Corporate credit rating feature importance: Does ESG matter?

Lachlan Michalski[a] and Rand Kwong Yew Low[a,b]

[a]UQ Business School, University of Queensland, Australia

[b]Bond Business School, Bond University, Australia

Abstract

We examine the inclusion of environmental, social, and governance (ESG) features, with commonly-used features such as accounting ratios, macroeconomic, and microeconomic variables, into multi-class corporate credit rating prediction. Tree-based ensemble feature importance and classification is investigated across quarterly United States (US) and global firms over the period 1982-2019. Using random forests and extremely randomized trees models, with mean decrease impurity, mean decrease accuracy and Tree SHapley Additive exPlanations feature importance methods, prediction accuracy is consistent across all methods. For US and global firms, Thomson Reuters Diversity and Inclusion Rating (TRDIR) metrics of diversity score, inclusion score, people development score, and score demonstrate high importance across all feature importance and tree-based ensemble methods. The constituents of these TRDIR metrics primarily comprise governance and social features. ESG features tend to be more important for investment-grade classes, whereas, solvency, profitability, financial soundness, and valuation financial ratios exhibit greater importance for speculative-grade classes. Overall, ESG features consistently and robustly matter for US and global firms corporate credit rating prediction.

1 Introduction

Corporate credit ratings are often the principal source of investor information for the quality and marketability of issued or issuing bonds (Pinches and Singleton, 1978; Pogue and Soldofsky, 1969). The determinants of corporate credit ratings are investigated extensively, however most studies differentiate with the features (variables) and models utilized. Financial, macroeconomic, and governance factors are highlighted as the main determinants (traditional features) of corporate credit ratings using conventional tools (Ashbaugh-Skaife et al., 2006; Blume et al., 1998; Ederington, 1985), with logistic regression (LR) primarily used for class prediction (Hajek and Michalak, 2013; Kamstra et al., 2001).

Environmental, social, and governance (ESG) features are commonly investigated in equities research (e.g., Fan and Michalski (2020) and Pedersen et al. (2020)), however, the presence within corporate credit research is limited (Attig et al., 2013). In 2017, Morningstar acquired a 40% stake in ESG rater, Sustainalytics, and acquired the remaining 60% in July 2020. Moody's purchased ESG provider, Vigeo Eiris in April 2019, whilst in November 2019, Standard & Poor's (S&P) Global acquired the ESG arm of RobecoSAM.

Hachenberg and Schiereck (2018) investigate whether green bonds are priced differently from conventional bonds, and illustrate their is no real marginal difference. Tang and Zhang (2020) demonstrate that the main advantage of green bonds is not cheaper debt financing, rather increased institutional ownership and improved stock liquidity after issuance. Moreover, Kiesel and Lücke (2019) find that ESG consideration is a significant determinant in the stock return and credit default swap (CDS) spread around the rating announcement, with corporate governance playing the most important role.

As green bond issuance and sustainability-linked loans can attract increased media exposure, in addition to impact investors, large, and smaller institutional investors holding these to satisfy their investment mandates, investigating which ESG features are drivers of corporate credit predictability has important real world implications in capital markets functioning (e.g., finding mispriced bonds).

Articles between December 2019 and November 2020 in *The Economist* highlight the importance of diversity inclusion, carbon emissions, resource reduction, and employee training for firms.[1]

[1] Articles include: Social unrest has fuelled a boom for the diversity industry (28 November 2020), How much can financiers do about climate change? (20 June 2020), and, Climate change has made ESG a force in investing (7

Consequently, to bring clarity to the ESG debate, we introduce ESG features into United States (US) and global firm samples with traditional features exhibited in the literature for corporate credit rating prediction. Identifying key ESG features can provide insight into which areas firms should actively seek on improving, particularly those with issued or looking at issuing corporate bonds.

In terms of classification, sophisticated classifiers such as support vector machine (SVM), neural networks and tree-based ensembles have displayed promising results for credit rating prediction (Jabeur et al., 2020; Jones et al., 2015; Ozturk et al., 2016). Interestingly, determining the importance of features on model performance still relies on conventional statistical tools. Consequently, we seek to bring consistency to the corporate credit ratings process by using tree-based ensemble methods in the feature selection and classification process.

Advocated by López de Prado (2020), mean decrease impurity (MDI) and mean decrease accuracy (MDA) introduced by Breiman (2001), and Tree SHapley Additive exPlanations (TreeSHAP) introduced by Lundberg et al. (2018) and Lundberg and Lee (2017) will be utilized for feature selection.[2] These methodologies are primarily used in statistical healthcare literature, and one application in finance using the University of California Irvine credit card dataset (Hall, 2018; Lundberg and Lee, 2017; Nembrini et al., 2018; Rodríguez-Pérez and Bajorath, 2020).[3]

We present several key contributions to the corporate credit rating literature. First, to the best knowledge of the authors, this is the first study to introduce ESG features in corporate credit rating prediction and feature importance research[4]. Second, we investigate how tree-based feature importance methods, constructed using MDI, MDA, and TreeSHAP, help to identify important explanatory features for multi-class corporate credit rating samples. Notably, this is the first known study that applies MDI, MDA, and TreeSHAP to the corporate credit rating literature. Shapley values, computed from tree-based methods, provide a coalitional game theory approach to explain the output of the classification model. As machine learning models are often touted as black boxes, along with regulatory restrictions in financial markets, explaining machine learning models is of significant importance. TreeSHAP offers a theoretically based method for interpreting the role input features play in explaining machine learning model outputs (Lundberg et al., 2018).

December 2019).

[2]See Shapley (1953) for the seminal introduction of Shapley values.

[3]See Appendix A for a detailed description of the implemented feature importance methods.

[4]Kiesel and Lücke (2019) analyze stock returns and CDS spreads changes to investigate whether ESG criterion have an impact on capital market reactions; however, their work does not include a predictive study.

Third, a comprehensive comparison between conventional and sophisticated classifiers will be utilized to evaluate corporate credit rating predictive performance. This contribution is divided into two components: (1) investment-grade and speculative-grade prediction (binary classification); and (2) individual classes (e.g., multi-class classification from AAA to D). This will enhance Jones et al.'s (2015) study on binary credit rating changes through the investigation of multi-class and binary credit rating prediction.

Our results show that industry, institutional, macroeconomic, sovereign credit ratings, analyst recommendations, and financial features are exhibited as high importance traditional features for corporate credit rating prediction using random forests (RF) and extremely randomized trees (ERT). Consistent across US and global firms, Thomson Reuters Diversity and Inclusion Rating (TRDIR) metrics of TRDIR diversity score (DS), TRDIR inclusion score (IS), TRDIR people development score (PDS), and TRDIR score (Score) demonstrate high importance across all feature importance and tree-based ensemble methods investigated.[5] The constituents of these metrics are primarily governance and social features, and is consistent with findings in Attig et al. (2013) and Kiesel and Lücke (2019). Environmental features also exhibit broad importance across US and global samples, particularly carbon emissions features. Using approximated TreeSHAP values, ESG features tend to be more important for cumulative investment-grade classes, whereas solvency, profitability, financial soundness, and valuation financial ratios exhibit greater cumulative importance for speculative-grade classes.

ESG features, particularly TRDIR metrics, do matter for US and global firms corporate credit rating prediction. Credit rating agencies should move towards mandatory inclusion of ESG features for corporate credit rating applications. ESG features allow for a thorough assessment of a company's creditworthiness, in addition to traditional features.

The paper is organized as follows. Section 2 describes the datasets. Section 3 details the methods used in feature selection and classification of corporate credit ratings. Section 4 presents and discusses the empirical results of our study and we conclude in Section 5.

[5]Score is the arithmetic mean of DS, IS, PDS as well as TRDIR controversies score.

2 Data

2.1 Features

To determine feature importance and credit rating predictability, four samples are used in this study. For US listed firms, US standard (US-STD) covers September 1982 to October 2019 (i.e., 37 years), whilst US ESG (US-ESG) is January 2002 to October 2019 (i.e., 17 years). For global firms (countries excluding the US), Global standard (GLB-STD) covers December 1995 to October 2019 (i.e., 24 years), while Global ESG (GLB-ESG) is December 2002 to October 2019. 168 (486) features for US-STD (US-ESG) are utilized, while 144 (476) features for GLB-STD (GLB-ESG) are used.

The dependent variable, long-term issuer credit rating, is retrieved for S&P from Refinitiv Eikon.[6] Quarterly equity return data and accounting data for US listed firms are retrieved from Center for Research in Security Prices (CRSP) and Compustat, respectively. Moreover, quarterly equity return data and accounting data for global listed firms are retrieved through Compustat global. Presently, only for US listed firms through subscription to Wharton Research Data Services (WRDS), 71 of the most commonly used financial ratios in empirical accounting and finance research can be retrieved in both an annual and quarterly format. As this study covers both US and global firms, 57 ratios are constructed for global listed firms.[7]

To form a number of the financial ratios, consensus analyst EPS estimates are used from Institutional Brokers' Estimate System (IBES). The implemented financial ratios in this study include profitability, valuation, capitalization, financial soundness, solvency, liquidity and efficiency.[8] A large proportion of financial ratios found to be important with common statistical tools (e.g., logistic regression and discriminant analysis) are evident in this study (Cantor and Packer, 1994; Ederington, 1985; Jones et al., 2015).

As credit rating downgrades occur more frequently than upgrades, we utilize quarterly financial ratios to avoid staleness in the data. Instead of investigating only corporate credit rating changes, we

[6] Moody's and Fitch credit ratings were also retrieved, however, due to the dearth of observations, only S&P ratings are investigated. Moreover, the use of only S&P is consistent with Jones et al. (2015).

[7] The discrepancy in start date for US and global samples is due to differing quarterly financial data availability in Compustat global. The difference in number of financial ratios for global and US samples is due to various financial features not available in quarterly format for global firms. Refer to appendices for a detailed description of financial ratio features implemented.

[8] Market capitalization and adjusted price are included as additional financial features. Global firms adjusted price and market capitalization are converted into United States dollar (USD). Moreover, a detailed description of the financial ratios are reported in Appendix D, Table D.1

forward fill the firm rating with quarterly financial ratios, expanding the total number of observations for each firm. For example, if a firms credit rating changes only once in five years, two firm level observations would be evident. However, with this expanded dataset, 20 firm level observations are captured. CUSIP codes for US firms and ISIN codes for global firms are utilized for merging the remaining independent features from Refintiv Eikon.

For firms to be included in this study, they must have a S&P credit rating and quarterly financial variables (CRSP/Compustat) to compute quarterly financial ratio features. In addition to financial ratios, Table 1 reports the traditional features used in corporate credit studies and included in this study. The remaining features are collected from Refinitiv Eikon and Federal Reserve Bank of St. Louis and include nine monthly institutional (Ashbaugh-Skaife et al., 2006; Bhojraj and Sengupta, 2003; Jones et al., 2015), 13 monthly equity analyst recommendations, 34 quarterly macroeconomic (Figlewski et al., 2012; Jones et al., 2015; Koopman et al., 2011), two sovereign S&P credit ratings (Almeida et al., 2017; Boot et al., 2006; Poon and Shen, 2020), three industry, one recession indicator (NBER for US and OECD for global) , and one financial quarter.[9]

A key contribution of this study is the inclusion of ESG features for credit rating predictability and feature importance. 318 (330) features are retrieved from Refintiv Eikon and utilized for the US-ESG (GLB-ESG) sample.[10] Table 2, 3, and 4 report the governance, environmental, and social features utilized, respectively. Table 5 displays aggregate pillar scores (i.e., ESG combined score), TRDIR metrics, and economic features.

As missing values and non-normality of input features can impact the performance of conventional and sophisticated classifiers, we implement the following steps. Financials Global Industry Classification Standard (GICS) sector is excluded from this study due to the proportion of missing firm level financial ratios. Missing values between quarters for the underlying financial features are grouped by their firm and filled with linear interpolation prior to forming the financial ratios.[11] Moreover, institutional and macroeconomic features are also forward filled with linear interpolation, whilst back filling macroeconomic features prior to their first observation per firm. Analyst recommendations are

[9]Institutional features are not implemented for global firms due to subscription restrictions. Detailed descriptions of the features can be found in the Appendix D, Table D.2 . Macroeconomic features that could not be retrieved from Refinitiv Eikon and the recession indicator were collected from Federal Reserve Bank of St. Louis.

[10]Descriptions of ESG features investigated can be found here, zeerovery.nl/blogfiles/DataStream_ESG_Glossary_Extranet.xlsx

[11]Some financial features are imputed with zero, for example, dividend rate, special items and extraordinary items and discontinued operations.

forward filled, and additional missing values are imputed with zero. If analyst recommendations are missing completely, this indicates that no analysts are following the firm. If values are still missing for financials, these are filled with the median of the country and rating class grouped together, while institutional features are imputed with zero - these firms exhibit no institutional ownership.

For the US-ESG sample, there are 217 ESG features that have less than 5% missing data and these are filled with the industry median. For ESG features up to 85% missing data, these are filled with the minimum industry value for each firm, resulting in a downward bias for ESG features. The same process is applied to global firms and 30 ESG features exhibit less than 5% missing values. As the creation of financial ratios can often produce unintended extreme outliers, we impose a 99% winsorization across all features exhibiting outliers. The Kolmogorov–Smirnov test is used to investigate the distribution of per corporate credit rating classes on each feature. Overall, non-normality is primarily exhibited across the feature set with log normal, Weibull, exponentiated Weibull, and gerneralized extreme value distributions represented.

As feature transformation can be applied to improve conventional and sophisticated model performance (Hastie et al., 2009; Jones et al., 2015), Yeo and Johnson (2000) transformation is utilized given the positive and negative values of the features in the dataset. If features are strictly positive, the Yeo-Johnson transformation is the same as the Box and Cox (1964) power transformation. However, for strictly negative or both positive and negative features, Yeo-Johnson transformation modifies the Box-Cox transformation.

Table 1: Traditional features

This table reports the traditional features used in this study. Financial ratios, institutional, equity analyst, macroeconomic, sovereign rating, country, industry, recession, and quarter are the feature types investigated for credit rating prediction feature importance.

Financial ratio	Acronym	Ratio type	Feature	Acronym	Feature type
Capitalization Ratio	capital_ratioq	Capitalization	Largest 5 Institutional Ownership Size	Top5InstOwn	Institutional
Common Equity/Invested Capital	equity_invcapq	Capitalization	Largest 10 Institutional Ownership Size	Top10InstOwn	Institutional
Long-term Debt/Invested Capital	debt_invcapq	Capitalization	Number of 5% Institutional Block Ownerships	NumInstBlockOwners	Institutional
Market Capitalization	mktcap (USD_mktcap)	Capitalization	Total Ownership by Institutional Blockholders	InstBlockOwn	Institutional
Total Debt/Invested Capital	totdebt_invcapq	Capitalization	Number of 13-F Institutional Owners	NumInstOwners	Institutional
Asset Turnover	at_turnq	Efficiency	Largest Institutional Ownership Size	MaxInstOwn	Institutional
Inventory Turnover	inv_turnq	Efficiency	Total Institutional Ownership	InstOwn	Institutional
Payables Turnover	pay_turnq	Efficiency	Ownership Concentration-Herfindahl-Hirschman Index	InstOwn_HHI	Institutional
Receivables Turnover	rect_turnq	Efficiency	Total Institutional Ownership % of Shares	InstOwn_Perc	Institutional
Sales/Stockholders Equity	sale_equityq	Efficiency	Recommendation - Mean (1-5)	Rec_Mean	Equity Analyst
Sales/Invested Capital	sale_invcapq	Efficiency	Recommendation - Number Of Total	Rec_Total	Equity Analyst
Sales/Working Capital	sale_nwcq	Efficiency	Recommendation - Median (1-5)	Rec_Median	Equity Analyst
Inventory/Current Assets	invt_actq	Financial Soundness	Recommendation - Low (1-5)	Rec_Low	Equity Analyst
Receivables/Current Assets	rect_actq	Financial Soundness	Recommendation - High (1-5)	Rec_High	Equity Analyst
Free Cash Flow/Operating Cash Flow	fcf_ocfq	Financial Soundness	Recommendation - Number Of Strong Buy	Rec_SBuy	Equity Analyst
Operating CF/Current Liabilities	ocf_lctq	Financial Soundness	Recommendation - Number Of Buy	Rec_Buy	Equity Analyst
Cash Flow/Total Debt	cash_debtq	Financial Soundness	Recommendation - Number Of Hold	Rec_Hold	Equity Analyst
Cash Balance/Total Liabilities	cash_ltq	Financial Soundness	Recommendation - Number Of Sell	Rec_Sell	Equity Analyst
Cash Flow Margin	cfmq	Financial Soundness	Recommendation - Number Of Strong Sell	Rec_SSell	Equity Analyst
Short-Term Debt/Total Debt	short_debtq	Financial Soundness	Recommendation - Number Of No Opinion	Rec_NoOpinion	Equity Analyst
Profit Before Depreciation/Current Liabilities	profit_lctq	Financial Soundness	Long Term Growth - Mean	LTG_Mean	Equity Analyst
Current Liabilities/Total Liabilities	curr_debtq	Financial Soundness	Number of Analysts	Num_Analyst	Equity Analyst
Total Debt/EBITDA	debt_ebitdaq	Financial Soundness	Gross Domestic Product	GDP	Macroeconomic
Long-term Debt/Book Equity	dltt_beq	Financial Soundness	Consumer Spending	consumer_spending	Macroeconomic
Interest/Average Long-term Debt	int_debtq	Financial Soundness	Government Spending	government_spending	Macroeconomic
Interest/Average Total Debt	int_totdebtq	Financial Soundness	Gross Fixed Capital Investment	gross_fixed_capital_investment	Macroeconomic
Long-term Debt/Total Liabilities	lt_debtq	Financial Soundness	Exports - Goods and Services	exports_goods_services	Macroeconomic
Total Liabilities/Total Tangible Assets	lt_ppentq	Financial Soundness	Imports - Goods and Services	imports_goods_services	Macroeconomic
Cash Conversion Cycle	cash_conversionq	Liquidity	Money Supply MO	money_supply_M0	Macroeconomic
Cash Ratio	cash_ratioq	Liquidity	Money Supply M1	money_supply_M1	Macroeconomic
Current Ratio	curr_ratioq	Liquidity	Money Supply M2	money_supply_M2	Macroeconomic
Quick Ratio (Acid Test)	quick_ratioq	Liquidity	Consumer Price Index	CPI_SA	Macroeconomic
Accruals/Average Assets	accrualq	Liquidity	Export Prices	export_prices	Macroeconomic
Research and Development/Sales	rd_saleq	Liquidity	Import Prices	import_prices	Macroeconomic
Adjusted Market Price	price_adj	Price	Terms of Trade	terms_of_trade	Macroeconomic
Effective Tax Rate	efftaxq	Profitability	Labor Force Survey	labor_force_survey	Macroeconomic
Gross Profit/Total Assets	gProfq	Profitability	Unemployment Rate	unemployment_rate	Macroeconomic
After-tax Return on Average Common Equity	aftret_eqq	Profitability	Retail Sales	retail_sales	Macroeconomic
After-tax Return on Total Stockholders' Equity	aftret_equityq	Profitability	Industrial Production	industrial_production	Macroeconomic
After-tax Return on Invested Capital	aftret_invcapxq	Profitability	Central Government Deficit	central_government_deficit	Macroeconomic
Gross Profit Margin	gpmq	Profitability	Government External Debt	government_external_debt	Macroeconomic
Net Profit Margin	npm	Profitability	Current Account Balance	current_account_balance	Macroeconomic
Operating Profit Margin After Depreciation	opmaqdq	Profitability	Current Account Balance % GDP	cab%gdp	Macroeconomic
Operating Profit Margin Before Depreciation	opmbdq	Profitability	Exports of Goods, Balance of Payment Basis	export_goods_bop	Macroeconomic
Pre-tax Return on Total Earning Assets	pretret_earnatq	Profitability	Imports of Goods, Balance of Payment Basis	imports_goods_bop	Macroeconomic
Pre-tax return on Net Operating Assets	pretret_noq	Profitability	Visible Trade Balance, Balance of Payment Basis	balance_on_trade_bop	Macroeconomic
Pre-tax Profit Margin	ptpmq	Profitability	Merchandise Exports	merchandise_exports	Macroeconomic
Return on Assets	roaq	Profitability	Merchandise Imports	merchandise_imports	Macroeconomic
Return on Capital Employed	roceq	Profitability	Visible Trade Balance	foreign_trade_balance	Macroeconomic
Return on Equity	roeq	Profitability	International Reserves	international_reserves	Macroeconomic
Total Debt/Equity	de_ratioq	Solvency	Resident Population	population	Macroeconomic
Total Debt/Total Assets	debt_atq	Solvency	Cash Rate Target	overnight_rate	Macroeconomic
Total Liabilities/Total Assets	lt_atq	Solvency	Government Bond Yield	government_bond	Macroeconomic
Total Debt/Capital	debt_capitalq	Solvency	Consumer Confidence Index	consumer_confidence	Macroeconomic
After-tax Interest Coverage	intcovq	Solvency	Business Survey Optimisim	business_confidence	Macroeconomic
Interest Coverage Ratio	intcov_ratioq	Solvency	Stock Market Index	stock_index	Macroeconomic
Dividend Payout Ratio	dprq	Valuation	S&P Sovereign Rating	sov_rating_rank	Sovereign rating
Forward P/E to 1-year Growth	PEG_1yrforwardq	Valuation	S&P Sovereign Outlook Rating	sov_outlook_rank	Sovereign rating
Forward P/E to Long-term Growth	PEG_ltgforwardq	Valuation	Country Unique Rating	loc_rank	Country
Trailing P/E to Growth	PEG_trailingq	Valuation	Developed Nation	developed_nation	Country
Book/Market	bmq	Valuation	GICS Sector	gsector	Industry
Shillers Cyclically Adjusted P/E Ratio	capeiq	Valuation	GICS Group	ggroup	Industry
Dividend Yield	div yieldq	Valuation	GICS Industry	gind	Industry
Enterprise Value Multiple	evmq	Valuation	NBER Recession Indicator	NBER_ri	Recession
Price/Cash flow	pcfq	Valuation	OECD Recession Indicator	OECD_ri	Recession
P/E (Diluted, Excl. EI)	pe_exiq	Valuation	Financial quarter	fqtr	Quarter
P/E (Diluted, Incl. EI)	pe_incq	Valuation			
Price/Operating Earnings (Basic, Excl. EI)	pe_op_basicq	Valuation			
Price/Operating Earnings (Diluted, Excl. EI)	pe_op_dilq	Valuation			
Price/Sales	psq	Valuation			
Price/Book	ptbq	Valuation			

Table 2: Governance features
This table reports 105 governance features retrieved from Refinitiv Eikon, pre-preprocessed, and utilized in this study. The subcategory (Category) is reported for each governance feature, and consists of: Board Functions, Board Structure, Compensation Policy, Management, Shareholder Rights, and Vision Strategy.

Governance features	Category	Governance features	Category
Audit Board Committee	Board Functions	Policy Executive Retention	Compensation Policy
Audit Committee Expertise	Board Functions	Shareholders Approval Stock Compensation Plan	Compensation Policy
Audit Committee Independence	Board Functions	Sustainability Compensation Incentives	Compensation Policy
Audit Committee Mgt Independence	Board Functions	Total Senior Executives Compensation	Compensation Policy
Audit Committee NonExecutive Members	Board Functions	Executive Compensation Controversies	Management
Auditor Tenure	Board Functions	Executive Compensation LT Objectives	Management
Average Board Tenure	Board Functions	Executive Compensation Policy	Management
Board Attendance	Board Functions	Executive Individual Compensation	Management
Board Functions Policy	Board Functions	Executive Members Gender Diversity, Percent	Management
Board Meeting Attendance Average	Board Functions	Executives Cultural Diversity	Management
Committee Meetings Attendance Average	Board Functions	Advance Notice for Shareholder Proposals	Shareholder Rights
Compensation Committee Independence	Board Functions	Advance Notice Period Days	Shareholder Rights
Compensation Committee Mgt Independence	Board Functions	Anti Takeover Devices Above Two	Shareholder Rights
Compensation Committee NonExecutive Members	Board Functions	Classified Board Structure	Shareholder Rights
Corporate Governance Board Committee	Board Functions	Company Cross Shareholding	Shareholder Rights
External Consultants	Board Functions	Confidential Voting Policy	Shareholder Rights
Nomination Committee Independence	Board Functions	Different Voting Right Share	Shareholder Rights
Nomination Committee Involvement	Board Functions	Director Election Majority Requirement	Shareholder Rights
Nomination Committee Mgt Independence	Board Functions	Elimination of Cumulative Voting Rights	Shareholder Rights
Nomination Committee NonExecutive Members	Board Functions	Equal Shareholder Rights	Shareholder Rights
Number of Board Meetings	Board Functions	Fair Price Provision	Shareholder Rights
Succession Plan	Board Functions	Golden Parachute	Shareholder Rights
Board Background and Skills	Board Structure	Limitation of Director Liability	Shareholder Rights
Board Cultural Diversity, Percent	Board Structure	Limitations on Removal of Directors	Shareholder Rights
Board Gender Diversity, Percent	Board Structure	Limited Shareholder Rights to Call Meetings	Shareholder Rights
Board Individual Re-election	Board Structure	Minimum Number of Shares to Vote	Shareholder Rights
Board Member Affiliations	Board Structure	Poison Pill	Shareholder Rights
Board Member Membership Limits	Board Structure	Policy Equal Voting Right	Shareholder Rights
Board Member Term Duration	Board Structure	Policy Shareholder Engagement	Shareholder Rights
Board Size	Board Structure	Pre-emptive Rights	Shareholder Rights
Board Size More Ten Less Eight	Board Structure	Public Availability Corporate Statutes	Shareholder Rights
Board Specific Skills, Percent	Board Structure	Shareholder Approval Significant Transactions	Shareholder Rights
Board Structure Policy	Board Structure	Shareholder Rights Policy	Shareholder Rights
Board Structure Type	Board Structure	Shareholders Vote on Executive Pay	Shareholder Rights
CEO Board Member	Board Structure	Staggered Board Structure	Shareholder Rights
CEO-Chairman Separation	Board Structure	State Owned Enterprise SOE	Shareholder Rights
Chairman is ex-CEO	Board Structure	Supermajority Vote Requirement	Shareholder Rights
Independent Board Members	Board Structure	Unlimited Authorized Capital or Blank Check	Shareholder Rights
Nomination Board Committee	Board Structure	Veto Power or Golden share	Shareholder Rights
Policy Board Diversity	Board Structure	Voting Cap	Shareholder Rights
Policy Board Experience	Board Structure	Voting Cap Percentage	Shareholder Rights
Policy Board Independence	Board Structure	Written Consent Requirements	Shareholder Rights
Policy Board Size	Board Structure	CSR Sustainability Committee	Vision Strategy
Strictly Independent Board Members	Board Structure	CSR Sustainability External Audit	Vision Strategy
Board Member Compensation	Compensation Policy	CSR Sustainability Report Global Activities	Vision Strategy
Board Member LT Compensation Incentives	Compensation Policy	CSR Sustainability Reporting	Vision Strategy
CEO Compensation Link to TSR	Compensation Policy	ESG Reporting Scope	Vision Strategy
Compensation Board Committee	Compensation Policy	Global Compact Signatory	Vision Strategy
Compensation Improvement Tools	Compensation Policy	GRI Report Guidelines	Vision Strategy
Highest Remuneration Package	Compensation Policy	Integrated Strategy in MDA	Vision Strategy
Non-Executive Board Members	Compensation Policy	Stakeholder Engagement	Vision Strategy
Policy Executive Compensation ESG Performance	Compensation Policy	UNPRI Signatory	Vision Strategy
Policy Executive Compensation Performance	Compensation Policy		

Table 3: Environmental features
This table reports 102 environment features retrieved from Refinitiv Eikon, pre-preprocessed, and utilized in this study. The subcategory (Category) is reported for each environment feature and consists of: Emission Reduction, Product Innovation, and Resource Reduction.

Environmental features	Category	Environmental features	Category
Accidental Spills	Emission Reduction	Animal Testing	Product Innovation
Biodiversity Impact Reduction	Emission Reduction	Animal Testing Cosmetics	Product Innovation
Climate Change Commercial Risks Opportunities	Emission Reduction	Animal Testing Reduction	Product Innovation
CO_2 Equivalent Emissions Direct, Scope 1	Emission Reduction	Eco-Design Products	Product Innovation
CO_2 Equivalent Emissions Indirect, Scope 2	Emission Reduction	Environmental Products	Product Innovation
CO_2 Equivalent Emissions Indirect, Scope 3	Emission Reduction	Environmental Project Financing	Product Innovation
CO_2 Equivalent Emissions Total	Emission Reduction	Equator Principles	Product Innovation
CO_2 Estimation Method	Emission Reduction	GMO Products	Product Innovation
Emissions Trading	Emission Reduction	Hybrid Vehicles	Product Innovation
EMS Certified Percent	Emission Reduction	Labeled Wood	Product Innovation
Environmental Assets Under Mgt	Emission Reduction	Noise Reduction	Product Innovation
Environmental Controversies	Emission Reduction	Nuclear	Product Innovation
Environmental Expenditures	Emission Reduction	Organic Products Initiatives	Product Innovation
Environmental Expenditures Investments	Emission Reduction	Product Environmental Responsible Use	Product Innovation
Environmental Investments Initiatives	Emission Reduction	Product Impact Minimization	Product Innovation
Environmental Partnerships	Emission Reduction	Renewable/Clean Energy Products	Product Innovation
Environmental Provisions	Emission Reduction	Sustainable Building Products	Product Innovation
Environmental Restoration Initiatives	Emission Reduction	Take-back and Recycling Initiatives	Product Innovation
Estimated CO_2 Equivalents Emission Total	Emission Reduction	Water Technologies	Product Innovation
Fossil Fuel Divestment Policy	Emission Reduction	Electricity Produced	Resource Reduction
Hazardous Waste	Emission Reduction	Electricity Purchased	Resource Reduction
Internal Carbon Pricing	Emission Reduction	Energy Produced Direct	Resource Reduction
ISO 14000 or EMS	Emission Reduction	Energy Purchased Direct	Resource Reduction
Non-Hazardous Waste	Emission Reduction	Energy Use Total	Resource Reduction
NOx and SOx Emissions Reduction	Emission Reduction	Env Supply Chain Partnership Termination	Resource Reduction
NOx Emissions	Emission Reduction	Environment Management Team	Resource Reduction
Ozone-Depleting Substances	Emission Reduction	Environment Management Training	Resource Reduction
Particulate Matter Emissions Reduction	Emission Reduction	Environmental Materials Sourcing	Resource Reduction
Policy Emissions	Emission Reduction	Environmental Supply Chain Management	Resource Reduction
Self-Reported Environmental Fines	Emission Reduction	Environmental Supply Chain Monitoring	Resource Reduction
SOx Emissions	Emission Reduction	Fresh Water Withdrawal Total	Resource Reduction
Staff Transportation Impact Reduction	Emission Reduction	Green Buildings	Resource Reduction
Total CO_2 Equivalent Emissions To Revenues USD	Emission Reduction	Land Environmental Impact Reduction	Resource Reduction
Total Energy Use To Revenues USD	Emission Reduction	Policy Energy Efficiency	Resource Reduction
Total Hazardous Waste To Revenues USD	Emission Reduction	Policy Environmental Supply Chain	Resource Reduction
Total Waste To Revenues USD	Emission Reduction	Policy Sustainable Packaging	Resource Reduction
VOC Emissions	Emission Reduction	Policy Water Efficiency	Resource Reduction
VOC Emissions Reduction	Emission Reduction	Renewable Energy Produced	Resource Reduction
VOC or Particulate Matter Emissions Reduction	Emission Reduction	Renewable Energy Purchased	Resource Reduction
Waste Recycled To Total Waste	Emission Reduction	Renewable Energy Use	Resource Reduction
Waste Recycled Total	Emission Reduction	Renewable Energy Use Ratio	Resource Reduction
Waste Recycling Ratio	Emission Reduction	Resource Reduction Policy	Resource Reduction
Waste Reduction Initiatives	Emission Reduction	Resource Reduction Targets	Resource Reduction
Waste Total	Emission Reduction	Targets Emissions	Resource Reduction
Water Discharged	Emission Reduction	Targets Energy Efficiency	Resource Reduction
Water Pollutant Emissions	Emission Reduction	Targets Water Efficiency	Resource Reduction
Water Pollutant Emissions To Revenues USD	Emission Reduction	Toxic Chemicals Reduction	Resource Reduction
Water Use To Revenues USD	Emission Reduction	Water Recycled	Resource Reduction
Agrochemical 5 % Revenue	Product Innovation	Water Withdrawal Total	Resource Reduction
Agrochemical Products	Product Innovation	e-Waste Reduction	Resource Reduction

Table 4: Social features

This table reports 115 social features retrieved from Refinitiv Eikon, pre-preprocessed, and utilized in this study. The subcategory (Category) is reported for each social feature and consists of: Community, Diversity Opportunity, Employment Quality, Health Safety, Product Responsibility, and Training Development.

Social features	Category	Social features	Category
Anti-competition Controversies	Community	Lost Days To Total Days	Health Safety
Anti-Competition Controversies Count	Community	Lost Time Injury Rate Contractors	Health Safety
Bribery, Corruption and Fraud Controversies	Community	Lost Time Injury Rate Employees	Health Safety
Business Ethics Controversies	Community	Lost Time Injury Rate Total	Health Safety
Corporate Responsibility Awards	Community	Lost Working Days	Health Safety
Crisis Management Systems	Community	Occupational Diseases	Health Safety
Diseases of the Developing World	Community	Policy Employee Health Safety	Health Safety
Donations Total	Community	Policy Supply Chain Health Safety	Health Safety
Extractive Industries Transparency Initiative	Community	Supply Chain Health Safety Improvements	Health Safety
Improvement Tools Business Ethics	Community	Supply Chain Health Safety Training	Health Safety
Intellectual Property Controversies	Community	Total Injury Rate Employees	Health Safety
Lobbying Contribution Amount	Community	Total Injury Rate Total	Health Safety
OECD Guidelines for Multinational Enterprises	Community	Ethical Trading Initiative ETI	Human Rights
Policy Bribery and Corruption	Community	Fundamental Human Rights ILO UN	Human Rights
Policy Business Ethics	Community	Human Rights Breaches Contractor	Human Rights
Policy Community Involvement	Community	Human Rights Contractor	Human Rights
Policy Fair Competition	Community	Human Rights Policy	Human Rights
Political Contributions	Community	Policy Child Labor	Human Rights
Total Donations To Revenues	Community	Policy Forced Labor	Human Rights
Whistleblower Protection	Community	Policy Freedom of Association	Human Rights
Day Care Services	Diversity Opportunity	Policy Human Rights	Human Rights
Diversity and Opportunity Controversies	Diversity Opportunity	Alcohol	Product Responsibility
Employees With Disabilities	Diversity Opportunity	Anti-Personnel Landmines	Product Responsibility
Flexible Working Hours	Diversity Opportunity	Armaments	Product Responsibility
HRC Corporate Equality Index	Diversity Opportunity	Cluster Bombs	Product Responsibility
New Women Employees	Diversity Opportunity	Contraceptives	Product Responsibility
Policy Diversity and Opportunity	Diversity Opportunity	Embryonic Stem Cell Research	Product Responsibility
Targets Diversity and Opportunity	Diversity Opportunity	Firearms	Product Responsibility
Women Employees	Diversity Opportunity	Gambling	Product Responsibility
Women Managers	Diversity Opportunity	Healthy Food or Products	Product Responsibility
Announced Layoffs	Employment Quality	ISO 9000	Product Responsibility
Announced Layoffs To Total Employees	Employment Quality	Obesity Risk	Product Responsibility
Management Departures	Employment Quality	Policy Customer Health Safety	Product Responsibility
Net Employment Creation	Employment Quality	Policy Data Privacy	Product Responsibility
Number of Employees from CSR reporting	Employment Quality	Policy Fair Trade	Product Responsibility
Salaries and Wages from CSR reporting	Employment Quality	Policy Responsible Marketing	Product Responsibility
Salary Gap	Employment Quality	Pornography	Product Responsibility
Strikes	Employment Quality	Product Access Low Price	Product Responsibility
Trade Union Representation	Employment Quality	Product Quality Controversies	Product Responsibility
Turnover of Employees	Employment Quality	Product Recall	Product Responsibility
Wages Working Condition Controversies	Employment Quality	Product Responsibility Monitoring	Product Responsibility
Wages Working Condition Controversies Count	Employment Quality	Product Sales at Discount to Emerging Markets	Product Responsibility
Accidents Total	Health Safety	Quality Mgt Systems	Product Responsibility
Contractor Fatalities	Health Safety	Responsible Marketing Controversies	Product Responsibility
Employee Accidents	Health Safety	Retailing Responsibility	Product Responsibility
Employee Engagement Voluntary Work	Health Safety	Six Sigma and Quality Mgt Systems	Product Responsibility
Employee Fatalities	Health Safety	Tobacco	Product Responsibility
Employee Health Safety Training Hours	Health Safety	Average Training Hours	Training Development
Employee Lost Working Days	Health Safety	Internal Promotion	Training Development
Employee Resource Groups	Health Safety	Management Training	Training Development
Employees Health Safety Controversies	Health Safety	Policy Career Development	Training Development
Employees Health Safety OHSAS 18001	Health Safety	Policy Skills Training	Training Development
Employees Health Safety Team	Health Safety	Supplier ESG training	Training Development
Health Safety Policy	Health Safety	Training and Development Policy	Training Development
Health Safety Training	Health Safety	Training Costs Per Employee	Training Development
HIV-AIDS Program	Health Safety	Training Costs Total	Training Development
Injuries To Million Hours	Health Safety	Training Hours Total	Training Development

Table 5: ESG features
This table reports 16 environmental, social, and governance (ESG) pillar features, five Thomson Reuters Diversity and Inclusion Rating (TRDIR) features, and 11 economic features retrieved from Refinitiv Eikon, pre-preprocessed, and utilized in this study. The subcategory (Category) is reported for each economic feature and consists of: Client Loyalty, Performance, and Shareholder Loyalty.

ESG features	TRDIR features	Economic features	Economic category
Community Score	TRDIR Controversies Score	Consumer Complaints Controversies	Client Loyalty
CSR Strategy Score	TRDIR Diversity Score	Consumer Complaints Controversies Count	Client Loyalty
Emissions Score	TRDIR Inclusion Score	Customer Satisfaction	Client Loyalty
Environment Pillar Score	TRDIR People Development Score	Employee Satisfaction	Performance
Environmental Innovation Score	TRDIR Score	Accounting Controversies	Shareholder Loyalty
ESG Combined Score		Earnings Restatement	Shareholder Loyalty
ESG Controversies Score		Insider Dealings Controversies	Shareholder Loyalty
ESG Score		Internal Audit Department Reporting	Shareholder Loyalty
Governance Pillar Score		Litigation Expenses	Shareholder Loyalty
Human Rights Score		Non-audit to Audit Fees Ratio	Shareholder Loyalty
Management Score		Profit Warnings	Shareholder Loyalty
Product Responsibility Score			
Resource Use Score			
Shareholders Score			
Social Pillar Score			
Workforce Score			

2.2 Summary statistics

Table 6 details the observations, firm count, number of investment-grade and speculative-grade corporate credit ratings for each country across the four samples utilized in this study. Panel A and B report standard and ESG sample statistics for developed nations, whilst Panel C and D report the aforementioned for developing nations. Countries are reported as developed or developing as defined by the United Nations (2020).

For the US-STD (US-ESG) sample, there are a total of 1,854 (768) firms with 85,480 (25,632) unique quarterly corporate credit rating observations with 44,445 (16,282) investment-grade and 41,035 (9,350) speculative-grade observations. The GLB-STD sample (GLB-ESG), that is global companies outside the US, has 27,892 (20,708) unique quarterly observations comprising 820 (619) firms with 17,942 (15,809) investment-grade and 9,950 (4,899) speculative-grade. In comparison, Jones et al. (2015) utilize 5,053 long-term issuer credit rating changes across 1983-2013.

Following the US, United Kingdom (UK), France, Germany, and Australia exhibit the highest number of unique observations within developed nations across standard and ESG samples. For developing nations, Brazil reports the highest unique firms, with Mexico, Russia, and South Korea comprising some of the top nations across the standard and ESG samples. The samples investigated include standard and ESG for US firms, and standard and ESG for all global firms. Across US and global firms, the percentage of investment-grade to speculative-grade ratings increases in the ESG sample. This indicates the ESG sample is tilted towards higher credit worthiness firms on average when compared to the standard sample.

Figure 1 illustrates the per class distribution of standard and ESG samples for US and global samples. The change in the proportion of investment-grade to speculative-grade issuer ratings for standard and ESG samples can be visualized across the distributions. Corporate credit ratings between A and B exhibit the majority of observations across the four samples, whilst a smaller percentage is evident in high investment-grade and high speculative-grade rating classes.

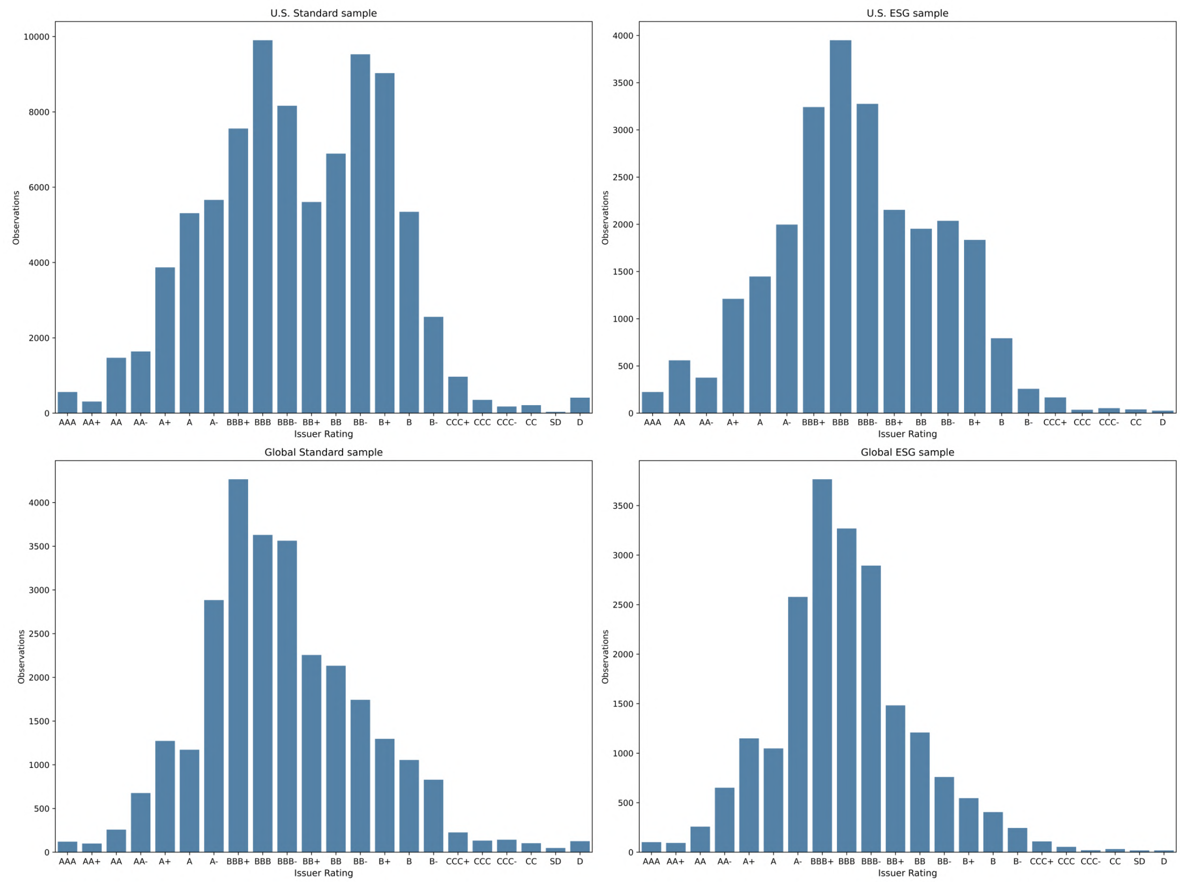

Figure 1: Per class rating distribution
This figure illustrates the per class corporate credit rating distribution across the United States (US) and global standard and environmental, social, and governance (ESG) samples.

Table 6: Country level summary statistics
This table demonstrates the country, observations, firm count, number of investment-grade, and speculative-grade corporate credit ratings for each country across the four samples utilized in this study. Panel A and B report standard and environmental, social, and governance (ESG) sample statistics for developed nations, whilst Panel C and D report the aforementioned for developing nations. Countries are reported as developed or developing as defined by the United Nations (2020).

Country	Obs	Firms	Investment-grade	Speculative-grade	Obs	Firms	Investment-grade	Speculative-grade
Panel A: Developed nations - standard					*Panel B: Developed nations - ESG*			
Austria	311	6	252	59	195	5	184	11
Australia	1 855	57	1 470	385	1 697	49	1 362	335
Belgium	238	7	183	55	230	7	180	50
Denmark	224	9	191	33	156	7	146	10
Finland	368	7	239	129	303	6	227	76
France	2 520	54	2 008	512	2 314	49	1 932	382
Germany	2 346	62	1 464	882	1 818	50	1 248	570
Greece	206	7	49	157	150	5	43	107
Ireland	202	6	109	93	167	5	106	61
Italy	1 112	27	780	332	827	18	712	115
Netherlands	908	26	663	245	691	17	598	93
New Zealand	389	12	366	23	292	9	291	1
Norway	380	10	246	134	366	9	240	126
Spain	865	28	618	247	694	25	601	93
Sweden	971	26	837	134	911	24	814	97
Switzerland	829	24	686	143	713	23	630	83
United Kingdom	3 365	112	2 332	1 033	3 066	101	2 210	856
United States	85 480	1 854	44 445	41 035	25 632	768	16 282	9 350
Panel C: Developing nations - standard					*Panel D: Developing nations - ESG*			
Argentina	412	10	1	411	71	8	0	71
Brazil	2 049	53	670	1 379	1 042	34	443	599
Chile	811	22	566	245	433	14	342	91
Colombia	146	4	129	17	94	4	93	1
Hong Kong	597	15	526	71	554	13	510	44
India	487	20	248	239	357	13	213	144
Indonesia	946	37	38	908	248	9	38	210
Israel	132	5	107	25	97	2	97	0
Kazakhstan	120	5	17	103	32	1	12	20
Malaysia	388	12	348	40	252	8	244	8
Mexico	1 224	36	529	695	448	17	394	54
Peru	244	10	78	166	91	8	20	71
Philippines	137	5	35	102	60	2	34	26
Russia	1 066	38	409	657	732	24	383	349
Singapore	377	14	307	70	253	7	247	6
Taiwan	333	11	330	3	291	9	291	0
Thailand	491	14	378	113	262	9	212	50
South Africa	152	5	76	76	123	5	65	58
South Korea	691	24	657	34	678	23	647	31

3 Research Method

To determine feature importance and credit rating predictive performance, various classifiers and metrics are utilized. To investigate this research problem, we must divide our method into three components: (1) classifiers and samples; (2) classification metrics; (3) feature importance and selection metrics.

3.1 Classifiers and samples

Table 7 reports the classifiers utilized to determine predictive accuracy. We employ five conventional classifiers which include: LR, linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and Naive Bayes – Gaussian (GNB) and Bernoulli (BNB). For the sophisticated classifiers, three are implemented and include: K-nearest-neighbors (KNN), SVM, and neural networks (MLP). For the tree-based methods, decision tree (DT), boosted DT – AdaBoost (AB), XGBoost (XGB), and Histogram-based gradient boost (HGB), and bagged DT - RF, and ERT are examined.

Table 7: Classifiers
This table reports the classifiers used for credit rating predictability. Conventional, sophisticated, tree-based, and tree-based ensemble classifiers are investigated.

Classifier	Acronym	Classifier type
Extremely randomized trees	ERT	Tree-based ensemble
Random forests	RF	Tree-based ensemble
XGBoost	XGB	Tree-based ensemble
Histogram gradient boost	HGB	Tree-based ensemble
AdaBoost	AB	Tree-based ensemble
Decision tree	DT	Tree-based
Multilayer perceptron	MLP	Sophisticated
Support vector machine	SVM	Sophisticated
K-nearest neighbours	KNN	Sophisticated
Logistic regression	LR	Conventional
Quadratic discriminant analysis	QDA	Conventional
Linear discriminant analysis	LDA	Conventional
Bernoulli Naive Bayes	BNB	Conventional
Gaussian Naive Bayes	GNB	Conventional

The classifiers used have been previously outlined/investigated in Hastie et al. (2009) and Jones et al. (2015). For classification problems, not all classifiers are inherently multi-class. This multi-class group contains algorithms such as KNN (Hastie et al., 2009), and classification and regression trees (Breiman et al., 1984) (e.g., DT, ERT, RF). For binary specific classifiers, for example, SVM, multi-class classification problems have to be reduced to binary ones. Commonly employed reduction techniques utilized include: one-vs-the-rest (Schölkopf et al., 2002) and one-vs-one (pairwise comparison) (Friedman, 1996; Hastie and Tibshirani, 1998; Kressel, 1998). As some classifiers are not inherently multiclass, one-vs-rest will be utilized to apply binary specific classifiers for multi-class applications (Hastie et al., 2009).

In order to investigate predictive performance of the aforementioned classifiers, stratified 75/25 (training/testing) allocation convention is implemented.[12] The test sample is utilized to evaluate out-of-sample (OOS) predictive performance, whilst the training set is used for model estimation. In addition, 5-fold cross-validation is utilized with the training set to estimate predictive performance metrics of the validation set. The temporal order of the data is not considered which ensures random observational dates are included in both the training and test sets.

Unlike Jones et al. (2015), all S&P long-term issuer credit rating observations are included as our classification application deals with class prediction, rather than binary changes. For ERT and RF classifiers, tree depth is evaluated and ranges between 15-20, while number of trees is investigated over 100-500 with reported results using 100. Stochastic average gradient is utilized to solve the LR optimization problem. Using the framework outlined in Gu et al. (2020) as a guide, the number of hidden layer neurons used in the MLP classifier are two-thirds of the size of the input layer and a total of two hidden layers are utilized.[13]

To investigate predictive performance, accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC ROC) will be utilized. Whilst AUC ROC is suitable only for binary classification problems, to assess this metric for our multi-class component of this study, we apply one-vs-rest (Schölkopf et al., 2002). Accuracy, precision, recall, and F1 score is the weighted harmonic mean for multi-class classification.[14]

3.2 Feature importance and selection metrics

Most researchers use p-values to evaluate the significance of explanatory features. However, López de Prado (2020) argues that p-values suffer from four major flaws (i.e., model assumptions, noisy

[12]Consistent training and testing datasets are investigated across samples and models, whilst consistent k-fold splits are used for cross-validated accuracy and area under the receiver operating characteristic curve.
[13]Rectified Linear Unit is used as the activation function.
[14]See Appendix A, for description of classification metrics utilized.

estimates, irrelevant probability, and in-sample estimation). For the investigated tree-based ensemble models, modern computational feature importance techniques can be utilized to provide insight into what features drive the model's prediction. Three methods to compute feature importance include: MDI, MDA and TreeSHAP. These methods will be utilized to recursively reduce the feature set from $n = 50, \ldots, 10$, with a 10 step reduction, to assess the balance between the number of features and predictive performance.[15]

Breiman (2001) introduce MDI that estimates the importance of a feature as the weighted information across all nodes (i.e., tree-based classifiers) where that feature was selected. The information gain that occurs at each split is the reduction in impurity. MDI is computed in-sample which results in computational efficiency. As the gini index is the most common impurity method utilized in classification trees, a known biased within the literature is that features with many split points (e.g., categorical or continuous variables) or high category frequencies are favored (Breiman et al., 1984; López de Prado, 2018; Nicodemus, 2011).

Breiman (2001) presents MDA that calculates importance by randomly permuting the features and computing the increase in OOS estimate of the accuracy loss (cross-validated). The tree-wise OOS estimate of the prediction error is computed with this permuted data. The difference between this estimate and the OOS error without permutation, averaged over all trees, is the MDA value of feature f (Zhang and Ma, 2012). MDA does not suffer from the MDI biases and is generally preferred within statistics literature (López de Prado, 2020; Nicodemus et al., 2010; Ziegler and König, 2014). However, for high dimensional data, MDA is computationally intensive (Calle and Urrea, 2011; Nembrini et al., 2018).

Shapley (1953) propose the Shapley value that is a method which assigns payouts to players depending on their contribution to the total payout. Players cooperate together and receive a certain profit from cooperation. Based on coalitional game theory and for any machine learning model, Shapley values are a solution for computing feature contributions for single predictions. The game is considered the prediction task for a single instance of a dataset, the gain is the actual prediction for this instance minus the average prediction for all instances, whilst the players are the features who collaborate to receive the gain. The Shapley value is defined using a value function of features in S and the feature value is the contribution to the payout, weighted and summed over all feature value

[15] See appendices (supplementary material) for a detailed explanation of MDI and MDA.

combinations:

$$\phi_j(v) = \sum_{S \subseteq \{x_1,\ldots,x_p\} \setminus \{x_j\}} \frac{|S|!(p-|S|-1)!}{p!}(v(S \cup \{x_j\} - v(S))), \tag{1}$$

where S is a subset of the features, x is the vector of feature values of the instance and p is the number of features. $v(S)$ is the prediction for feature values in set S that are marginalized over features that are not included in set S.

For a machine learning scenario where a training set $\{y^i, x^i\}_{i=1,...,n_{train}}$ of size n_{train} is used to train a model $f(x)$ with the objective of correctly predicting y. If we wish to explain the prediction from the model $f(x^*)$, for a specific feature vector $x = x^*$, Štrumbelj and Kononenko (2010, 2014) and Lundberg and Lee (2017) suggest using Shapley values. Using coalitional game theory to decompose an individual prediction into feature contributions, the single prediction can be considered the payout and features take the place of the players. The prediction of $f(x^*)$ is evaluated as follows:

$$f(x^*) = \phi_0 + \sum_{j=1}^{p} \phi_j^*, \tag{2}$$

where $\phi_0 = [f(x)]$ and ϕ_j^* is the ϕ_j for the prediction $x = x^*$. Therefore, Shapley values explain the difference between the prediction $y^* = f(x^*)$ and the global average prediction. This type of model is an additive feature attribution method and is the only type that satisfies the axioms to be considered a fair payout - efficiency, symmetry, dummy and additivity.[16]

In reality, computation of Shapley values is computationally expensive (NP-hard) as there is exponential possible coalitions of the features. Lundberg and Lee (2017) and Lundberg et al. (2018) propose SHapley Additive exPlanations (SHAP), KernelSHAP, TreeSHAP and DeepSHAP. Notably, TreeSHAP is a variant of SHAP that can be applied to tree-based models. SHAP values combine conditional expectations, rather than the marginal expectations, with the Shapley value to attribute ϕ_i values to each feature:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(M-|S|-1)!}{M!}[f_x(S \cup \{i\} - f_x(S))], \tag{3}$$

[16]See Aas et al. (2019) for an explanation of the desirable properties Shapley values exhibit.

where M is the the number of input features, N is the set of all input features, $f_x(S) = f(h_x(z')) = E[f(x) \mid x_s]$ where S is the set of non-zero indexes in a mapping between a binary pattern of missing features z' and the original function input space. $E[f(x) \mid x_s]$ is the expected value of the function conditioned on a subset S of the input features.

TreeSHAP estimates Shapley values using unique consistent and locally accurate attribution values in polynomial time instead of exponential for individual predictions.[17] Global feature importance can be estimated by summing the absolute Shapley values for each feature. In comparison to MDI and MDA, TreeSHAP provides local explanations, allowing the identification of features that might not be globally important, but can affect per class outcomes significantly. Additionally, TreeSHAP succinctly summarises the effect magnitude and direction (positive or negative) of features to model output. Depending on depth the of trees, the estimation is computationally efficient for small tree depth. For models requiring large tree depth, an approximation TreeSHAP method that considers only a single feature ordering can be used efficiently.

In addition to SHAP values, Lundberg et al. (2018) introduce SHAP interaction values. For individual model prediction, the SHAP interaction values follow similar axioms as SHAP values and allow the separate consideration of main and interaction effects. This separation can uncover important interactions captured by tree-based ensembles. SHAP interaction values can be obtained through the Shapley interaction index:

$$\phi_{i,j} = \sum_{S \subseteq N \setminus \{i,j\}} \frac{|S|!(M - |S| - 2)!}{2(M - 1!)} \nabla_{ij}(S), \tag{4}$$

where $i \neq j$, and

$$\begin{aligned} \nabla_{ij}(S) &= f_x(S \cup \{i,j\}) - f_x(S \cup \{i\}) - f_x(S \cup \{j\}) + f_x(S) \\ &= f_x(S \cup \{i,j\}) - f_x(S \cup \{i\}) - [f_x(S \cup \{i\}) - f_x(S)], \end{aligned} \tag{5}$$

In Equation 4, the SHAP interaction value between feature i and feature j is equally split between each feature such that $\phi_{i,j} = \phi_{j,i}$ and the total interaction effect is $\phi_{i,j} + \phi_{j,i}$.

[17] Refer to Lundberg et al. (2018) for a full explanation of the TreeSHAP algorithm that estimates attribution values and interaction effect values. Path dependent feature perturbation is utilized in our study. Model time runs at $O(TLD^2)$, where T = number of trees, L = number of leaves and D = depth of trees.

4 Results

4.1 Baseline performance

Table 8 reports the baseline prediction performance for the 14 classifiers implemented across US and global samples. Predictive performance with features in their respective units and Yeo and Johnson (2000) power transformed features are investigated for all features. Both methods produce consistent results with tree-based methods, however, we find that transformations improve the efficiency of the model estimation process for conventional classifiers. Only KNN (sophisticated) and LDA (conventional) exhibits test accuracy improvements with transformations. This finding is broadly consistent with Jones et al. (2015).

Panel A presents the performance for classifying individual corporate credit rating classes for US and global firms in the standard sample (multi-class). Tree-based ensemble methods report the highest test accuracy and test AUC utilizing all features. ERT, RF and XGB demonstrate the same rankings across both samples. ERT reports 95.01%, and 94.79% test accuracy for US and global, respectively, whilst RF estimates 93.38% for US and 94.05% for global correctly. The overall results indicate that tree-based methods outperform conventional classifiers by a significant margin across both samples.

Whilst popular sophisticated methods of SVM and MLP outperformed conventional methods, they nevertheless under performed tree-based methods. Conventional methods of LR, LDA, and QDA report 34.09% (91.14%), 30.55% (88.15%) and 29.98% (62.52%) for test accuracy (test AUC) for US, and consistent rankings are displayed for the global sample. Panel B presents the performance for classifying individual corporate credit rating classes for US and global firms in the ESG sample. Similar to Panel A results, ERT, RF, and XGB report the highest test accuracy across US and global samples, whilst classifier ranking also demonstrate consistency.[18] MLP, SVM, DT, and KNN also demonstrate high predictive performance for the ESG sample, followed by conventional methods of LR and LDA.

In addition to multi-class prediction, for robustness, two-class (investment-grade vs speculative-grade) was tested.[19] The classifier rankings on individual classes also transpire in the the two-class

[18]A comparable standard sample to the ESG sample for US and global was tested and also demonstrated consistency. The inclusion of ESG features improved accuracy, however, the values were insignificantly different.

[19]See Appendix B, Table B.1, for the two-class prediction performance.

classification with ERT, XGB, and RF exhibiting the highest predictive performance. The estimated cross-validation AUC across the two class problem resembles the performance displayed in Jones et al. (2015).

Overall, using all features in US and global samples across the standard and ESG samples, tree-based ensemble methods outperform the remaining sophisticated and conventional classifiers. Taking into account predictive performance and training time for the classifiers, ERT and RF are utilized to investigate feature importance.[20] Feature importance is examined for the individual class problem, and the following sections implement MDI, MDA, and TreeSHAP methods for the RF and ERT methods.

[20] The use of bagging is consistent with the arguments presented in López de Prado (2018). Bagging addresses overfitting, whilst boosting addresses underfitting. Moreover, bagging can be parallelized, while boosting requires sequential running.

Table 8: Baseline prediction accuracy

This table reports multi-class predictive performance of the 14 classifiers implemented across United States (US) and global samples with no dimensionality reduction applied to the feature set. The classifiers include: extremely randomized trees (ERT), random forests (RF), XGBoost (XGB), histogram gradient boost (HGB), AdaBoost (AB), decision tree (DT), multi-layer perceptron (MLP), support vector machine (SVM), k-nearest neighbours (KNN), logistic regression (LR), quadratic discriminant analysis (QDA), linear discriminant analysis (LDA), Bernoulli Naive Bayes (BNB), and Gaussian Naive Bayes (GNB). Cross-validated (CV) area under the curve (AUC) and CV Accuracy is calculated through 5-fold cross-validation on the training sample, whilst Test AUC, Test Accuracy and is performance on the test sample. Rank is sorted on test accuracy performance. Panel A reports the performance for US and global firms for classifying individual classes in the standard samples, whilst Panel B presents the performance for environmental, social, and governance (ESG) samples. Yeo and Johnson (2000) transformations are applied to the features in the results.

Model	US CV		US Test			Global CV		Global Test		
	AUC	Accuracy	AUC	Accuracy	Rank	AUC	Accuracy	AUC	Accuracy	Rank
Panel A: Standard sample										
ERT	99.7	93.6	99	95	1	99.7	94.2	99.8	94.8	1
RF	99.6	92.7	99.1	93.4	2	99.6	92.9	99.8	94.1	2
XGB	99.8	94.2	99.1	82.3	3	99.6	92.2	99.7	90.6	3
AB	96.4	74.2	87.6	77.2	4	98.6	82.0	99.2	82.4	4
MLP	96.4	70.2	98.3	75.3	5	96.5	72.4	98.3	76.7	5
SVM	93.9	58.1	96.7	61	6	96.7	71.1	98.3	74.6	6
DT	85.9	74.2	76.9	58.5	7	87.2	76.7	83.4	70.9	8
KNN	83.5	46.8	86.4	49.6	8	94.1	71.1	94.3	73.6	7
HGB	95.0	71.7	75.5	46.8	9	95.6	82.7	84.9	69.9	9
LR	85.6	34.5	91.1	34.1	10	86.5	42	92.2	42.5	10
LDA	83.6	31.2	88.2	30.6	11	84.1	37.4	88.9	38.2	12
QDA	72.4	19.0	62.5	30	12	87.2	43.8	61.5	41.4	11
GNB	65.4	5.3	59	19.2	13	72.1	15.3	59.1	21.6	14
BNB	72.7	17.1	80.5	17.1	14	74.3	21.8	83.9	21.7	13
Panel B: ESG sample										
ERT	99.8	96.4	99.6	96.9	1	99.6	94.5	98.3	95.9	1
RF	99.8	96.4	99.9	96.9	1	99.7	94.5	99.4	95.6	2
XGB	99.9	96.7	99.9	96	3	99.8	94.5	99.7	94.4	3
AB	99.7	94.9	99.8	94.6	4	99.6	93.4	99.7	93.5	4
MLP	99.6	93.4	99.7	94.2	5	99.3	91	99.4	92.8	5
SVM	99.6	92.8	99.8	93.6	6	99.5	90.6	99.7	91.6	6
DT	93.1	87.4	88.3	81.0	8	92.6	86.8	86.4	83.4	8
KNN	99.1	91.7	99.2	93.2	7	98.8	89.7	97.4	91.1	7
LR	96.9	77.4	97.6	78	9	97.8	82.3	98	83.2	9
HGB	96.1	84.3	77.33	71.6	10	93.2	80.0	72.2	74.5	10
LDA	94.2	66.7	96	64	11	94.8	71.3	94.1	72.4	11
QDA	95.1	88.2	55.9	12.1	14	93.1	85.1	56.7	12.2	13
GNB	68.4	13.2	50.5	13.8	13	65.6	10.8	50	12.1	14
BNB	74.7	26.9	84.3	26	12	74	30.4	83.4	30.3	12

4.2 Recursive feature reduction performance

Table 9, Table 10, and Table 11 reports the predictive performance of ERT and RF for the individual classes (multi-class) where features utilized are recursively reduced using MDI values, MDA values, and TreeSHAP values, respectively. As ERT and RF are invariant to transformations, transformations are not applied to the remaining tables reported.[21] Feature importance is first evaluated over the entire feature set and the top 50 features identified using MDI, MDA, and TreeSHAP importance scores are retained.[22].At each iteration, cross-validated AUC, cross-validated accuracy, test AUC and test accuracy is evaluated.

Panel A across Table 9, Table 10, and Table 11 reports the MDI, MDA, and TreeSHAP predictive performance of the US-STD sample for the two highest performing classifiers identified in Table 8. In Table 9, ERT demonstrates cross-validated accuracy (AUC) of 93.69% (99.66%) utilizing 50 features, with test accuracy (AUC) of 94.97% (99.04%). Test performance reports predictive performance at 94.86% accuracy using 30 features. RF exhibits consistent performance, with the highest test performance at 95.26% accuracy using both 50 and 40 features. The detraction of test performance using 20 features is minimal, displaying 94.74% and 94.65% accuracy for ERT and RF, respectively. Similarly, reduction of features using MDA (Table 10) in the US-STD sample demonstrates consistent predictive ability with MDI values (Table 9).

TreeSHAP (Table 11) reports a slight under performance using ERT (both TreeSHAP and TreeSHAP approximation), however, RF exhibits similar findings. Panel B (GLB-STD), Panel C (US-ESG), and Panel D (GLB-ESG) also report consistent findings using MDI and MDA recursive reduction. ERT has slightly higher test accuracy for MDI reduction. For example, with 20 features in the US-ESG sample, MDI test accuracy is 97.05% compared to MDA test accuracy of 95.94%. The attributing factor for this additional performance is likely due to MDI being an in-sample feature importance method, whilst MDA permutes features and estimates importance OOS.

TreeSHAP exhibits similar findings with MDI and MDA when using 50, 40, and 30 features. As the computation of TreeSHAP values are $O(TLD^2)$, given the feature size and depth of trees

[21] Yeo and Johnson (2000) transformations were applied for robustness and reports consistent results for the remaining tables in this study.

[22] Table 11 demonstrates predictive performance of recursively reduced features using Tree SHapley Additive exPlanations (TreeSHAP) for the AAA class and approximated TreeSHAP values are averaged across all ratings classes. The low performance in GLB-STD (Panel B) for ERT TreeSHAP in comparison to the approximated method is attributed to averaging across all classes in the approximation

utilized (between 13 and 18), iterative reduction utilizing approximated TreeSHAP values are also evaluated.[23][24] There is a detraction in performance when using 20 features (e.g., for US ESG sample, ERT test accuracy is 90.67% and 90.20% for the TreeSHAP approximation), however, the advantage of TreeSHAP is the ability to estimate per class feature importance. These insights are not obtained from MDI or MDA feature importance methods, highlighting the benefit TreeSHAP provides whilst exhibiting a detraction in performance.

[23] Refer to Lundberg et al. (2018) for a full explanation of the TreeSHAP algorithm that estimates attribution values and interaction effect values. Model time runs at $O(TLD^2)$, where T = number of trees, L = number of leaves and D = depth of trees.

[24] Approximated Shapley values use Saabas, which use conditional expectations, but only consider a single order of the features. As n! orderings are not considered, the approximation method does not uphold the consistency attributes of TreeSHAP. However, approximated TreeSHAP values are computationally efficient with the depth of trees utilized.

Table 9: MDI recursive prediction

This table demonstrates the performance of recursively reduced features using mean decrease impurity (MDI) feature importance rankings. Cross-validated (CV) area under the curve (AUC), CV accuracy, test AUC, and test accuracy for extremely randomized trees (ERT) and random forests (RF) are presented. CV performance is evaluated on the trained data (75%) and test performance is calculated on the remaining 25% data. Panel A and Panel B report the MDI performance for the standard United States (US-STD) and global (GLB-STD) samples, respectively. Panel C and Panel D display the environmental, social, and governance (ESG) United States (US-ESG) and global (GLB-ESG) MDI performance. The values $n = 50, \ldots, 10$, represent the recursively reduced feature set size utilized in the classification prediction, where features are eliminated based off MDI feature rankings. Values are reported in percentages.

Features	ERT CV		ERT Test		RF CV		RF Test	
	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy
Panel A: MDI US-STD								
50	99.7	93.7	99.0	95.0	99.7	93.9	99.5	95.3
40	99.7	93.6	99.0	95.0	99.7	93.9	99.5	95.3
30	99.7	93.7	98.9	94.9	99.7	93.9	99.5	95.2
20	99.6	93.5	98.9	94.7	99.6	93.3	99.5	94.7
10	99.5	92.7	98.9	94.2	99.1	89.1	99.3	90.7
Panel B: MDI GLB-STD								
50	99.8	95.0	99.8	95.4	99.7	94.3	99.9	94.9
40	99.8	95.1	99.7	95.5	99.7	94.2	99.7	94.9
30	99.7	95.0	99.6	95.6	99.7	94.2	99.7	95.0
20	99.7	95.0	99.6	95.2	99.7	93.8	99.5	94.8
10	99.7	94.1	99.6	95.0	99.6	93.2	99.5	94.1
Panel C: MDI US-ESG								
50	99.8	96.5	99.6	97.1	99.8	96.6	99.9	97.0
40	99.8	96.6	99.6	97.1	99.8	96.4	99.9	97.1
30	99.8	96.6	99.7	97.2	99.8	96.4	99.9	97.1
20	99.8	96.6	99.7	97.1	99.8	96.4	99.9	97.1
10	99.6	95.9	99.6	96.5	99.8	95.9	99.6	96.6
Panel D: MDI GLB-ESG								
50	99.6	95.1	98.2	96.2	99.7	95.0	98.7	95.9
40	99.6	95.2	98.5	96.3	99.7	95.1	98.7	95.8
30	99.6	95.0	98.6	96.1	99.7	95.0	98.6	95.8
20	99.6	95.2	98.3	96.3	99.7	94.8	98.6	95.7
10	99.5	94.3	97.7	95.4	99.6	94.4	98.5	95.7

Table 10: MDA recursive prediction

This table demonstrates the performance of recursively reduced features using mean decrease accuracy MDA feature importance rankings. CV Area under the curve (AUC), cross-validated (CV) accuracy, test AUC, and test accuracy for extremely randomized trees (ERT) and random forests (RF) are presented. CV performance is evaluated on the trained data (75%) and test performance is calculated on the remaining 25% data. Panel A and Panel B report the MDA performance for the standard United States (US-STD) and global (GLB-STD) samples, respectively. Panel C and Panel D display the environmental, social, and governance (ESG) United States (US-ESG) and global (GLB-ESG) MDA performance. The values $n = 50, \ldots, 10$, represent the recursively reduced feature set size utilized in the classification prediction, where features are eliminated based off MDA feature rankings. Values are reported in percentages.

Features	ERT CV		ERT Test		RF CV		RF Test	
	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy
Panel A: MDA US-STD								
50	99.7	94.2	98.9	95.4	99.8	94.4	99.2	95.7
40	99.7	94.2	99.0	95.4	99.8	94.5	99.9	95.6
30	99.7	93.4	98.9	94.6	99.7	93.9	99.5	95.4
20	99.6	93.3	98.8	94.5	99.6	93.3	99.5	94.7
10	99.2	89.4	98.6	91.1	98.5	83.7	99.3	85.9
Panel B: MDA GLB-STD								
50	99.7	94.9	99.6	95.5	99.7	94.2	99.8	94.9
40	99.7	95.0	99.7	95.5	99.7	94.3	99.7	94.9
30	99.7	94.8	99.6	95.4	99.7	94.2	99.7	95.0
20	99.7	94.5	99.6	95.3	99.7	94.0	99.7	94.7
10	98.9	89.6	97.8	90.2	99.5	91.6	99.4	93.1
Panel C: MDA US-ESG								
50	99.6	96.0	99.5	96.8	99.8	96.3	99.9	97.0
40	99.6	95.9	99.4	96.6	99.8	96.4	99.9	97.1
30	99.6	95.7	99.5	96.5	99.8	96.3	99.9	96.9
20	99.6	95.2	99.5	95.9	99.8	96.2	99.9	96.7
10	98.8	92.1	98.9	92.6	99.7	95.2	99.9	96.3
Panel D: MDA GLB-ESG								
50	99.6	94.6	98.2	96.0	99.7	95.1	98.6	96.2
40	99.5	94.6	98.1	95.9	99.7	95.1	98.6	96.0
30	99.5	94.4	97.8	95.8	99.7	94.4	98.3	95.8
20	99.1	93.2	97.4	94.1	99.7	94.1	98.5	95.5
10	98.8	90.5	96.4	91.9	99.6	93.7	98.0	95.2

Table 11: TreeSHAP recursive prediction

This table demonstrates predictive performance of recursively reduced features using Tree SHapley Additive exPlanations (TreeSHAP) for the AAA class and approximated TreeSHAP values are averaged across all ratings classes. Number of features utilized, Area under the curve (AUC), cross-validated (CV) accuracy, test AUC, and test accuracy for extremely randomized trees (ERT) and random forests (RF) are presented. CV is evaluated on 75/25 train and test split. Panel A and Panel B report the TreeSHAP performance for the standard United States (US-STD) and global (GLB-STD) samples, respectively. Panel C and D display the environmental, social, and governance (ESG) United States (US-ESG) and global (GLB-ESG) TreeSHAP performance. Values are reported in percentages.

Features	ERT TreeSHAP				ERT Approx TreeSHAP				RF TreeSHAP				RF Approx TreeSHAP			
	CV		Test		CV		Test		CV		Test		CV		Test	
	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy
Panel A: TreeSHAP US-STD																
50	99.2	89	99.4	90.9	99.1	88.5	99.2	90.2	99.7	94	99.9	95.2	99.7	93.5	99.7	94.7
40	99.2	88.9	99.1	90.5	99	88.3	99.2	89.9	99.7	93.6	99.5	94.8	99.7	93.6	99.8	94.8
30	99.1	88.1	99.4	89.5	98.9	87.4	99.2	89	99.7	93.6	99.5	94.7	99.6	93.2	99.7	94.4
20	98.7	85.4	99	86	98.7	85.5	99	87	99.5	91.1	99.4	92.6	99.5	91.1	99.6	92.6
10	98.5	83.1	98.7	84.4	97.2	77	98.1	78.4	98.2	80.6	98.8	83	98	80.3	98.9	82.1
Panel B: TreeSHAP GLB-STD																
50	92.6	54.2	95.9	55.6	99.7	94.3	99.7	95.1	99.7	94	99.8	94.6	99.7	94.1	99.8	94.9
40	92.4	53.8	95.7	54.1	99.7	94.3	99.7	95.1	99.7	94	99.8	94.7	99.7	94.1	99.8	94.8
30	91.9	52.3	95.4	54	99.7	94	99.7	94.8	99.6	93.1	99.8	94.1	99.7	94	99.8	94.6
20	91.4	51.3	95.3	52.2	99.6	93.3	99.6	94.2	99.6	92.6	99.8	93.6	99.7	93.5	99.7	94.2
10	90	46.6	94.6	47.9	99.3	90	99.4	91.2	88.4	40	93.2	39.6	99.3	90.2	99.5	91.2
Panel C: TreeSHAP US-ESG																
50	99.6	94.4	99.8	95	99.6	94.8	99.7	95.4	99.7	94.9	99.9	95.7	99.8	95.1	99.9	95.8
40	99.6	94.1	99.7	94.5	99.6	94.7	99.7	95.3	99.7	94.3	99.9	95	99.8	95	99.9	95.7
30	99.4	92.1	99.7	93.1	99.6	94.3	99.6	94.9	99.7	94	99.8	94.9	99.7	94.7	99.9	95.5
20	99.1	89.6	99.4	90.7	99.5	92.6	99.6	93.2	99.6	93.2	99.8	93.8	99.7	93.8	99.8	94.8
10	97.8	81.8	98.8	82.6	98.4	85.4	99	86.3	98.7	86.4	99.3	87.7	99.2	90.2	99.6	91.3
Panel D: TreeSHAP GLB-ESG																
50	99.6	94	99.1	94.8	99.6	94	99	95.1	99.7	94.5	98.9	95.8	99.7	94.3	99.6	95.4
40	99.6	93.8	98.4	94.6	99.6	94	98.6	95.1	99.7	94.6	99.7	95.5	99.7	94.3	99.5	95.3
30	99.5	93.3	98.7	94.1	99.5	93.6	98.7	94.6	99.7	94.3	98.8	95.2	99.7	94.1	99.3	95.2
20	99.4	92.5	98.3	93.6	99.4	92.6	98.6	93.5	99.7	93.6	98.8	94.7	99.6	93.8	99.4	94.8
10	96.6	74.3	96.7	74.6	98.8	88.2	98.3	89	99.6	92.9	98.7	94	99.4	91.9	99.1	93.1

4.3 Performance evaluation metrics

Table 12 reports predictive performance evaluation metrics for individual corporate credit rating classes (i.e., multi-class classfication) for an ERT using MDA.[25] On the balance between number of features and predictive performance in Table 10, precision, recall, F1 score, and support (number of observations) are evaluated on 20 features for US and global standard and ESG samples. Similar with the classification report for MDI, investment-grade classes (i.e., AAA to BBB-) report higher F1 scores on average in comparison to speculative-grade classes (i.e., BB+ to D). This finding is consistent across US and global standard and ESG samples.

On average, across the four samples, investment-grade corporate credit rating classes report higher precision values relative to speculative-grade classes. For US-STD, AA+, AAA, and D report the highest precision (support) values at 98.46% (67), 98.43% (127) and 97.80% (104), respectively. The classes with the largest number of observations (support) are BBB at 2,438, BB- at 2,397, and B+ at 2,261. The speculative-grade classes, BB+ and less, exhibit true positives ranging between 97.80% and 87.84%, with approximately half of the sample representation.[26] The misclassification of all classes is primarily within two classes. For example, for the BB- class, misclassification is typically two notches above (i.e., BB+ and BB) or two below (i.e., B+ and B). For all observations that are a true-positive, investment-grade rated classes again are classified correctly more than speculative-grade classes.

For the US-ESG sample, 6,408 observations are represented in the test sample. The proportion of observations is consistent for high investment-grade and high speculative-grade classes relative to the US-STD, where A+ to CCC+ classes represents 94.08% of the US-STD and 94.70% of the US-ESG. The overall F1 scores are higher across the classes with the additional ESG features and sample size reduction. Moreover, this is further reflected in the cumulative class prediction accuracy of 95.94%. For GLB-STD and GLB-ESG samples, investment-grade classes report higher F1 scores on average. Findings are consistent with the US, standard and ESG report cumulative accuracy of 95.32% and 94.09%, respectively.

[25] For brevity, only MDA classification report is displayed.

[26] The SD class reports precision of 50.00%, however only seven observations are represented in the test sample.

Table 12: MDA classification report

This table demonstrates the mean decrease accuracy (MDA) classification report for the individual corporate credit rating classes for an extremely randomized trees (ERT). Based on the balance between number of features and predictive performance in Table 10, Precision, Recall, F1 score and Support are evaluated on 20 features for United States (US) standard (US-STD), global standard (GLB-STD), US environmental, social and governance (ESG) (US-ESG) and global ESG (GLB-ESG) samples. Values are reported in percentages. Macro avg calculates the metric (e.g., F1 score) separated by class, while weighted avg calculates the metric for each class independently.

Credit Rating	US-STD				US-ESG				GLB-STD				GLB-ESG			
	Precision	Recall	F1 score	Support	Precision	Recall	F1 score	Support	Precision	Recall	F1 score	Support	Precision	Recall	F1 score	Support
AAA	98.4	98.4	98.4	127	100	100	100	550	100	100	100	340	100	100	100	270
AA+	98.5	95.5	97	670	0	0	0	0	100	95.6	97.8	230	93.3	77.8	84.8	180
AA	96.5	98.4	97.4	369	98	100	99	148	90.1	100	94.8	640	92.2	98.3	95.2	600
AA-	97.5	95.6	96.5	408	100	99	99.5	960	98.8	96.4	97.6	166	100	98.2	99.1	169
A+	97.3	97.8	97.5	995	99.7	99	99.4	309	98.4	96.4	97.4	309	98	97.1	97.6	309
A	97	97	97	1357	96.3	96.9	96.6	350	94.9	94.6	94.7	295	92.7	92	92.3	275
A-	95.6	96	95.8	1464	97.7	97.3	97.5	482	96.3	96.4	96.4	730	95.8	95.7	95.7	622
BBB+	95.2	96.3	95.8	1901	96.8	97.7	97.2	823	96.9	97.7	97.3	1090	95.6	96.3	95.9	914
BBB	96	95.7	95.8	2438	96.5	95.5	96	956	96.3	96.7	96.5	914	94.6	95.4	95	821
BBB-	94.8	95	94.9	2031	95.6	96.4	96	802	95.4	96.7	96	930	93.9	93.8	93.8	754
BB+	93.4	93.5	93.5	1406	93.8	93.9	93.8	543	94.1	93.9	94	527	91.6	91.3	91.5	358
BB	95.2	92.8	94	1721	92.2	93.1	92.7	509	95.5	90	92.7	521	92.2	90.3	91.2	288
BB-	93	94.9	93.9	2397	95	93.5	94.2	524	94.4	94.8	94.6	424	90.6	89	89.8	218
B+	92.4	93.4	92.9	2261	95.2	97.5	96.3	473	91	92.2	91.6	320	88.6	91.6	90.1	119
B	92.8	92.9	92.9	1341	95.3	91.9	93.5	197	90.9	93.7	92.3	255	86.2	87.1	86.6	93
B-	89.9	88	89	577	98.2	88.9	93.3	63	94.7	91.1	92.9	214	98.2	85.7	91.5	63
CCC+	90.4	84.6	87.4	233	94.4	91.9	93.2	37	85.7	87.5	86.6	48	84.8	93.3	88.9	30
CCC	87.8	77.4	82.3	84	90	100	94.7	90	84.6	94.3	89.2	35	77.8	100	87.5	14
CCC-	91.1	82	86.3	50	100	94.1	97	17	97.2	97.2	97.2	36	100	100	100	60
CC	90.9	80	85.1	50	100	100	100	11	92	76.7	83.6	30	77.8	70	73.7	10
SD	50	71.4	58.8	70	0	0	0	0	92.3	100	96	12	50	66.7	57.1	30
D	97.8	85.6	91.3	104	100	100	100	40	91.4	100	95.5	32	62.5	83.3	71.4	60
accuracy	94.5	94.5	94.5	94.5	95.9	95.9	95.9	95.9	95.3	95.3	95.3	95.3	94.1	94.1	94.1	94.1
macro avg	92.4	91	91.5	21388	96.7	96.3	96.5	6408	94.1	94.6	94.3	7009	88.9	90.6	89.5	5177
weighted avg	94.5	94.5	94.5	21388	96	95.9	95.9	6408	95.3	95.3	95.3	7009	94.1	94.1	94.1	5177

4.4 Important features

4.4.1 US firms

To highlight features that drive model performance, Table 13 reports the features comprising the 20 feature models in magnitude order for the US-STD and US-ESG sample.[27] In Panel A, for the US-STD sample, for ERT, the 20 features consist of financial, institutional, industry, and analyst recommendation features. GICS industry demonstrates to be the feature that decreases the impurity the most, whilst 15 financial features comprise the 20 with financial soundness ratios (five), efficiency (four), profitability (three), capitalization (one) and valuation (one). Other top explanatory features are inventory turnover, total liabilities/total tangible assets, number of 13-F institutional owners, gross profit margin, dividend yield and market capitalization.

For RF, 15 out of 20 features are the same as ERT that consist of 10 financial features. These 10 financial features comprise of financial soundness ratios (four), efficiency ratios (three), profitability (one), and capitalization, liquidity, and valuation. The remaining five features that are the same for both RF and ERT are GICS industry, and four institutional features. The five features that are different between ERT and RF are three valuation ratios, and two solvency ratios. Whilst price/operating earnings (diluted), forward P/E to 1-year growth, forward P/E to long-term growth, total debt/total assets, and interest coverage ratio are not present in the top 20 for ERT, they are included in the top 50. Similar findings are evident with the features in ERT not reported in RF.

In the US-ESG sample, traditional features evident across ERT and RF are consistent with those reported in the standard sample. For ERT and RF, 10 and six ESG features with varying units (binary, number or percentage) are displayed within the top 20 features, respectively. Thomson Reuters Diversity and Inclusion Rating (TRDIR) Diversity Score (DS), TRDIR People Development Score (PDS), TRDIR Inclusion Score (IS), and TRDIR Score (Score) exhibit high importance across ERT and RF.[28]

DS is a overall diversity score that is the arithmetic mean of the pillar scores. Eight factors comprise DS and include: board gender diversity, board member cultural diversity, women employees, new women employees, women executive employees, women managers, diversity process, and diversity

[27] Appendix E, Figure E.1 and E.2 illustrates the MDI values in magnitude order for US and global standard and ESG samples.

[28] The diversity, inclusion, and people development pillars methodology to calculate pillar scores is defined by Refinitiv: https://www.refinitiv.com/content/dam/marketing/en_us/documents/faqs/diversity-inclusion-index-faq.pdf

objectives. Chapple and Humphrey (2014) find some weak evidence of a negative correlation between having multiple women on the board and performance, however, that in some industries, diversity is positively correlated with performance. Reguera-Alvarado et al. (2017) suggests that investors in Spain do not penalize firms which increase their female board membership, and, that greater gender diversity may generate economic gains. Liu et al. (2014) report that the fraction of women directors has a positive impact on firm performance in China. The absolute number of female directors also matters. Moreover, Isidro and Sobral (2015) suggest that greater female representation on corporate boards of large European firms can increase firm value indirectly.

In terms of cultural diversity, Frijns et al. (2016) findings indicate that cultural diversity in boards, particularly independent directors, negatively affects firm performance measured by Tobin's Q and return on assets. However, the negative impact of cultural diversity on performance is mitigated by the complexity of the firm and the size of foreign sales and operations. Erhardt et al. (2003) report that the percentage of women and minorities on boards of directors for 127 large US companies is positively associated with financial firm performance. Given these impacts, the finding that an aggregated diversity score is important for predicting corporate credit ratings is consistent with the literature's overall support that gender and cultural diversity play a role in firm performance.

IS is an overall inclusion score and comprises: flexible working hours, day care services, employee with disabilities, HIV/AIDS program, and human rights campaign (HRC) corporate equality index. Schein et al. (1977) indicates that the flexible working hours has no adverse impact on productivity, and in cases, increases employee productivity. Lindsay et al. (2018) show that hiring people with disabilities included improvements in profitability, competitive advantage, inclusive work culture, and ability awareness. As the inclusion of diverse cultures, flexible working hours, and employees with disabilities is shown to impact firm performance, the finding that IS is an important predictor of corporate credit ratings is consistent with the literature.

PDS is an overall people development score and includes: internal promotion, management training, career development processes, skills training of employees, employee satisfaction, average training hours, and training cost per employee. Huselid (1995) argues that the use of internal promotion systems that focus on employee merit, and other forms of incentives intended to align of employees with those of shareholders. Economically and statistically significant impacts on turnover, productivity, and short- and long-term measures of corporate financial performance are reported.

Ngo et al. (1998) find that structural training and development and retention-oriented compensation are related to various measures of firm performance. Edmans (2011) report that employee satisfaction is positively correlated with shareholder returns. Consequently, the importance of PDS for corporate credit prediction is consistent with the literature on the impact of people development and corporate performance.

Score is the arithmetic mean of DS, IS, PDS, and TRDIR controversies score. This finding indicates that the interaction of the aforementioned pillar scores are important ESG features as they comprise the overall scores deemed highly important. Interestingly, the features comprising DS, IS, and PDS metrics are primarily social features and are situated in the training development, diversity opportunity, and health & safety category. Governance features (i.e, board member cultural diversity or board gender diversity) are located in the board structure sub category. The high importance of these metrics is consistent with the findings of Attig et al. (2013) and Kiesel and Lücke (2019). The authors find that credit rating agencies tend to award relatively high ratings to firms with good social performance. Furthermore, they find that Corporate Social Responsibility (CSR) factors that influence credit ratings are related to primary stakeholder management. The diversity features and employee satisfaction feature which comprise DS and PDS, is consistent with Attig et al. (2013) findings that diversity and employee relations are components that matter most in explaining firms' creditworthiness.

Additional important ESG features across ERT and RF include three governance features (i.e., average board tenure, advanced notice period, anti-takeover devices above two) and an environmental feature (i.e., estimated CO_2 equivalent emissions total). The aforementioned governance features are within the category of shareholder rights and board functions. Anderson et al. (2004) finds that board tenure is positively related to corporate bond yield spreads, whilst Hammoudeh et al. (2020) reports a significant time-varying causality from the CO_2 emission allowances price to green bonds. Consequently, the ESG features identified are consistent with findings within bond research.

In Panel B, MDA values are utilized to recursively reduce the US sample feature set.[29][30] As MDA is calculated by randomly permuting the features and computing the increase in OOS estimate

[29] Appendix E, Figure E.3 and E.4 illustrates the MDA values in magnitude order for US and global standard and ESG samples.

[30] For robustness, recursive reduction was applied to an investment-grade and speculative-grade classification task for ERT and RF. The features comprising the 10 feature model for ERT and RF across US and global samples is reported in Appendix E Table C.1.

of the accuracy loss, the 20 features differ to the MDI method (in-sample). ERT (RF) returns eight (16) of the same features as the MDI method in Panel A for the standard sample. Industry, institutional, macroeconomic, financial, and analyst features comprise the 20 across ERT and RF.

Whilst MDI importance reported no macroeconomic or analyst features for the US-STD sample, two analyst features (i.e., number of sell-side equity analysts, number of sell-side equity analyst recommendations) and four macroeconomic features (i.e., population, money supply M2, imports goods balance of payment and export goods balance of payment) comprise the top 20 for ERT. Financial ratios reported by both MDI and MDA are long-term debt/total liabilities, dividend yield, interest/average total debt, market capitalization, and gross profit margin. Additional financial features identified as highly important are common equity/invested capital and capital ratio.

For RF, 11 financial features are consistent across MDI and MDA which include dividend yield, market capitalization, price/operating earnings (diluted), forward P/E to 1-year growth, forward P/E to long-term growth, interest coverage ratio, long-term debt/total liabilities, inventory turnover, total liabilities/total tangible assets, inventory/current assets, and total debt/total assets. The four differing features include two macroeconomic and two financial features. Overall, these findings indicate that OOS MDA importance using ERT or RF for US-STD sample, variants of macroeconomic, financial, institutional, analyst, and industry features are important for corporate credit rating prediction. There is some disparity between the constituents of the 20 features for ERT and RF, however overall predictive performance is consistent regardless which method is utilized.

With the introduction of ESG features, a larger proportion of the top 20 features for ERT and RF are ESG features. This indicates that using MDA importance values, an OOS feature importance method, more ESG features matter in multi-class corporate credit rating prediction of US firms. Of the 15 ESG features identified, seven are governance, four are TRDIR scores, two are social and two environmental. Five of the seven governance features are within the shareholder rights category, and, in magnitude of importance, these include: elimination of cumulative voting rights, confidential voting policy, director election majority requirement, written consent requirements, and shareholder approval significant transactions.

Danielson and Karpoff (1998) demonstrate the uses of corporate governance provisions, including firms use of eliminating cumulative voting. Gordon and Pound (1993) examines how information and ownership structure affect voting outcomes on shareholder-sponsored proposals to change corporate governance structure. Proposal outcomes including the establishment of confidential voting or poison pill repeal is found to be a function of ownership by insiders, institutions, outside blockholder, and outside directors who are blockholders. The high importance of institutional features in the US-STD and US-ESG sample, along with shareholder rights features is consistent with the governance literature.

The social features are HRC corporate equality index - national benchmarking tool on corporate policies, practices and benefits pertinent to lesbian, gay, bisexual, transgender and queer employees, and employee engagement voluntary work - does the firm have a volunteer policy for employees. Plewa et al. (2015) report that corporate volunteering positively impacts CSR image, firm image, and, in turn, strengthens affective and cognitive loyalty of consumers to the firm. Improving the ability to retain consumers is important and it is not hard to see how firm perception can impact firms long-term corporate credit. The HRC corporate equality index feature is the score of the company in the equality index. Hossain et al. (2020) utilize the HRC corporate equality index and find a significantly positive relationship with US firm innovation. Given firm innovation impacts firm performance, the inclusion of employees of differing sexual orientation also appears to be a strong driver of multi-class corporate credit rating predictions for US firms.

In addition to HRC corporate equality index being found important itself, it is also a component of the IS. Similar to MDI importance values, TRDIR pillar scores of DS, PDS, IS, and Score encompass the top 20 for ERT. Environmental features found important include policy emissions - whether the firm have a policy to improve emissions reduction and targets emissions - whether the firm set targets or objectives to be achieved on emissions reduction. Stakeholder pressure for firms to reduce their environmental impact is forever evolving and unlikely to slowdown. Not only is reducing firms emissions good for the planet, it also resonates with the growing socially conscious investor.

Klassen and McLaughlin (1996) find that significant positive returns are measured for strong environmentally managed firms. Al-Najjar and Anfimiadou (2012) find that eco-efficient UK firms have higher market values than those lacking environmental strategies, whilst Lee and Min (2015) report a positive link between green research and development and financial performance. Moreover, Attig et al. (2013) show that environmental performance loads positively and significant on firms credit ratings. This finding of policy emissions and target emissions acting as a strong driver for an

ERT is consistent with the literature and the socially desirable objectives stakeholders portray.

When ESG features are introduced into the RF, seven ESG features comprise the 20 feature model. The aforementioned TRDIR variant features, two community focused social features of total donations to revenue and donations total and one environmental feature of CO_2 equivalent emissions direct scope 1 are demonstrated to be highly important. Dean (2003) suggests that irresponsible firms can increase their favor with consumers by pursuing either conditional or unconditional donations. The average firm was found to enhance its image by pursuing an unconditional donation, however, conditional donations do not damage firm image. Engaging with the community through, for example, donations to local charities, can build consumer loyalty or expand a firms consumer base.

Retaining or expanding the number of consumers is an objective that firm management should pursue in their goal of maximizing shareholder wealth. Consequently, the parallels between firm donations and corporate credit is consistent with both human intuition and the literature. Scope 1 emissions are direct emissions from firm-owned and operated resources. CO_2 emissions are process emissions and encompass this category. Consistent with the CO_2 feature found important in ERT, along with the aforementioned supporting literature, direct CO_2 emissions (scope 1) is an important driver of US corporate credit prediction.

In Panel C, TreeSHAP values for standard ERT (RF) reports 15 (18) traditional features consistent with those found using MDA in Panel B. Across the ESG sample and for the RF, six ESG features are found important and are the same as those identified in Panel B (excluding CO_2 equivalent emissions direct scope 1). We find that 17 features are consistent with the MDA importance method findings for the RF, with the three differential features being financial ratios. For ERT, 12 ESG features are included in the model and 10 are consistent with those in Panel B. The two additional ESG features are governance related (e.g., unlimited authorized capital or blank check, policy executive compensation ESG performance).

Overall, TreeSHAP and MDA feature importance identify largely consistent traditional and ESG features across the samples. MDI is broadly consistent with these two methods. The most notable finding is across ERT and RF and the three feature importance methods, TRDIR metrics are consistently reported as highly important features. The inclusion of ESG features, in addition to TRDIR variants, do matter for US firm corporate credit prediction. This finding for US firms is robust across three differing feature importance methods and two tree-based ensemble classifiers.

Table 13: Important features US sample

This table reports the features in the 20 feature model utilized in Table 9, 10, and 11 in rank order for the United States (US) standard and environmental, social, and governance (ESG) sample. After recursive reduction to 20 features, for extremely randomized trees (ERT) and random forests (RF), Panel A, Panel B, and Panel C report using mean decrease impurity (MDI), mean decrease accuracy (MDA), and Tree SHapley Additive exPlanations (TreeSHAP) methods, respectively. Text in bold represents ESG features.

Rank	ERT Standard Features	ERT ESG Features	RF Standard Features	RF ESG Features
Panel A: MDI				
1	GICS Industry	**TRDIR Diversity Score**	Price/Operating Earnings (Diluted)	Inventory Turnover
2	Inventory Turnover	**TRDIR People Development Score**	Forward P/E to Long-term Growth	**TRDIR Diversity Score**
3	Total Liabilities/Total Tangible Assets	**TRDIR Score**	Forward P/E to 1-year Growth	**TRDIR Score**
4	Market Capitalization	**TRDIR Inclusion Score**	Inventory Turnover	Market Capitalization
5	Inventory/Current Assets	GICS Industry	Market Capitalization	Total Liabilities/Total Tangible Assets
6	Gross Profit Margin	Total Institutional Ownership	Total Liabilities/Total Tangible Assets	Gross Profit Margin
7	Dividend Yield	Inventory Turnover	Dividend Yield	Cash Conversion Cycle
8	Number of 13-F Institutional Owners	Gross Profit Margin	Total Institutional Ownership % of Shares	Inventory/Current Assets
9	Receivables/Current Assets	**Estimated CO_2 Equivalents Emission Total**	Total Institutional Ownership	Total Institutional Ownership
10	Total Institutional Ownership	Number of 13-F Institutional Owners	GICS Industry	GICS Industry
11	Cash Conversion Cycle	GICS Group	Inventory/Current Assets	**TRDIR People Development Score**
12	Receivables Turnover	**Advance Notice Period Days**	Long-term Debt/Total Liabilities	**Estimated CO_2 Equivalents Emission Total**
13	Long-term Debt/Total Liabilities	Inventory/Current Assets	Receivables Turnover	Receivables Turnover
14	Total Institutional Ownership % of Shares	Total Liabilities/Total Tangible Assets	Number of 13-F Institutional Owners	Gross Profit/Total Assets
15	Asset Turnover	Market Capitalization	Interest Coverage Ratio	Number of 13-F Institutional Owners
16	Operating Profit Margin Before Depreciation	Largest 10 Institutional Ownership Size	Cash Conversion Cycle	Payables Turnover
17	Payables Turnover	Cash Conversion Cycle	Total Debt/Total Assets	Receivables/Current Assets
18	Interest/Average Total Debt	Long-term Debt/Total Liabilities	Receivables/Current Assets	**Average Board Tenure**
19	Largest 10 Institutional Ownership Size	Asset Turnover	Largest 10 Institutional Ownership Size	Total Debt/Total Assets
20	Gross Profit/Total Assets	**Anti Takeover Devices Above Two**	Gross Profit Margin	Interest Coverage Ratio
Panel B: MDA				
1	Number of 13-F Institutional Owners	**Elimination of Cumulative Voting Rights**	Market Capitalization	Market Capitalization
2	GICS Industry	**TRDIR Inclusion Score**	Dividend Yield	**TRDIR Score**
3	GICS Group	**TRDIR Score**	Price/Operating Earnings (Diluted)	Number of 13-F Institutional Owners
4	Market Capitalization	**TRDIR People Development Score**	Number of 13-F Institutional Owners	Dividend Yield
5	S&P Sovereign Rating	Market Capitalization	Forward P/E to Long-term Growth	**TRDIR Diversity Score**
6	GICS Sector	**TRDIR Diversity Score**	Forward P/E to 1-year Growth	**TRDIR Inclusion Score**
7	Dividend Yield	Number of 13-F Institutional Owners	Dividend Payout Ratio	Inventory Turnover
8	Long-term Debt/Total Liabilities	**CSR Sustainability Report Global Activities**	Interest Coverage Ratio	Total Institutional Ownership
9	Number of Analysts	**Confidential Voting Policy**	Total Institutional Ownership	**TRDIR People Development Score**
10	Recommendation - Total Number	**HRC Corporate Equality Index**	Long-term Debt/Total Liabilities	Total Debt/Total Assets
11	Interest/Average Total Debt	**Board Individual Re-election**	Inventory Turnover	Interest Coverage Ratio
12	Total Institutional Ownership % of Shares	GICS Industry	Total Institutional Ownership % of Shares	**Total Donations To Revenues**
13	Resident Population	Long-term Debt/Total Liabilities	GICS Industry	**Donations Total**
14	Common Equity/Invested Capital	**Director Election Majority Requirement**	Largest 10 Institutional Ownership Size	**CO_2 Equiv Emissions Direct, Scope 1**
15	Capitalization Ratio	**Written Consent Requirements**	Total Debt/Total Assets	Inventory/Current Assets
16	Total Institutional Ownership	**Policy Emissions**	Interest/Average Total Debt	GICS Industry
17	Money Supply M2	**Employee Engagement Voluntary Work**	Money Supply M2	Long-term Debt/Total Liabilities
18	Imports of Goods, Balance of Payment Basis	GICS Sector	Total Liabilities/Total Tangible Assets	Gross Profit Margin
19	Gross Profit Margin	**Shareholder Approval Significant Transactions**	Largest Institutional Ownership Size	Total Liabilities/Total Tangible Assets
20	Exports of Goods, Balance of Payment Basis	**Targets Emissions**	Shillers Cyclically Adjusted P/E	Dividend Payout Ratio
Panel C: TreeSHAP				
1	GICS Industry	Number of 13-F Institutional Owners	Dividend Yield	Market Capitalization
2	Number of 13-F Institutional Owners	**TRDIR Score**	Price/Operating Earnings (Diluted)	**TRDIR Score**
3	GICS Group	**TRDIR Inclusion Score**	Dividend Payout Ratio	Dividend Yield
4	Long-term Debt/Total Liabilities	**TRDIR People Development Score**	Forward P/E to 1-year Growth	Number of 13-F Institutional Owners
5	Market Capitalization	**TRDIR Diversity Score**	Market Capitalization	Inventory Turnover
6	Long-term Debt/Invested Capital	Market Capitalization	Interest Coverage Ratio	Interest Coverage Ratio
7	Dividend Yield	GICS Industry	Number of 13-F Institutional Owners	Total Debt/Total Assets
8	Common Equity/Invested Capital	**Elimination of Cumulative Voting Rights**	Forward P/E to Long-term Growth	**TRDIR Diversity Score**
9	GICS Sector	GICS Group	Long-term Debt/Total Liabilities	Dividend Payout Ratio
10	S&P Sovereign Rating	Long-term Debt/Total Liabilities	Total Institutional Ownership % of Shares	**TRDIR People Development Score**
11	Number of Analysts	**HRC Corporate Equality Index**	Total Institutional Ownership	Inventory/Current Assets
12	Capitalization Ratio	**Confidential Voting Policy**	Shillers Cyclically Adjusted P/E	Total Institutional Ownership
13	Interest/Average Total Debt	GICS Sector	Inventory Turnover	Common Equity/Invested Capital
14	Recommendation - Total Number	Capitalization Ratio	GICS Industry	**TRDIR Inclusion Score**
15	Short-Term Debt/Total Debt	**CSR Sustainability Report Global Activities**	After-tax Interest Coverage	**Donations Total**
16	Total Institutional Ownership % of Shares	**Unlimited Authorized Capital/Blank Check**	Total Debt/Total Assets	GICS Industry
17	Total Debt/Capital	**Director Election Majority Requirement**	Total Liabilities/Total Tangible Assets	**Total Donations To Revenues**
18	Return on Capital Employed	Total Institutional Ownership	Interest/Average Total Debt	After-tax Interest Coverage
19	Total Institutional Ownership	**Policy Executive Comp ESG Perf**	Common Equity/Invested Capital	Cash Conversion Cycle
20	Pre-tax Profit Margin	**Employee Engagement Voluntary Work**	Largest 10 Institutional Ownership Size	Long-term Debt/Total Liabilities

4.4.2 Global firms

Similar to Table 13, to highlight which features are driving model performance for global firms, Table 14 reports the features utilized in the ERT and RF in magnitude order for the GLB-STD and GLB-ESG sample.

In Panel A, when MDI values reduce the feature set, traditional features driving the ERT consists of nine financial, three industry, six macroeconomic features, one country identifier and one S&P sovereign rating. Market capitalization, GICS industry, labor force survey and population decrease impurity the most. These macroeconomic features comprised are in the order of labor force survey, population, government spending, consumer spending, Gross Domestic Product (GDP), country and retail sales.[31] Seven of the financial features in the GLB-STD are present in the equivalent US sample. Three efficiency ratios (asset turnover, inventory turnover, and receivable turnover), two financial soundness ratios (total liabilities/total tangible assets and inventory/current assets), one profitability ratio (gross profit margin), and market capitalization.

For RF, 11 of the features are consistent with ERT, comprising similarly of financial ratios and macroeconomic data. The differing features include six financial ratios: cash conversion, Shillers cyclically adjusted P/E Ratio, sales/stockholders equity, payables turnover, total debt/total assets, long-term debt/total liabilities and one macroeconomic feature: stock market index. Nine of the 15 financial features also encompass the US-STD sample, whilst none of the macroeconomic features are present. This is reasonable given only the US is investigated in the US sample, while 36 developed and developing nations are included in the global sample, allowing for broader dispersion of macroeconomic feature values.

Within the GLB-ESG sample, nine ESG features are included in the 20 feature ERT and RF models. Consistent with the US sample, DS, PDS, IS, and Score are identified as important drivers of corporate credit rating predictability. Women employees, number of employees from CSR reporting, salaries and wage from CSR reporting, CO_2 variants, trade union representation, employee satisfaction, water withdrawal total, and electricity purchased are the additional ESG features. Whilst women employees and employee satisfaction encompass the DS and PDS scores, high importance to their individual metrics are also found.

[31] Labor force survey is a large household sample survey providing results of labor participation.

The impact of women executives is researched extensively in the literature (e.g., Adams and Ferreira (2009), Liu et al. (2014)), however, the total percentage of women employees and the impact on firm performance is limited. As the number of women directors have significant firm impacts, parallels to why total percentage of women employees is identified as an important driver of corporate credit can be seen. In terms of employee satisfaction, human relations theories argue that employee satisfaction causes stronger corporate performance through improved recruitment, retention, and motivation. Edmans (2011), Guiso et al. (2015), and Huang et al. (2015) provide evidence of a positive significant relationship between employee satisfaction and firm performance. Again, the literature highlights the importance of employee satisfaction and the inclusion of this feature in the model is justified.

Porter and Kramer (2006) discuss links between competitive advantage and corporate social responsibility. The authors highlight that WholeFoods purchased renewable wind energy credits equal to 100% of its electricity use in all of its stores and facilities. At this time, they were the only Fortune 500 company to offset its electricity consumption entirely. This green mission WholeFoods and many organizations are seeking to achieve resonates with consumers and builds brand loyalty. The impact of electricity purchased, as well as water withdrawal total, are areas of environmental and social importance and should be strategically managed. The impact to firm perception, leading to firm performance and overall corporate credit can be seen.

Panel B displays the features driving ERT and RF when MDA feature importance values are used to recursively reduce the global sample to 20 features. For ERT (RF), the 12 (12) features identified as important in MDI are exhibited in MDA. Similar to MDA for US-STD sample, eight features are consistent with ERT and include: GICS industry, GICS group, GICS sector, market capitalization, S&P sovereign rating, gross profit margin, number of sell-side analysts covering the firm and total number of sell-side equity recommendations. Whether a firm comes from a developed nation demonstrates high importance for ERT predictability. For RF, features identified as important across MDI and MDA consist again of previously discussed features, e.g., market capitalization, Shillers cyclically adjusted P/E ratio, and GICS industry. The differing features MDA identifies includes interest coverage ratio and total debt/capital.

In the GLB-ESG sample, ERT identifies a larger proportion of ESG features (16 of 20) and includes six governance, four environmental, three social, two TRDIR variants, and one economic feature. The sub categories of the pillars are dispersed broadly, for example, for governance, three are shareholder rights (i.e., staggered board structure, limitation of director liability, and veto power or golden share) two are vision & strategy (i.e., global compact signatory, CSR sustainability external audit), and one is board structure (i.e., Chief Executive Officer (CEO) board member). These governance features are highlighted in the literature in having material impacts on firm value.

Cremers et al. (2017) suggests that staggered boards promote value creation for some firms through innovative projects, whilst Jiraporn and Liu (2008) demonstrate that firms with staggered boards have lower leverage. Brook and Rao (1994) find that the net benefit of limitation of director liability is larger for financially distressed firms than for other firms because outside directors are valuable to the troubled firm. Moreover, Bradley and Chen (2011) report that firms that provide limited liability for directors enjoy higher credit ratings and lower yield spreads. These previsions allow directors to pursue their own interests by adopting low-risk, self-serving operating strategies, which benefit corporate bondholders. Additional ESG features consistent with the literature include, for example, United Nations global compact signatory (Janney et al., 2009), environmental management system (EMS) certified percent (Nishitani, 2011) and product impact minimization (Miroshnychenko et al., 2017).

Using RF, of the 10 traditional features identified, nine are consistent with the standard sample. TRDIR variants (i.e., DS, Score, PDS, IS), employee accidents, water withdrawal total, salaries and wages from CSR reporting, water recycled, employee satisfaction, and CO_2 equivalent emissions total are the identified ESG features. Compared with ERT, for global firms, a smaller proportion of ESG features comprise the 20 features.

In Panel C, the recursive reduction to 20 features utilizing TreeSHAP values reports features largely similar with Panel B. The finding in the US-ESG sample that ESG features are consistent over MDA and TreeSHAP feature importance methods also transpires in the global sample. Similarly, the ESG features identified with MDI values are broadly consistent with Panel B and Panel C. Macroeconomic features, particularly, labor force survey and resident population are largely important in the global sample. Moreover, TRDIR feature variants of DS, PDS, IS, and Score are strong drivers for global firms corporate credit predictability.

Table 14: Important features global sample
This table reports the features in the 20 feature model utilized in Table 9, 10, and 11 in rank order for the GLB-STD and environmental, social and governance (ESG) sample. After recursive reduction to 20 features, for extremely randomized trees (ERT) and random forests (RF), Panel A, Panel B, and Panel C report using mean decrease impurity (MDI), mean decrease accuracy (MDA), and Tree SHapley Additive exPlanations (TreeSHAP) methods, respectively. Text in bold represents ESG features.

Rank	ERT Standard Features	ERT ESG Features	RF Standard Features	RF ESG Features
Panel A: MDI				
1	Market Capitalization	Market Capitalization	Market Capitalization	Market Capitalization
2	GICS Industry	**TRDIR Diversity Score**	Total Liabilities/Total Tangible Assets	Labor Force Survey
3	Labor Force Survey	**TRDIR People Development Score**	Labor Force Survey	Resident Population
4	Resident Population	Labor Force Survey	Inventory Turnover	Government Spending
5	GICS Group	**TRDIR Inclusion Score**	Adjusted Market Price	**TRDIR Diversity Score**
6	Total Liabilities/Total Tangible Assets	Resident Population	GICS Industry	Total Liabilities/Total Tangible Assets
7	Government Spending	**TRDIR Score**	Cash Conversion Cycle	**Salaries and Wages from CSR reporting**
8	Inventory/Current Assets	GICS Industry	Resident Population	Inventory Turnover
9	Consumer Spending	Government Spending	Inventory/Current Assets	**Number of Employees from CSR reporting**
10	Gross Domestic Product	**Women Employees**	Shillers Cyclically Adjusted P/E	Consumer Spending
11	Gross Profit Margin	**Number of Employees from CSR reporting**	Total Liabilities/Total Assets	**Women Employees**
12	Asset Turnover	Consumer Spending	Asset Turnover	**Water Withdrawal Total**
13	Country	Gross Domestic Product	Gross Profit Margin	**Electricity Purchased**
14	GICS Sector	Total Liabilities/Total Tangible Assets	Receivables Turnover	**CO_2 Equiv Emissions Direct, Scope 1**
15	Inventory Turnover	Inventory/Current Assets	Sales/Stockholders Equity	**TRDIR People Development Score**
16	Adjusted Market Price	GICS Group	Government Spending	**TRDIR Score**
17	S&P Sovereign Rating	**Estimated CO_2 Equivalents Emission Total**	Payables Turnover	Total Debt/Total Assets
18	Receivables Turnover	**Trade Union Representation**	Long-term Debt/Total Liabilities	Asset Turnover
19	Total Liabilities/Total Assets	Asset Turnover	Total Debt/Total Assets	Gross Profit Margin
20	Retail Sales	**Employee Satisfaction**	Stock Market Index	GICS Industry
Panel B: MDA				
1	Market Capitalization	Market Capitalization	Market Capitalization	Market Capitalization
2	GICS Industry	Labor Force Survey	Resident Population	**TRDIR Diversity Score**
3	GICS Group	**TRDIR Diversity Score**	Labor Force Survey	Resident Population
4	Developed Nation	Developed Nation	GICS Industry	Labor Force Survey
5	GICS Sector	**Global Compact Signatory**	Number of Analysts	GICS Industry
6	Recommendation - Total Number	**Staggered Board Structure**	Adjusted Market Price	**TRDIR Score**
7	Labor Force Survey	**Targets Emissions**	Recommendation - Total Number	**Employee Accidents**
8	Number of Analysts	**Limitation of Director Liability**	GICS Group	**TRDIR People Development Score**
9	S&P Sovereign Rating	**EMS Certified Percent**	Total Liabilities/Total Tangible Assets	**TRDIR Inclusion Score**
10	Country	**OECD Guidelines for Multinationals**	Gross Profit Margin	GICS Group
11	Adjusted Market Price	GICS Industry	Interest Coverage Ratio	**Water Withdrawal Total**
12	Inventory/Current Assets	**Product Impact Minimization**	Shillers Cyclically Adjusted P/E	**Salaries and Wages from CSR reporting**
13	Consumer Confidence Index	**CSR Sustainability External Audit**	Asset Turnover	Interest Coverage Ratio
14	Total Debt/Total Assets	**Targets Energy Efficiency**	Country	**Water Recycled**
15	Government Spending	**Healthy Food or Products**	Inventory Turnover	**Employee Satisfaction**
16	Total Debt/Capital	**Veto Power or Golden share**	Total Liabilities/Total Assets	**CO_2 Equivalent Emissions Total**
17	Gross Profit Margin	**TRDIR Inclusion Score**	Inventory/Current Assets	Adjusted Market Price
18	Resident Population	**Employee Satisfaction**	Money Supply MO	Number of Analysts
19	Interest Coverage Ratio	**CEO Board Member**	Government Spending	Consumer Spending
20	Consumer Spending	**Flexible Working Hours**	Stock Market Index	Recommendation - Total Number
Panel C: TreeSHAP				
1	Market Capitalization	Market Capitalization	Market Capitalization	Market Capitalization
2	GICS Sector	**TRDIR Diversity Score**	Labor Force Survey	Labor Force Survey
3	GICS Industry	GICS Industry	Resident Population	**TRDIR Diversity Score**
4	Recommendation - Total Number	**TRDIR People Development Score**	GICS Industry	Resident Population
5	Labor Force Survey	Labor Force Survey	Recommendation - Total Number	Recommendation - Total Number
6	GICS Group	**TRDIR Inclusion Score**	Number of Analysts	GICS Industry
7	Number of Analysts	**EMS Certified Percent**	Interest Coverage Ratio	**TRDIR People Development Score**
8	S&P Sovereign Rating	**CSR Sustainability External Audit**	Adjusted Market Price	Number of Analysts
9	Resident Population	GICS Group	GICS Group	Interest Coverage Ratio
10	Interest Coverage Ratio	**Targets Emissions**	Shillers Cyclically Adjusted P/E	Shillers Cyclically Adjusted P/E
11	Inventory/Current Assets	Recommendation - Total Number	Total Liabilities/Total Tangible Assets	Adjusted Market Price
12	Government Spending	**TRDIR Score**	Inventory Turnover	**Salaries and Wages from CSR reporting**
13	Adjusted Market Price	S&P Sovereign Rating	Interest/Average Total Debt	Dividend Payout Ratio
14	Country	Resident Population	Total Debt/Total Assets	**Employee Accidents**
15	Developed Nation	**Employee Satisfaction**	Dividend Payout Ratio	After-tax Interest Coverage
16	Total Debt/Capital	Developed Nation	Government Spending	**CO_2 Equivalent Emissions Total**
17	Total Debt/Total Assets	GICS Sector	Cash Conversion Cycle	**TRDIR Inclusion Score**
18	Total Liabilities/Total Tangible Assets	**Staggered Board Structure**	After-tax Interest Coverage	**TRDIR Score**
19	Recommendation - Number Of Hold	Number of Analysts	Gross Profit Margin	**CO_2 Equiv Emissions Indirect, Scope 2**
20	After-tax Interest Coverage	**Climate Change Commercial Risks**	Asset Turnover	**Number of Employees from CSR reporting**

4.4.3 Consistent US and global features

To investigate feature importance of corporate credit ratings, three feature importance methods and two tree-based ensemble methods are utilized across US and global samples. Table 15 reports in Panel A (B) the features consistent across the three feature importance methods for the US (global) standard and ESG samples. The features with normal text are identified with all three methods, whilst features with italic text are consistent across two methods.

For US firms in the standard sample, GICS industry, dividend yield, total institutional ownership % of shares, total institutional ownership, number of 13-F institutional owners, long-term debt/total liabilities, and market capitalization are demonstrated across MDI, MDA, and TreeSHAP importance methods for ERT and RF. For RF (ERT), we find that 15 (8) of the traditional features are consistent across the three methods including: price/operating earnings (diluted, excluding extraordinary items), forward P/E to 1-year growth ratio, forward P/E to long-term growth ratio, interest coverage ratio, total debt/total assets, inventory turnover, and largest 10 institutional ownership size.

In the US-ESG sample, for ERT, eight (17) features are consistent across three (two) feature importance methods. Moreover, for RF, 11 (20) features are consistent across three (two) methods. Market capitalization, GICS industry, number of 13-F institutional owners, DS, PDS, and Score are consistent across the three methods for ERT and RF. For US firms, 10 ESG features are found in at least two of the methods for ERT, compared with six in RF. For ERT, these include: HRC corporate equality index, elimination of cumulative voting rights, CSR sustainability report global activities, confidential voting policy, director election majority requirement, and employee engagement voluntary work. For RF, this includes total donations to revenue and donations total.

In Panel B, given there is no institutional or dividend yield features in the global sample, thus these features cannot demonstrate consistency with the US sample. In the GLB-STD sample, for an ERT, 11 traditional features are exhibited across the three methods, whilst 20 features are evident across at least two methods. Non financial features demonstrate greater consistency with GICS industry, GICS sector, GICS group, country, S&P sovereign rating, government spending, resident population, and labor force survey reported across the three methods. Financial ratios and analyst features are primarily consistent across at least two methods (e.g., total debt/total assets, interest coverage ratio, and number of analysts). Similar features demonstrate consistency in the RF across the feature importance methods; however, there is a greater proportion of financial features evident.

The introduction of ESG features into the global sample finds nine ESG features consistent across at least two methods for ERT and RF. Again, TRDIR feature variants (DS, IS, PDS, Score) are the only ESG features reported across both models. ERT specific ESG features include: employee satisfaction, CSR sustainability external audit, EMS certified percent, targets emissions, and staggered board structure. RF specific ESG features include: salaries and wages from CSR reporting, water withdrawal total, CO_2 equivalent emissions total, employee accidents, and number of employees from CSR reporting. Macroeconomic features such as resident population, and labor force survey, industry feature of GICS industry, and market capitalization are reported for across both models within the GLB-ESG sample.

Panel C reports the features that are evident across both US and global samples. Again, font in normal text are consistent across MDI, MDA and TreeSHAP methods, whilst italics text is reported across at least two methods. GICS industry and market capitalization are the only features evident across the standard and ESG samples for both ERT and RF. For ERT standard, GICS sector, S&P sovereign rating, GICS group, and gross profit margin are evident across all methods for US and global firms. Moreover, number of analysts and total number of recommendations are exhibited across at least two feature importance methods in both US and global samples. For RF standard, inventory turnover, total liabilities/total tangible assets, and total debt/total assets are consistent across three methods, whilst interest coverage ratio and Shillers cyclically adjusted P/E ratio is evident across at least two methods for US and global samples.

In terms of the ESG samples, DS, PDS, IS, and Score are the only ESG features that are consistent across ERT and RF and at least two feature importance methods for the US and global samples. DS is the only ESG feature exhibited across MDI, MDA and TreeSHAP. As this metric is comprised of diversity and cultural diversity features, this indicates that diversity metrics are an important driver of corporate credit rating predictability. Moreover, the importance of the TRDIR metrics of DS, PDS, IS, and Score, demonstrates that ESG features do matter in corporate credit rating prediction across US and global firms. As the TRDIR feature constituents are primarily social and governance based, this also supports the findings presented in Attig et al. (2013) and Kiesel and Lücke (2019).

Table 15: US and global important features

This table reports features consistent across extremely randomized trees (ERT) and random forests (RF) utilizing mean decrease impurity (MDI), mean decrease accuracy (MDA) and Tree SHapley Additive exPlanations (TreeSHAP) feature importance methods in Table 13 and 14. After recursive reduction to 20 features, important features for the United States (US) standard and environmental, social, and governance (ESG) samples are reported in Panel A, whilst the global standard and ESG samples are exhibited in Panel B. Text in normal font represents features found in the 20 feature model across all three feature importance methods - MDI, MDA, and TreeSHAP, while font in italics details features in the 20 feature model across two feature importance methods. Moreover, text in bold represents ESG features.

ERT Standard Features	ERT ESG Features	RF Standard Features	RF ESG Features
Panel A: US sample			
GICS Industry	**TRDIR Diversity Score**	Price/Operating Earnings (Diluted)	Inventory Turnover
Dividend Yield	**TRDIR People Development Score**	Number of 13-F Institutional Owners	GICS Industry
Interest/Average Total Debt	Market Capitalization	Dividend Yield	Total Institutional Ownership
Total Institutional Ownership % of Shares	Long-term Debt/Total Liabilities	Forward P/E to 1-year Growth Ratio	**TRDIR Score**
Total Institutional Ownership	**TRDIR Score**	Market Capitalization	Inventory/Current Assets
Number of 13-F Institutional Owners	GICS Industry	Inventory Turnover	Number of 13-F Institutional Owners
Long-term Debt/Total Liabilities	TRDIR Inclusion Score	Total Liabilities/Total Tangible Assets	Market Capitalization
Market Capitalization	Number of 13-F Institutional Owners	Total Institutional Ownership	**TRDIR People Development Score**
Gross Profit Margin	***HRC Corporate Equality Index***	Total Institutional Ownership % of Shares	**TRDIR Diversity Score**
Capitalization Ratio	***Elimination of Cumulative Voting Rights***	GICS Industry	Total Debt/Total Assets
GICS Sector	***CSR Sustainability Report Global Activities***	Long-term Debt/Total Liabilities	Interest Coverage Ratio
S&P Sovereign Rating	***Confidential Voting Policy***	Forward P/E to Long-term Growth Ratio	*Dividend Yield*
GICS Group	*Total Institutional Ownership*	Interest Coverage Ratio	***Total Donations To Revenues***
Common Equity/Invested Capital	***Director Election Majority Requirement***	Total Debt/Total Assets	*Total Liabilities/Total Tangible Assets*
Number of Analysts	***Employee Engagement Voluntary Work***	Largest 10 Institutional Ownership Size	*Dividend Payout Ratio*
Recommendation - Total Number	*GICS Sector*	*Dividend Payout Ratio*	*Long-term Debt/Total Liabilities*
	GICS Group	*Interest/Average Total Debt*	*Cash Conversion Cycle*
		Shillers Cyclically Adjusted P/E Ratio	*Gross Profit Margin*
			Donations Total
			TRDIR Inclusion Score
Panel B: Global sample			
Market Capitalization	Market Capitalization	Market Capitalization	Market Capitalization
Inventory/Current Assets	**TRDIR Inclusion Score**	Gross Profit Margin	**TRDIR Diversity Score**
GICS Industry	**TRDIR Diversity Score**	Resident Population	**Salaries and Wages from CSR reporting**
Country	GICS Industry	USD Adjusted Market Price	GICS Industry
USD Adjusted Market Price	**Employee Satisfaction**	GICS Industry	Labor Force Survey
S&P Sovereign Rating	Labor Force Survey	Labor Force Survey	**TRDIR Score**
GICS Sector	Resident Population	Inventory Turnover	**TRDIR People Development Score**
Government Spending	GICS Group	Shillers Cyclically Adjusted P/E Ratio	Resident Population
GICS Group	**CSR Sustainability External Audit**	Asset Turnover	Number of Analysts
Resident Population	**EMS Certified Percent**	Total Liabilities/Total Tangible Assets	Consumer Spending
Labor Force Survey	Targets Emissions	Government Spending	Recommendation - Total Number
Developed Nation	**TRDIR Score**	Total Debt/Total Assets	Interest Coverage Ratio
Total Debt/Total Assets	*Developed Nation*	GICS Group	**TRDIR Inclusion Score**
Total Debt/Capital	***TRDIR People Development Score***	*Recommendation - Total Number*	**Water Withdrawal Total**
Interest Coverage Ratio	***Staggered Board Structure***	*Number of Analysts*	***CO_2 Equivalent Emissions Total***
Number of Analysts		*Stock Market Index*	***Employee Accidents***
Gross Profit Margin		*Cash Conversion Cycle*	*USD Adjusted Market Price*
Consumer Spending		*Inventory/Current Assets*	***Number of Employees from CSR reporting***
Total Liabilities/Total Tangible Assets		*Total Liabilities/Total Assets*	
Recommendation - Total Number		*Interest Coverage Ratio*	
Panel C: Consistent US and global sample			
GICS Industry	GICS Industry	GICS Industry	GICS Industry
Market Capitalization	Market Capitalization	Market Capitalization	Market Capitalization
GICS Sector	**TRDIR Diversity Score**	Inventory Turnover	**TRDIR Diversity Score**
S&P Sovereign Rating	**TRDIR Inclusion Score**	Total Liabilities/Total Tangible Assets	**TRDIR People Development Score**
GICS Group	***TRDIR People Development Score***	Total Debt/Total Assets	**TRDIR Score**
Gross Profit Margin	***TRDIR Score***	*Interest Coverage Ratio*	*Interest Coverage Ratio*
Number of Analysts	*GICS Group*	*Shillers Cyclically Adjusted P/E Ratio*	***TRDIR Inclusion Score***
Recommendation - Total Number			

4.4.4 Investment- and speculative-grade

As each corporate credit rating class can have differentiating features driving the performance, approximated TreeSHAP values are estimated on a per class basis. As MDI and MDA values do not have per class functionality, the iterative process evaluated in Table 11 is used to reduce the feature set to $n = 20$ for all classes in Table 16 and Table 17. The individual class features are combined into their respective investment-grade (i.e., >= BBB-) or speculative-grade (i.e., <= BB+) class and are reported in cumulative count order.

In Table 16, Panel A highlights the aggregated features across all individual credit rating classes for US firms. For standard ERT, GICS industry and number of 13-F institutional owners are highly important features across all of the 22 credit rating classes (i.e., AAA to D).[32] As there are 10 investment-grade classes (i.e., AAA to BBB-) and 12 speculative-grade classes (i.e., BB+ to D), this indicates that the two aforementioned features are highly important across all investment- and speculative-grade classes. Similarly, GICS group, long-term debt/total liabilities, market capitalization, long-term debt/invested capital, and dividend yield are important across at least 19 credit rating classes. When ESG features are introduced, number of 13-F institutional owners, IS, Score, DS, PDS, and market capitalization are important drivers across ERT and RF.

Panel B and C report the recursively reduced TreeSHAP aggregated investment-grade and aggregated speculative-grade classes, respectively. In the standard sample, for ERT (RF), 14 (15) features are consistent across investment- and speculative-grade classes; however, the number of classes exhibiting importance varies. In ERT, higher importance is given to interest/average total debt and total number of analyst recommendations in the speculative-grade classes. Moreover, 13 financial ratios comprise the 20 for speculative-grade classes in comparison to eight for the investment-grade classes. Whilst capitalization and financial soundness ratios are consistent, profitability and solvency ratios are evident in the speculative-grade classes.

Similar findings are presented in the RF standard, with 16 financial ratios comprising the 20 for speculative-grade classes compared to 13 with investment-grade classes. However, the financial ratios driving the performance for RF differ to those in ERT, for example, valuation and efficiency

[32] Approximated TreeSHAP values are utilized to recursively reduce the feature set for each class to $n = 50, 40, 30, 20$. If a feature has an aggregation score (Rating Classes column) of 22, this indicates this feature is important across all rating classes when a $n = 20$ feature model is used.

ratios (e.g., forward P/E to 1-year growth ratio and inventory turnover) are highly important across both investment- and speculative-grade classes. The financial soundness, solvency, and capitalization ratios are consistent with ERT.

In the ESG sample, number of 13-F institutional owners, market capitalization, Score, PDS, and DS are highly important investment- and speculative-grade classes features for ERT and RF. In ERT, 14 (12) ESG features comprise the 20 for investment- (speculative-) grade classes, whilst RF exhibits 6 across both investment- and speculative-grade classes. Confidential voting policy, elimination of cumulative voting rights, and HRC corporate equality index are consistent ESG across all classes, whilst the additional ESG features are largely consistent with previously discussed features. For RF, IS is reported in the investment-grade classes, whilst DS, PDS, and Score are evident across investment- and speculative-grade classes. Similar to ERT, a larger proportion of financial ratios are important in the speculative-grade classes, with 9 of the top 12 features financial ratios. In comparison, 6 of the top 12 features for investment-grade classes are financial ratios.

Similar to the US, Panel A, B, and C in Table 17 reports the aggregated features across all individual classes, investment-grade classes, and speculative-grade classes, respectively, for the global dataset. Features highlighted in the overall count (Panel A) consistently appear in the speculative- and investment-grade counts, however, the order of importance differs. In the GLB-STD samples, for ERT (RF), the features are largely consistent with 15 (13) across the investment- and speculative-grades. Number of analysts and total number of analyst recommendations, market capitalization, labor force survey, GICS industry, GICS group, and S&P sovereign rating demonstrate high importance across all classes.

Comparable with the US findings, speculative-grade classes tend to exhibit a greater proportion of financial ratio features across the standard and ESG samples. For example, in the standard (ESG) sample, and using a RF, nine (two) financial ratios are evident in investment-grade classes, whilst 12 (7) are exhibited in the speculative-grade classes. IS is a highly important ESG feature for investment-grade classes across ERT and RF. However, IS is only reported as an important driver in four speculative-grade classes for an ERT, whilst, for RF, it is not evident in at least four speculative-grade classes. The importance of ESG is also more pronounced in investment-grade classes for global firms. For a RF (ERT), 11 (12) ESG features comprise the 20 for investment-grade, whilst 7 (9) are included in the speculative-grade classes.

Table 16: US sample investment- and speculative-grade

Approximated Tree SHapley Additive exPlanations (TreeSHAP) recursively reduce features to 20 for each rating class in the United States (US) standard and environmental, social, and governance (ESG) samples. Extremely randomized trees (ERT) and random forests (RF) models are utilized. Panel A reports the aggregated feature count across all classes, Panel B reports the aggregated feature count across investment-grade classes, whilst Panel C reports the aggregated feature count across speculative-grade classes. Aggregated feature count represents the number of classes which a specific feature appears in the individual 20 feature model (e.g., In Panel A, 22 in Rating Classes column symbolises GICS Industry is a top 20 important feature for all classes). Text in bold represents ESG features.

ERT Standard Features		ERT ESG Features		RF Standard Features		RF ESG Features	
Rating Classes		Rating Classes		Rating Classes		Rating Classes	
Panel A: TreeSHAP all classes							
22	GICS Industry	19	Number of 13-F Institutional Owners	22	Dividend Yield	20	Market Capitalization
22	Number of 13-F Institutional Owners	19	**TRDIR Score**	22	Price/Operating Earnings (Diluted)	20	**TRDIR Score**
20	GICS Group	18	**TRDIR Inclusion Score**	22	Dividend Payout Ratio	20	Dividend Yield
19	Long-term Debt/Total Liabilities	18	**TRDIR People Development Score**	22	Forward P/E to 1-year Growth Ratio	20	Number of 13-F Institutional Owners
19	Market Capitalization	18	**TRDIR Diversity Score**	22	Market Capitalization	18	Inventory Turnover
19	Long-term Debt/Invested Capital	18	Market Capitalization	22	Interest Coverage Ratio	18	Interest Coverage Ratio
19	Dividend Yield	17	GICS Industry	21	Number of 13-F Institutional Owners	17	Total Debt/Total Assets
17	Common Equity/Invested Capital	17	**Elimination of Cumulative Voting Rights**	20	Forward P/E to Long-term Growth Ratio	16	**TRDIR Diversity Score**
17	GICS Sector	13	GICS Group	18	Long-term Debt/Total Liabilities	16	Dividend Payout Ratio
17	S&P Sovereign Rating	10	Long-term Debt/Total Liabilities	18	Total Institutional Ownership % of Shares	14	**TRDIR People Development Score**
17	Number of Analysts	10	**HRC Corporate Equality Index**	17	Total Institutional Ownership	13	Inventory/Current Assets
17	Capitalization Ratio	10	**Confidential Voting Policy**	17	Shillers Cyclically Adjusted P/E ratio	11	Total Institutional Ownership
15	Interest/Average Total Debt	9	GICS Sector	17	Inventory Turnover	11	Common Equity/Invested Capital
14	Recommendation - Number Of Total	8	Capitalization Ratio	16	GICS Industry	11	**TRDIR Inclusion Score**
12	Short-Term Debt/Total Debt	7	**CSR Sustainability Report Global Activities**	12	After-tax Interest Coverage	10	**Donations Total**
12	Total Institutional Ownership % of Shares	6	**Unlimited Authorized Capital/Blank Check**	11	Total Debt/Total Assets	9	GICS Industry
10	Total Debt/Capital	6	**Director Election Majority Requirement**	10	Total Liabilities/Total Tangible Assets	9	**Total Donations To Revenues**
10	Return on Capital Employed	6	Total Institutional Ownership	10	Interest/Average Total Debt	9	After-tax Interest Coverage
9	Total Institutional Ownership	6	**Policy Executive Comp ESG Perf**	9	Common Equity/Invested Capital	8	Cash Conversion Cycle
8	Pre-tax Profit Margin	6	**Employee Engagement Voluntary Work**	9	Largest 10 Institutional Ownership Size	7	Long-term Debt/Total Liabilities
Panel B: TreeSHAP investment-grade							
10	GICS Industry	9	Number of 13-F Institutional Owners	10	Dividend Yield	9	Market Capitalization
10	GICS Group	9	**TRDIR Inclusion Score**	10	Interest Coverage Ratio	9	Dividend Yield
10	Long-term Debt/Total Liabilities	9	**TRDIR People Development Score**	10	Long-term Debt/Total Liabilities	9	Number of 13-F Institutional Owners
10	Market Capitalization	9	**TRDIR Diversity Score**	10	Number of 13-F Institutional Owners	9	Interest Coverage Ratio
10	Number of 13-F Institutional Owners	9	Market Capitalization	10	Price/Operating Earnings (Diluted)	9	**TRDIR Score**
10	GICS Sector	9	**TRDIR Score**	10	Market Capitalization	9	**TRDIR Inclusion Score**
10	S&P Sovereign Rating	8	GICS Industry	10	Forward P/E to 1-year Growth Ratio	8	Inventory Turnover
9	Capitalization Ratio	7	**Elimination of Cumulative Voting Rights**	10	Dividend Payout Ratio	8	**TRDIR Diversity Score**
9	Long-term Debt/Invested Capital	7	GICS Group	9	Total Institutional Ownership % of Shares	8	Dividend Payout Ratio
9	Dividend Yield	6	GICS Sector	9	GICS Industry	8	**TRDIR People Development Score**
8	Total Institutional Ownership % of Shares	5	Total Institutional Ownership	9	Total Institutional Ownership	8	Total Institutional Ownership
8	Short-Term Debt/Total Debt	5	Long-term Debt/Total Liabilities	8	Forward P/E to Long-term Growth Ratio	6	Total Debt/Total Assets
8	Common Equity/Invested Capital	5	**HRC Corporate Equality Index**	6	Shillers Cyclically Adjusted P/E ratio	5	**Donations Total**
6	Number of Analysts	3	**Product Impact Minimization**	6	Gross Profit/Total Assets	5	Cash Conversion Cycle
6	Total Institutional Ownership	3	**Animal Testing Reduction**	6	Largest 10 Institutional Ownership Size	5	Inventory/Current Assets
5	Recommendation - Number Of Total	3	**TRDIR Controversies Score**	5	Total Debt/EBITDA	4	Total Institutional Ownership % of Shares
4	Interest/Average Total Debt	3	**Confidential Voting Policy**	5	Inventory Turnover	4	**CO_2 Equiv Emissions Direct, Scope 1**
4	Largest 10 Institutional Ownership Size	3	**Director Election Majority Requirement**	5	Retail Sales	4	Gross Profit/Total Assets
4	Retail Sales	3	**Unlimited Authorized Capital/Blank Check**	5	Total Liabilities/Total Tangible Assets	4	Gross Profit Margin
4	Resident Population	2	**Advance Notice Period Days**	5	Government Spending	4	Largest 10 Institutional Ownership Size
Panel C: TreeSHAP speculative-grade							
12	GICS Industry	10	Number of 13-F Institutional Owners	12	Dividend Yield	11	Market Capitalization
12	Number of 13-F Institutional Owners	10	**TRDIR Score**	12	Forward P/E to 1-year Growth Ratio	11	**TRDIR Score**
11	Number of Analysts	10	**Elimination of Cumulative Voting Rights**	12	Inventory Turnover	11	Dividend Yield
11	Interest/Average Total Debt	9	**TRDIR Inclusion Score**	12	Interest Coverage Ratio	11	Total Debt/Total Assets
10	GICS Group	9	**TRDIR People Development Score**	12	Dividend Payout Ratio	11	Number of 13-F Institutional Owners
10	Long-term Debt/Invested Capital	9	**TRDIR Diversity Score**	12	Forward P/E to Long-term Growth Ratio	10	Inventory Turnover
10	Dividend Yield	9	Market Capitalization	12	Market Capitalization	9	Interest Coverage Ratio
9	Long-term Debt/Total Liabilities	9	GICS Industry	12	Price/Operating Earnings (Diluted)	9	Common Equity/Invested Capital
9	Market Capitalization	7	**Confidential Voting Policy**	11	Number of 13-F Institutional Owners	8	Dividend Payout Ratio
9	Recommendation - Number Of Total	6	**CSR Sustainability Report Global Activities**	11	Shillers Cyclically Adjusted P/E ratio	8	Inventory/Current Assets
9	Common Equity/Invested Capital	6	GICS Group	9	Total Institutional Ownership % of Shares	8	**TRDIR Diversity Score**
8	Capitalization Ratio	6	Capitalization Ratio	9	Interest/Average Total Debt	7	After-tax Interest Coverage
8	Total Debt/Capital	5	Long-term Debt/Total Liabilities	9	Total Debt/Total Assets	6	**TRDIR People Development Score**
7	Pre-tax Profit Margin	5	**HRC Corporate Equality Index**	8	Total Institutional Ownership	5	**Donations Total**
7	Return on Capital Employed	4	Dividend Yield	8	Long-term Debt/Total Liabilities	5	Forward P/E to Long-term Growth Ratio
7	S&P Sovereign Rating	4	**Employee Engagement Voluntary Work**	8	After-tax Interest Coverage	5	Long-term Debt/Total Liabilities
7	GICS Sector	4	**Policy Executive Comp ESG Perf**	7	GICS Industry	5	**Total Donations To Revenues**
5	Total Debt/Total Assets	4	Common Equity/Invested Capital	7	Inventory/Current Assets	5	GICS Industry
5	Net Profit Marginq	3	**Policy Water Efficiency**	7	Price/Operating Earnings (Basic)	5	**Average Board Tenure**
5	Return on Assets	3	**Green Buildings**	6	Common Equity/Invested Capital	4	Capitalization Ratio

Table 17: Global sample investment- and speculative-grade

Approximated Tree SHapley Additive exPlanations (TreeSHAP) recursively reduce features to 20 for each rating class in the global standard and environmental, social, and governance (ESG) samples. Extremely randomized trees (ERT) and random forests (RF) models are utilized. Panel A reports the aggregated feature count across all classes, Panel B reports the aggregated feature count across investment-grade classes, whilst Panel C reports the aggregated feature count across speculative-grade classes. Aggregated feature count represents the number of classes which a specific feature appears in the individual 20 feature model (e.g., In Panel A, 22 in Rating Classes column symbolises Market Capitalization is a top 20 important feature for all classes). Text in bold represents ESG features.

ERT Standard Features		ERT ESG Features		RF Standard Features		RF ESG Features	
Rating Classes		Rating Classes		Rating Classes		Rating Classes	
Panel A: TreeSHAP all classes							
22	Market Capitalization	20	Market Capitalization	22	Market Capitalization	22	Market Capitalization
22	GICS Sector	18	**TRDIR Diversity Score**	22	Labor Force Survey	21	Labor Force Survey
22	GICS Industry	17	GICS Industry	22	Resident Population	19	**TRDIR Diversity Score**
22	Recommendation - Number Of Total	15	**TRDIR People Development Score**	22	GICS Industry	17	Resident Population
22	Labor Force Survey	15	Labor Force Survey	21	Recommendation - Number Of Total	16	Recommendation - Number Of Total
22	GICS Group	13	**TRDIR Inclusion Score**	20	Number of Analysts	16	GICS Industry
21	Number of Analysts	13	**EMS Certified Percent**	19	Interest Coverage Ratio	15	**TRDIR People Development Score**
21	S&P Sovereign Rating	13	**CSR Sustainability External Audit**	18	USD Adjusted Market Price	14	Number of Analysts
19	Resident Population	12	GICS Group	17	GICS Group	13	Interest Coverage Ratio
17	Interest Coverage Ratio	11	**Targets Emissions**	16	Shillers Cyclically Adjusted P/E ratio	13	Shillers Cyclically Adjusted P/E ratio
16	Inventory/Current Assets	9	Recommendation - Number Of Total	15	Total Liabilities/Total Tangible Assets	11	USD Adjusted Market Price
16	Government Spending	9	TRDIR Score	15	Inventory Turnover	11	**Salaries and Wages from CSR reporting**
16	USD Adjusted Market Price	9	S&P Sovereign Rating	12	Interest/Average Total Debt	10	Dividend Payout Ratio
14	Country	9	Resident Population	12	Total Debt/Total Assets	10	**Employee Accidents**
13	Developed Nation	9	**Employee Satisfaction**	12	Dividend Payout Ratio	10	After-tax Interest Coverage
12	Total Debt/Capital	8	Developed Nation	12	Government Spending	9	CO_2 Equivalent Emissions Total
11	Total Debt/Total Assets	8	GICS Sector	12	Cash Conversion Cycle	9	**TRDIR Inclusion Score**
10	Total Liabilities/Total Tangible Assets	8	**Staggered Board Structure**	11	After-tax Interest Coverage	9	**TRDIR Score**
10	Recommendation - Number Of Hold	7	Number of Analysts	11	Gross Profit Margin	7	**CO_2 Equiv Emissions Indirect, Scope 2**
10	After-tax Interest Coverage	7	**Climate Change Commercial Risks**	10	Asset Turnover	6	**Number of Employees from CSR reporting**
Panel B: TreeSHAP investment-grade							
10	Market Capitalization	9	Market Capitalization	10	Market Capitalization	10	Market Capitalization
10	GICS Industry	9	**TRDIR Diversity Score**	10	Resident Population	10	Resident Population
10	Recommendation - Number Of Total	9	**TRDIR Inclusion Score**	10	GICS Industry	10	Labor Force Survey
10	Labor Force Survey	8	GICS Industry	10	Labor Force Survey	9	**TRDIR Diversity Score**
10	GICS Group	8	**EMS Certified Percent**	9	Total Liabilities/Total Tangible Assets	8	**TRDIR Inclusion Score**
10	S&P Sovereign Rating	7	Labor Force Survey	9	Inventory Turnover	8	**Salaries and Wages from CSR reporting**
10	Resident Population	6	GICS Group	9	Asset Turnover	7	GICS Industry
10	GICS Sector	6	S&P Sovereign Rating	9	Number of Analysts	6	**TRDIR People Development Score**
9	Number of Analysts	6	**Employee Satisfaction**	9	Recommendation - Number Of Total	6	Recommendation - Number Of Total
9	Country	5	**TRDIR People Development Score**	8	USD Adjusted Market Price	5	**Waste Recycled To Total Waste**
8	Developed Nation	5	**CSR Sustainability External Audit**	8	Interest Coverage Ratio	5	**Employee Accidents**
8	Inventory/Current Assets	5	**TRDIR Score**	8	GICS Group	5	Gross Profit Margin
8	Government Spending	4	**OECD Guidelines for Multinationals**	6	Gross Profit Margin	4	Number of Analysts
8	USD Adjusted Market Price	4	**Flexible Working Hours**	6	Country	4	**TRDIR Score**
7	Interest Coverage Ratio	4	**Trade Union Representation**	6	Sales/Stockholders Equity	4	**Water Recycled**
7	Gross Profit Margin	4	**Veto Power or Golden share**	6	Stock Market Index	4	**Environmental Expenditures**
6	Retail Sales	4	Country	5	GICS Sector	4	**CO_2 Equivalent Emissions Total**
6	Asset Turnover	4	**Staggered Board Structure**	5	Inventory/Current Assets	3	Government Spending
6	Total Liabilities/Total Tangible Assets	4	GICS Sector	5	Total Debt/Total Assets	3	Government External Debt
4	After-tax Interest Coverage	4	Developed Nation	4	Government Spending	3	**VOC Emissions**
Panel C: TreeSHAP speculative-grade							
12	Market Capitalization	11	Market Capitalization	12	Market Capitalization	12	Market Capitalization
12	Recommendation - Number Of Total	10	**TRDIR People Development Score**	12	GICS Industry	11	Labor Force Survey
12	Labor Force Survey	9	**TRDIR Diversity Score**	12	Labor Force Survey	11	Interest Coverage Ratio
12	GICS Group	9	**Targets Emissions**	12	Resident Population	10	**TRDIR Diversity Score**
12	GICS Sector	9	Recommendation - Number Of Total	12	Shillers Cyclically Adjusted P/E ratio	10	Recommendation - Number Of Total
12	Number of Analysts	9	GICS Industry	12	Recommendation - Number Of Total	10	Number of Analysts
12	GICS Industry	8	Labor Force Survey	11	Interest Coverage Ratio	10	Shillers Cyclically Adjusted P/E ratio
11	S&P Sovereign Rating	8	**CSR Sustainability External Audit**	11	Number of Analysts	10	Dividend Payout Ratio
10	Interest Coverage Ratio	7	Number of Analysts	10	USD Adjusted Market Price	9	USD Adjusted Market Price
9	Resident Population	7	**Climate Change Commercial Risks**	10	Dividend Payout Ratio	9	GICS Industry
8	Recommendation - Number Of Hold	6	Resident Population	9	Cash Conversion Cycle	9	**TRDIR People Development Score**
8	Total Debt/Total Assets	6	GICS Group	9	Interest/Average Total Debt	9	After-tax Interest Coverage
8	Total Debt/Capital	5	**EMS Certified Percent**	9	GICS Group	7	Resident Population
8	USD Adjusted Market Price	5	Interest Coverage Ratio	8	Government Spending	6	**CO_2 Equiv Emissions Indirect, Scope 2**
8	Government Spending	5	**Stakeholder Engagement**	8	After-tax Interest Coverage	5	Total Debt/Total Assets
8	Inventory/Current Assets	5	Government Spending	7	Total Debt/Total Assets	5	**CO_2 Equivalent Emissions Total**
7	Shillers Cyclically Adjusted P/E ratio	5	Consumer Spending	6	Inventory Turnover	5	**CO_2 Equiv Emissions Direct, Scope 1**
6	After-tax Interest Coverage	5	After-tax Interest Coverage	6	Total Liabilities/Total Tangible Assets	5	**TRDIR Score**
6	Interest/Average Total Debt	4	**TRDIR Score**	5	Gross Profit Margin	5	**Employee Accidents**
5	Developed Nation	4	**TRDIR Inclusion Score**	5	Interest/Average Long-term Debt	4	Total Liabilities/Total Assets

4.4.5 TreeSHAP feature drivers

Tree-based ensemble statistical classification models (i.e., RF, ERT, XGB) have not been widely adopted in financial institutions for credit risk management purposes due their lack of interpretability. Although sophisticated statistical learning models exhibit powerful prediction ability, financial regulators are reluctant to approve such models for use due to the opaqueness of the interactions between the independent (explanatory) variables with each other and their impact on the dependent (response) variable.[33] The introduction of SHAP values are useful as they allow for an interpretation of sophisticated statistical learning models with a clarity that traditionally has only been reserved for linear models (e.g., logistic regression).

We broadly define three rating classes as investment-grade (i.e., AAA to A+), high-yield (i.e., BBB+ to BBB-) and speculative-grade (i.e., CCC+ to D).[34] TreeSHAP is applied to estimate SHAP values to explain the aggregated local feature contributions for these classes as shown in Figure 2 and Figure 4. These figures can assist with disentangling the impact that feature inputs have into model outputs. By providing intuitive interpretation of the feature movements across classes, TreeSHAP allows for greater insights into the drivers of credit ratings compared to MDI and MDA.

For each rating class (i.e., investment-grade, high-yield, speculative), the y-axis shows the most important 20 features in descending order of importance. The absolute sum of individual SHAP values for each feature is taken to determine the importance rank. The x-axis represents the SHAP values of each feature and its impact on per class model output. The color reveals whether the feature had a high (i.e., red) or low (i.e., blue) value for the individual observation, and the x-axis shows whether the effect of the local value caused a higher or lower prediction. When a feature's distribution exhibits long tails, this highlights that the feature is a strong predictor for individual observations. When a feature's distribution exhibits significant clustering, this demonstrates that a feature has consistently similar impact on the per class model output across the individual observations. Moreover, when the distribution of a feature exhibits clustering around a similar range on the x-axis, this is indicative of significant interactions between these features.

[33]The article "Derisking machine learning and artificial intelligence" published by McKinsey & Company in February 2019 (https://www.mckinsey.com/business-functions/risk/our-insights/derisking-machine-learning-and-artificial-intelligence) discusses that methodologies such as local interpretable model-agnostic explanations or Shapley values help to ensure interpretability of machine learning models for credit risk applications.

[34]Plotting all individual corporate credit rating classes is possible; however, for succinctness, the three illustrated classes exhibit high predictability, whilst reporting high number of observations in the test sample.

Given highly important traditional features remain prevalent in the ESG sample, US-ESG and GLB-ESG SHAP summary plots are presented in Figure 2 and 4, respectively. A RF is utilized as this method, when used with TreeSHAP, outperforms ERT in the test sample. In Figure 2, for the investment-grade class in the US, high (low) values of market capitalization positively (negatively) impact the predictability of this class due to the larger proportion of red values on the positive range of the SHAP value. High values of number of 13-F institutional owners, Score, donations total, and gross profit/total assets also positively impact the probability of the class being investment-grade. Small values of interest coverage ratio, dividend payout rate, and dividend yield tend to decrease the predictability. In the tails in the SHAP summary plot, inventory turnover, gross profit margin, and PDS are less important at an aggregate level; however, for specific individual observations, this feature is a strong predictor of the investment-grade class. This granularity in explanations provides insights beyond the reach of MDI and MDA for tree interpretability.

Vertical spread in the SHAP summary plot is indicative of feature interactions. SHAP interaction values can provide additional model interpretability for input features. These values can be interpreted as the difference between the SHAP values for feature i when feature j is present and the SHAP values for feature i when feature j is absent. The main effects of each feature are visualised on the diagonal, whilst the interaction effects are the off-diagonal plots.[35] Vertical dots centred around 0 indicates feature i and feature j exhibit little to no interaction. Horizontal dispersion demonstrates that feature interaction effects impact the model output.

Figure 3 illustrates the SHAP interaction summary plot for the investment-grade class.[36] Other than dividend payout ratio and inventory turnover, high values of the remaining eight features exhibit interaction effects. For example, high values of number of 13-F institutional owners (i.e, a large number of institutional owners) demonstrate interaction effects with TRDIR score. To illustrate this further, Figure E.5 in Appendix E displays six features interacting with TRDIR Score.

For the high-yield class, BBB-, high values of Score and number of 13-F institutional owners tend to negatively impact BBB- prediction, with low values of market capitalization, institutional ownership, interest coverage ratio, CO_2 equivalent emissions (Scope 1 and Scope 2), and inventory turnover increasing prediction ability. High values of PDS, GICS industry, asset turnover, and

[35]The summary plot of the main effects for each feature is consistent with the SHAP summary plot.

[36]Due to visual limitations, the SHAP interaction plot demonstrates 10 important features of the 20 utilized in the model.

equity/invested capital are also important for BBB- class prediction. Feature points with similar TreeSHAP values cluster together to help visualize the aggregated local contributions density. For example, IS and gross profit margin largely cluster between SHAP values of -0.025 and 0.

The top drivers of the speculative-grade class include market capitalization, dividend yield, number of 13-F institutional owners, interest coverage ratio, and Score. These features positively impact class prediction with low values. The tails in interest/total debt, average board tenure and consumer spending indicate that whilst it is less important than market capitalization as far as aggregated model behavior is concerned, for specific individuals, this feature is a strong predictor. The number of financial features encompassing the 20 feature model are more prevalent in the speculative-grade class. Solvency, capitalization, valuation and financial soundness ratios are identified as the most important. Institutional features exhibit larger positive importance for the higher rated classes, with profitability, efficiency and valuation financial ratios displaying the same behavior.

For the global sample (Figure 4), high values of market capitalization, Score, DS, and IS all positively impact the investment-grade class. Low values of government foreign debt, population, sale/equity, labor force survey and employee accidents also improve A+ class prediction. high-yield class (i.e., BBB-) prediction improves with high values of labor force survey, population and low values of market capitalization. GICS industry and environmental management system certified percent are examples that cluster around mean SHAP values of 0, indicating feature interactions.

ESG features exhibit the highest aggregate importance in the investment-class with 13 features evident. This indicates that ESG features such as DS, employee satisfaction, average training hours and water recycled are highly important. Lower values of ESG features also often increase the likelihood of the investment-class. The speculative-class exhibits a greater proportion of financial features relative to the investment-grade and high yield class. This indicates that for firms considered in distress, financial features, particularly solvency ratios, are highly important. The associated values of these ratios, e.g., total liabilities/total assets, often are larger for firms that are potentially facing insolvency. Low values of Shillers cyclically adjusted P/E ratio, interest coverage ratio, market capitalization and adjusted market price also improve class predictability. Last, ESG features - environment pillar score, and PDS, as well as number of analysts following the firm, decrease the likelihood of being the high speculative-class when these features exhibit high values.

Figure 2: US-ESG SHAP summary plot
This figure illustrates the SHapley Additive exPlanations (SHAP) summary dot plot for United States (US) environmental, social, and governance (ESG) sample and has features ordered in descending importance on the y-axis, with local SHAP values on the x-axis. The color reveals if the feature was high (red) or low (blue) and the x-axis shows if the effect of the local value caused a higher or lower prediction. Random forests (RF) is utilized with 100 trees and depth of 13.

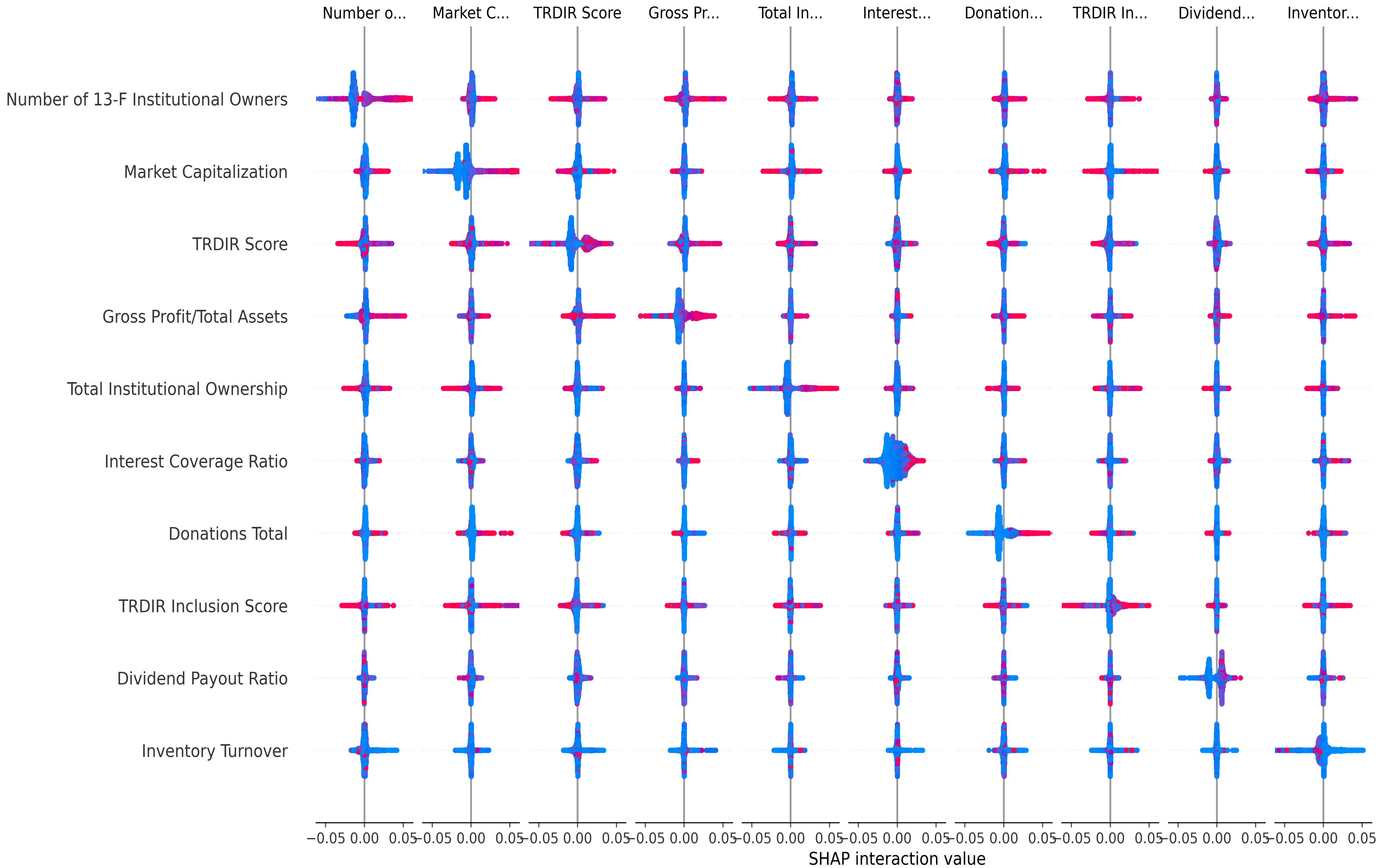

Figure 3: US-ESG SHAP interaction summary plot
This figure illustrates the SHapley Additive exPlanations (SHAP) summary dot plot for United States (US) environmental, social, and governance (ESG) sample and has features ordered in descending importance on the y-axis, with local TreeSHAP values on the x-axis. The color reveals if the feature was high (red) or low (blue) and the x-axis shows if the effect of the local value caused a higher or lower prediction. Random forests (RF) is utilized with 100 trees and depth of 13.

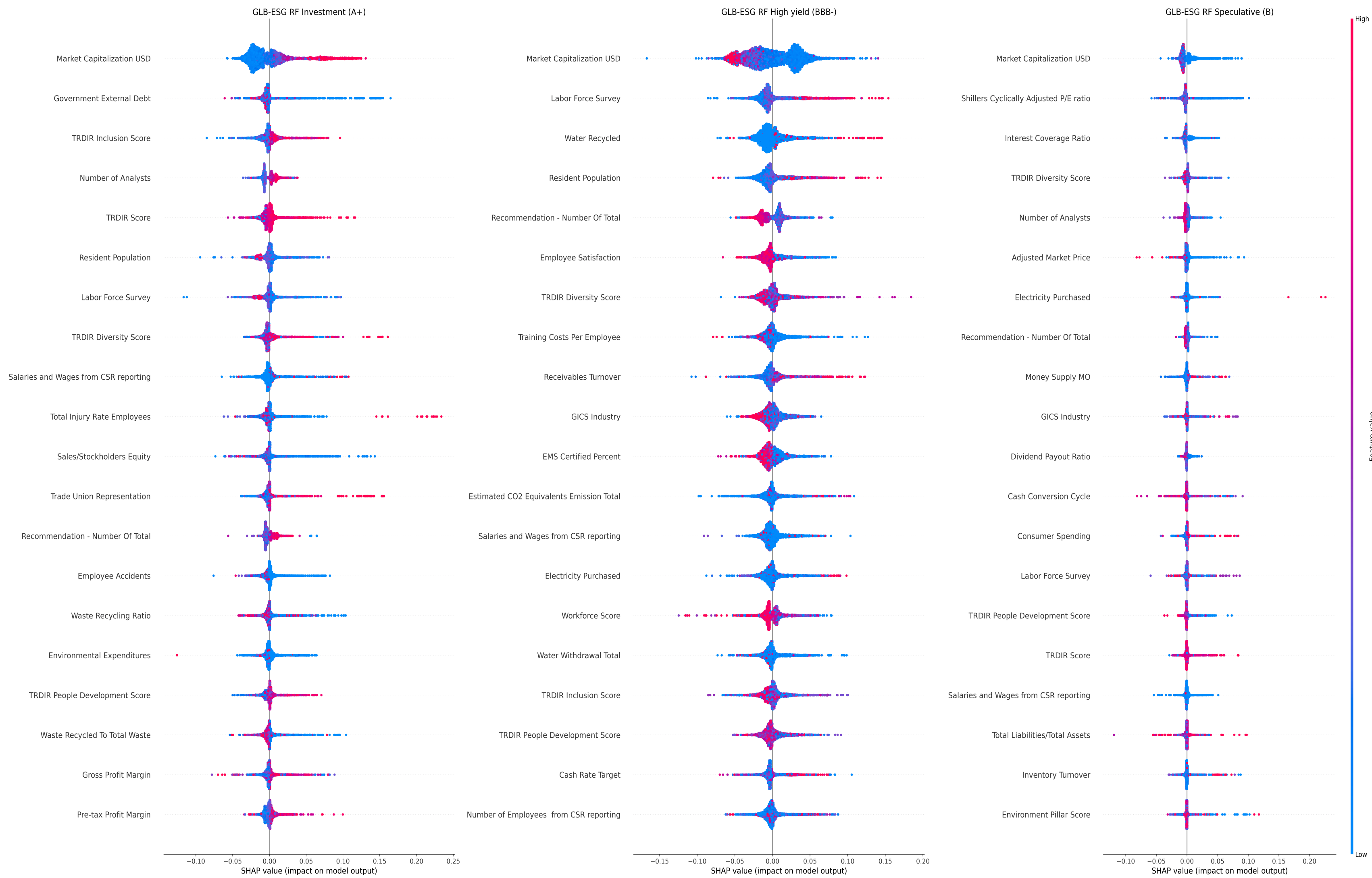

Figure 4: GLB-ESG SHAP summary plot
This figure illustrates the SHapley Additive exPlanations (SHAP) summary dot plot for global environmental, social and. governance (ESG) sample and has features ordered in descending importance on the y-axis, with local SHAP values on the x-axis. The color reveals if the feature was high (red) or low (blue) and the x-axis shows if the effect of the local value caused a higher or lower prediction. Random forests (RF) is utilized with 100 trees and depth of 14.

5 Conclusion

After a comprehensive comparison of 14 classifiers between 1982 and 2019, bagged tree-based ensemble methods, extremely randomized trees (ERT) and random forest (RF), demonstrate the highest out-of-sample (OOS) multi-class classification predictability relative to conventional, sophisticated, and boosted tree-based methods. This finding is consistent across standard and environmental, social, and governance (ESG) samples for United States (US) and global firms.

We investigate a range of feature importance methods (i.e., mean decrease impurity (MDI), mean decrease accuracy (MDA), Tree SHapley Additive exPlanations (TreeSHAP)) and their ability to recursively reduce the feature set for corporate credit rating prediction using ERT and RF. In the standard sample, recursively reducing features to 20 features, the important drivers of US (global) firm corporate credit rating predictability include market capitalization, number of 13-F institutional owners, total institutional ownership, total institutional % of shares, GICS industry, and long-term debt/total liabilities (market capitalization, GICS industry, labor force survey, resident population, government spending, and USD adjusted market price).

In the ESG sample, consistent across the three feature importance methods, and for US and global firms, Thomson Reuters Diversity and Inclusion Rating (TRDIR) metrics of TRDIR Score, TRDIR People Development Score, TRDIR Inclusion Score, and TRDIR Diversity Score are highly important ESG features for RF and ERT models. The social factors comprising the aforementioned metrics are consistent with the findings presented in Attig et al. (2013). Moreover, in the ESG sample, institutional features are still highly important for US firms, whilst macroeconomic features are informative corporate credit rating predictors for global firms. The high importance of institutional owners in the presence of ESG features for US firms resembles Tang and Zhang (2020) findings that green bond issuance increases institutional ownership.

Whilst MDI and MDA identifies features that provide greater OOS multi-class prediction, TreeSHAP provides insights into which features are driving the prediction of per class corporate credit ratings. Aggregated investment-grade and speculative-grade classes using approximated TreeSHAP values show that financial ratios tend to be more important drivers of speculative-grade corporate credit rating prediction. Alternatiely, ESG features exhibit greater aggregated importance for investment-grade classes.

ESG features, particularly TRDIR metrics, do matter for US and global firms multi-class corporate credit rating prediction. Consequently, we recommend that the relevant industry bodies and authorities to move towards the mandatory inclusion of ESG features for corporate credit rating applications. Credit rating agencies, for example, S&P, Fitch, and Moody's should require firms requesting a corporate credit rating to provide broad ESG related information to allow a thorough assessment of a company's creditworthiness.

References

Aas, K., Jullum, M., and Løland, A. (2019). Explaining individual predictions when features are dependent: More accurate approximations to shapley values. *arXiv preprint arXiv:1903.10464*.

Adams, R. B. and Ferreira, D. (2009). Women in the boardroom and their impact on governance and performance. *Journal of Financial Economics*, 94(2):291–309.

Al-Najjar, B. and Anfimiadou, A. (2012). Environmental policies and firm value. *Business Strategy and the Environment*, 21(1):49–59.

Almeida, H., Cunha, I., Ferreira, M. A., and Restrepo, F. (2017). The real effects of credit ratings: The sovereign ceiling channel. *The Journal of Finance*, 72(1):249–290.

Anderson, R. C., Mansi, S. A., and Reeb, D. M. (2004). Board characteristics, accounting report integrity, and the cost of debt. *Journal of Accounting and Economics*, 37(3):315–342.

Ashbaugh-Skaife, H., Collins, D. W., and LaFond, R. (2006). The effects of corporate governance on firms' credit ratings. *Journal of Accounting and Economics*, 42(1-2):203–243.

Attig, N., El Ghoul, S., Guedhami, O., and Suh, J. (2013). Corporate social responsibility and credit ratings. *Journal of Business Ethics*, 117(4):679–694.

Bhojraj, S. and Sengupta, P. (2003). Effect of corporate governance on bond ratings and yields: The role of institutional investors and outside directors. *The Journal of Business*, 76(3):455–475.

Blume, M. E., Lim, F., and MacKinlay, A. C. (1998). The declining credit quality of us corporate debt: Myth or reality? *The Journal of Finance*, 53(4):1389–1413.

Boot, A. W., Milbourn, T. T., and Schmeits, A. (2006). Credit ratings as coordination mechanisms. *The Review of Financial Studies*, 19(1):81–118.

Box, G. E. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2):211–243.

Bradley, M. and Chen, D. (2011). Corporate governance and the cost of debt: Evidence from director limited liability and indemnification provisions. *Journal of Corporate Finance*, 17(1):83–107.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and regression trees*. CRC press.

Brook, Y. and Rao, R. K. (1994). Shareholder wealth effects of directors' liability limitation provisions. *Journal of Financial and Quantitative Analysis*, 29(3):481–497.

Calle, M. L. and Urrea, V. (2011). Letter to the editor: stability of random forest importance measures. *Briefings in Bioinformatics*, 12(1):86–89.

Cantor, R. and Packer, F. (1994). The credit rating industry. *Quarterly Review*, 19(2):1–26.

Chapple, L. and Humphrey, J. E. (2014). Does board gender diversity have a financial impact? evidence using stock portfolio performance. *Journal of Business Ethics*, 122(4):709–723.

Cremers, K. M., Litov, L. P., and Sepe, S. M. (2017). Staggered boards and long-term firm value, revisited. *Journal of Financial Economics*, 126(2):422–444.

Danielson, M. G. and Karpoff, J. M. (1998). On the uses of corporate governance provisions. *Journal of Corporate Finance*, 4(4):347–371.

Dean, D. H. (2003). Consumer perception of corporate donations effects of company reputation for social responsibility and type of donation. *Journal of Advertising*, 32(4):91–102.

Ederington, L. H. (1985). Classification models and bond ratings. *Financial Review*, 20(4):237–262.

Edmans, A. (2011). Does the stock market fully value intangibles? employee satisfaction and equity prices. *Journal of Financial Economics*, 101(3):621–640.

Erhardt, N. L., Werbel, J. D., and Shrader, C. B. (2003). Board of director diversity and firm financial performance. *Corporate Governance: An International Review*, 11(2):102–111.

Fan, J. H. and Michalski, L. (2020). Sustainable factor investing: Where doing well meets doing good. *International Review of Economics & Finance*, 70:230–256.

Figlewski, S., Frydman, H., and Liang, W. (2012). Modeling the effect of macroeconomic factors on corporate default and credit rating transitions. *International Review of Economics & Finance*, 21(1):87–105.

Friedman, J. H. (1996). Another approach to polychotomous classification. *Technical Report, Statistics Department, Stanford University.*

Frijns, B., Dodd, O., and Cimerova, H. (2016). The impact of cultural diversity in corporate boards on firm performance. *Journal of Corporate Finance*, 41:521–541.

Gordon, L. A. and Pound, J. (1993). Information, ownership structure, and shareholder voting: Evidence from shareholder-sponsored corporate governance proposals. *The Journal of Finance*, 48(2):697–718.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273.

Guiso, L., Sapienza, P., and Zingales, L. (2015). The value of corporate culture. *Journal of Financial Economics*, 117(1):60–76.

Hachenberg, B. and Schiereck, D. (2018). Are green bonds priced differently from conventional bonds? *Journal of Asset Management*, 19(6):371–383.

Hajek, P. and Michalak, K. (2013). Feature selection in corporate credit rating prediction. *Knowledge-Based Systems*, 51:72–84.

Hall, P. (2018). On the art and science of machine learning explanations. *arXiv preprint arXiv:1810.02909.*

Hammoudeh, S., Ajmi, A. N., and Mokni, K. (2020). Relationship between green bonds and financial and environmental variables: A novel time-varying causality. *Energy Economics*, 92:1–7.

Hastie, T. and Tibshirani, R. (1998). Classification by pairwise coupling. In *Advances in Neural Information Processing Systems*, pages 507–513.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction.* Springer Science & Business Media.

Hossain, M., Atif, M., Ahmed, A., and Mia, L. (2020). Do lgbt workplace diversity policies create value for firms? *Journal of Business Ethics*, 167:775–791.

Huang, M., Li, P., Meschke, F., and Guthrie, J. P. (2015). Family firms, employee satisfaction, and corporate performance. *Journal of Corporate Finance*, 34:108–127.

Huselid, M. A. (1995). The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of Management Journal*, 38(3):635–672.

Isidro, H. and Sobral, M. (2015). The effects of women on corporate boards on firm value, financial performance, and ethical and social compliance. *Journal of Business Ethics*, 132(1):1–19.

Jabeur, S. B., Sadaaoui, A., Sghaier, A., and Aloui, R. (2020). Machine learning models and cost-sensitive decision trees for bond rating prediction. *Journal of the Operational Research Society*, 71(8):1161–1179.

Janitza, S., Tutz, G., and Boulesteix, A.-L. (2016). Random forest for ordinal responses: prediction and variable selection. *Computational Statistics & Data Analysis*, 96:57–73.

Janney, J. J., Dess, G., and Forlani, V. (2009). Glass houses? market reactions to firms joining the un global compact. *Journal of Business Ethics*, 90(3):407–423.

Jiraporn, P. and Liu, Y. (2008). Capital structure, staggered boards, and firm value. *Financial Analysts Journal*, 64(1):49–60.

Jones, S., Johnstone, D., and Wilson, R. (2015). An empirical evaluation of the performance of binary classifiers in the prediction of credit ratings changes. *Journal of Banking & Finance*, 56:72–85.

Kamstra, M., Kennedy, P., and Suan, T. (2001). Combining bond rating forecasts using logit. *Financial Review*, 36(2):75–96.

Kiesel, F. and Lücke, F. (2019). Esg in credit ratings and the impact on financial markets. *Financial Markets, Institutions & Instruments*, 28(3):263–290.

Klassen, R. D. and McLaughlin, C. P. (1996). The impact of environmental management on firm performance. *Management Science*, 42(8):1199–1214.

Koopman, S. J., Lucas, A., and Schwaab, B. (2011). Modeling frailty-correlated defaults using many macroeconomic covariates. *Journal of Econometrics*, 162(2):312–325.

Kressel, U. (1998). Pairwise classification and support vector machines. *Advances in Kernel Methods: Support Vector Learning*, pages 255–268.

Lee, K.-H. and Min, B. (2015). Green R&D for eco-innovation and its impact on carbon emissions and firm performance. *Journal of Cleaner Production*, 108:534–542.

Li, X., Wang, Y., Basu, S., Kumbier, K., and Yu, B. (2019). A debiased mdi feature importance measure for random forests. *arXiv preprint arXiv:1906.10845*.

Lindsay, S., Cagliostro, E., Albarico, M., Mortaji, N., and Karon, L. (2018). A systematic review of the benefits of hiring people with disabilities. *Journal of Occupational Rehabilitation*, 28(4):634–655.

Liu, Y., Wei, Z., and Xie, F. (2014). Do women directors improve firm performance in china? *Journal of Corporate Finance*, 28:169–184.

Lundberg, S. M., Erion, G., and Lee, S.-I. (2018). Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888*.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874*.

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

López de Prado, M. (2020). *Machine Learning for Asset Managers*. Cambridge University Press.

Miroshnychenko, I., Barontini, R., and Testa, F. (2017). Green practices and financial performance: A global outlook. *Journal of Cleaner Production*, 147:340–351.

Nations, U. (2020). World economic situation and prospects.

Nembrini, S., König, I. R., and Wright, M. N. (2018). The revival of the gini importance? *Bioinformatics*, 34(21):3711–3718.

Ngo, H.-Y., Turban, D., Lau, C.-M., and Lui, S.-Y. (1998). Human resource practices and firm performance of multinational corporations: Influences of country origin. *International Journal of Human Resource Management*, 9(4):632–652.

Nicodemus, K. K. (2011). Letter to the editor: On the stability and ranking of predictors from random forest variable importance measures. *Briefings in Bioinformatics*, 12(4):369–373.

Nicodemus, K. K., Malley, J. D., Strobl, C., and Ziegler, A. (2010). The behaviour of random forest permutation-based variable importance measures under predictor correlation. *BMC Bioinformatics*, 11(1):1–13.

Nishitani, K. (2011). An empirical analysis of the effects on firms' economic performance of implementing environmental management systems. *Environmental and Resource Economics*, 48(4):569–586.

Ozturk, H., Namli, E., and Erdal, H. I. (2016). Modelling sovereign credit ratings: The accuracy of models in a heterogeneous sample. *Economic Modelling*, 54:469–478.

Pedersen, L. H., Fitzgibbons, S., and Pomorski, L. (2020). Responsible investing: The esg-efficient frontier. *Journal of Financial Economics*.

Pinches, G. E. and Singleton, J. C. (1978). The adjustment of stock prices to bond rating changes. *The Journal of Finance*, 33(1):29–44.

Plewa, C., Conduit, J., Quester, P. G., and Johnson, C. (2015). The impact of corporate volunteering on csr image: A consumer perspective. *Journal of Business Ethics*, 127(3):643–659.

Pogue, T. F. and Soldofsky, R. M. (1969). What's in a bond rating. *Journal of Financial and Quantitative Analysis*, 4(2):201–228.

Poon, W. P. and Shen, J. (2020). The roles of rating outlooks: the predictor of creditworthiness and the monitor of recovery efforts. *Review of Quantitative Finance and Accounting*, 55:1–29.

Porter, M. E. and Kramer, M. R. (2006). The link between competitive advantage and corporate social responsibility. *Harvard Business Review*, 84(12):78–92.

Reguera-Alvarado, N., de Fuentes, P., and Laffarga, J. (2017). Does board gender diversity influence financial performance? evidence from spain. *Journal of Business Ethics*, 141(2):337–350.

Rodríguez-Pérez, R. and Bajorath, J. (2020). Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. *Journal of Computer-Aided Molecular Design*, 34:1013–1026.

Schein, V. E., Maurer, E. H., and Novak, J. F. (1977). Impact of flexible working hours on productivity. *Journal of Applied Psychology*, 62(4):463–465.

Schölkopf, B., Smola, A. J., and Bach, F. (2002). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.

Shapley, L. S. (1953). A value for n-person games. *Contributions to the Theory of Games*, 2(28):307–317.

Strobl, C., Boulesteix, A.-L., Zeileis, A., and Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8(1):25.

Tang, D. Y. and Zhang, Y. (2020). Do shareholders benefit from green bonds? *Journal of Corporate Finance*, 61:1–18.

Yeo, I.-K. and Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, 87(4):954–959.

Zhang, C. and Ma, Y. (2012). *Ensemble machine learning: methods and applications*. Springer.

Ziegler, A. and König, I. R. (2014). Mining data with random forests: current options for real-world applications. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 4(1):55–63.

Štrumbelj, E. and Kononenko, I. (2010). An efficient explanation of individual classifications using game theory. *The Journal of Machine Learning Research*, 11:1–18.

Štrumbelj, E. and Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems*, 41(3):647–665.

Appendices

A Classification metrics

Accuracy is the proportion of true predictions among the total number of predictions. It is a valid choice of evaluation when the dataset is well balanced and not skewed or imbalanced. This metric is suited for both binary and multi-class classification problems. When using multi-class classification (predicting multiple bond rating classes) weighted accuracy is computed, where weights are proportional to the number of times the class appears in the sample. Accuracy is defined as:

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + FN + TN)}, \tag{6}$$

where TP is the number of positive class correctly predicted, TN is the number of negative class correctly predicted, FP is the number of incorrectly predicted positive class, whilst FN is the incorrectly predicted negative class.

Precision is what proportion of predicted positives is truly positive. Precision is defined as follows:

$$\text{Precision} = \frac{(TP)}{(TP + FP)}, \tag{7}$$

Recall is the proportion of actual positives correctly classified and is defined as:

$$\text{Recall} = \frac{(TP)}{(TP + FN)}, \tag{8}$$

F1 score is the harmonic mean of precision and recall and is bounded between 0 and 1, where 1 represents perfect precision and recall. Similar to accuracy, precision and recall, for multi-class classification, F1 score is the weighted harmonic mean of precision and recall. F1 score is defined as:

$$\text{F1 score} = \frac{(Precision \cdot Recall)}{(Precision + Recall)}, \tag{9}$$

A receiver operating characteristic curve (ROC) curve is a plot that displays the diagnostic ability of a binary classifier as its discrimination threshold is varied. The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. TPR is the proportion of true predictions (recall) and FPR is false predictions, defined as:

$$\text{FPR} = \frac{(FP)}{(FP + TN)}, \tag{10}$$

AUC ROC indicates how well the probabilities from the positive classes are separated from the negative classes. The AUC provides an aggregate measure of performance across all possible classification thresholds. AUC is scale-invariant and measures how well predictions are ranked, rather than their absolute values. Furthermore, it is threshold invariant, similar to log loss.

B Two-class prediction

Table B.1: Two-class prediction accuracy
This table reports two-class predictive performance of the 14 classifiers across United States (US) and global samples with no dimensionality reduction to the features. Classifiers include: extremely randomized trees (ERT), random forests (RF), XGBoost (XGB), histogram gradient boost (HGB), AdaBoost (AB), decision tree (DT), multi-layer perceptron (MLP), support vector machine (SVM), k-nearest neighbours (KNN), logistic regression (LR), quadratic discriminant analysis (QDA), linear discriminant analysis (LDA), Bernoulli Naive Bayes (BNB), and Gaussian Naive Bayes (GNB). Cross-validated (CV) area under the curve (AUC) and CV Accuracy is calculated through 5-fold cross-validation on the training sample, whilst Test AUC, Test Accuracy and is performance on the test sample. Rank is sorted on test accuracy performance. Panel A reports the performance for US and global firms for classifying investment-grade or speculative-grade classes in the standard samples, whilst Panel B presents the performance for environmental, social, and governance (ESG) samples. Yeo and Johnson (2000) transformations are applied to the features in the results.

Model	US CV		US Test			Global CV		Global Test		
	AUC	Accuracy	AUC	Accuracy	Rank	AUC	Accuracy	AUC	Accuracy	Rank
Panel A: Standard sample										
ERT	99.9	98.4	99.9	98.8	1.0	99.9	98.6	99.9	98.8	1.0
RF	99.8	97.7	99.8	98.1	3.0	99.7	97.4	99.7	98.0	3.0
XGB	99.9	98.8	99.8	98.2	2.0	99.8	98.4	99.8	98.4	2.0
AB	96.3	89.5	96.1	89.0	8.0	94.3	87.3	92.7	85.2	10.0
MLP	99.3	96.1	99.5	97.0	4.0	98.9	95.4	99.1	96.2	5.0
HGB	99.6	96.8	99.4	95.5	5.0	99.7	97.5	99.5	97.2	4.0
SVM	98.1	93.4	98.2	93.8	6.0	98.4	94.3	98.5	95.0	6.0
DT	93.7	93.7	93.8	93.8	7.0	92.4	93.0	88.8	89.0	8.0
KNN	94.2	87.1	94.6	87.9	9.0	97.7	93.2	97.8	94.1	7.0
LR	95.2	87.7	95.3	87.8	10.0	93.0	86.1	92.5	85.7	9.0
LDA	94.8	87.2	95.0	87.2	11.0	93.0	85.8	92.3	85.0	11.0
QDA	88.1	75.3	88.4	75.7	14.0	87.2	80.4	86.8	79.8	12.0
GNB	87.1	75.7	87.3	76.0	13.0	83.5	77.2	83.3	77.2	13.0
BNB	86.4	77.6	86.7	77.7	12.0	83.9	76.9	83.2	76.1	14.0
Panel A: ESG sample										
ERT	99.9	99.4	100.0	99.5	1.0	99.9	99.2	100.0	99.6	1.0
RF	99.9	99.2	100.0	99.3	3.0	99.9	99.1	100.0	99.3	2.0
XGB	100.0	99.4	100.0	99.5	2.0	99.9	99.2	100.0	99.3	3.0
AB	95.6	88.5	95.5	88.5	11.0	95.4	89.9	95.1	89.0	11.0
MLP	99.8	98.8	99.9	98.9	4.0	99.8	98.7	99.9	99.0	4.0
HGB	99.9	99.0	99.9	98.8	6.0	99.9	99.1	99.9	99.0	5.0
SVM	99.8	98.6	99.9	98.8	5.0	99.9	98.8	99.9	99.0	5.0
DT	95.4	95.8	93.8	94.4	8.0	94.8	96.2	93.8	95.3	8.0
KNN	99.7	98.1	99.9	98.6	7.0	99.5	98.5	99.8	98.6	7.0
LR	97.5	92.3	97.5	92.2	9.0	98.0	94.2	98.2	94.5	9.0
LDA	96.8	91.2	96.6	90.8	10.0	96.9	92.9	97.2	93.1	10.0
QDA	98.3	78.8	50.0	62.8	13.0	98.8	97.5	50.0	76.7	12.0
GNB	82.5	59.8	50.0	62.8	13.0	78.2	56.8	50.0	76.7	12.0
BNB	82.2	72.1	82.3	72.2	12.0	77.2	70.8	77.1	70.4	14.0

C Feature importance

C.1 Mean decrease impurity (MDI)

The importance of a feature is computed as the weighted information gain ($\Delta g[t, f]$) across all nodes where that feature was selected. The information gain that occurs at each split is simply the reduction in impurity. Tree-based estimators can be used to compute impurity-based feature importance. Introduced by Breiman (2001), MDI is bounded between 0 and 1, and all combined importance's sum to 1. A subset is purest when it contains only one class (e.g. only AAA), whilst it is most impure when the classes follow a uniform distribution (e.g. AAA, AA+, AA etc.). For ensembles of trees, for example, RF, the mean and variance of MDI values can also be computed for each feature across all trees. The mean and variance estimates can be utilized in testing the significance of a feature against a user-defined null hypothesis. Notably, MDI is computed in-sample which results in computationally efficiency, however, substitution effects arise when two features share predictive information. This drawback halves the importance for two identical features as they are chosen at random with equal probability.

To define MDI, we are given an observation matrix of size N and features F, $\{X_f\}_{f=1,\ldots,F}$ one label per observation. As a tree-based classification algorithm splits at each node t (internal node), for an associated feature X_f, if labels in node t are below a threshold , X_f is placed in the left child node, whilst the remaining are placed in the right child node. As each tree T is grown using a two-step recursive procedure for each node t, first, a subset $M \subset [F]$ of features is chosen uniformly at random, then, the optimal split is determined by maximizing:

$$\Delta g[t, f] = impurity[t] - \frac{N^{(left)}}{N_t} impurity[t^{(left)}] - \frac{N^{(right)}}{N_t} impurity[t^{(right)}], \tag{11}$$

where $impurity[t]$ the before split impurity of labels at node t, $impurity[t^{(left)}]$ is the impurity of labels in the left sample, and $impurity[t^{(right)}]$ is the impurity of labels in the right sample. At each node t, the classifier evaluates $\Delta g[t, f]$ for various features in $\{X_f\}_{f=1,\ldots,F}$, determines the optimal threshold that maximizes $\Delta g[t, f]$ for each of them, and selects the feature f with the highest $\Delta g[t, f]$ (Li et al., 2019; López de Prado, 2020). The algorithm continues splitting the subset M until either a maximum acceptable limit is achieved, or no additional information gain can be displayed. Two common information gain criteria g, that are utilized to measure the impurity of a node are gini index and entropy. Gini index is defined as:

$$\text{Gini index} = 1 - \sum_{i=1}^{c} p_i^2, \tag{12}$$

where p_i is the proportion of the subset that belongs to class c for a particular node t. Whilst entropy is defined as:

$$\text{Entropy} = -\sum_{i=1}^{c} p_i log_2(p_i), \tag{13}$$

where p_i is the proportion of the subset that belongs to class c for a particular node t.

C.2 Mean decrease accuracy (MDA)

As the gini index is the most common impurity method utilized in classification trees, a known biased within the literature is that features with many split points (e.g. categorical or continuous variables) or high category frequencies are favored (Breiman et al., 1984; Nicodemus, 2011; Strobl et al., 2007). MDA does not suffer from these biases and is generally preferred within statistics literature (Nicodemus et al., 2010; Ziegler and König, 2014). However, for high dimensional data, MDA is computationally intensive (Calle and Urrea, 2011; Nembrini et al., 2018). MDA is calculated by randomly permuting the features and computing the increase in OOS estimate of the accuracy loss (cross-validated). The tree-wise OOS estimate of the prediction error is computed with this permuted data. The difference between this estimate and the OOS error without permutation, averaged over all trees, is the MDA value of feature f (Zhang and Ma, 2012).

$$\text{OOSerror} = \frac{1}{n} \sum (y_r - y_p)^2, \tag{14}$$

where n represents the number of all samples, y_r is the true value of sample i, and y_p is the predictive value of the ith tree.

$$\text{y} = \frac{1}{c} \sum_{i=1}^{c} y_i, \tag{15}$$

In this equation, $c = n_{tree}$ represents the number of trees in for example, RF, and y_i is the predictive value of the ith tree.

$$MDA_i(f) = OOSerror_i^i - OOSerror_i, \tag{16}$$

where $OOSerror_i$ is the error when calculating the importance value of feature f based on the ith tree before permuting, whilst $OOSerror_i^i$ is calculated after the values of feature f in the OOS dataset are randomly rearranged (permuted) and the remaining features are unchanged. This process is repeated for each tree. The final MDA value of feature f can be obtained by averaging the MDA values of each tree:

$$MDA(f) = \frac{1}{c}\sum_{i=1}^{c} MDA_i(f), \tag{17}$$

where $c = n_{tree}$ represents the tree number. If a feature is important, then its values of different subsets will be dissimilar. After feature f values are randomly rearranged on the OOS data, the discrimination of different subsets will be reduced. Whilst MDA values are not bounded, a feature is identified as important if it has a larger positive effect on the prediction performance (Breiman, 2001; Janitza et al., 2016).

Table C.1: US and global important features

This table reports a 10 feature model across extremely randomized trees (ERT) and random forests (RF) for investment- and speculative-grade classification. Utilizing mean decrease accuracy (MDA) recursively reduced features to 10, important features for the United States (US) standard and environmental, social, and governance (ESG) samples are reported in Panel A, whilst the global standard and ESG samples are exhibited in Panel B. Test sample is used for reducing in features. Test prediction accuracy for ERT (RF) in the US standard is 98.5% (95.8%) and global standard is 98.3% (97.4%). Moreover, test prediction accuracy for ERT (RF) in the US ESG is 98.8% (97.6%) and global ESG is 98.9% (98.7%).

ERT Standard Features	ERT ESG Features	RF Standard Features	RF ESG Features
Panel A: US sample			
Number of 13-F Institutional Owners	Number of 13-F Institutional Owners	Dividend Yield	Dividend Yield
Dividend Yield	Elimination of Cumulative Voting Rights	Market Capitalization	Market Capitalization
Market Capitalization	Dividend Yield	Number of 13-F Institutional Owners	Number of 13-F Institutional Owners
Long-term Debt/Total Liabilities	Money Supply M1	Forward P/E to 1-year Growth	Interest Coverage Ratio
Interest/Average Total Debt	TRDIR Inclusion Score	Total Institutional Ownership	Dividend Payout Ratio
GICS Industry	Resource Use Score	Retail Sales	Common Equity/Invested Capital
GICS Group	Policy Water Efficiency	Gross Domestic Product	Total Debt/Total Assets
GICS Sector	TRDIR Score	Interest/Average Total Debt	Total Donations To Revenues
Number of Analysts	GICS Industry	Long-term Debt/Total Liabilities	Total Institutional Ownership
Retail Sales	Market Capitalization	Money Supply M2	GICS Industry
Panel B: Global sample			
Market Capitalization	Market Capitalization	Market Capitalization	Market Capitalization
Labor Force Survey	Targets Emissions	Resident Population	Resident Population
Developed Nation	Stakeholder Engagement	Labor Force Survey	Labor Force Survey
Number of Analysts	Climate Change Commercial Risks Opportunities	Recommendation - Number Of Total	CO2 Equivalent Emissions Total
Recommendation - Number Of Total	CSR Sustainability External Audit	Adjusted Market Price	Shillers Cyclically Adjusted P/E
GICS Industry	Shareholders Vote on Executive Pay	Number of Analysts	Dividend Payout Ratio
GICS Group	Labor Force Survey	Interest/Average Total Debt	Water Withdrawal Total
GICS Sector	Compensation Committee Independence	Interest Coverage Ratio	CSR Strategy Score
Resident Population	Government Spending	GICS Industry	TRDIR People Development Score
Adjusted Market Price	TRDIR Diversity Score	Dividend Payout Ratio	Salaries and Wages from CSR reporting

D Traditional features

D.1 Financial features

Table D.1: Financial features
This table reports the financial features constructed utilizing quarterly financial data retrieved from Center for Research in Security Prices (CRSP) and Compustat through subscription to Wharton Research Data Services (WRDS). SAS code published on WRDS' website for constructing financial ratios for United States (US) firms was replicated and applied to global listed firms in this study. 71 financial ratios are utilized for US firms, whilst 57 ratios are used for global firms.

Feature	Feature full name	Description	Ratio type	Source	US	Global
capital_ratioq	Capitalization Ratio	Total Long-term Debt as a fraction of the sum of Total Long-term Debt, Common/Ordinary Equity and Preferred Stock	Capitalization	CRSP/Compustat	Yes	Yes
equity_invcapq	Common Equity/Invested Capital	Common Equity as a fraction of Invested Capital	Capitalization	CRSP/Compustat	Yes	No
debt_invcapq	Long-term Debt/Invested Capital	Long-term Debt as a fraction of Invested Capital	Capitalization	CRSP/Compustat	Yes	No
totdebt_invcapq	Total Debt/Invested Capital	Total Debt (Long-term and Current) as a fraction of Invested Capital	Capitalization	CRSP/Compustat	Yes	No
at_turnq	Asset Turnover	Sales as a fraction of the average Total Assets based on the most recent two periods	Efficiency	CRSP/Compustat	Yes	Yes
inv_turnq	Inventory Turnover	COGS as a fraction of the average Inventories based on the most recent two periods	Efficiency	CRSP/Compustat	Yes	Yes
pay_turnq	Payables Turnover	COGS and change in Inventories as a fraction of the average of Accounts Payable	Efficiency	CRSP/Compustat	Yes	Yes
rect_turnq	Receivables Turnover	Sales as a fraction of the average of Accounts Receivables	Efficiency	CRSP/Compustat	Yes	Yes
sale_equityq	Sales/Stockholders Equity	Sales per dollar of total Stockholders' Equity	Efficiency	CRSP/Compustat	Yes	Yes
sale_invcapq	Sales/Invested Capital	Sales per dollar of Invested Capital	Efficiency	CRSP/Compustat	Yes	No
sale_nwcq	Sales/Working Capital	Sales per dollar of Working Capital, defined as difference between Current Assets and Current Liabilities'	Efficiency	CRSP/Compustat	Yes	Yes
invt_actq	Inventory/Current Assets	Inventories as a fraction of Current Assets	Financial Soundness	CRSP/Compustat	Yes	Yes
rect_actq	Receivables/Current Assets	Accounts Receivables as a fraction of Current Assets	Financial Soundness	CRSP/Compustat	Yes	Yes
fcf_ocfq	Free Cash Flow/Operating Cash Flow	Free Cash Flow as a fraction of Operating Cash Flow, where Free Cash Flow is defined as the difference between Operating Cash Flow and Capital Expenditures	Financial Soundness	CRSP/Compustat	Yes	Yes
ocf_lctq	Operating Cash Flow/Current Liabilities	Operating Cash Flow as a fraction of Current Liabilities	Financial Soundness	CRSP/Compustat	Yes	Yes
cash_debtq	Cash Flow/Total Debt	Operating Cash Flow as a fraction of Total Debt	Financial Soundness	CRSP/Compustat	Yes	Yes
cash_ltq	Cash Balance/Total Liabilities	Cash Balance as a fraction of Total Liabilities	Financial Soundness	CRSP/Compustat	Yes	Yes
cfmq	Cash Flow Margin	Income before Extraordinary Items and Depreciation as a fraction of Sales	Financial Soundness	CRSP/Compustat	Yes	Yes
short_debtq	Short-Term Debt/Total Debt	Short-term Debt as a fraction of Total Debt	Financial Soundness	CRSP/Compustat	Yes	Yes
profit_lctq	Profit Before Depreciation/Current Liabilities	Operating Income before DA as a fraction of Current Liabilities	Financial Soundness	CRSP/Compustat	Yes	Yes
curr_debtq	Current Liabilities/Total Liabilities	Current Liabilities as a fraction of Total Liabilities	Financial Soundness	CRSP/Compustat	Yes	Yes
debt_ebitdaq	Total Debt/EBITDA	Gross Debt as a fraction of EBITDA	Financial Soundness	CRSP/Compustat	Yes	Yes
dltt_beq	Long-term Debt/Book Equity	Long-term Debt to Book Equity	Financial Soundness	CRSP/Compustat	Yes	Yes
int_debtq	Interest/Average Long-term Debt	Interest as a fraction of average Long-term debt based on most recent two periods	Financial Soundness	CRSP/Compustat	Yes	Yes
int_totdebtq	Interest/Average Total Debt	Interest as a fraction of average Total Debt based on most recent two periods	Financial Soundness	CRSP/Compustat	Yes	Yes
lt_debtq	Long-term Debt/Total Liabilities	Long-term Debt as a fraction of Total Liabilities	Financial Soundness	CRSP/Compustat	Yes	Yes
lt_ppentq	Total Liabilities/Total Tangible Assets	Total Liabilities to Total Tangible Assets	Financial Soundness	CRSP/Compustat	Yes	Yes
cash_ratioq	Cash Conversion Cycle	Inventories per daily COGS plus Account Receivables per daily Sales minus Account Payables per daily COGS	Liquidity	CRSP/Compustat	Yes	Yes
curr_ratioq	Current Ratio	Current Assets as a fraction of Current Liabilities	Liquidity	CRSP/Compustat	Yes	Yes
quick_ratioq	Quick Ratio (Acid Test)	Current Assets net of Inventories as a fraction of Current Liabilities	Liquidity	CRSP/Compustat	Yes	Yes
accrualq	Accruals/Average Assets	Accruals as a fraction of average Total Assets based on most recent two periods	Liquidity	CRSP/Compustat	Yes	Yes
rd_saleq	Research and Development/Sales	RD expenses as a fraction of Sales	Liquidity	CRSP/Compustat	Yes	No
efftaxq	Effective Tax Rate	Income Tax as a fraction of Pretax Income	Profitability	CRSP/Compustat	Yes	Yes
GProfq	Gross Profit/Total Assets	Gross Profitability as a fraction of Total Assets	Profitability	CRSP/Compustat	Yes	Yes

aftret_eqq	After-tax Return on Average Common Equity	Net Income as a fraction of average of Common Equity based on most recent two periods	Profitability	CRSP/Compustat	Yes	Yes
aftret_equityq	After-tax Return on Total Stockholders' Equity	Net Income as a fraction of average of Total Shareholders' Equity based on most recent two periods	Profitability	CRSP/Compustat	Yes	Yes
aftret_invcapxq	After-tax Return on Invested Capital	Net Income plus Interest Expenses as a fraction of Invested Capital	Profitability	CRSP/Compustat	Yes	No
gpmq	Gross Profit Margin	Gross Profit as a fraction of Sales	Profitability	CRSP/Compustat	Yes	Yes
npm	Net Profit Margin	Net Income as a fraction of Sales	Profitability	CRSP/Compustat	Yes	Yes
opmaqdq	Operating Profit Margin After Depreciation	Operating Income After Depreciation as a fraction of Sales	Profitability	CRSP/Compustat	Yes	Yes
opmbdq	Operating Profit Margin Before Depreciation	Operating Income Before Depreciation as a fraction of Sales	Profitability	CRSP/Compustat	Yes	Yes
pretret_earnatq	Pre-tax Return on Total Earning Assets	Operating Income After Depreciation as a fraction of average Total Earnings Assets (TEA) based on most recent two periods, where TEA is defined as the sum of Property Plant and Equipment and Current Assets	Profitability	CRSP/Compustat	Yes	Yes
pretret_noq	Pre-tax return on Net Operating Assets	Operating Income After Depreciation as a fraction of average Net Operating Assets (NOA) based on most recent two periods, where NOA is defined as the sum of Property Plant and Equipment and Current Assets minus Current Liabilities	Profitability	CRSP/Compustat	Yes	Yes
ptpmq	Pre-tax Profit Margin	Pretax Income as a fraction of Sales	Profitability	CRSP/Compustat	Yes	Yes
roaq	Return on Assets	Operating Income Before Depreciation as a fraction of average Total Assets	Profitability	CRSP/Compustat	Yes	Yes
roceq	Return on Capital Employed	Earnings Before Interest and Taxes as a fraction of average Capital Employed based on most recent two periods, where Capital Employed is the sum of Debt in Long-term and Current Liabilities and Common/Ordinary Equity	Profitability	CRSP/Compustat	Yes	Yes
roeq	Return on Equity	Net Income as a fraction of average Book Equity based on most recent two periods, where Book Equity is defined as the sum of Total Parent Stockholders' Equity and Deferred Taxes and Investment Tax Credit	Profitability	CRSP/Compustat	Yes	Yes
de_ratioq	Total Debt/Equity	Total Liabilities to Shareholders' Equity (common and preferred)	Solvency	CRSP/Compustat	Yes	Yes
debt_atq	Total Debt/Total Assets	Total Debt as a fraction of Total Assets	Solvency	CRSP/Compustat	Yes	Yes
lt_atq	Total Liabilities/Total Assets	Total Liabilities as a fraction of Total Assets	Solvency	CRSP/Compustat	Yes	Yes
debt_capitalq	Total Debt/Capital	Total Debt as a fraction of Total Capital, where Total Debt is defined as the sum of Accounts Payable and Total Debt in Current and Long-term Liabilities, and Total Capital is defined as the sum of Total Debt and Total Equity (common and preferred)	Solvency	CRSP/Compustat	Yes	Yes
intcovq	After-tax Interest Coverage	Multiple of After-tax Income to Interest and Related Expenses	Solvency	CRSP/Compustat	Yes	Yes
intcov_ratioq	Interest Coverage Ratio	Multiple of Earnings Before Interest and Taxes to Interest and Related Expenses	Solvency	CRSP/Compustat	Yes	Yes
dprq	Dividend Payout Ratio	Dividends as a fraction of Income Before Extra. Items	Valuation	CRSP/Compustat	Yes	Yes
PEG_1yrforwardq	Forward P/E to 1-year Growth ratio	Price-to-Earnings, excl. Extraordinary Items (diluted) to 1-Year EPS Growth rate	Valuation	CRSP/Compustat	Yes	No
PEG_ltgforwardq	Forward P/E to Long-term Growth ratio	Price-to-Earnings, excl. Extraordinary Items (diluted) to Long-term EPS Growth rate	Valuation	CRSP/Compustat	Yes	No
PEG_trailingq	Trailing P/E to Growth ratio	Price-to-Earnings, excl. Extraordinary Items (diluted) to 3-Year past EPS Growth	Valuation	CRSP/Compustat	Yes	No
bmq	Book/Market	Book Value of Equity as a fraction of Market Value of Equity	Valuation	CRSP/Compustat	Yes	Yes
capeiq	Shillers Cyclically Adjusted P/E Ratio	Multiple of Market Value of Equity to 5-year moving average of Net Income	Valuation	CRSP/Compustat	Yes	Yes
div yieldq	Dividend Yield	Indicated Dividend Rate as a fraction of Price	Valuation	CRSP/Compustat	Yes	No
evmq	Enterprise Value Multiple	Multiple of Enterprise Value to EBITDA	Valuation	CRSP/Compustat	Yes	Yes
pcfq	Price/Cash flow	Multiple of Market Value of Equity to Net Cash Flow from Operating Activities	Valuation	CRSP/Compustat	Yes	Yes
pe_exiq	P/E (Diluted, Excl. EI)	Price-to-Earnings, excl. Extraordinary Items (diluted)	Valuation	CRSP/Compustat	Yes	No
pe_incq	P/E (Diluted, Incl. EI)	Price-to-Earnings, incl. Extraordinary Items (diluted)	Valuation	CRSP/Compustat	Yes	No
pe_op_basicq	Price/Operating Earnings (Basic, Excl. EI)	Price to Operating EPS, excl. Extraordinary Items (Basic)	Valuation	CRSP/Compustat	Yes	No
pe_op_dilq	Price/Operating Earnings (Diluted, Excl. EI)	Price to Operating EPS, excl. Extraordinary Items (Diluted)	Valuation	CRSP/Compustat	Yes	No
psq	Price/Sales	Multiple of Market Value of Equity to Sales	Valuation	CRSP/Compustat	Yes	Yes
ptbq	Price/Book	Multiple of Market Value of Equity to Book Value of Equity	Valuation	CRSP/Compustat	Yes	Yes
mktcap	Market capitalization	CRSP (Compustat) Unadjusted Price x Publicly Held Shares	Size feature	CRSP/Compustat	Yes	Yes
price_adj	Market price	CRSP (Compustat) Adjusted Price	Price feature	CRSP/Compustat	Yes	Yes

D.2 Additional traditional features

Table D.2: Additional traditional features
This table reports the remaining traditional features utilized for United States (US) and global firms. Panel A reports institutional, Panel B analyst, Panel C macroeconomic, Panel D industry, Panel E sovereign credit rating, Panel F recession indicator, and Panel G country and financial quarter. TR denotes Thomson Reuters.

Feature	Feature full name	Description	Source	US	Global
Panel A: Institutional					
Top5InstOwn	Largest 5 institutional ownership size	Largest 5 institutional ownership size	TR Institutional (13f) Holdings	Yes	No
Top10InstOwn	Largest 10 institutional ownership size	Largest 10 institutional ownership size	TR Institutional (13f) Holdings	Yes	No
NumInstBlockOwners	Number of 5% institutional block ownerships	Number of 5% institutional block ownerships	TR Institutional (13f) Holdings	Yes	No
InstBlockOwn	Total ownership by institutional blockholders	Total ownership by institutional blockholders	TR Institutional (13f) Holdings	Yes	No
NumInstOwners	Number of 13-F institutional owners	Number of 13-F institutional owners	TR Institutional (13f) Holdings	Yes	No
MaxInstOwn	Largest institutional ownership size	Largest institutional ownership size	TR Institutional (13f) Holdings	Yes	No
InstOwn	Total institutional ownership	Total institutional ownership	TR Institutional (13f) Holdings	Yes	No
InstOwn_HHI	Ownership concentration-Herfindahl-Hirschman index	Ownership concentration-Herfindahl-Hirschman index	TR Institutional (13f) Holdings	Yes	No
InstOwn_Perc	Total institutional ownership, % of shares outstanding	Total institutional ownership, % of shares outstanding	TR Institutional (13f) Holdings	Yes	No
Panel B: Analyst					
Rec_Mean	Recommendation - Mean (1-5)	Recommendation numeric mean based on the standard scale of strong buy (1), buy (2), hold (3), sell (4) and strong sell (5).	Refinitiv Eikon	Yes	Yes
Rec_Total	Recommendation - Number Of Total	The total number of contributor recommendations provided	Refinitiv Eikon	Yes	Yes
Rec_Median	Recommendation - Median (1-5)	Recommendation numeric median based on the standard scale of strong buy (1), buy (2), hold (3), sell (4) and strong sell (5).	Refinitiv Eikon	Yes	Yes
Rec_Low	Recommendation - Low (1-5)	The lowest broker numeric recommendation of all recommendations included in the summary calculation.	Refinitiv Eikon	Yes	Yes
Rec_High	Recommendation - High (1-5)	The highest broker numeric recommendation of all recommendations included in the summary calculation.	Refinitiv Eikon	Yes	Yes
Rec_SBuy	Recommendation - Number Of Strong Buy	Number of strong buys. The publisher strongly rates this instrument as one that should be bought	Refinitiv Eikon	Yes	Yes
Rec_Buy	Recommendation - Number Of Buy	Number of buys. The publisher rates this instrument as one that should be bought	Refinitiv Eikon	Yes	Yes
Rec_Hold	Recommendation - Number Of Hold	Number of holds. The publisher strongly rates this instrument as one that should be held if already owned	Refinitiv Eikon	Yes	Yes
Rec_Sell	Recommendation - Number Of Sell	Number of sells. The publisher rates this instrument as one that should be sold	Refinitiv Eikon	Yes	Yes
Rec_SSell	Recommendation - Number Of Strong Sell	Number of strong sells. The publisher strongly rates this instrument as one that should be sold	Refinitiv Eikon	Yes	Yes

Rec_NoOpinion	Recommendation - Number Of No Opinion	Number of no opinion recommendation. The publisher covers the company but currently as no opinion on the stock	Refinitiv Eikon	Yes	Yes
LTG_Mean	Long Term Growth - Mean	The statistical average of all broker estimates determined to be on the majority accounting basis. Long-term growth is an estimate of the compound average rate of EPS growth an analyst expects over a period of three to five years.	Refinitiv Eikon	Yes	Yes
Num_Analyst	Number of Analysts	Number of sell-side analysts covering the security	Refinitiv Eikon	Yes	Yes
Panel C: Macroeconomic					
GDP	Gross Domestic Product	Gross Domestic Product, Standardized, Constant Prices, SA, USD, 2010 Prices	Refinitiv Eikon	Yes	Yes
GDP_growth%	Gross Domestic Product Growth	Gross Domestic Product, % Quarter On Quarter, Standardized, SA	Refinitiv Eikon	Yes	Yes
consumer_spending	Consumer Spending	Consumer Spending, Standardized, Constant Prices, SA, USD, 2010 Prices	Refinitiv Eikon	Yes	Yes
consumer_spending%	Consumer Spending Growth	Consumer Spending, % Quarter On Quarter, Standardized, SA	Refinitiv Eikon	Yes	Yes
government_spending	Government Spending	Government Consumption, Standardized, Constant Prices, SA, USD, 2010 Prices	Refinitiv Eikon	Yes	Yes
government_spending%	Government Spending Growth	Government Consumption, % Quarter On Quarter, Standardized, SA	Refinitiv Eikon	Yes	Yes
gross_fixed_capital_investment	Gross Fixed Capital Investment	Gross Fixed Capital Investment, Standardized, Constant Prices, SA, USD, 2010 Prices	Refinitiv Eikon	Yes	Yes
gross_fixed_capital_investment%	Gross Fixed Capital Investment Growth	Gross Fixed Capital Investment, % Quarter On Quarter, Standardized, SA	Refinitiv Eikon	Yes	Yes
exports_goods_services	Exports - Goods and Services	Exports - Goods and Services, Standardized, Constant Prices, SA, USD, 2010 Prices	Refinitiv Eikon	Yes	Yes
exports_goods_services%	Exports - Goods and Services Growth	Exports - Goods and Services, % Quarter On Quarter, Standardized, SA	Refinitiv Eikon	Yes	Yes
imports_goods_services	Imports - Goods and Services	Imports - Goods and Services, Standardized, Constant Prices, SA, USD, 2010 Prices	Refinitiv Eikon	Yes	Yes
imports_goods_services%	Imports - Goods and Services Growth	Imports - Goods and Services, % Quarter On Quarter, Standardized, SA	Refinitiv Eikon	Yes	Yes
money_supply_M0	Money Supply MO	Money Supply Money Supply M0, Standardized, SA, USD	Refinitiv Eikon	Yes	Yes
money_supply_M0%	Money Supply MO Growth	Money Supply Money Supply M0, Standardized, SA, Change y/y	Refinitiv Eikon	Yes	Yes
money_supply_M1	Money Supply M1	Money Supply Money Supply M1, Standardized, SA, USD	Refinitiv Eikon	Yes	Yes
money_supply_M1%	Money Supply M1 Growth	Money Supply Money Supply M1, Standardized, SA, Change y/y	Refinitiv Eikon	Yes	Yes
money_supply_M2	Money Supply M2	Money Supply Money Supply M2, Standardized, SA, USD	Refinitiv Eikon	Yes	Yes
money_supply_M2%	Money Supply M2 Growth	Money Supply Money Supply M2, % Year On Year, Standardized, SA, Change y/y	Refinitiv Eikon	Yes	Yes
CPI_SA	Consumer Price Index	Consumer Price Index, Standardized, SA, Index, 2010 = 100	Refinitiv Eikon	Yes	Yes
CPI_SA%	Consumer Price Index Growth	Consumer Price Index, % Month On Month, Standardized, SA	Refinitiv Eikon	Yes	Yes
export_prices	Export Prices	Export Prices, Standardized, Index, 2010 = 100	Refinitiv Eikon	Yes	Yes
export_prices%	Export Prices Growth	Export Prices, % Quarter On Quarter, Standardized	Refinitiv Eikon	Yes	Yes
import_prices	Import Prices	Import Prices, Standardized, Index, 2010 = 100	Refinitiv Eikon	Yes	Yes
import_prices%	Import Prices Growth	Import Prices, % Quarter On Quarter, Standardized	Refinitiv Eikon	Yes	Yes
terms_of_trade	Terms of Trade	Terms of Trade, Standardized, Index, 2010 = 100	Refinitiv Eikon	Yes	Yes
terms_of_trade%	Terms of Trade Growth	Terms of Trade, % Quarter On Quarter, Standardized	Refinitiv Eikon	Yes	Yes
labour_force_survey	Labor Force Survey	Employment, Labour Force Survey, Standardized, SA	Refinitiv Eikon	Yes	Yes
labour_force_survey%	Labor Force Survey Growth	Employment, Labour Force Survey, % Quarter On Quarter, Standardized, SA	Refinitiv Eikon	Yes	Yes
unemployment_rate	Unemployment Rate	Unemployed Rate, Standardized, SA	Refinitiv Eikon	Yes	Yes
unemployment_rate%	Unemployment Rate Growth	Unemployed Rate, Month On Month, Standardized, SA	Refinitiv Eikon	Yes	Yes
retail_sales	Retail Sales	Retail Sales, Standardized, SA, Index, 2010 = 100	Refinitiv Eikon	Yes	Yes
retail_sales%	Retail Sales Growth	Retail Sales, Standardized, SA	Refinitiv Eikon	Yes	Yes
industrial_production	Industrial Production	Industrial Production Index, Standardized, SA, Index, 2010 = 100	Refinitiv Eikon	Yes	Yes
industrial_production%	Industrial Production Growth	Industrial Production Index, Standardized, SA	Refinitiv Eikon	Yes	Yes
central_government_deficit	Central Government Deficit	Central Government Deficit / Surplus, Standardized, USD	Refinitiv Eikon	Yes	Yes
central_government_deficit%	Central Government Deficit Growth	Central Government Deficit / Surplus, Month On Month, Standardized, Absolute Change, USD	Refinitiv Eikon	Yes	Yes
government_external_debt	Government External Debt	Government External Debt, Standardized, USD	Refinitiv Eikon	Yes	Yes
government_external_debt%	Government External Debt Growth	Government External Debt, % Quarter On Quarter, Standardized	Refinitiv Eikon	Yes	Yes
current_account_balance	Current Account Balance	Current Account Balance, Standardized, SA, USD	Refinitiv Eikon	Yes	Yes
current_account_balance%	Current Account Balance Growth	Current Account Balance, Quarter On Quarter, Standardized, SA, Absolute Change, USD	Refinitiv Eikon	Yes	Yes
cab%gdp	Current Account Balance % GDP	Current Account Balance AS A Percentage of Gross Domestic Product, Standardized, SA	Refinitiv Eikon	Yes	Yes
cab%gdp%	Current Account Balance % GDP Growth	Current Account Balance AS A Percentage of Gross Domestic Product, Quarter On Quarter, Standardized, SA	Refinitiv Eikon	Yes	Yes

export_goods_bop	Exports of Goods, Balance of Payment Basis	Exports of Goods, Balance of Payments Basis, Standardized, USD	Refinitiv Eikon	Yes	Yes
export_goods_bop%	Exports of Goods, Balance of Payment Basis Growth	Exports of Goods, Balance of Payments Basis, % Month On Month, Standardized	Refinitiv Eikon	Yes	Yes
imports_goods_bop	Imports of Goods, Balance of Payment Basis	Imports of Goods, Balance of Payments Basis, Standardized, USD	Refinitiv Eikon	Yes	Yes
imports_goods_bop%	Imports of Goods, Balance of Payment Basis Growth	Imports of Goods, Balance of Payments Basis, % Month On Month, Standardized	Refinitiv Eikon	Yes	Yes
balance_on_trade_bop	Visible Trade Balance, Balance of Payment Basis	Visible Trade Balance, Balance of Payments Basis, Standardized, USD	Refinitiv Eikon	Yes	Yes
balance_on_trade_bop%	Visible Trade Balance, Balance of Payment Basis Growth	Visible Trade Balance, Balance of Payments Basis, Month On Month, Standardized, Absolute Change, USD	Refinitiv Eikon	Yes	Yes
merchandise_exports	Merchandise Exports	Merchandise Exports, Standardized, SA, USD	Refinitiv Eikon	Yes	Yes
merchandise_exports%	Merchandise Exports Growth	Merchandise Exports, % Month On Month, Standardized, SA	Refinitiv Eikon	Yes	Yes
merchandise_imports	Merchandise Imports	Merchandise Imports, Standardized, SA, USD	Refinitiv Eikon	Yes	Yes
merchandise_imports%	Merchandise Imports Growth	Merchandise Imports, % Month On Month, Standardized, SA	Refinitiv Eikon	Yes	Yes
foreign_trade_balance	Visible Trade Balance	Visible Trade Balance, Standardized, SA, USD	Refinitiv Eikon	Yes	Yes
foreign_trade_balance%	Visible Trade Balance Growth	Visible Trade Balance, Month On Month, Standardized, SA, Absolute Change, USD	Refinitiv Eikon	Yes	Yes
international_reserves	International Reserves	Official International Reserves, Standardized, USD	Refinitiv Eikon	Yes	Yes
international_reserves%	International Reserves Growth	Official International Reserves, % Month On Month, Standardized	Refinitiv Eikon	Yes	Yes
population	Resident Population	Resident Population, Total	Refinitiv Eikon	Yes	Yes
population%	Resident Population Growth	Resident Population, Total	Refinitiv Eikon	Yes	Yes
overnight_rate	Cash Rate Target	Cash Target Rate, End of Period	Refinitiv Eikon	Yes	Yes
overnight_rate%	Cash Rate Target Growth	Cash Target Rate	Refinitiv Eikon	Yes	Yes
government_bond	Government Bond Yield	Government Bond Yield	Refinitiv Eikon	Yes	Yes
government_bond%	Government Bond Yield Growth	Government Bond Yield	Refinitiv Eikon	Yes	Yes
consumer_confidence	Consumer Confidence Index	Consumer confidence index	Refinitiv Eikon	Yes	Yes
consumer_confidence%	Consumer Confidence Index Growth	Consumer confidence index	Refinitiv Eikon	Yes	Yes
business_confidence	Business Survey Optimisim	Business survey optimisim	Refinitiv Eikon	Yes	Yes
business_confidence%	Business Survey Optimisim Growth	Business survey optimisim	Refinitiv Eikon	Yes	Yes
stock_index	Stock Market Index	Country stock price index	Refinitiv Eikon	Yes	Yes
stock_index%	Stock Market Index Growth	Country stock price index	Refinitiv Eikon	Yes	Yes
Panel D: Industry					
gsector	GICS sector	First level in the hierarchy of the GICS	Refinitiv Eikon	Yes	Yes
ggroup	GICS group	Second level in the hierarchy of the GICS	Refinitiv Eikon	Yes	Yes
gind	GICS industry	Third level in the hierarchy of the GICS	Refinitiv Eikon	Yes	Yes
Panel E: Sovereign credit rating					
sov_rating_rank	SP sovereign rating	Code representing the outlook rank	Refinitiv Eikon	Yes	Yes
sov_outlook_rank	SP sovereign outlook rating	Code representing the outlook trend for a country	Refinitiv Eikon	Yes	Yes
Panel F: Recession indicator					
NBER_ri	NBER Recession Indicator	NBER based Recession Indicators for the United States from the Period following the Peak through the Trough	St. Louis Fed	Yes	No
OCED_ri	OECD Recession Indicator	OECD based Recession Indicators for the United States from the Peak through the Trough	St. Louis Fed	No	Yes
Panel G: Country and financial quarter					
loc_rank	Country	Unique count of each nation	Refinitiv Eikon	No	Yes
developed_nation	Developed Nation	Binary feature representing developed or developing nation	United Nations	Yes	Yes
fqtr	Financial quarter	Quarter of financials	Compustat	Yes	Yes

E Feature importance plots

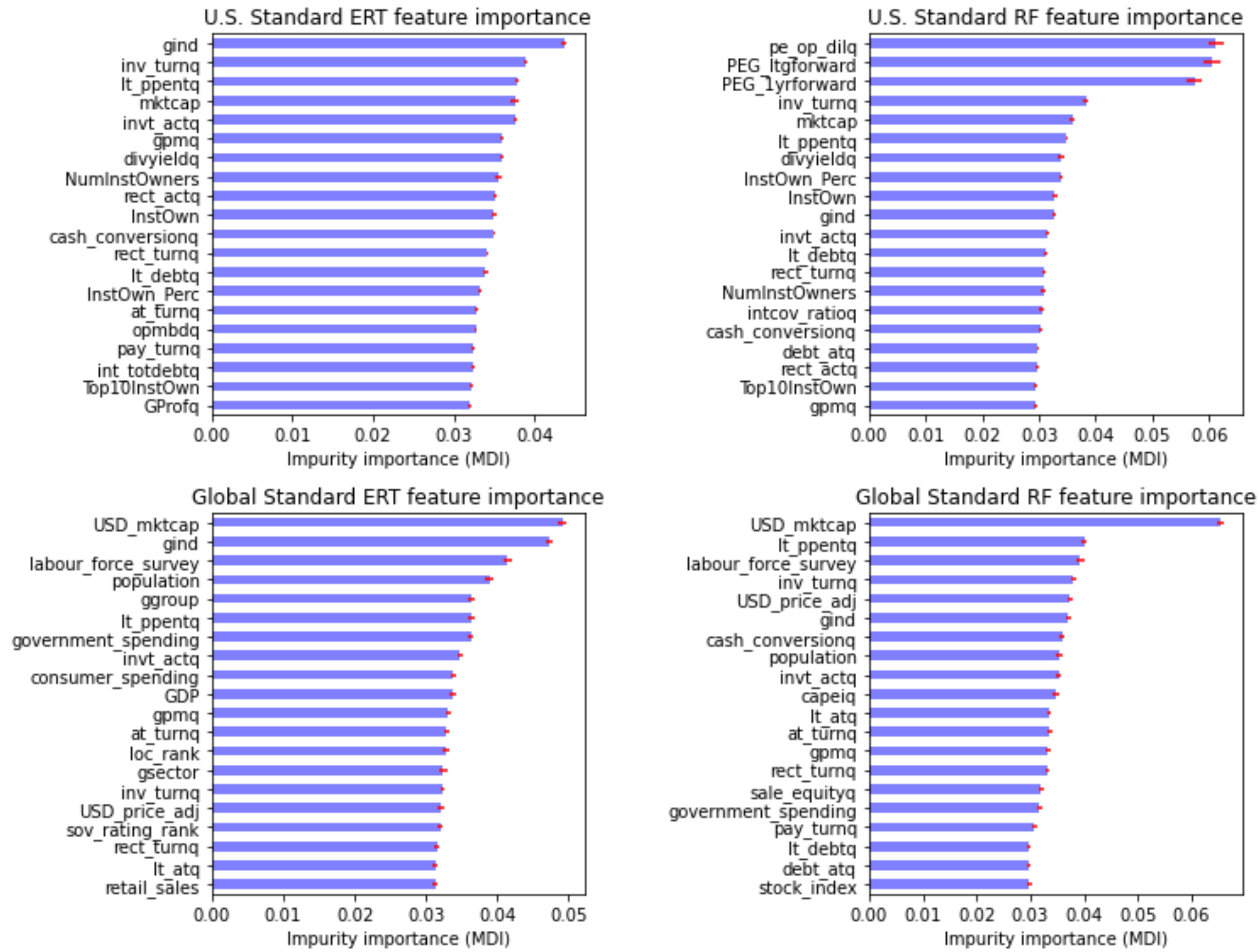

Figure E.1: MDI standard feature importance
This figure illustrates mean decrease impurity (MDI) feature importance of the recursively reduced 20 feature extremely randomized trees (ERT) and random forests (RF) models. Standard United States (US) and global samples are utilized. MDI prediction performance is recursively investigated in Panel A of Table 13 (14) for US (global) and the reported figure uses the features where $n = 20$ features.

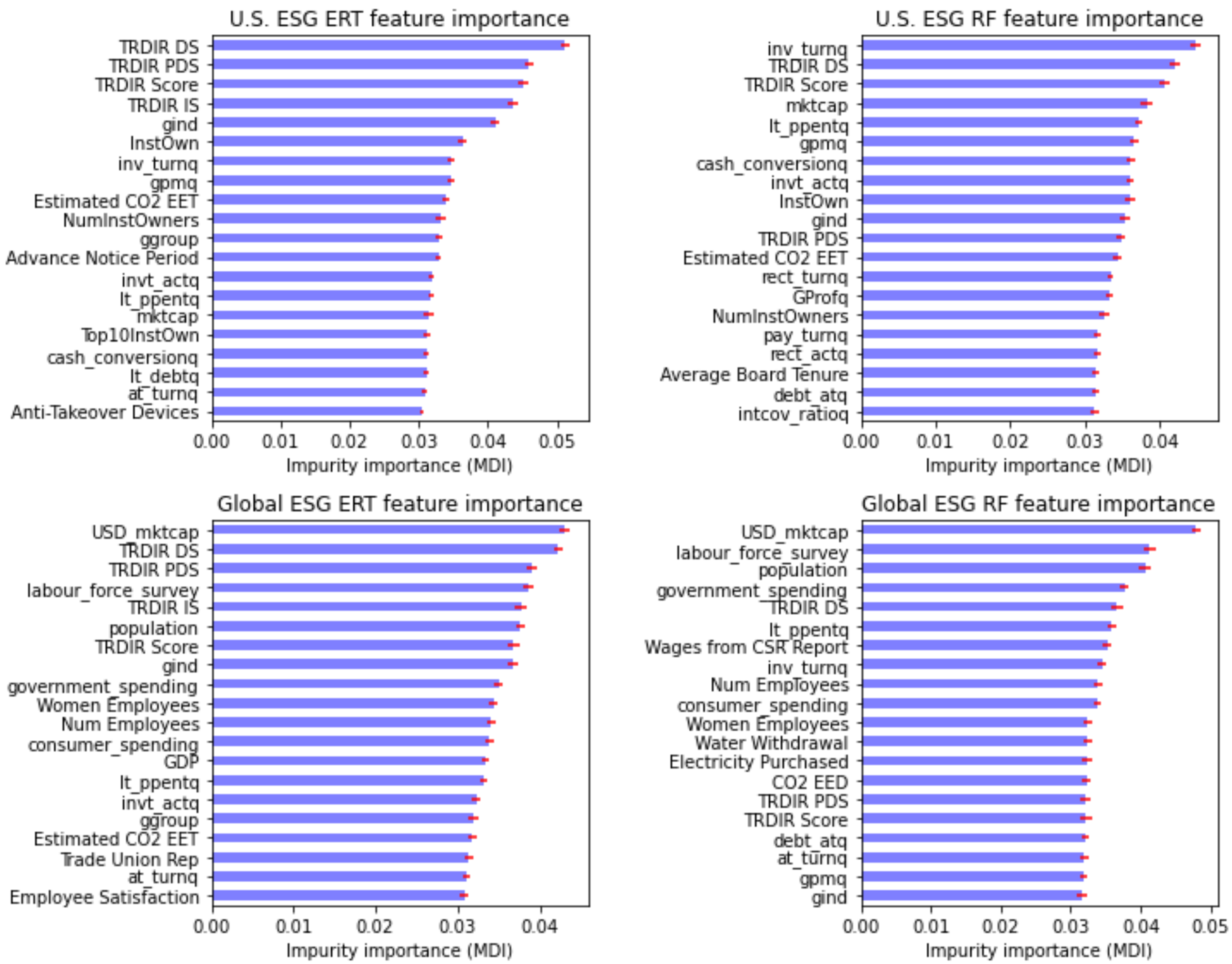

Figure E.2: MDI ESG feature importance
This figure illustrates mean decrease impurity (MDI) feature importance of the recursively reduced 20 feature extremely randomized trees (ERT) and random forests (RF) models. Environmental, social, and governance (ESG) United States (US) and global samples are utilized. MDI prediction performance is recursively investigated in Panel A of Table 13 (14) for US (global) and the reported figure uses the features where $n = 20$ features.

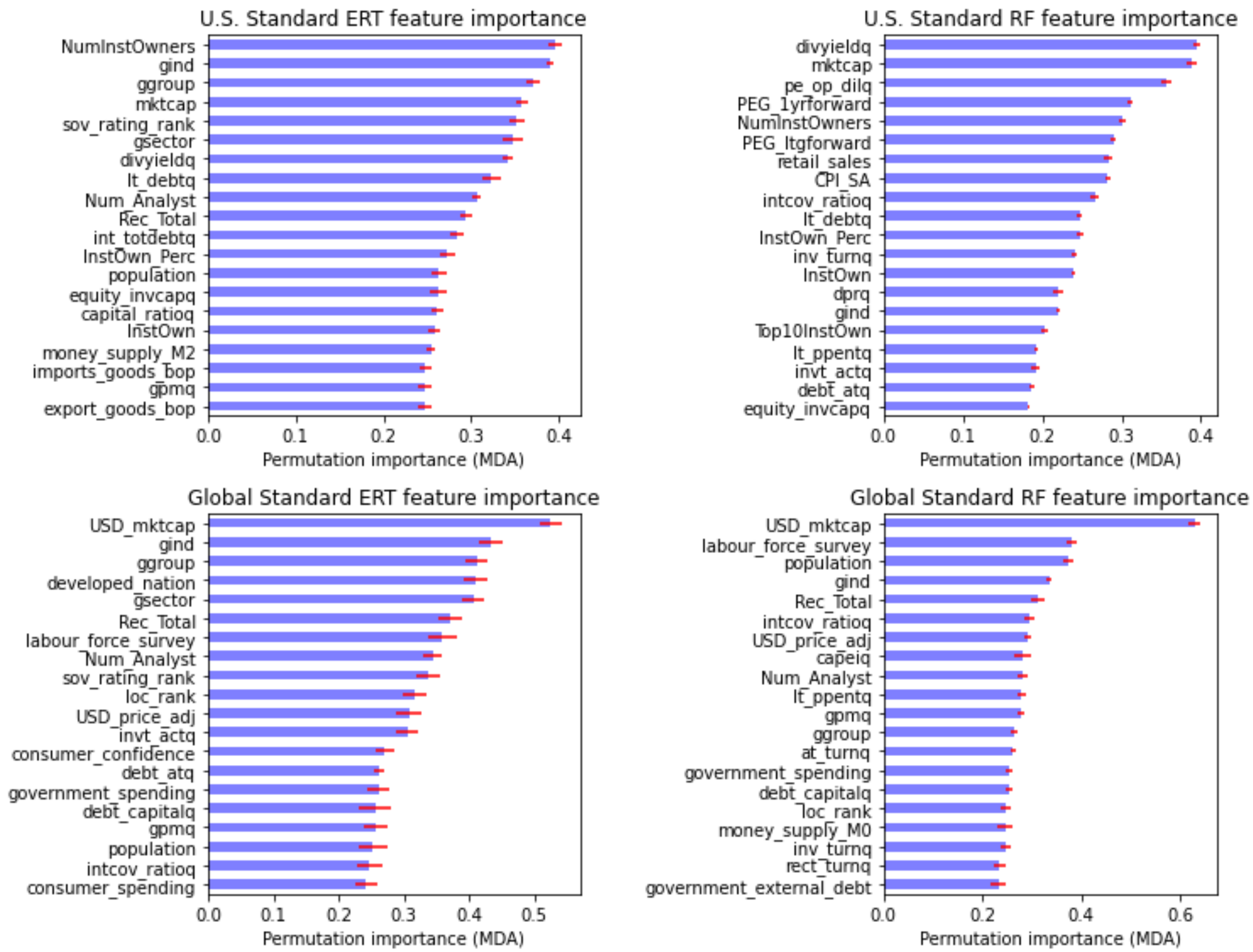

Figure E.3: MDA standard feature importance
This figure illustrates mean decrease accuracy (MDA) feature importance of the recursively reduced 20 feature extremely randomized trees (ERT) and random forests (RF) models. Standard United States (US) and global samples are utilized. MDA prediction performance is recursively investigated in Panel B of Table 13 (14) for US (global) and the reported figure uses the features where $n = 20$ features.

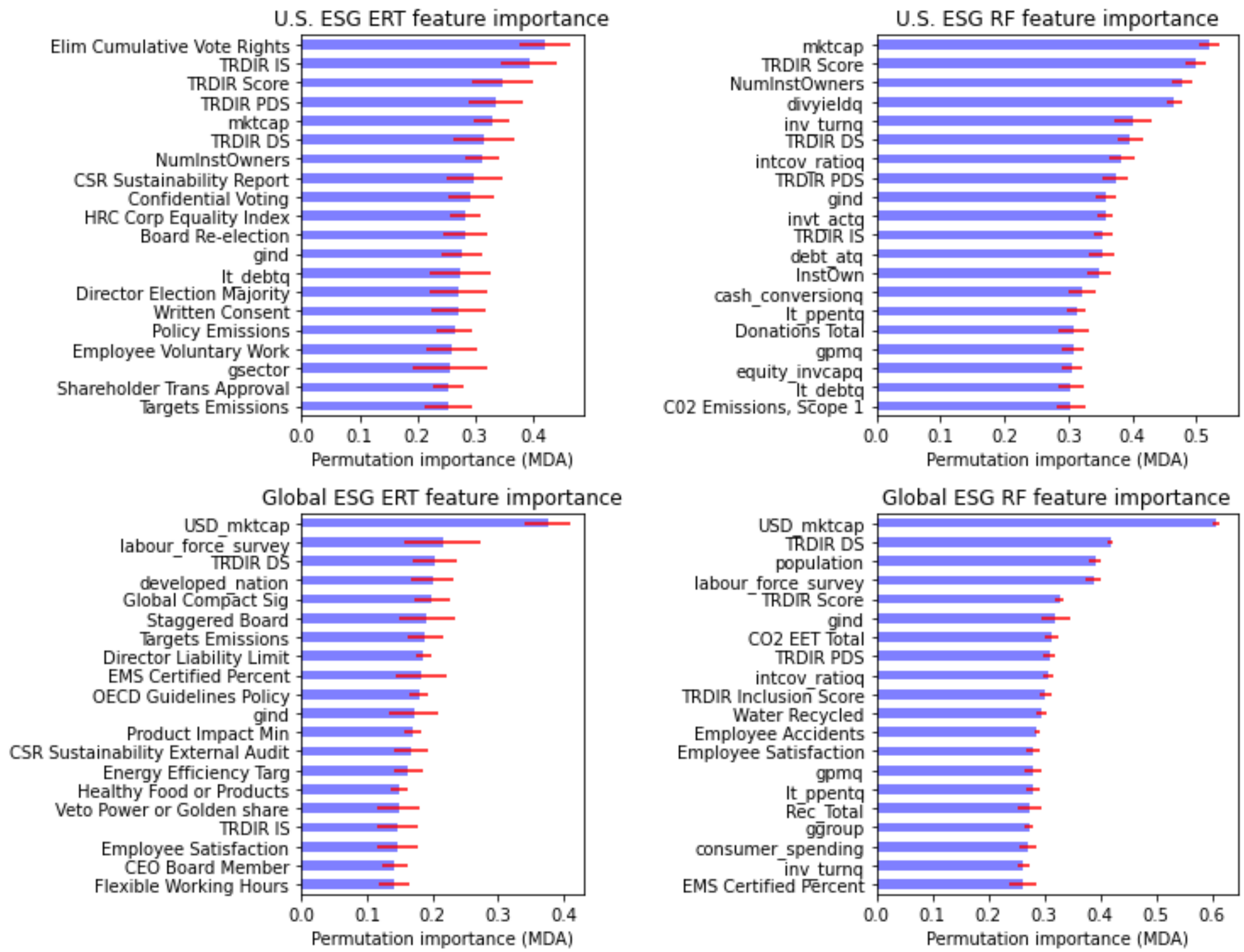

Figure E.4: MDA ESG feature importance

This figure illustrates mean decrease accuracy (MDA) feature importance of the recursively reduced 20 feature extremely randomized trees (ERT) and random forests (RF) models. Environmental, social, and governance (ESG) United States (US) and global samples are utilized. MDA prediction performance is recursively investigated in Panel B of Table 13 (14) for US (global) and the reported figure uses the features where $n = 20$ features.

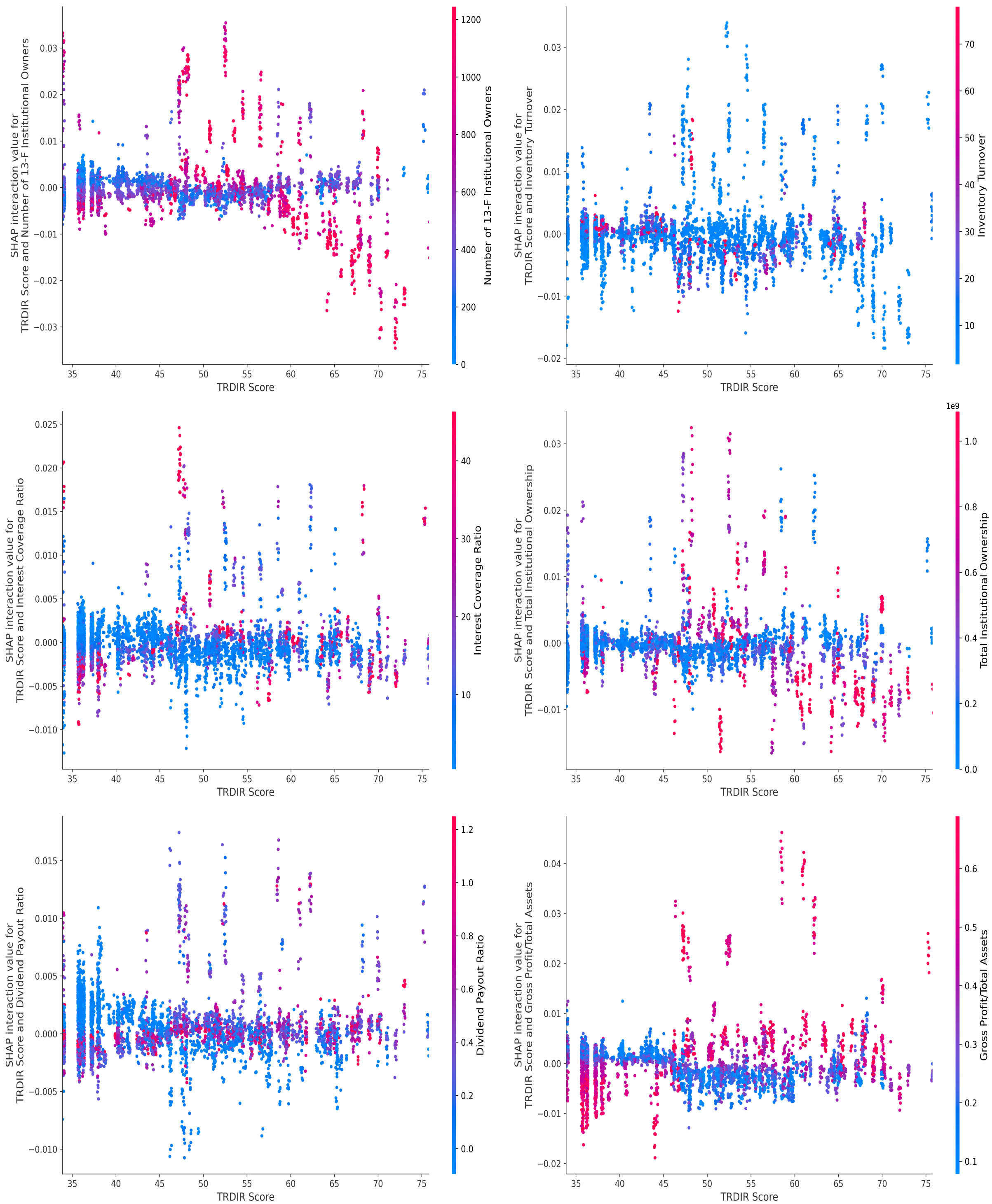

Figure E.5: US-ESG SHAP feature interactions
This figure illustrates SHapley Additive exPlanations (SHAP) interaction value effects for Thomson Reuters Diversity and Inclusion Rating (TRDIR) Score and varying features. TRDIR Score is situated on the x-axis and the interacting feature on the right-hand side y-axis. The color reveals if the feature value was high (i.e., red) or low (i.e., blue) for the interacting feature. The position of the observation on the plot corresponds to the SHAP interaction value between the two features (left-hand side y-axis). Random forests is utilized with 100 trees and depth of 13 for the United States environmental, social, and governance sample.