Firms' profitability and ESG score: a machine learning approach

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

[1]University of Salerno
[2]Sapienza University of Rome

July 24, 2022

Abstract

Corporate social responsibility (CSR) has a potential impact on firms performance, for instance enhancing firm reputation, increasing innovation capabilities, customer loyalty and customer satisfaction could help improve financial performance. However, the literature provides only limited evidence of the relationship between non-financial 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 deepen the relationship between ESG score and EBIT through machine learning interpretability toolboxes such as partial dependence plots and individual conditional expectation, which help to visualize the functional relationship between the predicted response and one or more features, and the Shapley value allowing 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, Machine Learning, Interpretability tools.

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 increasingly 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.

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 GSIA (2014). 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 (GSIA, 2021).

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 1,200 per cent increase in total Assets Under Management(AUM) between 2012 and 2018,by reaching $1,018 million by the end of 2018, besides being the youngest ESG strategy.
The thematic approach is about identifying a 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" (Naffa and Fain, 2020).
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 Moskowitz (1972); Bragdon and Marlin (1972); Bowman and Haire (1975). 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 Naffa and Fain (2020) 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.

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 (Morana and Sbrana (2019); Monasterolo (2020)). The sudden changes in climate change has fostered the introduction of new policies and regulation, this generates mispricing of climate affecting asset price volatility and financial stability (Monasterolo, 2020). 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, for instance, 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 (Monasterolo, 2020).

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 countries of the European region. 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 covers 422 companies (about 70% of the total) that have been enclosed in the index throughout 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[1] 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 weights of the three pillars being 34% for E, 35.5% for S and 30.5% for G (Thomson Reuters, 2020a). Companies are classified according to the Thomson Reuters Business' Classification that is an owned industry classification system operated by Thomson Reuters (Thomson Reuters, 2020b). The industry sector proportions related to our dataset are shown in Table 1. We observe that about 50% of the analyzed companies belongs to the financials, industrials, and consumer cyclicals sectors.

[1]Environmental: resource use, emissions, innovation; Social: workforce, human rights, community, product responsibility; Governance: management, shareholders, CSR strategy.

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 1: Industry sectors' proportion of the dataset

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 aggregate ESG score of the companies included in our sample in the years 2011-2020 is 64.27.

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+	

Table 2: Conversion from a percentile score to a letter grade. Source: Refinitiv ESG.

In Table 3, we provide a description of the variables included in our model:

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 3: Variables' description

In Table 4, we report the yearly mean values of the ESG score by economic sectors. We note that firms belonging to the Energy sector have the highest ESG score, while those in the financial sector show the lowest ESG score. Overall, we observe that the average ESG score increased about 16 points from 2011 to 2020, moving from 57.72 to 73.86.

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

Table 4: Mean values of the ESG score by economic sectors. Years 2011-2020.

Fig. 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.

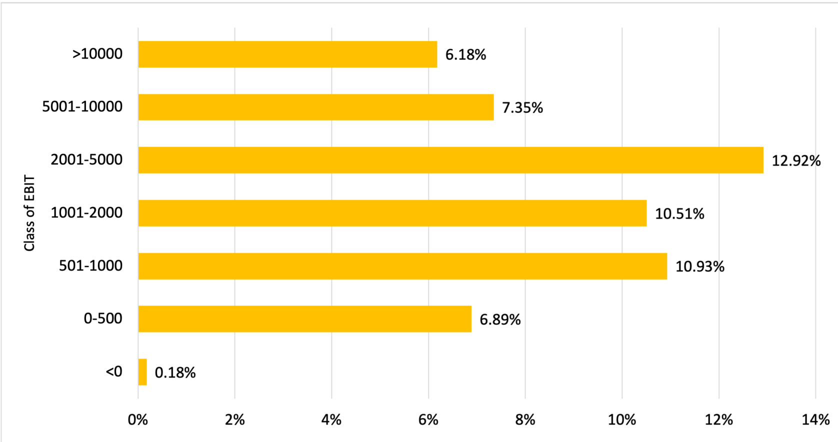

Figure 1: Percentage distribution of EBIT values (in Million Euros). Years 2011-2020.

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.

$EBIT$	$ESG.Score(mean)$
<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

Table 5: Mean values of the ESG score by classes of EBIT values (in Million Euros). Years 2011-2020.

3 Regression models

Given a set of features, $X_1, X_2, ..., X_p$ 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 the features $\mathbf{X}$:

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

where ϵ is the error term.
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, $R_1, R_2, ..., R_J$, 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 (Breiman, 2001). 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 Friedman (2001) 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. Each tree is calibrated on the results of the previous trees to improve the current fit.

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[2]: $EBIT \sim Year + Net.Sales + ESG.Score + Sector + PE + ROE + DY + I1 + I2 + I3$.

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 (PDP)

One of the most used model agnostic tool is the PDP proposed by Friedman (2001). 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.

Accumulated Local Effect plots (ALE)

The accumulated local effects (ALE) plot (Apley, 2020) 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.

Individual Conditional Expectation (ICE)

Goldstein et al. (2015) 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.

[2]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.

Ceteris-paribus (CP) 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. We use the CP profiles as implemented in the *DALEX* R package for R (Biecek, 2018).

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.

SHAP (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. We consider the unified framework based on the Shapley value proposed by Lundberg and Lee (2017), the Shapley value can split an individual prediction among all contributed features, providing a full explanation of why a given variabe has received a specific EBIT. SHAP (SHapley Additive exPlanations) is a relatively recent approach and compute the importance of a predictor by comparing what the model predicts with and without that predictor. SHAP averages across all possible combinations of variable contributions.

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 (Gosiewska and Biecek, 2019) 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. The authors proved that the SHAP value is an average over Break Down contributions for all possible ordering of variables.

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 as a measure of 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 by means of the main suitable methods and metrics of the machine learning interpretability. They help to meaning their internal logic and inner workings of the proposed models that are hidden to the user, in order to fully understand the rationale behind their predictions. In particular, we implement model-agnostic methods previously described that allow to harness 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 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 - 7 support the common finding that ensemble methods (BAG, RF, GB) outperform individual techniques (DT).

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

Table 6: R^2 values

Model	DT	BAG	RF	GB	GLM
RMSE-train	1,808	2,023	1,980	897	2,330
MAE-train	727	831	823	541	1,179
RMSE-test	2,580	2,145	**2,102**	2,104	2,891
MAE-test	1,003	844	**831**	965	1,284

Table 7: RMSE and MAE of EBIT predicted values.

Fig. 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.

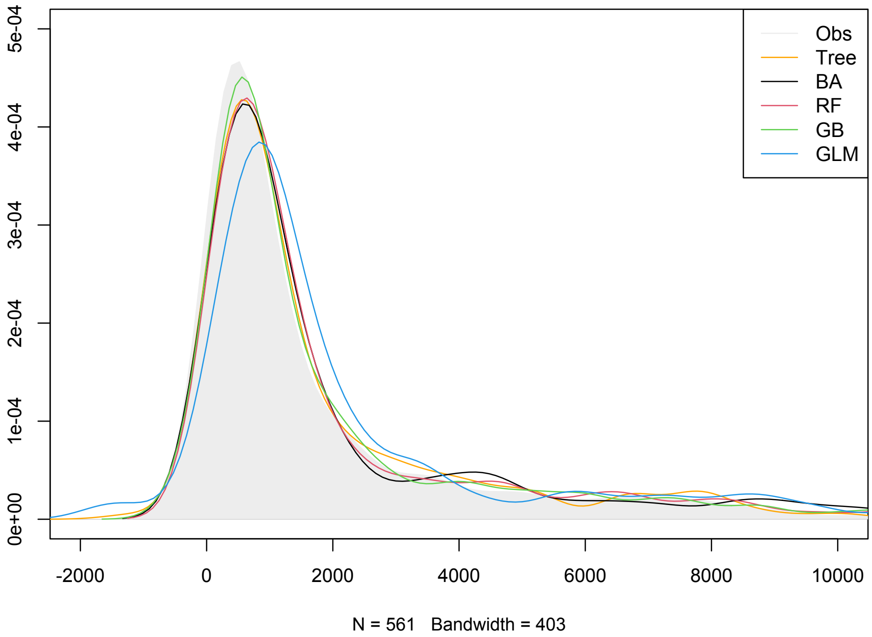

Figure 2: Density functions of observed values and models' estimated values.

In Fig. 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.

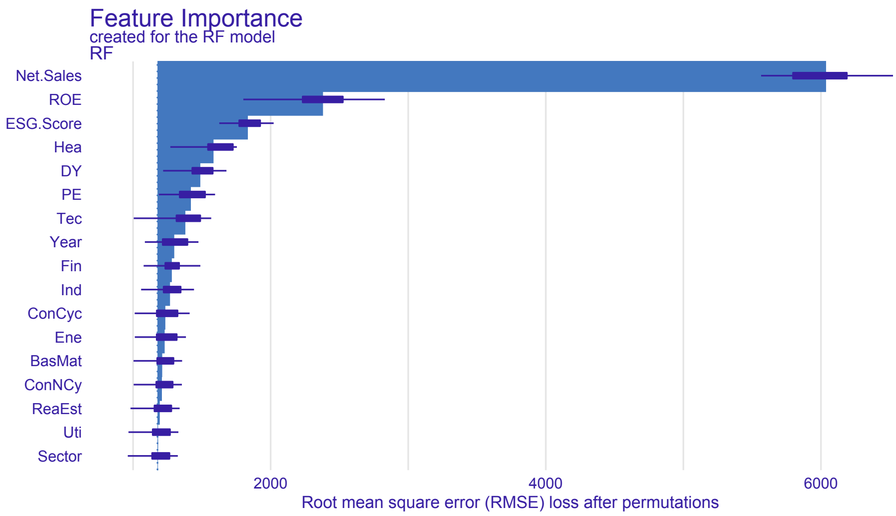

Figure 3: Variable importance according to the RF model.

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 Fig. 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. The U-shape of the PDP for the ESG.Score could confirm that the insight of lower ESG scores may not necessarily be value-adding, but rather charging the company with other expenses.

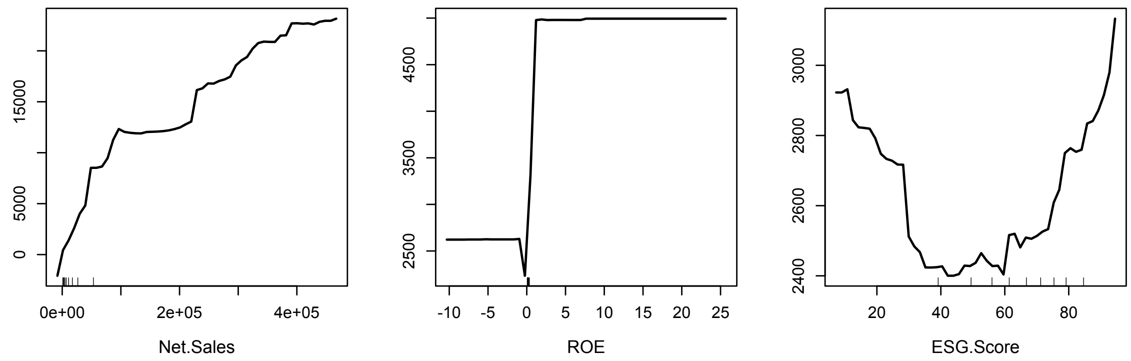

Figure 4: PDP for the main predictors: *Net.Sales*, *ROE* and *ESG.Score*.

Fig. 5 provides the ALE plot for the EBIT prediction model by the ESG score, obtained using the R package *ALEPlots* (Apley, 2018). Marks on x-axis indicate the ESG score distribution, showing how relevant a region is for interpretation. Overall, 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 the interpretation In the region 0-30 of the ESG score, the EBIT prediction decreases with increasing ESG score.

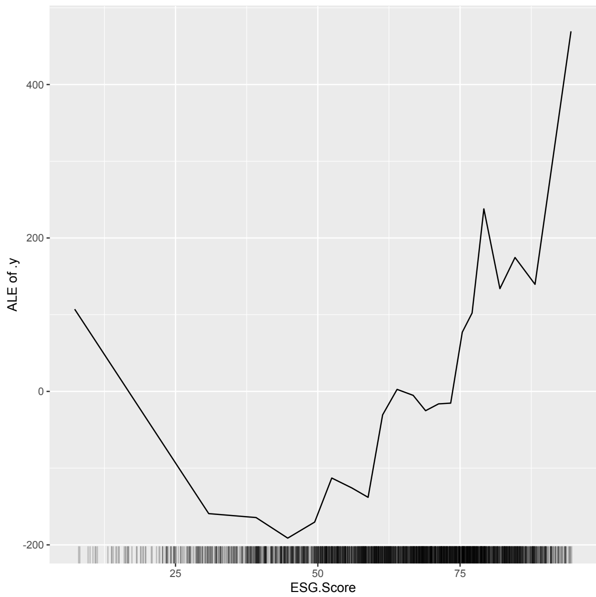

Figure 5: ALE plot for the EBIT prediction model by the ESG score.

In Fig. 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 (Goldstein et al., 2015). 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 Fig. 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 Fig. 6, sets the individual ICE lines to 0 at ESG score 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.

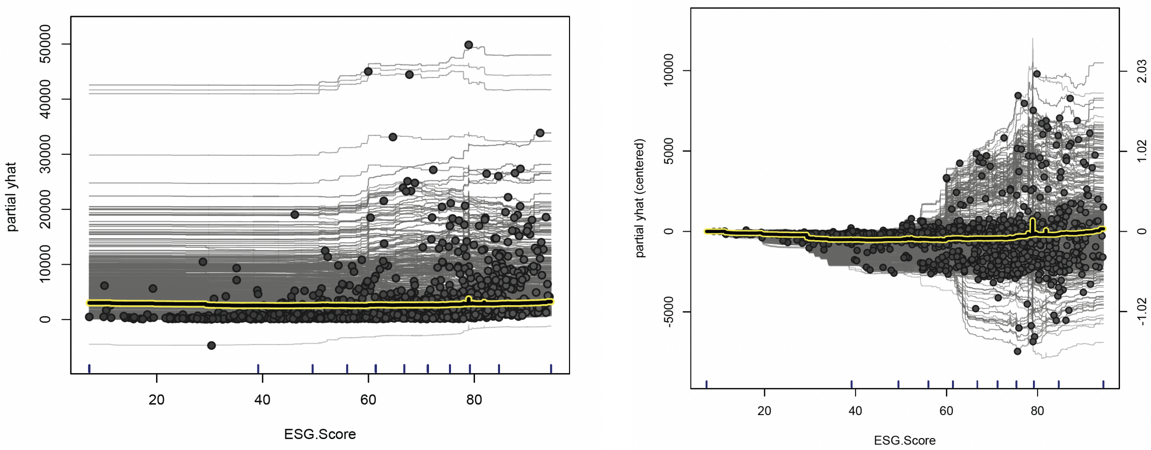

Figure 6: ICE plot (left panel) and centered ICE plot (right panel) for the ESG score. The yellow line represents the PDP of the ESG score. The right vertical axis of the right panel displays changes in the fitted model over the baseline as a fraction of the target variable's observed range.

Fig. 7 presents CP profiles for the explanatory variable *ESG.Score* for 100 randomly selected observations from our dataset. Overall, we note that profiles are not parallel, indicating non-additive effects of explanatory variables. 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. Its shape does not capture, for instance, the shape of the group of four CP profiles shown at the top of the panel.

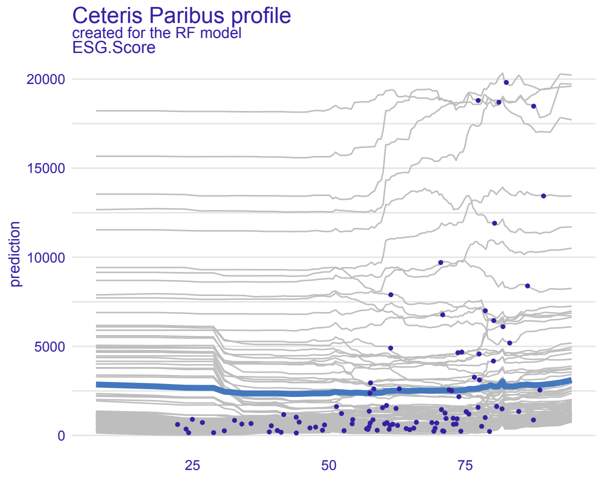

Figure 7: CP interpretation; feature: $ESG.Score$. Grey lines show the CP profiles for 100 randomly selected observations (dark blue dots). The blue line shows the mean of the CP profiles, which offers an estimate of the PD profile.

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. Fig. 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).

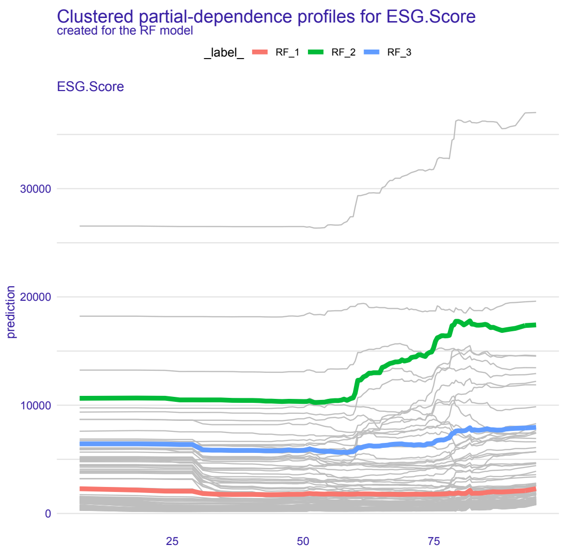

Figure 8: Clustered partial-dependence profiles for the $ESG.Score$.

Fig. 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%.

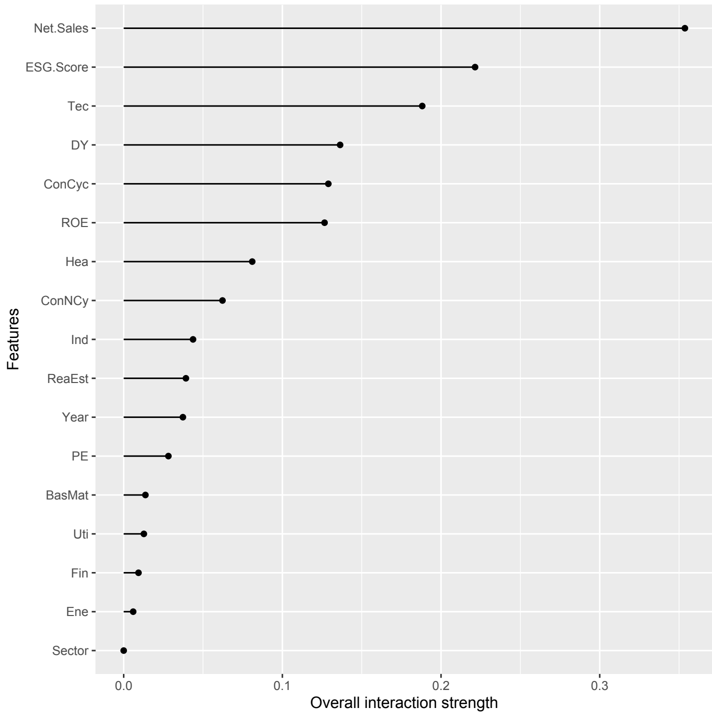

Figure 9: Feature interaction - Each of the input features with all other features for predicting EBIT values.

In Fig. 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.

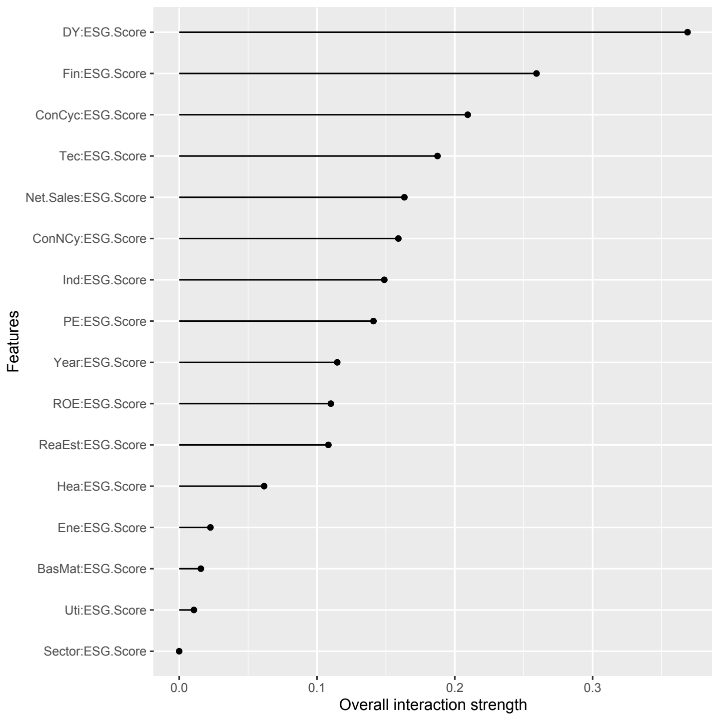

Figure 10: Two-way $ESG.Score$ interactions with the other features in predicting EBIT values.

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. Fig. 11) illustrates the SHAP attributions and Fig. 12 the break-down plots with interactions for two different data points: the first one corresponding to a negative EBIT value (-8,311) and the second one to a high positive value (53,683). From the left panel of Fig. 11, we can observe that the most important variable is *Net.Sales* (= -8,190) that decreases the EBIT prediction by 6,029. 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 Fig. 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 3,849. Note that the object of the SHAP function can be reused to explain all the data points.

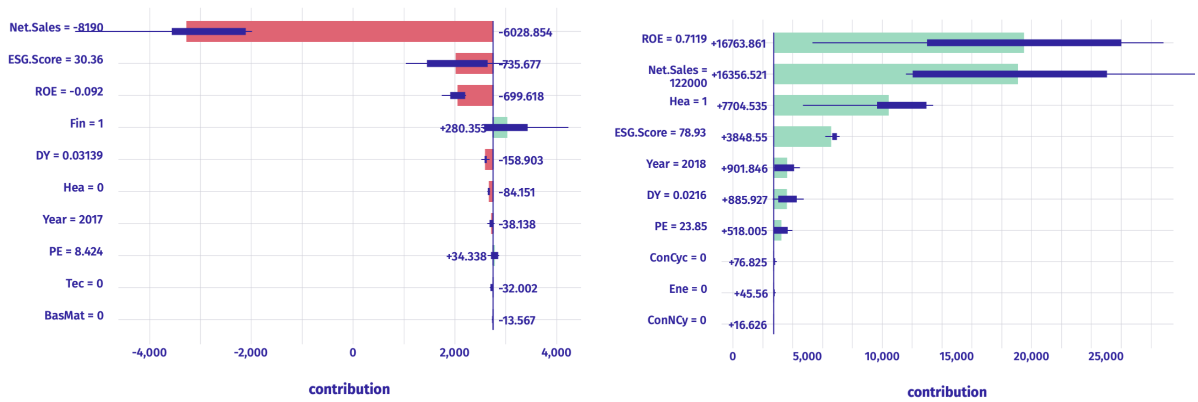

Figure 11: SHAP values. Data points: EBIT=-8,311 (left), EBIT=53,683 (right). Red (green) bars show a negative (positive) contribution of the predictors.

Relating the break-down plots, the left panel shows that RF predicts for the selected data point (EBIT = -8,311) a value equal to about -4,715, which is lower than the average model prediction (2,635). The most important variable is *Net.Sales* (= -8,190) that decreases the EBIT prediction by 4,685. The second most important variable is *ESG.Score* (= 30.36) that decreases the EBIT prediction by 1,670. 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 (2,635). 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 9,362. The third most important variable is *Hea* (= 1) that decreases the prediction by 9,149. The contribution of the other variables is less important.

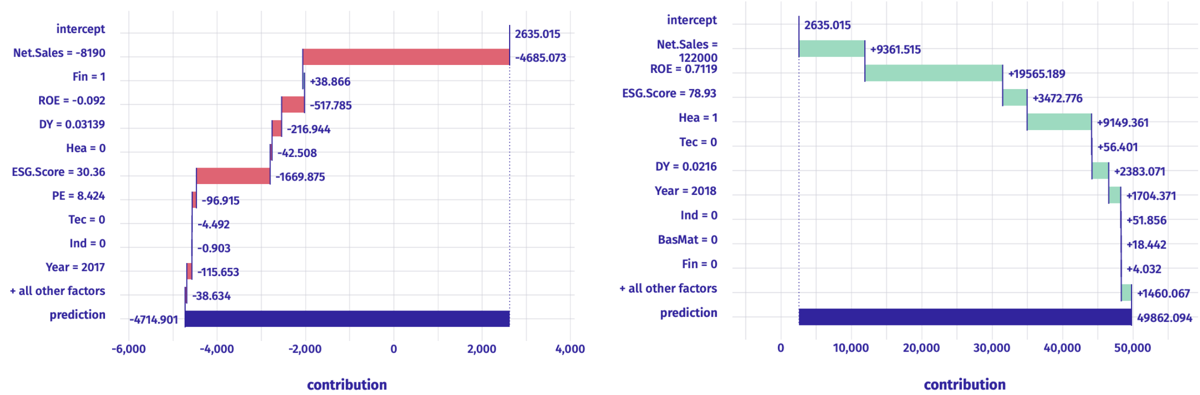

Figure 12: Break-down plot with interactions. Data points: EBIT=-8,311 (left), EBIT=53,683 (right). The blue bar shows the difference between the model's prediction for the selected observation and the average model prediction. Other bars show the contributions of variables. Red (green) bars show a negative (positive) contribution of the variables. The order of variables on the y-axis corresponds to their sequence.

6 Concluding remarks

Investors are paying increasing attention to the ESG factors, there is a wider recognition among investors 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 Refinitive 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.

References

Apley, D. (2018). ALEPlot: Accumulated Local Effects (Ale) Plots and Partial Dependence (Pd) Plots. `https://CRAN.R-project.org/package=ALEPlot`.

Apley, D., Zhou, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. J. R. Statist. Soc. B, 82, Part 4, 1059–1086.

Biecek, P. (2018). DALEX: Explainers for Complex Predictive Models in R. Journal of Machine Learning Research, 19 (84), 1–5.

Bowman, E.H., Haire, M., 1975. A Strategic Posture toward Corporate Social Responsibility. California Management Review 18, 49–58. https://doi.org/10.2307/41164638

Bragdon, J.H., Marlin, J., 1972. Is Pollution Profitable. Risk Management 19, 9–18.

Breiman, L., Friedman, J. et al., Classification and regression trees. (1984)

Breiman, L. (1996). Bagging predictors. Machine learning, 24(2): 123–140.

Breiman, L. (1999). Using adaptive bagging to debias regressions. Technical Report 547, Statistics Dept. UCB.

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

D'Amato, V., D'Ecclesia, R.L., Levantesi, S. (2021). ESG score prediction through random forest algorithm. Computational Management Science. DOI: https://doi.org/10.1007/s10287-021-00419-3. Published on line 2 December 2021.

Friedman, J. (2001). Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics, 29, 1189–1232.

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.

Gosiewska, A., Biecek, P. (2019). iBreakDown: Uncertainty of Model Explanations for Non-additive Predictive Models. https://arxiv.org/abs/1903.11420v1.

GSIA. Global Sustainable Investment Review 2014. GSIA; 2014. Available: http://www.gsi-alliance.org/members-resources/global-sustainable-investment-review-2014/

Global Sustainable Investment Alliance 2021, Global Sustainable investment review 2020

Liaw, A. (2018). Package *randomforest*. Available on line at https://cran.r-project.org/web/packages/randomForest/randomForest.pdf

Lundberg, S.M., Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems. Vol. 30. Curran Associates. https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf.

Monasterolo I., 2020, Climate Change and the Financial System, Annual Review of Resource Economics

Morana C, Sbrana G. 2019. Climate change implications for the catastrophe bonds market: an empirical analysis. Econ. Model. 81:274–94

Moskowitz, M., 1972. Choosing socially responsible stocks. Business and Society Review 1, 71–75.

Naffa H., Fain M., 2020, Performance measurement of ESG-themed megatrend investments in global equity markets using pure factor portfolios methodology, PLoS ONE 15, e0244225, https://doi.org/10.1371/journal.pone.0244225

Quinlan, J.R. (1986). Induction of decision trees. Machine Learning, 1: 81–106.

Rai, A. (2020). Explainable AI: from black box to glass box. J. of the Acad. Mark. Sci. 48, 137-141.

Shapley, L.S.: A value for n-person games. Contrib. Theory Games 2(28), 307-317.

Simpson, W.G., Kohers, T. (2002). The link between corporate social and financial performance: Evidence from the banking industry. Journal of business ethics, 35: 97-109.

Thomson Reuters (2020). Environmental, social and governance (ESG) scores. Available from Refinitiv at: Refinitiv.https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/refinitiv-esg-scores-methodology.pdf.

Thomson Reuters (2020).TRBC Sector Classification. Available from Refinitiv at: https://www.refinitiv.com/content/dam/marketing/en_us/documents/quick-reference-guides/trbc-business-classification-quick-guide.pdf