Received: 26 July 2022 | Revised: 18 January 2023 | Accepted: 27 February 2023

DOI: 10.1002/asmb.2758

SPECIAL ISSUE PAPER

WILEY

Firms' profitability and ESG score: A machine learning approach

Valeria D'Amato[1] | Rita D'Ecclesia[2] | Susanna Levantesi[2]

[1]Department of Pharmacy, University of Salerno, Via Giovanni Paolo II, 132 Fisciano, 84084, Italy

[2]Department of Pharmacy, Sapienza University of Rome, Piazzale Aldo Moro, 5 Roma, 00185, Italy

Correspondence
Susanna Levantesi, Department of Pharmacy, Sapienza University of Rome, Piazzale Aldo Moro, 5, 00185 Roma, Italy.
Email: susanna.levantesi@uniroma1.it

Abstract

Corporate social responsibility (CSR) is found to impact firms' performance, for instance, enhancing reputation, increasing innovation capabilities, customer loyalty, and customer satisfaction help improve financial performance. However, the literature provides limited evidence of the relationship between CSR indicators, such as the ESG score, and the firm's profitability, which is often measured by the earnings before interest and taxes (EBIT). We investigate this issue by analyzing a sample of about 400 companies constituting the EuroStoxx-600 index, from 2011 to 2020, using different machine learning models. The novelty of our contribution lies in assessing whether the ESG score has a significant influence on the firms' profitability. Specifically, we investigate the relationship between ESG score and EBIT using machine learning interpretability toolboxes such as partial dependence plots and individual conditional expectation. Tools which help to measure the functional relationship between the predicted response and one or more features, while the Shapley value allows to examine the contribution of the feature to the prediction. Our findings show that the model can reach high levels of accuracy in detecting EBIT and that the ESG score is a promising predictor, compared to other traditional accounting variables.

KEYWORDS

ESG investments, firm's performance, interpretability tools, machine learning

1 | INTRODUCTION

ESG adoption is becoming a crucial issue, driven by client demand and a desire to make an impact. Investors and banks are moving away from basic screening methods towards more targeted and sophisticated strategies. One common strategy is integrating ESG into the investment process, the business as usual process. Investors are taking a holistic approach as they look to comprehensively embed ESG into the investment process rigorous approach.

The increasing sophistication of ESG investors makes them recognize that companies with good sustainable credentials are more likely to outperform. Fewer investors point to sacrificing returns as an adoption hurdle. And more are now investing in ESG with the specific and sole aim of generating alpha. Furthermore, investors largely agree that investment returns and sustainable impact go hand in hand, so firms increasingly recognize the economic value of embedding ESG criteria in their activities.

Furthermore, it has been proved that ESG investments can be a chance to reduce the systemic risk for investors. Cerqueti et al.[1] evaluate the impact of portfolio liquidation on equity mutual funds with various ESG ratings. The authors notice that the relative market value loss of the high ESG ranked funds is lower than the loss experienced by the low ESG ranked ones in periods with lower volatility. Cerqueti et al.[2] also investigate if the ESG investments mitigate the risk of contagion among equity mutual funds, finding that their vulnerability to contagion decreases when the level of ESG compliance rises.

Several firms are already integrating environmental, social, and governmental considerations and risks into their governance, strategies, operations, and risk management. For the market to become mainstream, practices cannot continue to be assessed, based only on financial performance indicators. A wide-scale of ESG investment strategies exist, from exclusionary screening to impact/community investing, from best-in-class investment selection to norms-based screening; from ESG integration to sustainability-themed investing and engagement and voting on sustainability matters, as classified by Eurosif, following the Sustainable and Responsible investment (SRI) approaches introduced in 2012.

The taxonomy of the seven representative ESG investing strategies has been also codified in Reference 3. It is not exhaustive, being potentially unlimited the number of ESG-based Investment strategies that investors may develop and implement. Nevertheless, the aforementioned classification has become a global standard both in academia and among professionals.

From 2016 to 2020, on one side Sustainability-themed investing, ESG integration and engagement, and voting on sustainability matters have all experienced remarkable growth. On the other side, norms-based screening, positive screening as well as negative screening have all recorded a more variable trajectory.[4]

In particular, the strategy devoted to the integration of ESG in investment decisions has exceptional popularity, extensively promoted as a driver of long term financial performance. However thematic investing is the most used strategy with a 1200% increase in total Assets Under Management (AUM) between 2012 and 2018, by reaching $1018 million by the end of 2018, besides being the youngest ESG strategy.

The thematic approach is about identifying particular trends or themes specifically related to sustainability, such as clean energy, green technology, or sustainable agriculture, solar energy and so on. According to the UNCTAD definition, ESG-themed strategies include investments primarily focused on only one ESG pillar (environment, social or governance), "alternatively, they track a 'quasi sector', such as energy efficiency or food security".[5] The thematic style unhampered by individual countries is inherently global in nature and generally referred to a long-term horizon. It introduces a new perspective in operational and management processes, concerning the whole of an organization or a firm, involving all the operations addressed to a specific sustainability theme.

The debate about the performance measurement of sustainable investing has at least 50 years of history, starting from References 6-8. The stream of literature on the topic is characterized by contradictory views on the ESG and corporate financial performance relationship. Nonetheless, to the best of our knowledge, the academic literature does not still analyze the single ESG investment styles and their relationships and differences in terms of profitability, except in Reference 5 where the authors study the risk-adjusted financial performance of ESG-themed megatrend investment strategies in global equity markets. The research does not consider ESG scores of portfolio firms, emphasizing the Sustainable Development Goals (SDG)-related business models.

Finally, ESG is becoming an increasingly important topic also for commodity companies. Nevertheless, to make commodity investments in an ESG-centered world and support responsible investors' work to incorporate ESG into traditional macro commodity and multi-asset strategies is not a trivial task. Commodities play a dual role in the sustainability field, driving the transition, which transfers into a growing demand for commodities that promote clean energy, as well as sustainably sourced and ethically produced commodities. The global financial crisis of late 2008 has led to intense scrutiny of the foundational beliefs and structures that underpin current global markets and investment models. As part of this reexamination, questions about the efficacy of integration and related trends have intensified. The global financial crisis has fostered the commitment to responsible investments and ESG integration mainly for energy companies and focused products or services. National and International financial regulators (EU, European Securities, and Market Authority-ESMA, European insurance and banking bodies such as EIOPA and EBA) developed a range of practices associated with forms of sustainable finance* with an increasing focus on ESG taxonomies, approaches, and marketing to investors. Policy-makers also have contributed to strengthening practices concerning sustainable finance in several ways, including but not limited to Taxonomies to clarify meaning; Issuer disclosures of E, S, and G in both corporate and financial services sectors; Policy

*Recently, in mid-2020, they also issued a consultation paper seeking input on proposed ESG disclosure standards for financial market participants, advisers and products.[9]

development across Europe, the US, and Japan. Companies in different sectors have all moved toward a more responsible and sustainable business, financial companies, industries, and basic materials seem to have been the most active in the last decade. For instance, the outcome of the UN Climate Change Conference COP26 in November 2021 is having deep implications for producers and suppliers of commodities who are under pressure to reduce their carbon footprint whilst remaining profitable businesses throughout the value chain.

In this work, due to its impressive growth, we focus on ESG-themed investments by properly considering the ESG scores for explaining the profitability, being not trivial the virtuous circle between ESG investments and the firms' success. We show that only a massive investment in sustainability and ESG criteria, which can be measured by higher ESG scores, leads to enhancing the strength of a company's balance sheet. On the contrary, according to our findings, weak efforts in binding ESG elements into an investment strategy do not create extra profits. Our outcomes can be consistently framed in light of the new theories on the expectations of market participants about the implementation of the climate policies (climate sentiments). According to a new strand of literature, the climate sentiment discounted in market expectations contributes to create or destroy the investment profits. Indeed, some authors recognized that investors and financial markets are not yet pricing climate-related risks (and opportunities) in the value of financial contracts.[10,11] The sudden changes in climate has fostered the introduction of new policies and regulation, this generates mispricing of climate risks affecting asset price volatility and financial stability.[11] Broadly speaking, we could codify a sort of ESG sentiment, that contributes to the profitability of the investments.

Currently ESG ratings assigned to financial investments could contribute to the profitability of the firm business. As a matter of fact, banks and other institutions play a role of transmitters of political economic impulses on environmental issues by the implementation of adequate set of incentives to support lending to green projects. The introduction of Green Supporting Factors (GSFs) in the agenda of the international bank system involves a decrease in Basel III capital regulatory requirements for exposures with low-carbon firms. The lower risk weights for loans to low-carbon firms corresponds to lower interest rates and low-carbon firms' capital cost. Indeed, "the change in interest rate can affect the relative prices of low-carbon (carbon-intensive) goods and the level and composition of the final demand of the economy. Being more price competitive, the demand for low-carbon capital goods increases and so do the profits for the low-carbon firms". Lower (higher) interest rates determine lower (higher) prices, which in turn have an impact on demand, firms' investments, and then profits in the sectors.[11]

In our research, we develop a regression model to predict the EBIT of a firm by using both balance sheet information and the global ESG score. To the best of our knowledge, our study is the first one to define an EBIT prediction model that includes the ESG score among the predictors. In addition, we provide a contribution in the methodological approach by means of a comparison between a traditional statistical technique (generalized linear models), machine learning approach (Decision trees), and ensemble methods (Bagging, Random forest and Gradient Boosting). This allows us to evaluate and, in case, confirm the common opinion that ensemble methods often outperform individual techniques. Our analysis shows that the ESG score has a significant effect in the operating profit.

The rest of the paper is organized as follows. Section 2 discusses the data, Section 3 analyses the regression models. Section 4 provides a toolkit for the Machine Learning Interpretability. In Section 5, the main outcomes are illustrated. Finally Section 6 concludes.

2 | DATASET DESCRIPTION

We study the constituents of the Euro-Stoxx 600 Index, which represents large, mid and small capitalization companies across 17 European countries. We gather the ESG scores and balance sheet information of the constituents of the Euro-Stoxx 600 index by the Thomson Reuters Refinitiv ESG (Refinitiv ESG, henceforth) in the years 2011–2020. The final sample includes 422 companies (about 70% of the total) for which data on ESG scores were available for the selected period. The Refinitiv ESG database assigns a ESG measure to over 450 company-defining a score for each component: Environment-E, Social-S, and Governance-G. The companies are aggregated into 10 categories and are discounted for materially important ESG controversies. A combination of the 10 categories[†] provides the final ESG score, which is a reflection of the company's ESG performance based on publicly reported information in the three ESG pillars with the

[†] Environmental: Resource use, emissions, innovation; Social: Workforce, human rights, community, product responsibility; Governance: Management, shareholders, CSR strategy.

TABLE 1 Industry sectors' proportion of the dataset.

Sector	Abbreviation	Proportion (%)
Basic materials	BasMat	10.7%
Consumer cyclicals	ConCyc	16.4%
Consumer non-cyclicals	ConNCy	8.5%
Energy	Ene	4.3%
Financials	Fin	18.2%
Healthcare	Hea	7.3%
Industrials	Ind	16.8%
Real estate	ReaEst	4.0%
Technology	Tec	8.1%
Utilities	Uti	5.7%

TABLE 2 Conversion from a percentile score to a letter grade.

Score range	Grade	Description
$0.000 \leq score \leq 0.083$	D–	Poor relative ESG performance and insufficient degree of transparency in reporting material ESG data publicly
$0.083 < score \leq 0.166$	D	
$0.166 < score \leq 0.250$	D+	
$0.250 < score \leq 0.333$	C–	Satisfactory relative ESG performance and moderate degree of transparency in reporting material ESG data publicly
$0.333 < score \leq 0.416$	C	
$0.416 < score \leq 0.500$	C+	
$0.500 < score \leq 0.583$	B–	Good relative ESG performance and above- average degree of transparency in reporting material ESG data publicly
$0.583 < score \leq 0.666$	B	
$0.666 < score \leq 0.750$	B+	
$0.750 < score \leq 0.833$	A–	Excellent relative ESG performance and high degree of transparency in reporting material ESG data publicly
$0.833 < score \leq 0.916$	A	
$0.916 < score \leq 1$	A+	

Source: Refinitiv ESG.

weights of each pillar being 34% for E, 35.5% for S, and 30.5% for G.[12] Companies are classified according to the Thomson Reuters Business' Classification that is an owned industry classification system operated by Thomson Reuters.[13] The distribution of the 422 companies among the various economic sectors is shown in Table 1. We observe that about 18.2% of the analyzed companies belongs to the financials sectors, 16.8% industrials, 16.4% consumer cyclicals, and 10.7% basic materials.

The ESG score ranges between a minimum score (0) and a maximum score (100), and is available both in percentage (from 0% to 100%) and in letter summarized in four macro-classes representing the percentile of the distribution (see Table 2). The mean value of the ESG score for the companies included in our sample in the years 2011–2020 is 64.27.

In Table 3, we provide a description of the variables included in our model. In Table 4, we report the yearly mean values of the ESG score by economic sectors. Overall, we observe that the average ESG score increased about 16.14 points from 2011 to 2020, moving from 57.72 to 73.86. We note that each sector reports an increasing ESG score since 2011, however some sectors are by far more dynamic than others: Financials (+22.24), Health (+19.95), Basic materials (+19.12),

TABLE 3 Variables' description.

Variable	Description
EBIT	Earnings before interest and taxes, computed as total revenues for the fiscal year minus total operating expenses plus operating interest expense, unusual expense/income and non-recurring items, supplemental, total for the same period. This definition excludes non-operating income and expenses
ESG.Score	Measure of the overall corporate social responsibility
Year	2011–2020
Sector	Categorical variable indicating the company's industry sector
Net.Sales	Sales receipts for products and services, less cash discounts, trade discounts, excise tax, and sales returns and allowances
PE	Price-to-earnings, computed as the ratio of fiscal period price close to earnings per share excluding extraordinary items
ROE	Return on equity, profitability ratio calculated by dividing a company's net income by total equity of common shares (percentage values)
DY	Dividend yield, calculated as the dividends paid per share to the primary common shareholders for the fiscal period divided by the historical price close (percentage values)

TABLE 4 Mean values of the ESG score by economic sectors.

Year	BasMat	ConCyc	ConNCy	Ene	Fin	Hea	Ind	ReaEst	Tec	Uti	All
2011	57.19	59.05	60.40	70.22	41.71	59.96	54.95	57.08	54.21	60.63	57.72
2012	60.87	59.24	62.05	69.13	44.19	60.56	55.97	59.34	57.79	61.83	59.06
2013	61.97	58.71	63.39	72.30	43.80	62.07	56.27	59.60	56.77	60.67	59.52
2014	63.95	58.98	66.37	65.08	43.93	62.99	57.67	64.03	56.67	59.64	60.23
2015	67.95	62.50	66.17	69.20	50.74	63.86	61.66	68.42	58.66	64.06	63.31
2016	67.03	64.33	65.89	72.56	50.39	67.56	63.68	68.52	60.68	64.95	64.68
2017	68.90	66.20	68.87	74.46	54.41	73.31	64.83	67.89	62.77	63.18	66.75
2018	71.45	68.40	70.21	76.03	55.67	73.95	66.86	69.30	63.73	66.98	68.69
2019	73.64	69.34	71.75	69.13	58.23	77.21	70.37	71.15	68.21	73.34	70.69
2020	76.31	72.23	74.21	77.63	63.95	79.91	72.61	71.32	70.07	76.77	73.86

Note: Years 2011–2020.

and Industry (+17.6) are the sectors which report a larger increase in ESG score during the 2011–2020 period. The Energy sector by contrast is the one which shows the smaller increase (only +7.41 points) but it is also the sector which had the highest ESG score in 2011 (70.22), far above the average score (57.72). The energy sector, includes companies involved in the exploration and development of oil or gas reserves, oil and gas drilling, and refining. The energy industry also includes integrated power utility companies such as renewable energy and coal and it is the sector which started to pay attention to corporate responsible criteria earlier than other sectors. Since 2010, there has been a growing interest in impact studies to analyze strategies to reduce the vulnerability of the energy sector to climate change. The global energy sector aims to be transformed into a low carbon energy supply system in response to climate change mitigation and related policies (e.g., the 2015 Paris Agreement under the United Nations Framework Convention on Climate Change), it also needs to adapt to climate change and its effects to ensure that energy supplies remain secure and reliable. This can explain why the companies of the Energy sector reported a High ESG score (70.22) in 2011 and only after 2015 have shown an improvement of the score reaching 77.63 in 2020, still the highest score among the companies we analyze. Companies in the financial sector result the most dynamic ones, starting from an ESG score of 41.71 in 2011 and reporting an increase of 22.2 points, this is mainly due to the active interventions of Regulatory authorities and European Institutions. Just to mention few actions by the EU Committee and the ECB aimed to push financial companies to become more socially responsible: in 2016 the European Commission set up a High-level expert group on sustainable finance comprised of 20 senior experts from European and international institutions. Its role was to provide advice to the Commission on how to steer the flow of public and private capital toward sustainable investments, identify the steps that financial institutions

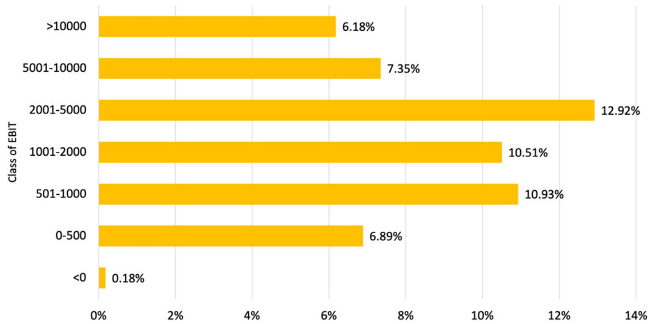

FIGURE 1 Percentage distribution of earnings before interest and taxes (EBIT) values (in million Euros). Years 2011–2020.

TABLE 5 Mean values of the ESG score by classes of earnings before interest and taxes (EBIT) values (in million Euros).

EBIT	ESG.Score (mean)
<0	57.02
0 – 500	61.33
501 – 1000	69.22
1001 – 2000	69.99
2001 – 5000	71.64
5001 – 10000	77.66
> 10000	76.26

Note: Years 2011–2020.

and supervisors should take to protect the stability of the financial system from risks related to the environment deploy these policies on a pan-European scale. The Member States expert group (MSEG) on sustainable finance was created in April 2018 with the aim to assist the European Commission in implementing EU legislation and policies related to sustainable finance. The recent introduction of the EU Taxonomy Regulation (Regulation (EU) 2020/852) is also providing criteria for financial institutions to identify and manage the environment risks. Particularly active sectors in terms of ESG commitment are the basic materials and the industry. The basic materials is an industry category made up of businesses engaged in the discovery, development, and processing of raw materials. The sector includes companies engaged in mining and metal refining, chemical products, and forestry products. All physical goods are made up of a combination of basic materials processed to create a finished good. Basic materials companies are the first stage in the supply chain of various goods, discovering and extracting natural resources. Many of the materials that basic materials companies produce are considered commodities, gold, coal and metals are heavily contributing in this sector. Gold or coal as mined commodities, are naturally exposed to environmental, social and governance (ESG) risks. Mining, by its nature, is physically disruptive: the energy intensive extraction of gold or coal can lead to water pollution, loss of biodiversity, and highly toxic emissions. Regulatory developments in this space have addressed these issues and are bringing further transparency to the supply chain. Organizations such as the OECD (Organization for Economic Co-operation and Development) have sought to establish best practice, providing guidelines for extraction companies addressing worker safety, human rights, the environment and corporate governance; extending to mining operations, supply chains and other business relationships. These standards seek to improve the behavior of the mining companies and drive positive change.

Figure 1 shows the EBIT percentage distribution. The percentage of firms having a negative EBIT value is very low (0.18%) so the sample collects firms with positive EBIT.

Looking at the average values of the ESG score by EBIT classes (Table 5), we can see that ESG score rises when EBIT increases, showing a non linear pattern.

3 | REGRESSION MODELS

Given a set of features, $X_1, X_2, \dots, X_p$ belonging to the predictor space $\mathbb{X}$, a generic regression model aims at estimating the relationship between a target variable Y, and the vector of p features $\mathbf{X}$:

$$Y = f(\mathbf{X}) + \epsilon \tag{1}$$

where ϵ is the error term. The generic observation is denoted by $\{y_i, \mathbf{x}_i\}$, where $\mathbf{x}_i = \left(x_{i1}, x_{i2}, \dots, x_{ip}\right)$, and $i = 1, 2, \dots, n$ with n is the total number of observations.

In our model, *EBIT* is the target variable Y, and Year, ESG.Score, PE, Net.Sales, DY, ROE and Sector are the features $\mathbf{X}$.

To estimate function $f(.)$ we use a machine learning approach, and apply both, individual techniques (decision trees) and ensemble methods (bagging, random forest, and gradient Boosting) to compare to traditional statistical techniques as the generalized linear model. The ensemble methods aim to combine the predictions of different estimators to improve the generalization capacity and the robustness of a single estimator. They are usually categorized into average methods and boosting methods. The former (e.g., bagging and random forest) build different estimators independently and calculate the average of their predictions. On average, the ensemble estimator is often better than any single estimator as it has a lower variance. The latter (e.g., gradient boosting) sequentially build basic estimators to achieve a bias reduction. The ensemble estimator is obtained by a combination of different weak estimators. In the following, we provide a brief description of the models used.

Decision trees. The decision trees (DT) algorithm splits the predictor space $\mathbb{X}$ into J distinct and non-overlapping regions, $R_1, R_2, \dots, R_J$, providing the same prediction for all the observations falling into R_j. The DT estimator is: $\hat{f}^{DT}(\mathbf{X}) = \sum_{j\in J} \hat{y}_{R_j} \mathbb{1}_{\{\mathbf{x}\in R_j\}}$, where $\mathbb{1}_{\{.\}}$ is the indicator function. Regions $(R_j)_{j\in J}$ are identified by minimizing the residual sum of squares $\sum_{j\in J}\sum_{i\in R_j}(y_i - \hat{y}_{R_j})^2$. The target variable $\hat{y}_{R_j}$ is estimated by the average values of the variable belonging to the same region R_j.

Bagging and random forest. The bagging was designed to improve machine learning algorithms' stability and accuracy. This algorithm creates multiple bootstrap samples from the training data and fits a weak learner for each sample. Finally, it aggregates the weak learners by averaging their outputs. Compared to bagging, the random forest (RF) peculiarity is the way of considering the predictors. At each split, the algorithm selects a random subset of predictors as candidates for the subdivision from the final set of predictors, thus preventing the predominance of strong predictors in the subdivisions of each tree. The idea behind RF is inserting a random perturbation into the learning system to differentiate the trees and combine their predictions using an aggregation technique.[14] The RF estimator is: $\hat{f}^{RF}(\mathbf{X}) = \frac{1}{B}\sum_{b=1}^{B}\hat{f}^{DT}(\mathbf{X}|b)$, where B is the number of the bootstrap sample and $\hat{f}^{DT}(\mathbf{X}|b)$ is the DT estimator over the $b \in B$ sample.

Gradient boosting. Gradient Boosting (GB) is an algorithm proposed by Reference 15, which uses fixed-sized DT as weak predictive models (typically, trees with a small number of splits). The prediction is obtained with a sequential approach and not parallelizing the tree build process as in RF. The decision tree at the m-th iteration (with $m = 1, 2, \dots, M$), $f_m^{DT}(\mathbf{X}) = \sum_{j\in J_m} \hat{y}_{R_{jm}} \mathbb{1}_{\{\mathbf{x}\in R_{jm}\}}$, where R_{jm} is the terminal region, is calibrated on the residuals of the tree at previous step to improve the current fit. The GB model updating rule is $f_m(\mathbf{X}) = f_{m-1}(\mathbf{X}) + \sum_{j=1}^{J_m}\gamma_{jm}\mathbb{1}_{\{\mathbf{x}\in R_{jm}\}}$, where γ_{jm} are computed by solving the optimization model $\gamma_{jm} = arg\,min_{\gamma}\sum_{X_j\in R_{jm}} L\left(y_{1j}, f_{m-1}(X_j) + \gamma\right)$, given a specified loss function $L(\cdot)$. The GB final estimation is $\hat{f}(\mathbf{X}) = f_M(\mathbf{X})$.

GLM. The GLM generalizes linear regression by relating the linear model to the response variable through a link function $g(\cdot)$. Therefore, denoting $\eta = g(E(Y))$ the linear predictor, the following equation describes how the mean of the response variable depends on the linear predictor: $\eta = X\beta$, where β is the vector of the regression coefficients that need to be estimated. We assume that Y is distributed as a Gaussian and the link function is an identity, so that: $\eta = E(Y)$. We formulate a model that includes three features' interactions: $I1 = Sector * ESG.Score$, $I2 = Net.Sales * ESG.Score$ and $I3 = Sector * Net.Sales$. Therefore, in this case we obtain the following regression model[‡]: $EBIT \sim Year + Net.Sales + ESG.Score + Sector + PE + ROE + DY + I1 + I2 + I3$.

‡The ANOVA test applied to the GLM with and without these interactions led to accept the model with interactions. The interactions have been chosen using the *interactions* R package, which allows for conducting and interpreting analysis of statistical interaction in regression models.

4 | MACHINE LEARNING INTERPRETABILITY

The increasing shift away from parametric models, such as GLMs, and towards non-parametric and non-linear machine learning models such as random forests, gradient boosting and others has accentuated the need and importance of machine learning interpretability. The complex non-linear machine learning algorithms do not have intelligible parameters and are hence often considered black boxes. To understand how a model operates we need to explain the various stages to know how it works and which decision rules it takes. Model-agnostic (the model's structure is irrelevant) interpretation methods clear up the predictive power of the machine learning models. Several techniques have been identified to prevent Machine Learning models from becoming "back boxes". These include techniques known as local interpretation techniques, as LIME, the Shapley values, SHAP (SHapley Additive exPlanations) the partial dependent plot and surrogate models (i.e., simpler, interpretable models that are trained to approximate the prediction of a more complex algorithm and are used to explain the relationship among data). LIME is Local Interpretable Model-agnostic Explanation, a technique that identifies the features that contribute most to an individual classification through a local approximation performed on slightly modified versions of the original observations; Shapley values measure how much each feature contribute to a prediction based on a large number of comparisons between pairs of alternative feature sets, while SHAP combines features from LIME and Shapley. In this paper we are using a set of techniques described in the following sections.

Partial dependence plots

One of the most used model agnostic tool is the PDP proposed by Reference 15. It shows the marginal effect of one or two features entering into the set of the predicted outcome averaged over the joint values of the other input features. Let X_S be the feature of interest, and X_C be the other features in the model, where C is the complement set of S. The partial dependence function of f on X_S is defined as $f_S^{PD}(x_S) = \mathbb{E}_{X_c}\left[f(x_S, X_C)\right] = \int f(x_S, X_C)dP(X_C)$. where $\mathbb{E}_{X_c}$ is the marginal expectation over the features in set C that corresponds to the integral over the predictions weighted by the probability distribution $P(X_C)$. The PD function is generally estimated by $\hat{f}_S^{PD}(x_S) = \frac{1}{n}\sum_{i=1}^{n} f(x_S, x_{iC})$, where i is a generic observation.

Accumulated local effect plots

The accumulated local effects (ALE) plot[16] shows how the prediction changes locally when the feature is varied. It addresses the bias arising in PD when the selected feature is highly correlated with other features by averaging over a conditional distribution (instead of over a marginal distribution as in PDP). Therefore, ALE plots are unbiased, and still work when predictions are correlated. The ALE function of x_S is defined as $\hat{f}_S^{(ALE)}(x_S) = \int_{z_{0,S}}^{x_S} \mathbb{E}_{X_C|X_S=x_s}\left[\hat{f}^S(X_s, X_c)|X_S = z_S\right] dz_S - c$, with $z_{0,S}$ be a value close to the smallest observation on X_S, and c be a constant. Note that $\hat{f}^S(x_s, x_c) = \frac{\partial \hat{f}(x_S, x_C)}{\partial x_S}$ is the local partial derivative on x_S that describes the local effect (or the change) of X_S on the model prediction.

Individual conditional expectation

Goldstein et al.[17] proposed an extension of PDP, named ICE, which disaggregates the output of PDPs by providing a certain number of estimated conditional expectation curves. Instead, PDP plots give the feature' average partial effect on the predicted response. It is considered a very useful tool for the identification of interactions. Therefore, ICE graphically represents the n estimated conditional expectation curves, where each curve depicts the model prediction as a function of feature X_S, conditional on X_C. Considering the estimated prediction function $\hat{f}$, for each value of X_C, X_{iC} $(i = 1, \ldots, n)$, an ICE line is defined as a single $\hat{f}(x_S, x_{iC})$ evaluated at X_S. Then, on the axis of abscissas, X_S is fixed and the X_C varies across n observations.

Ceteris-paribus profiles

An interesting extension of PDP and ICE plots is the methodology of ceteris-paribus (CP) profiles. CP assesses the influence of a selected feature by assuming that the values of all the other features remain unchanged. Based on the "ceteris paribus" principle ("other things held constant" or "all else unchanged"), it aims to understand how changes in the values of a feature affect the model's predictions. The CP profile shows the dependence of the conditional expectation of the target variable on the values of the selected feature. Let $\mathbf{x}_*$ be a vector with arbitrary values, and $\mathbf{x}_{*s}$ be its s-th element that is the value of feature X_S. The one-dimensional CP profile for model f, feature X_S, and point of interest $\mathbf{x}_*$ is defined as $\hat{f}^{CP}_{\mathbf{x}_{*s}}(z) = \hat{f}\left(\mathbf{x}_{*S|=z}\right)$, where $\mathbf{x}_{*S|=z}$ a vector in which all coordinates are equal to their values in $\mathbf{x}_*$, except of the s-th coordinate, whose value is set equal to z. Therefore, $\mathbf{x}_{*S|=z} = (x_{*1}, \dots, x_{*s-1}, z, x_{*s+1}, \dots, x_{*p})$.

We use the CP profiles as implemented in the *DALEX* R package for R.[18]

Feature interaction

We can also measure how strongly features interact with each other. The interaction measure regards how much of the variance of the model's estimation of the target variable is explained by the interaction. The measure is between 0 (no interaction) and 1 (= 100% of variance of the estimated target variable due to interactions). For each feature, we measure how much they interact with any other feature. Moreover, we also specify a feature and measure all it's two-way interactions with all other features. Feature interaction is based on the H-statistic proposed by Reference 19. The H-statistic for the interaction between, for example, feature X_S and X_K is: $H^2_{sk} = \frac{\sum_{i=1}^{n}\left[f^{PD}_{sk}(x_{is}, x_{ik}) - f^{PD}_{s}(x_{is}) - f^{PD}_{k}(x_{ik})\right]^2}{\sum_{i=1}^{n} PD^2_{sk}(x_{is}, x_{ik})}$.

Shapley additive explanations

An alternative method for unfolding individual predictions originates from the coalitional game theory through the Shapley value. It is assumed that, for one observation, the feature values play a game together, in which they get the prediction as a payoff (the model output). The Shapley value shows how to fairly allocate the payoff among the input features or, in other words, splits an individual prediction among all contributed features, providing a full explanation of why a given variable has received a specific output value. We consider the unified framework based on the Shapley value proposed by Reference 20, the SHapley Additive exPlanations (SHAP). We can express the prediction model f as $f(\mathbf{X}) = \phi_0 + \sum_{j=1}^{P} \phi_j X'_j$, where $X = h(X')$ with $h(\cdot)$ a mapping function relating X to X', and ϕ_0 indicating a constant value when all inputs are missing. The generic ϕ_j is the weight against the feature contribution summation for the output of the model for overall feature combinations. The only solution to the previous expression providing some desirable properties (local accuracy, missingness, and consistency) is the following: $\phi_j = \sum_{Z' \subseteq X'} \frac{|Z'|!(P-|Z'|-1)!}{P!}\left[f(Z') - f(Z' \setminus X_j)\right]$, where P is the number of features, Z' is a subset of X', and $Z' \setminus X_j$ is Z' when excluding feature X_j. This solution satisfies a set of properties, and allows the model to match the output of f for the simplified input x'. According to,[20] ϕ_j can be estimated by posing $f(Z') = \mathbb{E}\left[f(Z)|Z_S\right]$, where S is the set of non-zero indices in Z', known as SHAP values.

iBreak down

Break down is a model agnostic tool that essentially describes the contributions of each variable to the final prediction of a model. iBreakDown[21] is a successor of the breakDown package that is able to capture local interactions and generates non-additive explanations with interactions visualized by waterfall plots. While the SHAP values average over all possible orderings leading to additive contributions, iBreakDown analyzes the different orders to identify interactions in the model.[21] proved that the SHAP value is an average over Break Down contributions for all possible ordering of variables. Considering a model $f(X)$ and an instance of interest x_*, we denote the contribution from feature X_S as Δ_S and the joint contribution from the pair of features (x_S, x_Z), where $\Delta_S = \mathbb{E}\left[f(X)|x_S = x_{*S}\right] - \mathbb{E}\left[f(X)\right]$ and $\Delta_{SZ} = \mathbb{E}\left[f(X)|x_S = x_{*S}, x_Z = x_{*Z}\right] - \mathbb{E}\left[f(X)\right]$. The interaction contribution for each pair of features (x_S, x_Z) is therefore $\Delta^I_{SZ} = \Delta_{SZ} - \Delta_S - \Delta_Z$.

5 | RESULTS

In this section, we set up a regression model to predict the profitability of a company by including the global ESG score among the predictors. We consider the EBIT (which is expressed in Million Euros throughout the paper) as a measure of

TABLE 6 R^2 values.

Model	DT	BAG	RF	GB	GLM
R^2	73.18%	87.90%	**88.39%**	88.36%	78.03%

Note: The bold values indicate the best performance.

TABLE 7 Root mean square error (RMSE) and mean absolute error (MAE) of earnings before interest and taxes (EBIT) predicted values.

Model	DT	BAG	RF	GB	GLM
RMSE-train	1808	2023	1980	897	2330
MAE-train	727	831	823	541	1179
RMSE-test	2580	2145	**2102**	2104	2891
MAE-test	1003	844	**831**	965	1284

Note: The bold values indicate the best performance.

the firm's profit that, as the name suggests, represents the profit before taking into consideration the amount of interest and taxes paid for by the company. We provide a comparison of the outcomes under the traditional statistical technique of GLMs, machine learning approach (Decision trees), and ensemble methods (Bagging, Random forest, and Gradient Boosting). To ensure algorithmic fairness and to identify potential bias/problems in the training data, we offer explanations through the main suitable methods and metrics of machine learning interpretability. They help to mean the internal logic and inner workings of the proposed models hidden to the user, to understand the rationale behind their predictions fully. In particular, we implement the previously described model-agnostic methods that allow harnessing the predictive power of machine learning models while gaining insights into the black-box model. The main results show a higher contribution to the company's profitability as the ESG score increases.

5.1 | Model's prediction performance

The prediction performance of each model is evaluated according to the R-squared (see Table 6) and traditional error measures, such as the root mean square error (RMSE) and the mean absolute error (MAE), which are reported in Table 7 for both the train (80% of the data) and the test sample (20% of the data). Overall, RF algorithm provides the highest capacity to predict EBIT ($R^2 = 88.39\%$), closely followed by GB. Our results reported in Tables 6 and 7 support the common finding that ensemble methods (BAG, RF, GB) outperform individual techniques (DT). Figure 2 shows the density function of the observed values compared to the density function of the values predicted by machine learning algorithms and GLM. RF (red curve) provides the best fitting, followed by GB (green curve) that shows a very similar prediction's performance. However, these two algorithms work differently, GB best catches the expected value of the observations, while RF best captures EBIT higher values. GLM seems unbiased as regards the expected value of the observations. Indeed, the data show a remarkable positive asymmetry that is not well grasped by the linear regression.

In Figure 3, we depict the variable importance according to the best model, the RF. As we expected, the most important variable in explaining EBIT is *Net.Sales*, followed by *ROE*, and then by the *ESG.Score*. We are interested in understanding how the ESG score affects the company's profitability. That is, while some strategies that involve higher ESG scores may positively determine a firm's profit, other investment styles which correspond, on the contrary, to lower ESG scores may not be necessarily value-adding, but rather only burden the firm with extra costs.

5.2 | Model-agnostic methods for the interpretability of the prediction results: A focus on the ESG score

In this section, we deal with the interpretability of the results by using the model-agnostic methods previously described. We focus the analysis on the predictions provided by RF that showed the best performance on our dataset.

In Figure 4, we illustrate the PDP for the three main predictors, *Net.Sales*, *ROE* and *ESG.Score*. The PDP for the net sales shows an increasing trend, as well as for the ROE predictor, which reaches a plateau. For the PDP of ESG score, we find that as a firm's ESG score increases, its profitability declines at first, reaching minimum values at intermediate ESG

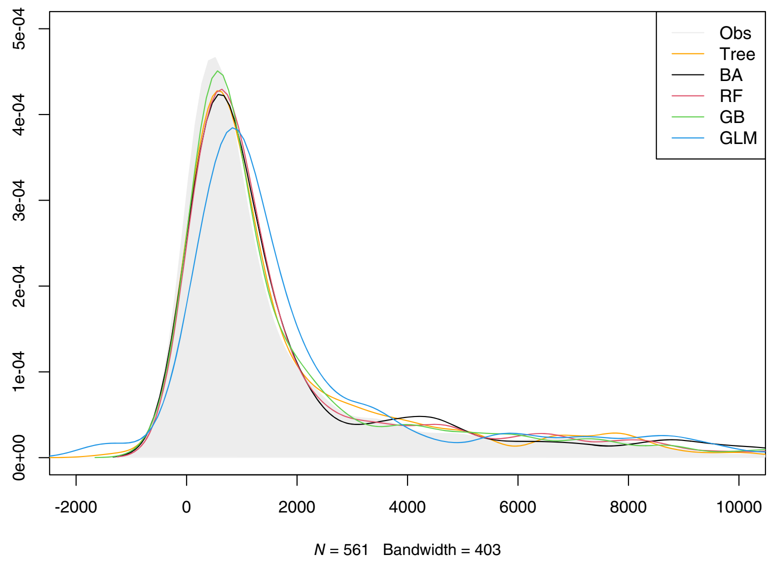

FIGURE 2 Density functions of observed values and models' estimated values.

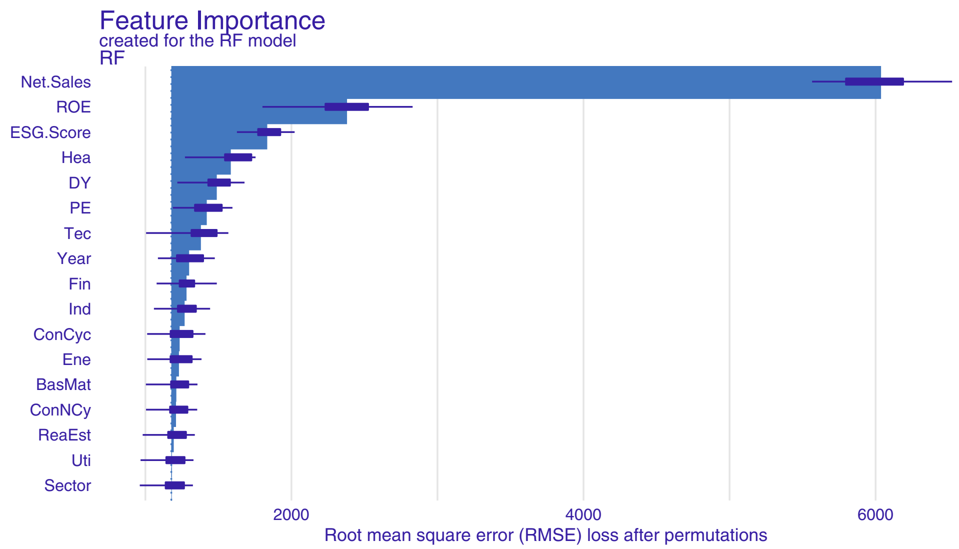

FIGURE 3 Variable importance according to the RF model.

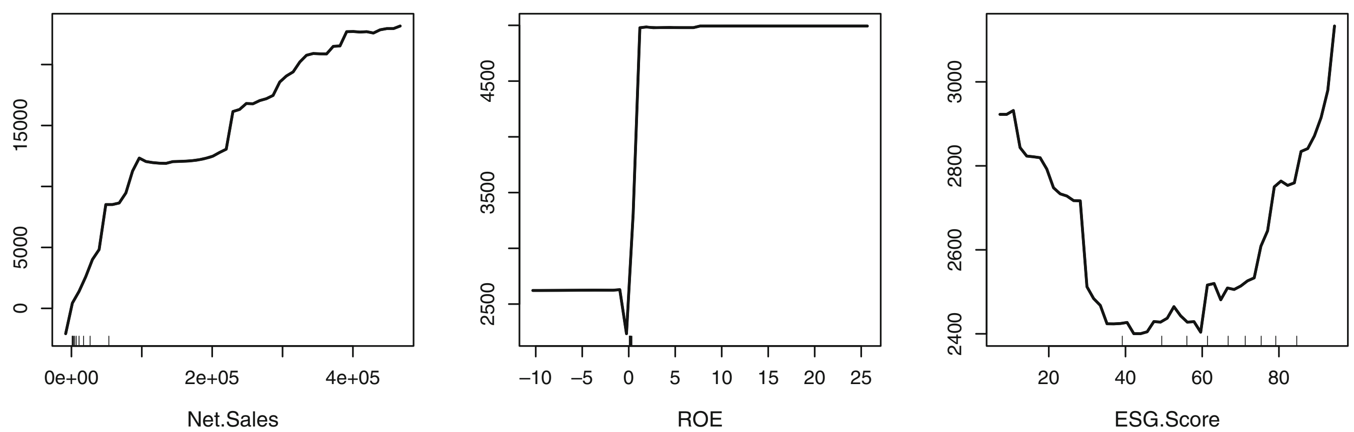

FIGURE 4 Partial dependence plots (PDP) for the main predictors: *Net.Sales*, *ROE*, and *ESG.Score*.

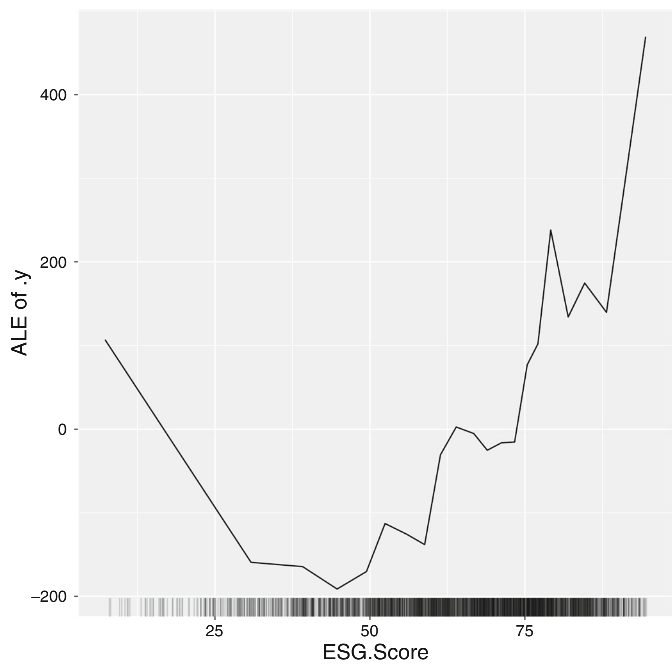

FIGURE 5 Accumulated local effects (ALE) plot for the earnings before interest and taxes (EBIT) prediction model by the ESG score.

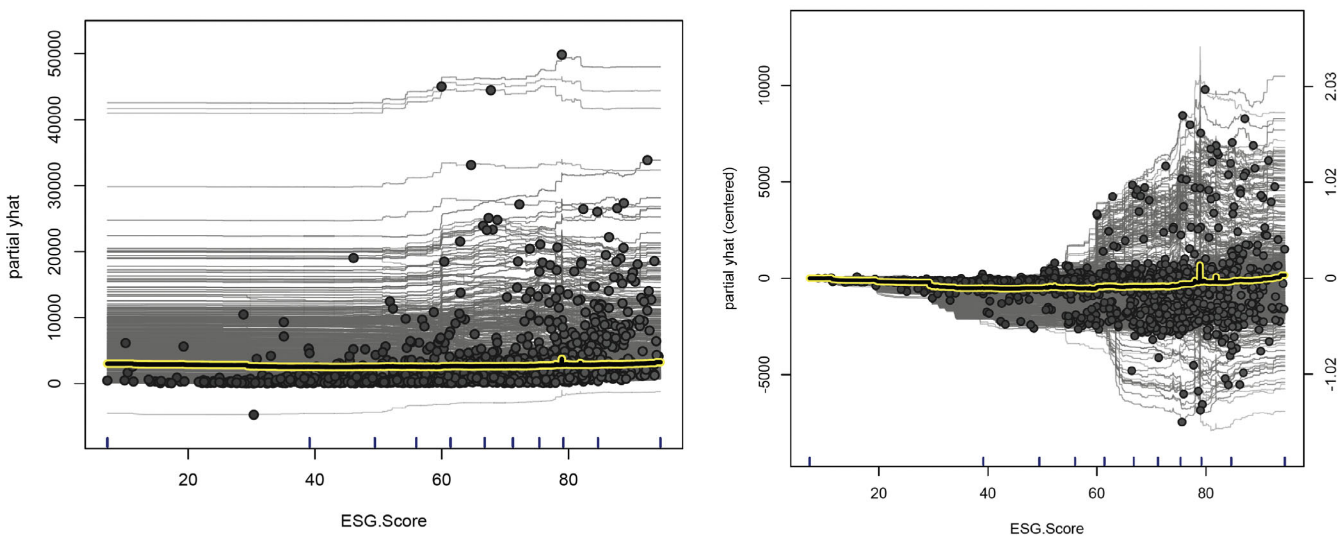

FIGURE 6 Individual conditional expectation (ICE) plot (left panel) and centered ICE plot (right panel) for the ESG score. The yellow line represents the partial dependence plots (PDP) of the ESG score. The right vertical axis of the right panel displays changes in the fitted model over the baseline as a fraction of the target variable's observed range.

score levels (the minimum is reached at an ESG score of 43), and then increases continuously until it reaches a maximum ESG score of 95. However, the U-shaped relationship is not symmetrical. Those firms with the highest ESG scores have significantly higher EBIT values than firms with the lowest ESG scores. A similar picture has been obtained by Reference 22 using different data. Therefore, our results suggest that it is more profitable for a firm to be highly socially responsible than only partially recurring to socially responsible investments. Moreover, the EBIT value is higher for firms with an ESG score of, for example, 30 than for firms with an ESG score of 40. The U-shape of the PDP of the ESG score could confirm that the insight of intermediate ESG scores (range of 40–60) may not be value-adding but rather charging the company with other expenses.

Figure 5 provides the ALE plot for the EBIT prediction model by the ESG score, obtained using the R package *ALEPlots*.[23] Marks on x-axis indicate the ESG score distribution, showing how relevant a region is for interpretation. Overall,

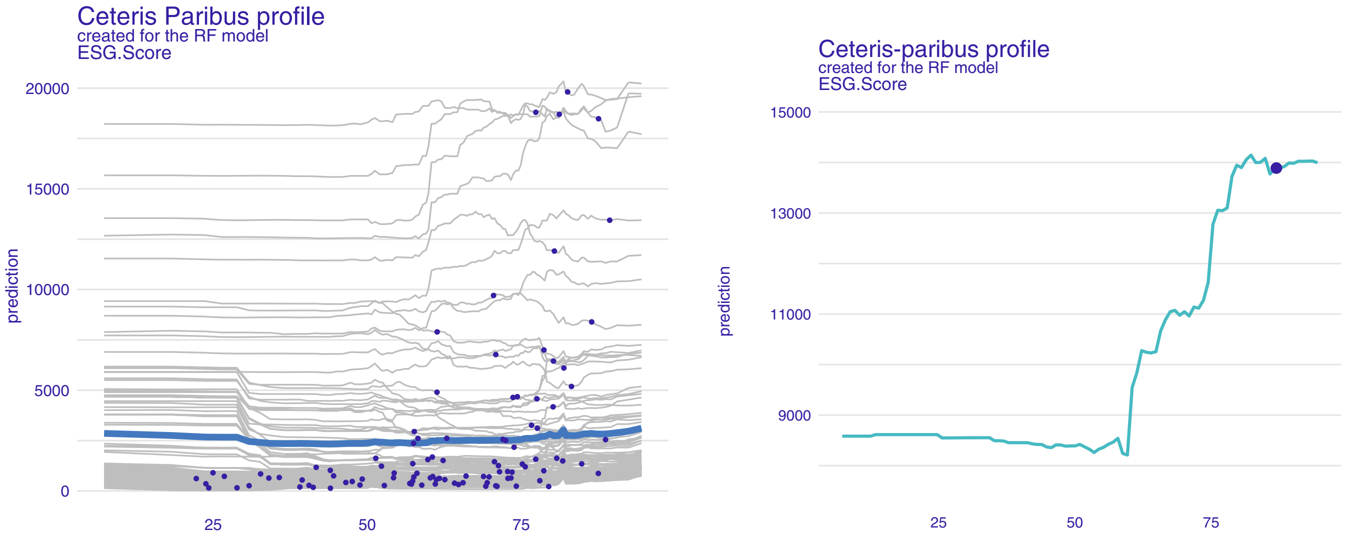

FIGURE 7 Ceteris-paribus (CP) interpretation; feature: *ESG.Score*. Left panel: CP profiles (grey lines) for 100 randomly selected observations (dark blue dots). The blue line shows the mean of the CP profiles, which offers an estimate of the PD profile. Right panel: The CP profile (turquoise) of a single observation (dark blue dot) with the following features: *year* = 2015, *EBIT* = 13,890, *ESG.Score* = 86.97, *Net.Sales* = 89,469, *PE* = 26.18, *ROE* = 0.17, *DY* = 0.03.

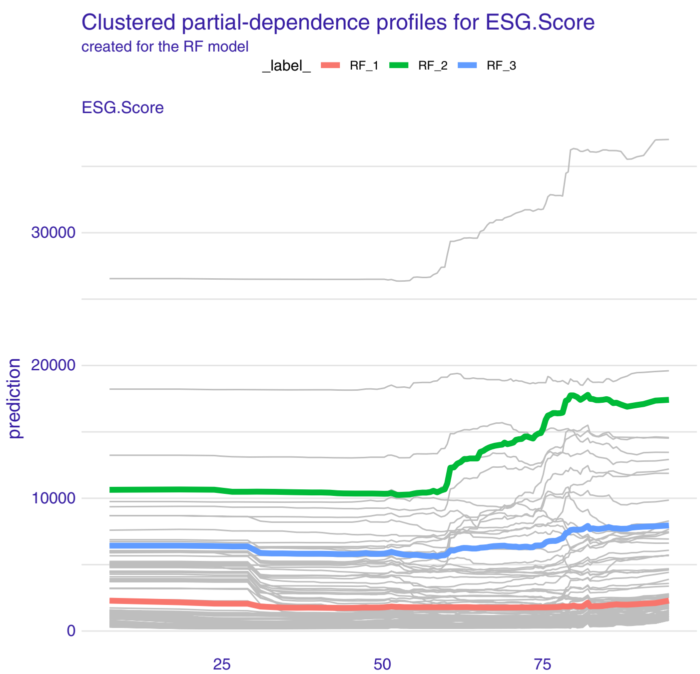

FIGURE 8 Clustered partial-dependence profiles for the *ESG.Score*.

we can see that the ESG score has a relevant influence on the EBIT prediction. Region 50–85 of the ESG score, where the EBIT prediction rises with increasing ESG score, is the most relevant for interpretation. In the region 0–30 of the ESG score, the EBIT prediction decreases with increasing ESG score.

In Figure 6, we depict the ICE plot (left panel) and the centered-ICE plot (right panel) for the *ESG.Score* feature. Generally, ICE plots highlight the variation in the fitted values across the range of a feature, suggesting where and to what extent heterogeneities might exist.[17] Each of the grey lines represents the conditional expectation for a single observation (the point from which the curve originates). We limit the ICE curves to 60% of the observations to not overcrowd the resulting plot. From the left panel of Figure 6, we note that EBIT values show a differentiation over the range 60–90 of the ESG score. The centered-ICE plot, reported in the right panel of Figure 6, sets the individual ICE lines to 0 at ESG score

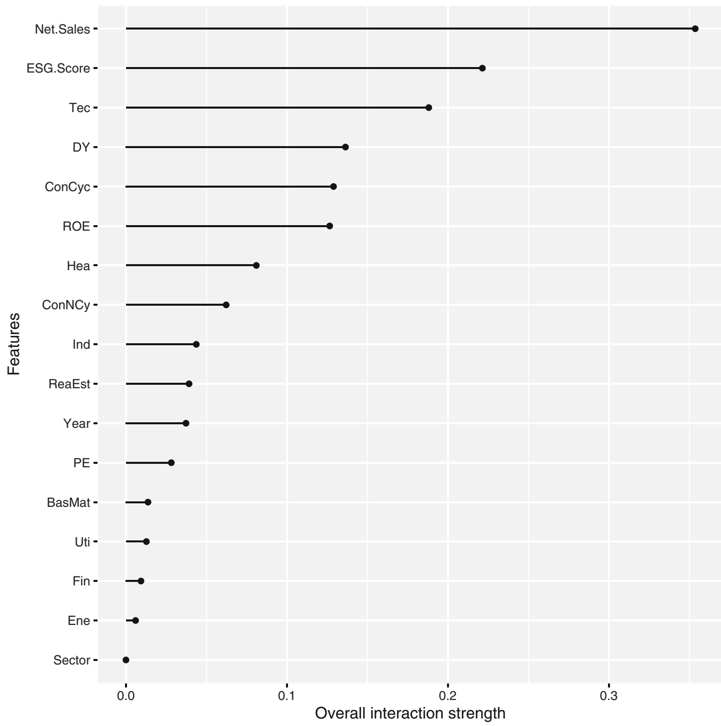

FIGURE 9 Feature interaction–Each of the input features with all other features for predicting earnings before interest and taxes (EBIT) values.

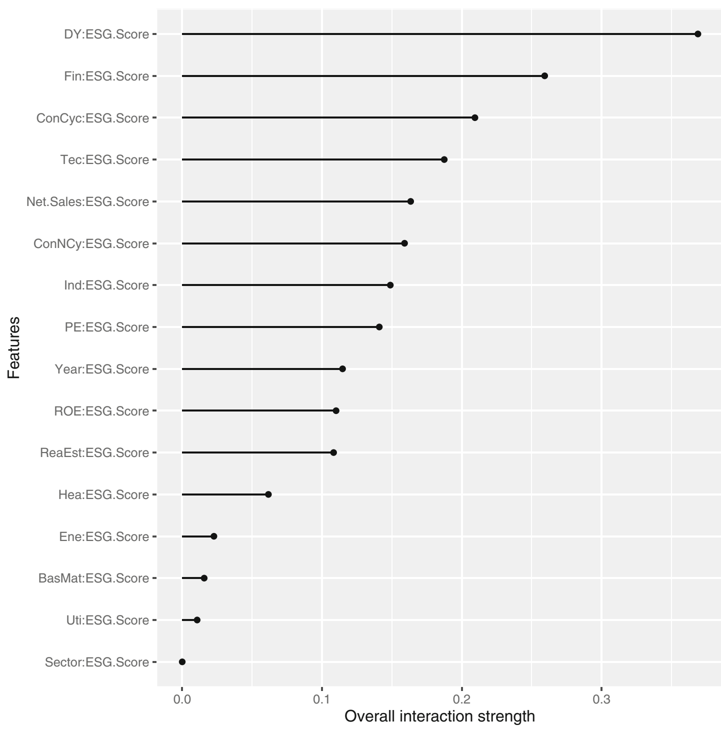

FIGURE 10 Two-way *ESG.Score* interactions with the other features in predicting earnings before interest and taxes (EBIT) values.

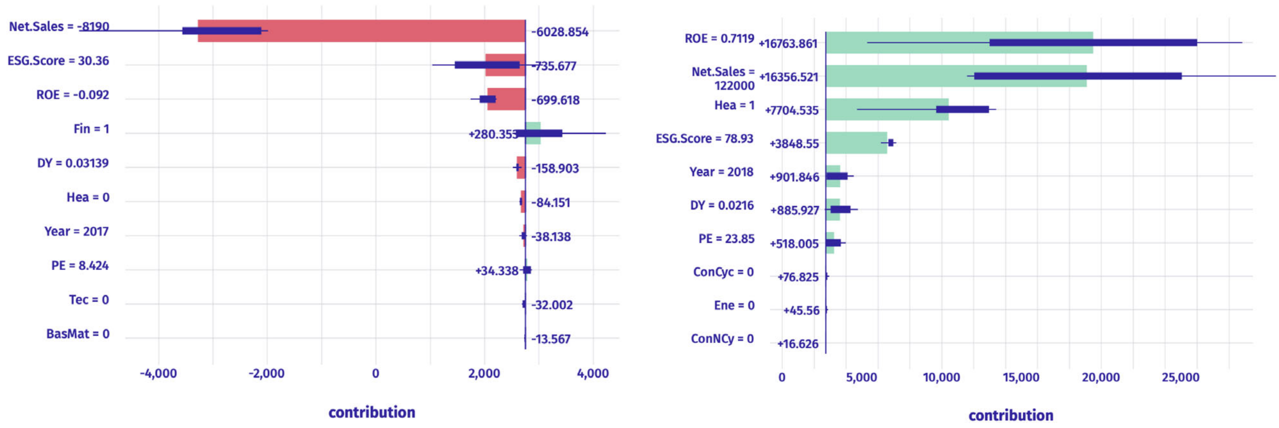

FIGURE 11 Shapley additive explanations (SHAP) values. Data points: EBIT = −8311 (left), EBIT = 53,683 (right). Red (green) bars show a negative (positive) contribution of the predictors.

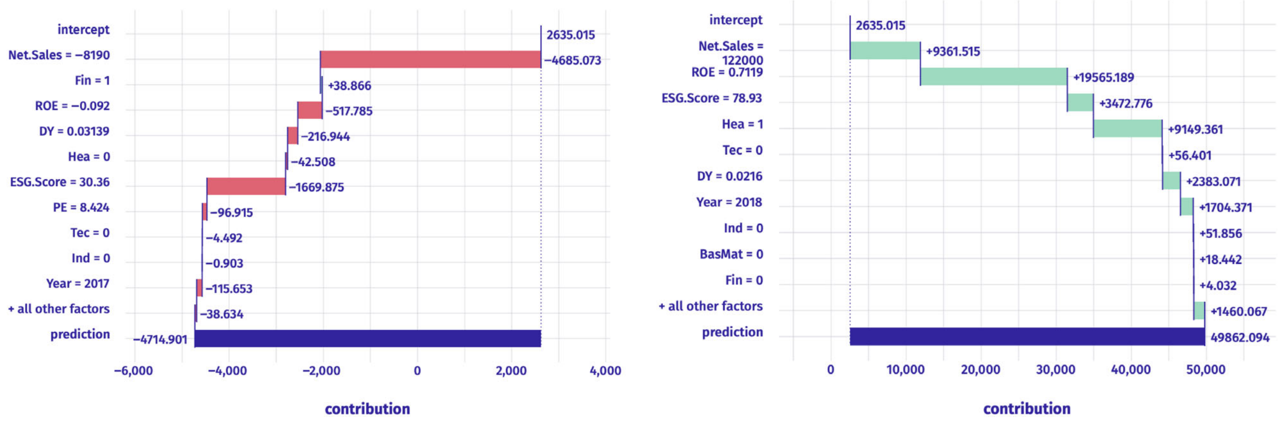

FIGURE 12 Break-down plot with interactions. Data points: EBIT = −8311 (left), EBIT = 53,683 (right). The blue bar shows the difference between the model's prediction for the selected observation and the average model prediction. Other bars show the contributions of variables. Red (green) bars show a negative (positive) contribution of the variables. The order of variables on the y-axis corresponds to their sequence.

0, favoring the comparisons across the different ICE lines. The predictions for most of the constituents of the Euro-Stoxx 600 Index remain unchanged until the ESG score is lower than 60. For ESG score values higher than 60 we have different dynamics of the profitability of the firms included in our sample; in some cases the profitability sharply increases in others decreases.

The left panel of Figure 7 presents CP profiles for the explanatory variable *ESG.Score* for 100 randomly selected observations from our dataset. As introduced in Section 4, CP Profiles are aimed to show model predictions around selected points in the feature space. On the axis of abscissas, we place the ESG score values, while, on the axis of ordinates, the prediction EBIT value of observations where only the value of the ESG score changes (all the other features remain unchanged). Overall, we note that profiles are not parallel, indicating non-additive effects of explanatory variables. It is worth observing that the profiles are step functions with some variability. Part of the profiles suggests an approximately linear relationship between the ESG score and the predicted EBIT value. The blue line shows the mean of the CP profiles, which offers an estimate of the PD profile. To better understand how the CP technique works, one may analyze the model predictions around a single instance of interest. For example, we consider the observation represented by a company operating in the Consumer Non-Cyclicals sector, which, in the year 2015, shows an EBIT value of 13,890, an ESG score of 86.97, and the following other features: Net Sales = 89,469, PE = 26.18, ROE = 0.17, DY = 0.03. From the right panel of Figure 7, we observe that if the company had an ESG score of less than 60, its predicted EBIT would be constant. If the company had an ESG score higher than 60 but less than 75, its predicted EBIT would increase by 26%. And if its ESG score were higher than 75, the EBIT prediction would remain almost constant.

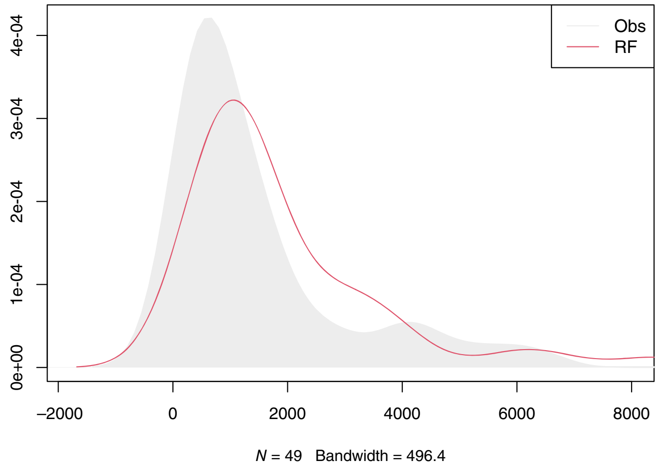

FIGURE 13 Density functions of observed values and models' estimated values. Year 2020.

TABLE 8 Root mean square error (RMSE) and mean absolute error (MAE) of earnings before interest and taxes (EBIT) predicted values.

Model	RF
RMSE-train	4179
MAE-train	1868
RMSE-test	5954
MAE-test	2321

Note: Year 2020.

The average value of CP profiles is a good summary if profiles are parallel. If not, we can cluster the profiles and calculate the average separately for each cluster. Figure 8 illustrates the clustered partial-dependence (PD) profiles for the *ESG.Score*. Profiles could be split into three clusters: one for a group of firms with a remarkable increase in the predicted EBIT for an ESG score higher than 60 (with the average represented by the green line), one with a slight increase of the predicted EBIT for an ESG score higher than 60 (with the average represented by the blue line), and one with almost constant predicted EBIT values (with the average represented by the red line).

Figure 9 provides the measure of how strongly the features interact with each other in predicting EBIT values. The net sales have the highest interaction effect with all other features, followed by the ESG score. The feature interaction tool measures how much of the variance of the model's estimated target variable is explained by the interaction. The interaction of *Net.sales* with the other features explains about 40% of variance of the estimated EBIT values, while that of *ESG.Score* about 22%.

In Figure 10, we illustrate how much the feature *ESG.Score* interacts with any other feature. We find that the most important interaction of the *ESG.Score* is with the *DY*, followed by *financial* sector.

In the following, we show the SHAP attributions and the break-down plots related to the model's prediction. They show which variables are most important for a specific instance. Figure 11 illustrates the SHAP attributions and Figure 12 the break-down plots with interactions for two different data points: the first one corresponding to a negative EBIT value (−8311) and the second one to a high positive value (53,683). From the left panel of Figure 11, we can observe that the most important variable is *Net.Sales* (= −8190) that decreases the EBIT prediction by 6029. The second most important variable is *ESG.Score* (= 30.36) that increases the EBIT prediction by 736. The third most important variable is *ROE* (= −0.092) that decreases the EBIT prediction by 700. The average contribution of all the variables depicted in the figure is significant. Looking at the right panel of Figure 11, we find that the most important variable is *ROE* (= 0.71) that increases the prediction by 16,764. The second most important variable is *Net.Sales* (= 122,000) that increases the prediction by 16,357. The third most important variable is *Hea* (= 1) that decreases the prediction by 7,705. Also *ESG.Score* (=78.93) is noteworthy, as it increases the EBIT prediction by 3849. Note that the object of the SHAP function can be reused to explain all the data points.

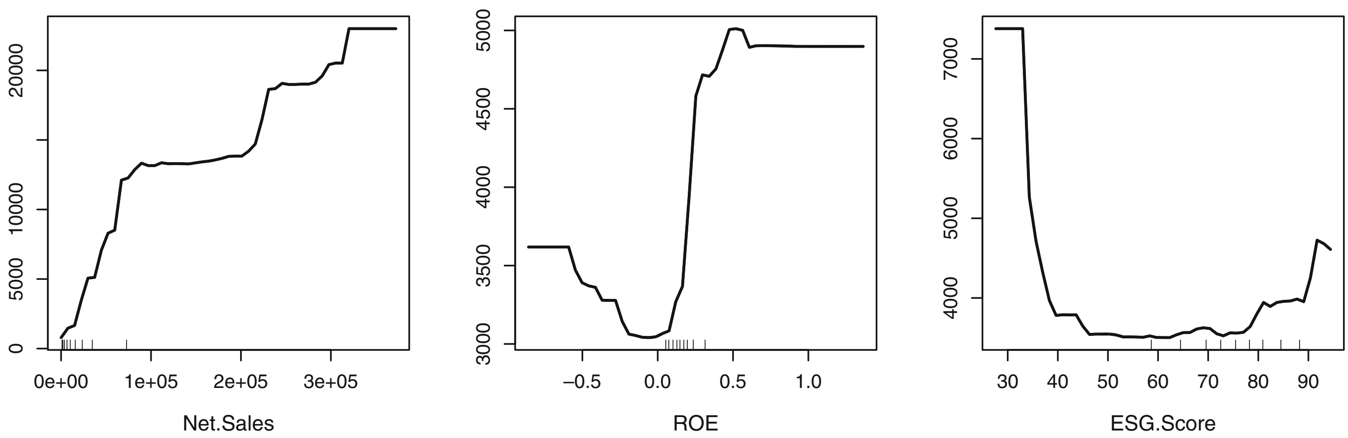

FIGURE 14 Partial dependence plots (PDP) for the main predictors: *Net.Sales*, *ROE*, and *ESG.Score*. Year 2020.

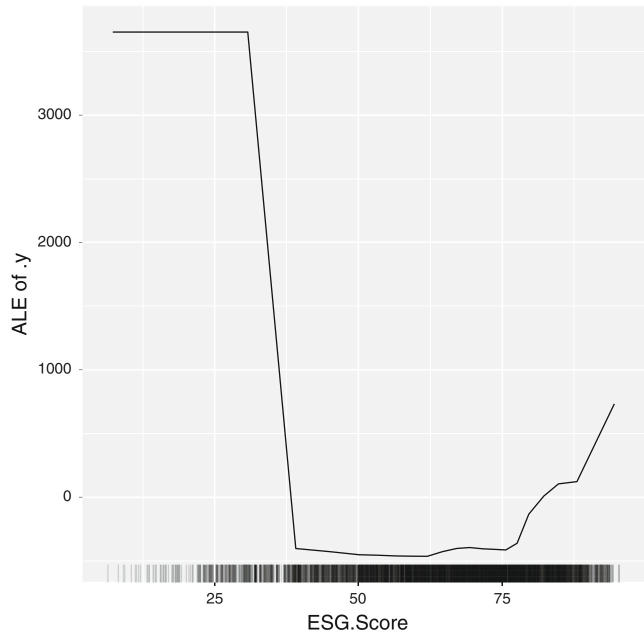

FIGURE 15 Accumulated local effect (ALE) plot for the earnings before interest and taxes (EBIT) prediction model by the ESG score. Year 2020.

Relating the break-down plots, the left panel shows that RF predicts for the selected data point (EBIT = −8311) a value equal to about −4715, which is lower than the average model prediction (2635). The most important variable is *Net.Sales* (= −8190) that decreases the EBIT prediction by 4685. The second most important variable is *ESG.Score* (= 30.36) that decreases the EBIT prediction by 1670. The third most important variable is *ROE* (= −0.092) that decreases the prediction by 518. The contribution of the other variables is less important. The right panel shows that RF predicts for the selected data point (EBIT = 53,683) a value equal to about 49,862, which is higher than the average model prediction (2635). The most important variable is *ROE* (= 0.71) that increases the EBIT prediction by 19,565. The second most important variable is *Net.Sales* (= 122,000) that increases the EBIT prediction by 9362. The third most important variable is *Hea* (= 1) that decreases the prediction by 9149. The contribution of the other variables is less important.

5.3 | Focus on the year 2020 data

In 2018 the European Commission started implementing the Action Plan on sustainable growth.[24] Since that date, sustainability regulations are becoming stricter. Therefore, it may be worth repeating part of the analysis (only the RF model, and PD and ALE plots) by considering only data from the year 2020. The RF algorithm applied to the year 2020 data shows much less predictive power than the entire dataset (the R^2 value is 55.43% compared to 88.39%). In Figure 13, we show the density functions of the observed and the predicted values that highlight the model's low accuracy for predicting EBIT. The level of the prediction errors reported in Table 8 is definitely higher than values in Table 7. Machine learning is generally used on large datasets, while the size of the year 2020 sample (10% of the original dataset) is probably too small to obtain good predictive performance. In particular, the year 2020 sample collects only 5.9% of firms with an ESG score smaller than 49 (of which 0.4% is smaller than 30), while 38.6% is in the 50–74 range and 55.5% higher than 75.

However, taking into account this important limitation, we observe a peculiar behavior of the relationship between EBIT and ESG score in the year 2020 data (see the right panel in Figure 14). Though the shape of the PDPs of the ESG score for the year 2020 data and the entire dataset differ, we note that, for scores higher than 75, the EBIT value rises with increasing ESG scores in both datasets. In the complete dataset, we find similar behavior in the region 50–75 of the ESG score. While, in the case of 2020 data, we observe that an increase in ESG score does not affect the EBIT value, which remains steady. The PDPs for net sales and ROE (Figure 14, left and central panels) are very similar to the corresponding ones from the whole dataset (Figure 4, left and middle panels).

Looking at the ALE plot, depicted in Figure 15, we observe that region 50–85 of the ESG score remains the most relevant for interpretation also in the analysis only based on the year 2020. Focusing on the results from the year 2020 data, we can speculate that to increase the EBIT value company has to be heavily involved in sustainable investments. The insight that low ESG scores may not improve the profit margin seems to be committed.

6 | CONCLUDING REMARKS

Investors are paying increasing attention to the ESG factors, as there is wide recognition that companies with good sustainable credentials are more likely to outperform. In our analysis, we focus on the role of the ESG score on the firms' profitability and not only on the financial performance.

High firm profitability will translate into better financial performance and therefore provide interesting outcomes for investors and asset managers. We find that the ESG score has an impact on the firm's profitability measured by the EBIT of the company.

Precisely we show that to have an impact on the EBIT, the company has to be quite active toward sustainability and invest to change the business model to comply with ESG criteria. This translates into higher ESG scores, usually higher than 60 according to Refinitiv ESG score.

Companies with low ESG score, can be considered less committed toward the sustainability goal and make weak efforts in binding ESG elements into an investment strategy, this does not create an extra profit margin as highlighted by our results.

DATA AVAILABILITY STATEMENT

Research data are not shared.

ACKNOWLEDGMENT

Open Access Funding provided by Universita degli Studi di Roma La Sapienza within the CRUI-CARE Agreement.

ORCID

Susanna Levantesi https://orcid.org/0000-0002-4644-4358

REFERENCES

1. Cerqueti R, Ciciretti R, Daló A, Nicolosi M. ESG investing: a chance to reduce systemic risk. *J Finan Stab*. 2021;54:100887.
2. Cerqueti R, Ciciretti R, Daló A, Nicolosi M. Mitigating contagion risk by ESG investing. *Sustain*. 2022;14:3805.
3. GSIA. *Global Sustainable Investment Review 2014*. GSIA; 2014 Available: http://www.gsi-alliance.org/members-resources/global-sustainable-investment-review-2014/
4. GSIA, *Global Sustainable Investment Review 2020*. GSIA; 2021. Available: https://www.gsi-alliance.org/wp-content/uploads/2021/08/GSIR-20201.pdf
5. Naffa H, Fain M. Performance measurement of ESG-themed megatrend investments in global equity markets using pure factor portfolios methodology. *PLoS ONE*. 2020;15:e0244225. doi:10.1371/journal.pone.0244225
6. Moskowitz M. Choosing socially responsible stocks. *Bus Soc Rev*. 1972;1:71-75.
7. Bragdon JH, Marlin J. Is pollution profitable. *Risk Manag*. 1972;19:9-18.

8. Bowman EH, Haire M. A strategic posture toward corporate social responsibility. *California Manag Rev*. 1975;18:49-58. doi:10.2307/41164638
9. ESAs. Joint Consultation Paper on ESG Disclosures. Draft regulatory technical standards with regard to the content, methodologies and presentation of disclosures pursuant to Article 2a, 4(6) and (7), Article 8(3), Article 9(5), Article 10(2) and Article 11(4) of Regulation (EU) 2019/2088' JC (2020) 16. The final report of the joint ESAs was published on 2 February 2021 (JC 2021 03); 2020.
10. Morana C, Sbrana G. Climate change implications for the catastrophe bonds market: an empirical analysis. *Econ. Model.* 2019;81:274-294.
11. Monasterolo I. Climate change and the financial system. *Ann Rev Resour Econ*. 2020;12(1):299-320.
12. Thomson Reuters. *Environmental, Social and Governance (ESG) Scores*. Refinitiv; 2020 Available from Refinitiv at: https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/refinitiv-esg-scores-methodology.pdf
13. Thomson Reuters. *TRBC Sector Classification*. Refinitiv; 2020. Available from Refinitiv at: https://www.refinitiv.com/content/dam/marketing/en_us/documents/quick-reference-guides/trbc-business-classification-quick-guide.pdf
14. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5-32.
15. Friedman J. Greedy function approximation: a gradient boosting machine. *Ann Stat*. 2001;29:1189-1232.
16. Apley D, Zhou J. Visualizing the effects of predictor variables in black box supervised learning models. *J. R. Stat. Soc. B*. 2020;82:1059-1086.
17. Goldstein A, Kapelner A, Bleich J, Pitkin E. Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation. *J. Comput. Gr. Stat.* 2015;24:44-65.
18. Biecek P. DALEX: explainers for complex predictive models in R. *J Mach Learn Res*. 2018;19(84):1-5.
19. Friedman JH, Popescu BE. Predictive learning via rule ensembles. *Ann Appl Stat*. 2008;2(3):916-954.
20. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*. Vol 30. Curran Associates; 2017 https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf
21. Gosiewska A, Biecek P. iBreakDown: uncertainty of model explanations for non-additive predictive models; 2019 https://arxiv.org/abs/1903.11420v1
22. Barnett ML, Salomon RM. Does it pay to be really good? Addressing the shape of the relationship between social and financial performance. *Strat Manag J*. 2012;33:1304-1320.
23. Apley D. ALEPlot: accumulated local effects (Ale) plots and partial dependence (Pd) plots; 2018 https://CRAN.R-project.org/package=ALEPlot
24. European Commission. *Action Plan: Financing Sustainable Growth*. European Commission; 2018. Retrieved from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0097

How to cite this article: D'Amato V, D'Ecclesia R, Levantesi S. Firms' profitability and ESG score: A machine learning approach. *Appl Stochastic Models Bus Ind*. 2024;40(2):243-261. doi: 10.1002/asmb.2758