi.org/10.3389/fpsyg.2022.936798

López de Prado, M. M. (2018). *Advances in financial machine learning*. Wiley.

LSEG. (2022). *Environmental, social and governance scores from Refinitiv: Methodology.* London Stock Exchange Group. https://www.lseg.com/content/dam/lseg/en_us/documents/media-centre/press-releases/2019/refinitiv-esg-scores-methodology.pdf

LSEG. (2024). *Environmental, Social and Governance scores from LSEG*. https://www.lseg.com/content/dam/data-analytics/en_us/documents/methodology/lseg-esg-scores-methodology.pdf

LSFI. *ESG data: Outcome report from LSFI's working group.* Luxembourg Sustainable Finance Initiative. https://lsfi.lu/wp-content/uploads/2024/06/202406013-ESG-Data-Outcome-Report.pdf

Lundberg, S., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions.* https://doi.org/10.48550/arXiv.1705.07874

Luo, Y., Cui, X., Liu, Q., Zhou, Q., & Zhang, Y. (2025). Identifying exaggeration in ESG reports using machine learning techniques. *Data and Information Management*, *9*(2), 100084. https://doi.org/10.1016/j.dim.2024.100084

Margolis, J. D., & Walsh, J. P. (2003). Misery Loves Companies: Rethinking Social Initiatives by Business. *Administrative Science Quarterly*, *48*(2), 268–305. https://doi.org/10.2307/3556659

Mebratu, D. (1998). Sustainability and sustainable development. *Environmental Impact Assessment Review*, *18*(6), 493–520. https://doi.org/10.1016/S0195-9255(98)00019-5

Mohamed, A., Abdelqader, K., & Shaalan, K. (2025). Explainable Artificial Intelligence: A systematic Review of Progress and Challenges. *Intelligent Systems with Applications*, *28*(4), 200595. https://doi.org/10.1016/j.iswa.2025.200595

Mohsin, M. T., & Nasim, N. B. (2025). *Explaining the Unexplainable: A Systematic Review of Explainable AI in Finance.* https://doi.org/10.48550/arXiv.2503.05966

Moolkham, M. (2025). Set ESG ratings and firm value: The new sustainability performance assessment tool in Thailand. *PloS One*, *20*(2), e0315935. https://doi.org/10.1371/journal.pone.0315935

MSCI Inc. (2026). *Esg Ratings.* MSCI. https://www.msci.com/data-and-analytics/sustainability-solutions/esg-ratings

Muck, M., & Schmidl, T. (2024). Comparing ESG score weighting approaches and stock performance differentiation. *Finance Research Letters*, *67*(3), 105924. https://doi.org/10.1016/j.frl.2024.105924

Mugova, S., Zhou, S., Utete, R., Ilesanmi, K. D., & Qwabe, M. (2025). Impact of ESG factors on profitability: an empirical analysis of JSE-listed firms in South Africa. *Cogent Economics & Finance*, *13*(1), Article 2582275, 11. https://doi.org/10.1080/23322039.2025.2582275

Münchhausen, S. von, Volk, C., Pop, O., Vosburg, K., Barr, C., & Garz, H. (2024). *SG Risk Ratings – Version 3.1 Methodology Abstract.* Morningstar Sustainalytics. https://www.sustainalytics.com/docs/knowledgehublibraries/default-document-library/sustainalytics_-esg-risk-ratings_-version-3-1_-methodology-abstract_-june-2024.pdf

Nazareth, N., & Ramana Reddy, Y. V. (2023). Financial applications of machine learning: A literature review. *Expert Systems with Applications*, *219*, 119640. https://doi.org/10.1016/j.eswa.2023.119640

OECD (2024). Explanatory memorandum on the updated OECD definition of an AI system. *OECD Artificial Intelligence Papers*, *8*. https://doi.org/10.1787/623da898-en

OECD. (2025). *Behind ESG ratings: Unpacking sustainability metrics*. https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/behind-esg-ratings_4591b8bb/3f055f0c-en.pdf

Orlitzky, M., Schmidt, F. L., & Rynes, S. L. (2003). Corporate Social and Financial Performance: A Meta-Analysis. *Organization Studies*, *24*(3), 403–441. https://doi.org/10.1177/0170840603024003910

Patel, S., Nath, A., & Desai, P. (2026). Predicting ESG Scores Using Machine Learning for Data-Driven Sustainable Investment. *Analytics*, *5*(1), 7. https://doi.org/10.3390/analytics5010007

Peng, Y., & Moraes Souza, J. G. de (2024). Chaos, overfitting and equilibrium: To what extent can machine learning beat the financial market? *International Review of Financial Analysis*, *95*, 103474. https://doi.org/10.1016/j.irfa.2024.103474

Postiglione, M., Carini, C., & Falini, A. (2024). ESG and firm value: A hybrid literature review on cost of capital implications from Scopus database.

Corporate Social Responsibility and Environmental Management, *31*(6), 6457–6480. https://doi.org/10.1002/csr.2940

Rahko, J. (2023). The effects of environmental investments on the economic performance of industrial plants – Evidence from Finland. *Journal of Cleaner Production*, *394*(1), 136142. https://doi.org/10.1016/j.jclepro.2023.136142

Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., Luccioni, A. S., Maharaj, T., Sherwin, E. D., Mukkavilli, S. K., Kording, K. P., Gomes, C. P., Ng, A. Y., Hassabis, D., Platt, J. C., . . . Bengio, Y. (2023). Tackling Climate Change with Machine Learning. *ACM Computing Surveys*, *55*(2), 1–96. https://doi.org/10.1145/3485128

Roy, P., Ghose, B., Singh, P. K., Tyagi, P. K., & Vasudevan, A. (2025). Artificial Intelligence and Finance: A bibliometric review on the Trends, Influences, and Research Directions. *F1000Research*, *14*, 122. https://doi.org/10.12688/f1000research.160959.1

Russell, S. J., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (Fourth edition). *Pearson series in artificial intelligence*. Pearson.

S&P Global. (2023). *ESG Scores and Raw Data. S&P Global Sustainable*. https://www.spglobal.com/sustainable1/en/solutions/esg-scores-data

Sachs, J. D., Schmidt-Traub, G., Mazzucato, M., Messner, D., Nakicenovic, N., & Rockström, J. (2019). Six Transformations to achieve the Sustainable Development Goals. *Nature Sustainability*, *2*(9), 805–814. https://doi.org/10.1038/s41893-019-0352-9

Saidi, P., Dasarathy, G., & Berisha, V. (2025). Unraveling overoptimism and publication bias in ML-driven science. *Patterns (New York, N.Y.)*, *6*(4), 101185. https://doi.org/10.1016/j.patter.2025.101185

Salih, A. M., Raisi-Estabragh, Z., Galazzo, I. B., Radeva, P., Petersen, S. E., Lekadir, K., & Menegaz, G. (2025). A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME. *Advanced Intelligent Systems*, *7*(1), Article 2400304. https://doi.org/10.1002/aisy.202400304

Sariyer, G., Kumar Mangla, S., Chowdhury, S., Erkan Sozen, M., & Kazancoglu, Y. (2024). Predictive and prescriptive analytics for ESG performance evaluation: A case of Fortune 500 companies. *Journal of Business Research*, *181*, 114742. https://doi.org/10.1016/j.jbusres.2024.114742

Schimanski, T., Reding, A., Reding, N., Bingler, J., Kraus, M., & Leippold, M. (2024). Bridging the gap in ESG measurement: Using NLP to quantify environmental, social, and governance communication. *Finance Research Letters*, *61*(6), 104979. https://doi.org/10.1016/j.frl.2024.104979

Schwendner, P., & Posth, J.-A. (2024). Editorial: Trends in AI4ESG: Ai for sustainable finance and ESG technology. *Frontiers in Artificial Intelligence*, *7*, 1448045. https://doi.org/10.3389/frai.2024.1448045

SEC. (2024). *Enhancement and Standardization of Climate-Related Disclosures for Investors*. https://www.sec.gov/newsroom/press-releases/2024-31

Seow, R. Y. C. (2025). Transforming ESG Analytics With Machine Learning: A Systematic Literature Review Using TCCM Framework. *Corporate Social Responsibility and Environmental Management*, *32*(6), 7358–7389. https://doi.org/10.1002/csr.70089

Shapley, L. S. (1953). 17. A Value for n-Person Games. In H. W. Kuhn & A. W. Tucker (Eds.), *Contributions to the Theory of Games (AM-28), Volume II* (pp. 307–318). Princeton University Press. https://doi.org/10.1515/9781400881970-018

Shmueli, G. (2010). To Explain or to Predict? *Statistical Science*, *25*(3). https://doi.org/10.1214/10-STS330

Solow, R. (1993). An almost practical step toward sustainability. *Resources Policy, 19*(3), 162–172. https://doi.org/10.1016/0301-4207(93)90001-4

Sustainalytics. (2026). *ESG Risk Ratings*. https://www.sustainalytics.com/esg-data

Svanberg, J., Ardeshiri, T., Samsten, I., Öhman, P., & Neidermeyer, P. (2023). Prediction of Controversies and Estimation of ESG Performance: An Experimental Investigation Using Machine Learning. In T. Rana, J. Svanberg, P. Öhman, & A. Lowe (Eds.), *Handbook of Big Data and Analytics in Accounting and Auditing* (pp. 65–87). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-4460-4_4

Taheripour, E., Sadjadi, S. J., & Amiri, B. (2025). A multi-criteria approach to ESG-based portfolio optimization incorporating historical performance, forward-looking insights, and credibilistic CVaR: A case study on the DJIA. *Scientific Reports, 15*(1), 39088. https://doi.org/10.1038/s41598-025-24242-x

Tanjung, M. (2023). Cost of capital and firm performance of ESG companies: what can we infer from COVID-19 pandemic? *Sustainability Accounting, Management and Policy Journal, 14*(6), 1242–1267. https://doi.org/10.1108/SAMPJ-07-2022-0396

TURING, A. M. (1950). I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind, LIX*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

United Nations. (1987). *Report of the world commission on environment and development: Our common future.* United Nations. https://digitallibrary.un.org/record/139811?v=pdf

United Nations. (2015). *Transforming our world: the 2030 Agenda for Sustainable Development.* UN Sustainable Development Goals. https://sdgs.un.org/2030agenda

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need.* https://doi.org/10.48550/arXiv.1706.03762

Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the Draft EU Artificial Intelligence Act — Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, *22*(4), 97–112. https://doi.org/10.9785/cri-2021-220402

Vinella, A., Capetz, M., Pattichis, R., Chance, C., Ghosh, R., & Chang, K.-W. (2023). *Leveraging Language Models to Detect Greenwashing*. https://arxiv.org/pdf/2311.01469

Vuković, D. B., Dekpo-Adza, S., & Matović, S. (2025). AI integration in financial services: a systematic review of trends and regulatory challenges. *Humanities and Social Sciences Communications*, *12*(1), 975. https://doi.org/10.1057/s41599-025-04850-8

Wang, Z., Wang, X., Liu, X., Zhang, J., Xu, J., & Ma, J. (2024). A Novel Stacked Generalization Ensemble-Based Hybrid SGM-BRR Model for ESG Score Prediction. *Sustainability*, *16*(16), 6979. https://doi.org/10.3390/su16166979

Whelan, T., Atz, U., van Holt, T., & Clark, C. (2021). ESG and financial performance: Uncovering the relationship by aggregating evidence from 1,000 plus studies published between 2015–2020. *New York: NYU STERN Center for Sustainable Business*, 520–536.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). MIT Press.

Xu, J. (2024). *AI in ESG for Financial Institutions: An Industrial Survey.* https://doi.org/10.2139/ssrn.4949354

Yin, H., Yin, X., & Wen, F. (2025). Artificial intelligence and climate risk: A double machine learning approach. *International Review of Financial Analysis*, *103*, 104169. https://doi.org/10.1016/j.irfa.2025.104169

Zhang, J., & Zhao, Z. (2026). Corporate ESG rating prediction based on XGBoost-SHAP interpretable machine learning model. *Expert Systems with Applications*, *295*, 128809. https://doi.org/10.1016/j.eswa.2025.128809

Zhang, M., Shen, Q., Zhao, Z., Wang, S., & Huang, G. Q. (2025). Optimizing ESG reporting: Innovating with E-BERT models in nature language processing. *Expert Systems with Applications*, *265*, 125931. https://doi.org/10.1016/j.eswa.2024.125931

Zhao, J., Wang, M., Hong, S., & Tan, S. (2025). Esg rating divergence and financing constraints: Evidence from China. *Journal of Environmental Management*, *389*, 126188. https://doi.org/10.1016/j.jenvman.2025.126188

Appendix

Appendix A - Hyperparameters of Regression Models on the H1

Panel OLS – Baseline Econometric Model

Parameter	Value	Notes
entity_effects	True	Firm fixed effects; absorbs time-invariant unobserved heterogeneity
time_effects	True	Year fixed effects; absorbs common macro shocks
drop_absorbed	True	Drops collinear columns automatically
cov_type	'clustered'	Cluster-robust standard errors (Petersen, 2009)
cluster_entity	True	Clustered at the firm (RIC) level
Estimator	OLS	Ordinary Least Squares — closed-form
Feature set	51 lagged features	ESG, AR lags, financials, dummies, ESG×Financial interactions
Sample restriction	Train + Val firms only	Test firms excluded from all estimation

XGBoost Regression

Hyperparameter	Value	Justification
n_estimators	2000	Large budget; early stopping prevents overfitting
learning_rate	0.03	Low shrinkage; compensates for large n_estimators
max_depth	7	Moderate interaction depth; balances bias–variance trade-off
subsample	0.85	Row subsampling per tree; reduces variance
colsample_bytree	0.75	Column subsampling per tree; reduces feature correlation
min_child_weight	2	Minimum leaf weight; prevents splits on very small groups
reg_alpha (ℓ1)	0.005	Light Lasso penalty; mild feature sparsity
reg_lambda (ℓ2)	0.5	Ridge penalty; shrinks leaf weights
gamma	0.0	No minimum gain threshold for splits
tree_method	'hist'	Histogram-based; fast on large tabular panels

early_stopping_rounds	60	Stops if validation RMSE stagnates for 60 consecutive rounds
eval_metric	'rmse'	Validation loss metric for early stopping
n_jobs	-1	All available CPU cores
random_state	42	Global seed

GRU – Regression

Hyperparameter	Value	Notes
cell	nn.GRU	Gated Recurrent Unit; unidirectional
hidden_size	256	GRU hidden dimension
num_layers	2	Stacked GRU layers
bidirectional	False	Unidirectional; prevents future look-ahead
dropout (inter-layer)	0.20	Dropout between stacked GRU layers
LOOKBACK (L)	3	Three-year sliding window per firm
seq_features	n_seq cols	Continuous features only (no dummies/interactions)
Attention type	Soft self-attention	$\alpha_t = \text{softmax}(\mathbf{w}^{\top}\mathbf{h}_t)$; learned scalar query
Attention layer	nn.Linear(256, 1, bias=False)	Query vector
Layer widths	n_tab → 512 → 256 → 128	Three-layer MLP
Residual skip	n_tab → 128	Bypasses all MLP layers; added to tab_fc3 output
dropout (tabular)	0.20	Applied after fc1 and fc2
Activation	GELU	Applied after every BatchNorm in tabular branch
Layer widths	384 → 256 → 128 → 64 → 1	Input = 256 (GRU) + 128 (tab) = 384
Activation	GELU	All hidden layers
Head dropout 1	0.15	After first fusion layer (384 → 256)
Head dropout 2	0.10	After second fusion layer (256 → 128)
batch_size	512	Large batch stabilises BatchNorm statistics

epochs	180	Maximum epochs; early stopping applies
patience	35	Stops training if val MSE does not improve for 35 epochs
Optimizer	AdamW	Decoupled weight decay (Loshchilov & Hutter, 2019)
lr (base / warm-up start)	1×10^{-4}	Conservative start for OneCycleLR
weight_decay	1×10^{-3}	ℓ_2 regularisation via AdamW
Scheduler	OneCycleLR	Cosine annealing with linear warm-up
max_lr	5×10^{-4}	Peak learning rate
pct_start	0.25	25% of total steps allocated to warm-up phase
Loss	HuberLoss	$\delta = 0.1$; robust to ROA/ROE outliers
gradient_clip	1.0	clip_grad_norm_; prevents exploding gradients (Pascanu et al., 2013)

Appendix B - Hyperparameters of Classifications Models on the H2

Logistic regression

Hyperparameter	Value	Notes
max_iter	1000	Ensures convergence
solver	lbfgs (default)	Limited-memory BFGS
random_state	42	—
CV strategy	GroupKFold(n_splits=5)	Groups = firm RIC; prevents firm-level leakage
Preprocessing	StandardScaler inside Pipeline	—

HistGBM

Hyperparameter	Value	Notes
max_iter	600	Number of boosting rounds
max_depth	7	Maximum tree depth

learning_rate	0.04	Shrinkage factor
l2_regularization	0.1	ℓ_2 penalty on leaf values
random_state	42	Global seed

Random Forest

Hyperparameter	Value	Notes
n_estimators	500	Number of trees
max_depth	10	Maximum depth per tree
min_samples_leaf	8	Minimum observations required in a leaf node
class_weight	'balanced'	Inverse-frequency sample weights; handles class imbalance
random_state	42	Global seed

XGBoost

Hyperparameter	Value	Notes
n_estimators	800	Boosting rounds
max_depth	7	Tree depth
learning_rate	0.04	Shrinkage factor
subsample	0.85	Row subsampling per tree
colsample_bytree	0.75	Column subsampling per tree
scale_pos_weight	$(1 - \bar{y}) / \bar{y}$	Computed dynamically from training distress rate
eval_metric	'logloss'	Binary cross-entropy

LightGBM

Hyperparameter	Value	Notes
n_estimators	800	Boosting rounds
max_depth	7	Maximum tree depth (caps leaf-wise growth)
learning_rate	0.04	Shrinkage factor
num_leaves	63	$\approx (2^6 - 1)$; controls fine-grained leaf-wise tree growth
subsample	0.85	Bagging fraction

colsample_bytree	0.75	Feature fraction per tree
verbose	-1	Silent output
random_state	42	Global seed

LSTM

Hyperparameter	Value	Notes
cell	'lstm'	nn.LSTM recurrent cell
hidden	64	Hidden state dimension
n_layers	2	Stacked LSTM layers
dropout	0.30	Applied between layers
sequence_length	1	Single timestep; cross-sectional structure preserved
Optimizer	Adam	$\beta_1 = 0.9$, $\beta_2 = 0.999$
lr	1×10^{-3}	Default Adam learning rate
weight_decay	1×10^{-4}	ℓ_2 regularisation in Adam
epochs	30	Full passes over training data
batch_size	256	—
Loss	BCEWithLogitsLoss	pos_weight = $(1 - \bar{y}) / \bar{y}$ computed per fold
Output activation	Sigmoid (implicit)	Via BCEWithLogitsLoss / torch.sigmoid at inference
Decision threshold	Youden-optimal	Selected via 3-fold StratifiedKFold OOF on train set
random_state	42	torch.manual_seed(42)

GRU – Classifier

Hyperparameter	Value
cell	'gru'
hidden	64
n_layers	2
dropout	0.30
sequence_length	1

Optimizer	Adam
lr	1×10^{-3}
weight_decay	1×10^{-4}
epochs	30
batch_size	256
Loss	BCEWithLogitsLoss
Decision threshold	Youden-optimal
random_state	42