ORIGINAL PAPER

ESG score prediction through random forest algorithm

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

Abstract

Environment-related risks affect assets in various sectors of the global economy, as well as social and governance aspects, giving birth to what is known as ESG investments. Sustainable and responsible finance has become a major aim for asset managers who are regularly dealing with the measurement and management of ESG risks. To this purpose, Financial Institutions and Rating Agencies have created an ESG score aimed to provide disclosure on the environment, social, and governance (corporate social responsibilities) metrics. CSR/ESG ratings are becoming quite popular even if highly questioned in terms of reliability. Asset managers do not always believe that markets consistently and correctly price climate risks into company valuations, in these cases ESG ratings, when available, provide an important tool in the company's fundraising process or on the shares' return. Assuming we can choose a reliable set of CSR/ESG ratings, we aim to assess how structural data- balance sheet items- may affect ESG scores assigned to regularly traded stocks. Using a Random Forest algorithm, we investigate how structural data affect the Thomson Reuters Refinitiv ESG scores for the companies which constitute the STOXX 600 Index. We find that balance sheet data provide a crucial element to explain ESG scores.

Keywords Machine Learning · ESG risks · Firm performance.

JEL Classifications G14 · C22

1 Introduction

The Environmental, Social and corporate Governance-ESG, or Socially Responsible Investments (SRIs), rating aims to help investors to identify and quantify the ESG risks and opportunities. ESG investments are playing a key role in the asset management

✉ Valeria D'Amato
   vdamato@unisa.it

[1] Department of Pharmacy, University of Salerno, Fisciano, Italy

[2] Department of Statistics, Sapienza University of Rome, Rome, Italy

for several reasons: the sustainability challenge introduced by Governments and Regulatory bodies; the changing preferences of investors and the growing availability of data. The main challenge nowadays is to make the production processes sustainable, so companies are exposed to new risk factors, i.e. flood risk and sea level rise, data security, demographic shifts and regulatory requirements, as a consequence modern investors have to deal with the measurement and management of these new risks. Investor preferences are adapting to new drivers, millennials and women are asking more of their investments. According to a recent study of Bank of America Corporation, in the next two to three decades US-domiciled ESG investments will double the size of the U.S. equity market. The increasing availability and better quality of data provides useful tools to explore the ESG investing. Artificial Intelligence natural language processing and big data can quickly identify hidden risks and opportunities that may be missing from traditional ESG analysis.

Recent regulatory directives require mandatory disclosure of sustainable activity in some countries as China, Denmark, Malaysia and South Africa, only on a voluntary basis in others. However the lack of uniform metrics in the measurement of ESG efforts does not allow an accurate comparison among the various countries. Additional ESG regulations are on the way prompting managers to integrate ESG considerations into their disclosure and investment processes. The European Sustainable Investment Forum (Eurosif) defines "Sustainable and responsible investment (SRI) as a long-term oriented investment approach which integrates Environment-Social-Governance factors in the research, analysis and selection process of securities within an investment portfolio. It combines fundamental analysis and engagement with an evaluation of ESG factors in order to better capture long-term returns for investors, and to benefit society by influencing the behavior of companies." Between 2018 and 2020, total U.S.-domiciled sustainably invested assets under management, both institutional and retail, grew 42%, to $17.1 trillion, up from $12 trillion, according to the Forum for Sustainable and Responsible Investment's 2020 trends report. This number represents 33% of the $51.4 trillion in total U.S. assets now under professional management. Active strategies represent the majority of ESG-related assets under management, at 75% in the U.S. and 82% in Europe. With passive ESG strategies capturing about 60% of new asset inflows in the U.S. in 2019. An increasing role has been played by green bonds, largely issued by corporations, so far they have provided the same risk-return profiles of its conventionally counterpart (Hachenberg and Schiereck 2018). However, according to a recent study by Barclays, pricing of green bonds does not reflect the quality of the bond (Zerbib 2019), we observe a small positive premium difference between conventional bonds and green bonds. A more accurate scrutiny from investors will make this to change.

In this context asset managers look for some assessment of sustainability for guidance and benchmarking (Joliet and Titova 2018; Weber 2013; Hartzmark and Sussman 2019). Several studies suggest that companies with robust ESG practices display a lower cost of capital, lower volatility and few instances of bribery, corruption and fraud, the opposite happening for companies with poor ESG practices (Lins et al. 2017; Chava 2011; Lansilahti 2012; Bhagat et al. 2008; Cremers et al. 2005; Deutsche Bank 2012). Contradicting results are provided by Arribas et al. (2019a), which analyze the concept of "socially responsible company" both from the perspective of the retail investors (and

how they perceive it) and the point of view of investment funds' managers to explore if a match exists between them. They find that the unexpected performance of sustainable and conventional mutual funds are mainly due to the methodologies applied to assign the ESG score. The need for a reliable ESG score brought various financial institutions to play. Fitch Ratings launched ESG Relevance Scores (ESG.RS) for 1,534 corporate issuers in January 2019, and has since released more than 143,000 ESG.RS for over 10,200 issuers and transactions. MSCI introduced a set of tools for ESG analysis: ratings, indexes and analytics. MSCI ratings cover 75,000 companies and more than 650,000 equity and fixed income securities globally and are based on the exposure of each company to industry specific ESG risks and their ability to manage those risks relative to peers. The Bloomberg ESG Data Service collects, checks and standardizes information from a variety of sources about 11,500 companies in 83 countries. It considers 800 metrics covering all the aspects of ESG, from emissions to the percentage of women employees. Such scores are divided into three different classes of disclosures: (i) Environmental (E), (ii) Social (S) and iii) corporate Governance (G). The ESG score is obtained by analyzing different features such as emissions, environmental product innovations, human rights and the companies' structure. It ranges from 0.1 to 100, where 100 represents the highest score attributed to a company that invests in corporate social responsibility (CSR) projects. Given the various criteria adopted to build the ESG score its accuracy is still widely questioned: the lack of standardization, credibility of information, transparency and independence implies bias and tradeoffs. ESG scores (Searcy and Elkhawas 2012) are also supported by the Dow Jones Sustainability Index (DJSI), or the MSCI ESG Indexes, which are used in the process of ESG investing. The fundamental issue relates to the ability of this tool to effectively discriminate between responsible and irresponsible firms.

How firm-level attributes affect the CSR participation has been a major goals for several researchers who have investigated the relationship between companies' characteristics, i.e. balance sheet and income statement information, and CSR performance as in Drempetic et al. (2020), Garcia et al. (2020), and Lin et al. (2019). Most of these studies have been performed assuming the firm data being heterogeneous, as well as vague and uncertain (Garcia et al. 2020). To deal with the heterogeneity of data, the use of Rough Set Theory is proposed, which allows to extract the information from this context unlike the traditional set theory. The theory of slack resources is often revoked as regards the analysis of the financial characteristics of companies and the impact on their ESG rating. According to the slack mechanism, the profitability is expected to have a positive impact on the ESG score: those companies with the greatest resources are precisely those who can afford the necessary investments to improve the ESG score (Drempetic et al. 2020). Other authors (Lin et al. 2019) represent the bidirectional linkages between corporate social responsibility (CSR) and corporate financial performance (CFP) by using the prospective and retrospective approaches, by implementing a Panel Vector Autoregression in Generalized Method of Moments (GMM) context. Finally, the influence of firm size, a company's available resources for providing ESG data, and the availability of a company's ESG data on the company's sustainability performance are positively correlated as stressed in Drempetic et al. (2020).

We believe that the ESG ratings, when available, still affect business and finance strategies and they may represent a crucial element in the company's fund raising process or on shares returns. In this paper we want to relate ESG scores to structural information of the company using a novel approach, the Random Forest algorithm. Using the Thomson Reuters Refinitiv ESG scores, we investigate the roles of structural variables as balance-sheet data on the ESG scores of the constituents of the STOXX Europe 600 index. We find that balance sheet items represent a powerful tool to explain the ESG score.

The remainder of this paper is organized as follows. Section 2 presents a brief review of the literature, Sect. 3 outlines the methodology we propose, describing the regression tree architecture, the random forest algorithm and the variable importance. In Sect. 4 we describe the empirical framework by illustrating the results and implications. Finally, Sect. 5 concludes.

2 Literature review

Most of the research on SRI studies the business case for sustainability rather than the sustainability case for business (Winn et al. 2012). The fact that environment-related and social risks can strand assets in different sectors of the global economy is now evident. This caused increasing attention and intervention by financial supervisors and central banks stimulating a bulk of research in this area. The need to properly measure and manage exposure to environment, social and governance related risks requires to deal with the poor availability of consistent, comparable and trusted data; costs of data and accessing resources to conduct analysis. According to Benabou et al. (2010), the involvement in social actions represents a voluntary action undertaken for the sake of social interest. Most research focuses on the financial return of SRI compared to mostly conventional benchmarks, only few studies measure the impacts for sustainable development (i.e. Cohen and Winn (2007) and Boiral and Paillé (2012)). Many studies showed a financial outperformance of SRI (Mahjoub and Khamoussi 2012; Mahler et al. 2009; Trucost and Mercer 2010; Nakao et al. 2007; Weber et al. 2010; Derwall et al. 2005; Van de Velde et al. 2005), others showed an underperformance of SRI (Makni et al. 2009; Renneboog et al. 2008; Simpson and Kohers 2002; Angel et al. 1997), and still others no meaningful differences (Belghitar et al. 2014; Hamilton et al. 1993; Statman 2000; Bauer et al. 2005; Bello 2005; Kreander et al. 2005; Utz et al. 2014), compared to conventional benchmarks.

A very large number of researches deal with the impact of CSR investments on economic growth or on corporate financial performances, however, we should mention that a reliable set of universally recognized measurement of CSR activities is still lacking. The current inadequacy of disclosures about ESG risks and opportunities outside the company's operational boundary has been stressed by the World Business Council for Sustainable Development WBCSD (2019). The role of ESG ratings and their reliability has been widely discussed by (Berg et al. 2019), who highlights the confusion created by ESG ratings. First, ESG performance is unlikely to be properly reflected in corporate stock and bond prices; second, the divergence frustrates the ambition of companies to improve their ESG performance; third, the divergence of rating poses

a challenge for empirical research. Chatterji et al. (2016) find that ratings from different providers differ dramatically, showing that information received from rating agencies is quite noisy. MSCI researchers (Melas et al. 2018) focused on understanding how ESG characteristics have led to financially significant effects. They assume that company when borrow from central banks create three "transmission channels" within a standard discounted cash flow (DCF) model: i) the cash-flow channel, ii) the idiosyncratic risk channel and iii) the valuation channel. The former two channels are transmitted through corporations' idiosyncratic risk profiles, whereas the latter transmission channel is linked to companies' systematic risk profiles. They show that ESG has an effect on valuation and performance of many of the companies they analyzed. In this context, investigating the relationship existing between structural data as balance sheet data and the existing ESG scores may provide useful information to assess the accuracy of the score.

Multivariate approaches have been used to analyze the relationship between the corporate sustainability performance and the financial performance of firms. Several studies use statistical methods to predict corporate financial performance based on corporate sustainability performance. For instance, Weber et al. (2008) employs ESG criteria to predict accounting indicators, using logistic regression (Anderson 1994), such as EBITDA margin (EBITDA margin), Return on Assets (ROA), and Return on Equity (ROE) as well as financial market indicators, such as Total Returns (TR). The results indicate that the statistical approach is useful to show that ESG performance can explain corporate financial performance with regard to EBITDA margin, ROA, and ROE. However, it cannot predict TR, because there might be too many other important influences on TR (Cerin et al. 2001) or that the shareholders do not integrate sustainability performance into the price of the company shares, as suggested by Schaltegger et al. (2000). Another study by Weber (2017) investigates the connection between ESG performance and financial performance of Chinese banks used panel regression to analyze the impact of ESG metrics on financial performance over time. He uses a time-lagged approach to analyze cause-and-effect between ESG and financial performance. Compared to the methods above, time-lagged panel regression delivered better information about cause-and-effect. There is no doubt that the integration of ESG data is useful for financial decision makers (Monk et al. 2019). However, we believe that statistical analyses certainly help to integrate ESG into lending and investment decision making, Artificial Intelligence techniques might be able to contribute to a better integration. Until now the discussion about whether and how ESG should be integrated into financial decision making is ongoing. Previous studies found a trade-off between the ESG and financial performance (Bauer et al. 2005), more recent studies find robust relationship between ESG and financial performance (Cui et al. 2018; Friede et al. 2015). These mixed results are usually explained due to the heterogeneity of ESG ratings (Berg et al. 2019). AI may improve the collection and data as well as its analysis (In et al. 2019), it provides methods to analyze mixed data. In contrast to financial data, ESG data can be text data, categorical data, and quantitative. Many statistical methods are not able to process different types of data. AI methods, such as machine learning or neural networks, however, are able to process different types of data. Furthermore, these methods are able to recognize patterns without assuming a certain distribution of the data, such as normal distribution. Since,

ESG evaluation usually does not follow statistical distributions, AI methods might be better suited to simulate human decision making that statistical methods. Machine learning applied to large set of data makes possible to deliver new kinds of analytics. Wang et al. (2012) use the AdaBoost algorithm to forecast equity returns and Wang et al. (2014) show that using different training windows provides better performance. Batres-Estrada (2015) and Takeuchi and Lee (2013) use the deep learning approach to forecast financial time series. Moritz and Zimmermann (2016) use tree-based models to predict portfolio returns. A slightly different approach is used by Alberg and Lipton (2017) who propose to forecast company fundamentals (e.g., earnings or sales) rather than returns. They find that the signal-to-noise ratio is higher when forecasting fundamentals, allowing them to use more complex machine learning models. Gu et al. (2020) forecast individual stock returns with a large set of firm characteristics and macro variables. Since they use total returns rather than market excess returns as the dependent variable, they jointly forecast the cross-section of expected returns and the equity premium and find that nonlinear estimators have better accuracy when compared to OLS regressions. The various studies all show how machine learning models succeed in uncovering non-linear patterns. We focus mainly on the cross-section of ESG scores and use firm fundamentals. We find that many machine learning algorithms can outperform linear regression.

3 The regression model based on Random Forest

Let consider a generic regression model used to estimate the relationship between a target (or response) variable, Y, and a set of predictors (or features), $X_1, X_2, ..., X_p$:

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

where $\mathbf{X} = X_1, X_2, ..., X_p$ is the features' vector and ϵ is the error term. As a measure of the distance between the model's prediction and the target variable one usually refers to the expected (squared) prediction error, $E(Y - \hat{Y})^2$, defined by the sum of squared residuals, where $\hat{Y}$ is the estimate of the target Y. Such error can be decomposed into two errors: the reducible error, which arises from the mismatch between f and $\hat{f}$, and the irreducible error, which is essentially the noise:

$$E(Y - \hat{Y})^2 = \underbrace{E[f(\mathbf{X}) - \hat{f}(\mathbf{X})]^2}_{Reducible} + \underbrace{Var(\epsilon)}_{Irreducible} \tag{2}$$

The aim of machine learning models is to estimate function f minimizing the reducible error.

Recent researches are using more often models that can dynamically learn from past data. Ordinary regression techniques are not successful mostly because financial data is inherently noisy, in many cases, the presence of multi-collinearity affects the results, and relationships between factors and returns can be variable, non-linear, and/or contextual. Hence, the estimation of dynamic relationships between potential features and the target variable proves quite tricky. In this context, machine learning

can offer a smart approach that naturally combines many weak sources of information into a synthesized ESG score, stronger than any of its sources. A variety of machine learning algorithms have been elaborated, such as regression trees, random forests, gradient boosting machines, artificial neural networks, and support vector machines. They proved to be capable to uncover complex patterns and hidden relationships that are often hard or impossible to identify by linear analysis, and in presence of multi-collinearity, they are more effective than linear regression and allow for precisely classifying observations. At the top of the classifiers' hierarchy is the random forest classifier that belongs to the family of ensemble methods. It is able to reduce the prediction error and at the same time monitoring the variance.

3.1 The standard random forest

Random forest is a supervised learning algorithm which can be used for both classi-fication and regression tasks. We are going to refer to the Breiman's (2001) original algorithm, being other aggregating random decision trees ambiguously related to the random forest where there are no specification on how the trees are obtained. In partic-ular, we focus on the regression tree setting. According to the nonparametric regression estimation general framework, let be $\mathbf{X} \in \Xi \subset \mathbb{R}^p$ an input random vector which is observed. The output values are numerical and we assume that the training set is inde-pendently drawn from the distribution of the random vector Y, X. The tree growth depends on a random vector Θ such that the tree predictor $h(x, \Theta_k)$ takes on numerical values. The algorithm is addressed to predict the response $Y \subset \mathbb{R}$ by estimating the regression function $m(x) = \mathbb{E}[Y|X = x]$. We assume $D_n = ((\mathbf{X}_1, Y_1), ..., (\mathbf{X}_n, Y_n))$ the training sample of independent random variables distributed as pair (X, Y). We can express the mean-squared generalization error for any numerical predictor $h(x)$ as in the following

$$\mathbb{E}_{\mathbf{X}, Y} = (Y - h(\mathbf{X}))$$

The average over k, $k = 1, ..., n$ trees $\{h(\mathbf{x}, \Theta_k)\}$ corresponds to the random forest predictor (Breiman 2001). The ensemble method for combining the prediction from multiple machine learning algorithms together to make more accurate projections than the individual model is bagging (bootstrap aggregation), as introduced in Breiman (1996) and then improved by the adaptive version (Breiman 1999) which reduces bias and operates effectively in classification as well as in regression. The idea behind the bagging is to generate many bootstrap samples and average the predictors. In order to improve the weak or unstable learners, the bootstrap role consists in choosing n times from n points with replacement to compute the individual tree estimates. Typically in decision tree setting, the algorithm shows high variance. The application of the bootstrap logic is used as a way to reduce the variance of a base estimator, the decision tree, by introducing randomization into its construction procedure and then making an ensemble out of it. The bagging improves the accuracy by changing the predictor due to the perturbation in learning set, caused by the randomization introduced by the bootstrap mechanism.

In the architecture of the random forest, the estimator of the target variable $\hat{y}_{R_j}$ is function of the regression tree estimator, $\hat{f}^{tree}(\mathbf{X}) = \sum_{j \in J} \hat{y}_{R_j} \mathbf{1}_{\{\mathbf{X} \in R_j\}}$, $\mathbf{1}_{\{.\}}$ being the indicator function and $(R_j)_{j \in J}$ the region of the predictor space which is divided into J distinct and non-overlapping $R_1, R_2, ..., R_J$ and obtained by minimizing the Residual Sum of Square:

$$\sum_{j=1}^{J} \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2$$

By denoting B the number of bootstrap samples and $\hat{f}^{tree}(\mathbf{X}|b)$ the decision tree estimator on the sample $b \in B$, we can express the random forest estimator as follows:
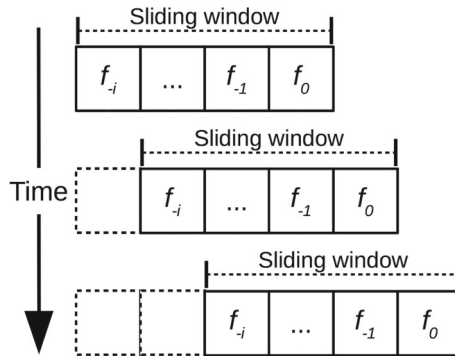
$$\hat{f}^{RF}(\mathbf{x}) = \frac{1}{B} \sum_{b=1}^{B} \hat{f}^{tree}(\mathbf{X}|b) \tag{3}$$

As the number of trees grows, it does not always mean the performance of the forest is significantly better than previous forests (fewer trees), so that some authors (as in Oshiro et al. (2012)) identify an optimal threshold from which increasing the number of trees would bring no significant performance gain, and would only increase the computational cost.

3.2 The temporal dynamic random forest (TDRF)

RF is an ensemble of tree-based techniques for classification or regression, working by constructing a multitude of decision trees at training time and outputting the average prediction of the individual trees. In this section, we aim at representing the random forest algorithm in a dynamic perspective, in order to measure the temporal effects in our approach. The temporal dynamics of tree models is a branch of the predictive machine learning algorithms. Nevertheless, how to restructure the time series to have a supervised learning problem is a big issue in social sciences since there is a need to preserve the temporal order in an algorithm that is intrinsically static respect to time. The literature on the dynamic RF, introduced by Bernard et al. (2012), is very scarce. In Bernard et al. (2012) the idea of a new Random Forest induction algorithm arises from a combination of the weights updating as in the boosting algorithm with the Random Forest. The other studies on the topic are mainly due to extensions or applications of this approach (as in Nami and Shajari 2018; Biau and Scornet 2016; Xu and Chen 2017). Nevertheless, the dynamic approach is absolutely not related to the temporal evolution of a phenomenon as classified in Econometrics and Statistics. According to the literature the dynamic RF represents an adaptive procedure that iteratively reproduces forests, in order to decrease the forecasting errors and to improve the accuracy by using the re-sampling principle of boosting, as opposite to the static approach. In light of these considerations, we formulate a dynamic RF from a temporal point of view. We develop a novel approach called Temporal Dynamic Random Forest (TDRF) using the suggestions coming from best practices (as for instance in Brownlee

(2020)) and the dynamic algorithm as in Carannante and D'Amato (2021). We reframe the time series problem as a supervised learning problem for machine learning, by means of a rolling window method where the prior time step is used to predict the next time step. In statistics and time series analysis, this is called a lag or lag method. In particular, we restructure the data to look like a supervised learning problem, so starting from our given time series dataset: the previous time steps are considered as input variables and use the next time step as output variable as shown in the following plot:

We create the features for each target variable, corresponding to the previous time step to which the target variable refers. The data restructure preserves the temporal order between the observations and it continues to be preserved when using this dataset to train a supervised model. As in best practices (Dietterich 2002; Bontempi et al. 2013; Brownlee 2020), we find that the first value in a time series is not predicted by the previous value so that the value under consideration is removed. We do not have a known next value to predict for the last value of the series. While training the supervised model we can remove also this value. The number of previous time steps represents the window width or the size of the lag. This rolling window method can turn any time series dataset into a supervised learning problem and it is flexible to handle time series with more than one observation at each time step, as in a multivariate setting. It is noteworthy that our approach by reformulating the problem using a lag-based rolling windows, in a multivariate framework, relies on the underlying common concept of the VAR (vector autoregression) models, where the system of simultaneous equations which are used to model the dynamics of each variable, includes the variable's lagged values. Machine learning are often able to integrate both, quantitative and qualitative data into their analyses so they can be used as efficient tools to analyze critical issues in business decisions given i) their ability to detect complex nonlinear relationships between variables, ii) their self-organization and self-learning skills; iii) their error tolerance capabilities (Aydin et al. 2015; Tu 1996; Oztemel 2003). The TDRF can be set up in a flexible way, being suitable for:

- a direct approach, where a separate model is built to forecast each future time;
- a recursive approach, where a single model is developed to make one-step forecasts, and the model is used recursively where prior forecasts are used as input for subsequent forecasts.

3.3 Model explainability

ML algorithms are frequently considered as black-boxes[1] because the process between input and output is opaque. Such a problem is sometimes a barrier to the adoption of machine learning models. A field of literature focusing on the explainability of artificial Intelligence is becoming increasingly important with the aim to interpret and explain individual model predictions to decision-makers, end-users, and regulators. To understand how a model operates we need to unfold the various steps in order to know how it works and which decisions it takes. We deal with the model interpretability considering a measure of the feature importance and the partial dependency plot. A common form of model explanations are based on feature attributions, so a score (attribution) is ascribed to each feature in proportion to the feature's contribution to the prediction. Recently there has been a surge in feature attribution methods, with methods based on Shapley values from cooperative game theory being prominent among them

Shapley values (Shapley 1953) provide a mathematically fair and unique method to attribute the payoff of a cooperative game to the players of the game. Recently, there have been a number of Shapley-value-based methods for attributing an ML model's prediction to input features. In applying the Shapley values method to ML models, the key step is to setup a cooperative game whose players are the input features and whose payoff is the model prediction. Due to its strong axiomatic guarantees, the Shapley values method is emerging as the de facto approach to feature attribution.

The ESG risk assessment by institutions and supervisors is still at its early stages. Currently, the incorporation of ESG risks into institutions' business strategies, processes, and risk management reflects in-house approaches and practices. The quality of the ESG risk measurements is significantly sensitive to the inconsistency and the lack of data, "the scarcity of relevant, comparable, reliable and user-friendly data"(EBA Consultation Paper 2020). Main risk-based approaches that take into account the likelihood and the severity of the materialization of ESG risks are based on the use of historical data which do not always reflect the actual risks, this is mainly due to the fact that historical data are processed with standard models. EBA (2020a) suggests that ESG factors are not captured by classical models and most of the ESG risks are non-linear. The non-linearity and the complexity that it creates stimulate the implementation of Artificial Intelligence and modern methods of learning by experience. The response to the EBA consultation by the European Banking Federation (EBF 2021), which is the voice of the European banking sector, alert the different stakeholders to the impossibility of evaluating the consistency of whatever methodologies in the short-term horizon, by recommending a long-term interval (they identify the temporal window from 3 to 10 years). They point out the current uncertainty around methodologies and data. The use of Artificial Intelligence in EU banks is becoming quite common, notably within 2 years, 12% of the EU banks have moved from pilot testing and development to the implementation of AI tools (EBA 2020b). An EU regulatory framework for AI is expected and planned for 2021. From the point of

---

[1] In science, computing, and engineering, a black box is a device, system, or object which can be viewed in terms of its inputs and outputs, without any knowledge of its internal workings. Its implementation is opaque or "black". Source: Investopedia.

view of the consumers' confidence, the transparency of the methodologies adopted by the financial system is insistently required. General Data Protection Regulation, applicable as of May 25th, 2018 in all member states, establishes that the consumers have the right not to be subject to a decision based solely on automated processing so that an explainability need emerges. The explainability involves the disclosure of the underlying logic of decisions, the highlighting of the strengths and weaknesses of the decision process, the interpretation of the model, the outcome explanation being interpretable local predictor, and model inspection rules which provide a visual or textual representation for improving the forecasting output. However, we need to be aware of the he trade-off between accuracy and explainability as shown in the the literature (for instance in Gunning et al. (2019), Hacker et al. (2020), Rai (2020)).

In particular, as regards Shapley-value-based explanations are considered by some researchers (Aas et al. 2019) the only method compliant with legal regulation such as the General Data Protection Regulation's "right to an explanation", others argue that Shapley values do not provide explanations that suit human-centric goals of explainability (Adler et al. 2018; Merrick and Taly 2019) as stressed in Kumar et al. (2020).

3.3.1 Variable importance

The calculation of feature importance measures is a crucial step towards ML algorithm interpretability, it provides information on the features' contribution to the explanation of the target variable. Feature importance considers the relative influence of each feature by calculating the number of times a feature is selected for splitting in the tree building process, weighted by the squared error improvement resulting from each split, and averaged over all trees. It offers more intuition into the algorithm learning process.

In RF algorithms, a large number of trees can make the understanding of the prediction functioning difficult. Therefore, Breiman (2001) proposed a weighted impurity measure, which gives the importance of a feature in the RF prediction rule. Specifically, it assesses the importance of a feature X_m in predicting the target variable Y, for all nodes t averaged over all N_T trees in the forest. Among the variants of the feature importance measures, we consider the Gini importance ratio (often called Mean Decrease Gini, MDG), which computes the importance of each feature X_m as the sum over the number of splits including the feature, proportionally to the number of samples it splits. It is obtained by assigning the Gini index to the impurity $i(t)$ index:

$$MDG(X_m) = \frac{1}{N_T} \sum_{T} \sum_{t \in T : v(s_t) = X_m} p(t) \Delta i(s_t, t) \qquad (4)$$

where $v(s_t)$ is the variable used in splitting s_t and $\Delta i(s_t, t)$ is the impurity decrease of a binary split s_t dividing node t into a left node t_l and a right node t_r.

Denoting the sample size by N, the proportion of samples reaching t by $p(t) = \frac{N_t}{N}$, the proportion of samples reaching the left node t_l by $p(t_l) = \frac{N_{t_l}}{N}$ and the right node t_r by $p(t_r) = \frac{N_{t_r}}{N}$, the impurity decrease $\Delta i(s_t, t)$ is given by:

$$\Delta i(s_t, t) = i(t) - \frac{N_{t_l}}{N_t} \cdot i(t_l) - \frac{N_{t_r}}{N_t} \cdot i(t_r) \tag{5}$$

3.3.2 Partial dependence plot

The partial dependence plot (PDP) shows the marginal effect of one or two features belonging to the set S on the predicted outcome of a machine learning model averaged over the joint values of the other features given by the algorithm. It is used to analyze whether the relationship between the target variable and a feature is linear, non-linear, monotonic or not. Let x_s be the feature of interest for which the partial dependence function should be plotted, and $x_{i,C}$ be the other (complementary) features considered in the model. The partial dependence function is defined as

$$\hat{f}_S(x_S) = \frac{1}{n} \sum_{i=1}^{n} f(x_S, x_{i,C}) \tag{6}$$

Partial dependence is obtained by marginalizing the model outcome over the distribution of the features belonging to set C. In doing so, the function $\hat{f}_S$ shows the relationship between the feature of interest and the predicted outcome.

4 Empirical Analysis

We apply the random forest algorithm to the balance sheet items of the companies constituent the STOXX Europe 600 Index in order to identify the main drivers of the ESG scores. The algorithm's performance is compared to those of a traditional Generalized Linear Model (GLM).

4.1 Data

Our analysis focuses on the constituents of the STOXX Europe 600 Index, which represents large, mid and small capitalization companies across 17 countries of the European region. Because of its wide market exposure, the STOXX Europe 600 index is usually considered the European equivalent of the S&P 500 index reflecting the dynamics of the US stock market. We collect balance sheet information and ESG scores for 67% (401 to 600 companies) of the constituents of the STOXX Europe 600 index from the Thomson Reuters Refinitiv ESG (Refinitiv ESG, henceforth) over the period 2009–2019. Data on single pillars are also provided. The final sample is made of 401 companies which are the companies that have been included in the index throughout the chosen 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 (Fig. 1), which is a reflection of the company's ESG performance based on publicly reported information in the three ESG pillars with the weights of
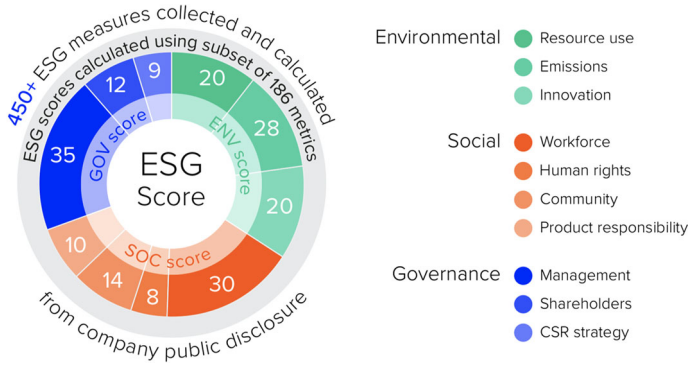
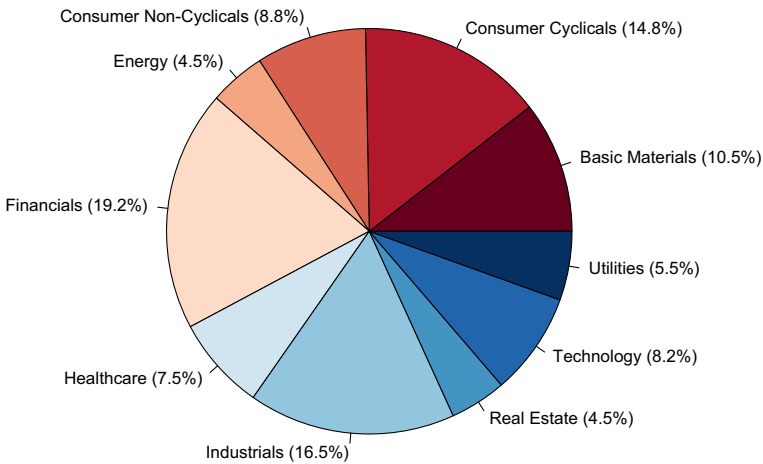**Fig. 1** ESG categories. Source: Thomson Reuters Refinitiv ESG scores

Fig. 2 Industry sector of the companies included in our analysis

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 depicted in Fig. 2.

The ESG score ranges between 0 (minimum score) and 100 (maximum score). The companies included in our sample have a 63.10 average ESG score in the decade 2009–2019. The dynamics of the average ESG score by sector is represented in Fig. 3, while the related standard deviation in Fig. 4.

We analyze the following set of balance sheet items which represent our model' features:
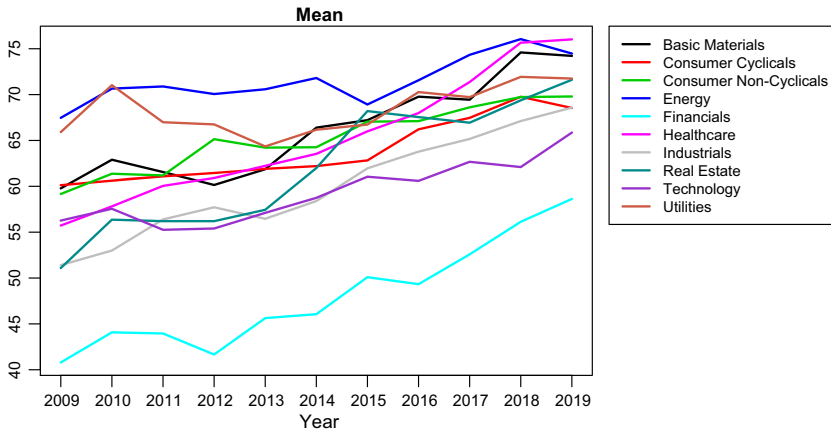
– Year: 2009–2019

Fig. 3 Main statistics of the ESG score distribution by year: average value
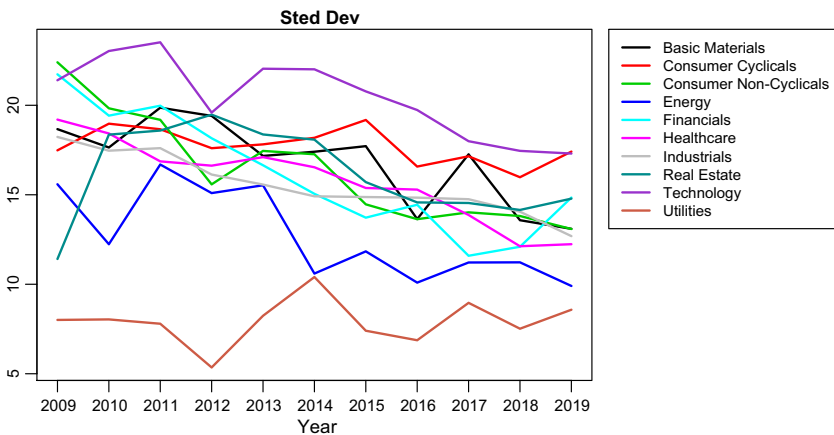

Fig. 4 Main statistics of the ESG score distribution by year: standard deviation

– Sector: categorical variable indicating the company's industry sector (transformed in dummy variables)
– ESG.Score: Refinitiv ESG score
– E.Score: Refinitiv Environmental score
– S.Score: Refinitiv Social score
– G.Score: Refinitiv Governance score
– Net.Sales: sales receipts for products and services, less cash discounts, trade discounts, excise tax, and sales returns and allowances
– 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
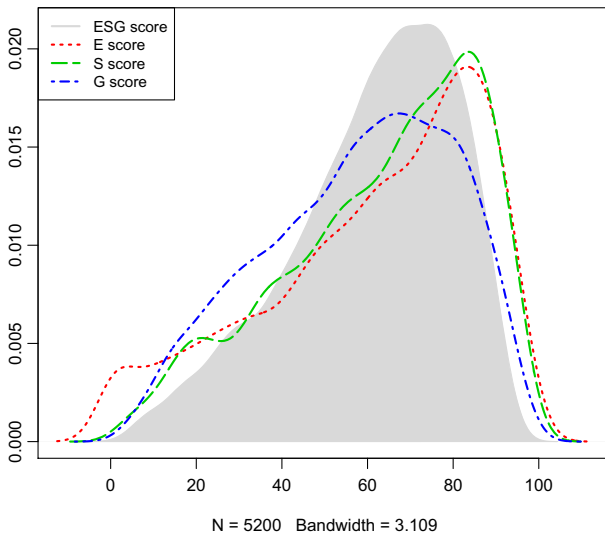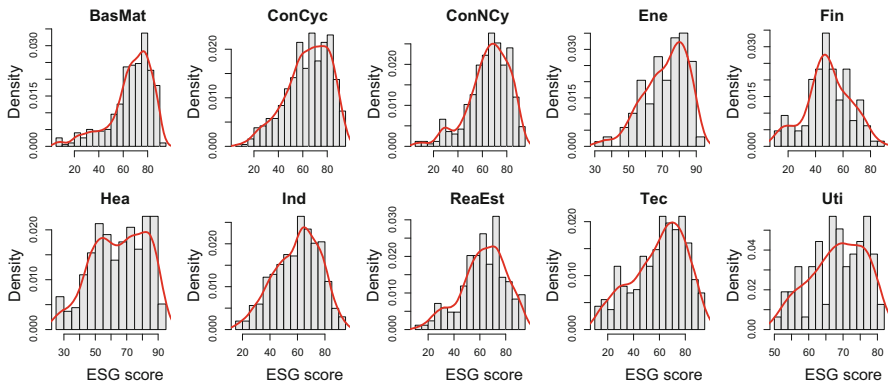
Fig. 5 Density functions of ESG variables

Fig. 6 Density functions of ESG.Score by sector

– 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)

Figure 5 shows the density functions of the *ESG.Score* and its components E,S,G related to our dataset. While, in Fig. 6 the density functions are illustrated by sector.

To analyze the relationship between each pair of variables in the dataset, we plot the variables' correlation in the correlogram depicted in Fig. 7 (positive correlations in blue and negative correlations in red). The color intensity is proportional to the correlation
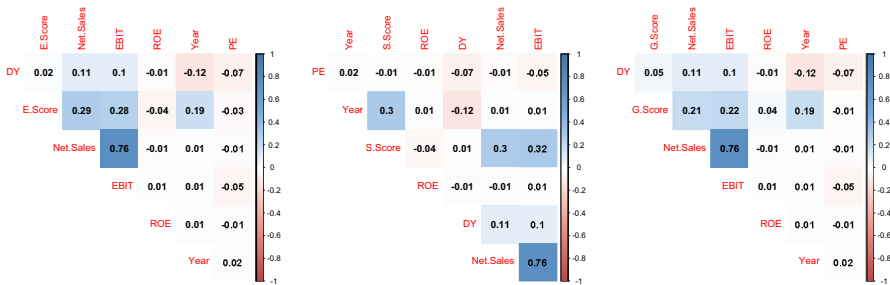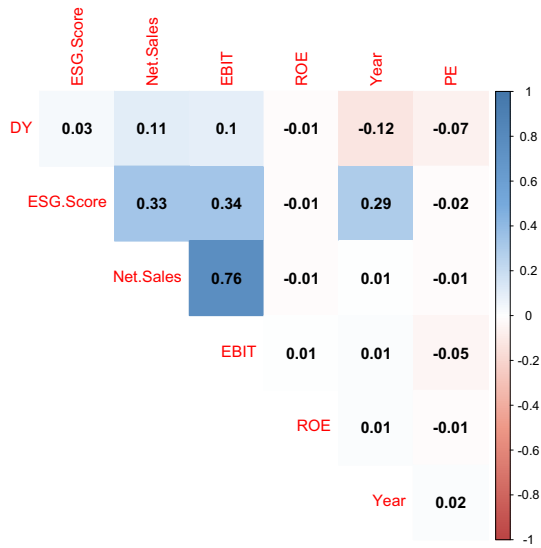
Fig. 7 Correlogram of the data set

Fig. 8 Correlogram of the ESG score components. Left side: $E.Score$, center: $S.Score$, right: $G.Score$

coefficient. As easily understandable, we find strong positive correlations between $Net.Sales$ and $EBIT$ (0.76), which are in turn correlated with the ESG score (0.33 and 0,34, respectively). Negative correlations are very low. The correlogram catches the linear dependence between a set of variables, while machine learning algorithms are able to discover non-linear patterns and hidden correlations.

We also show in Fig. 8 three other correlograms, one for each ESG score component (E, S and G), which help to shed light on the importance of each pillar in determining the ESG score.

4.2 RF estimation of the ESG score

Our model is formulated as follows:

$$Y \sim Year + Net.Sales + EBIT + PE + ROE + DY + Sector \qquad (7)$$

Table 1 RSS and MSR obtained by the RF algorithm

Indicator	ESG.Score	E.Score	S.Score	G.Score
RSS	66.62%	68.14%	63.23%	38.51%
MSR	124.40	205.90	200.81	284.12

Table 2 Sectors' abbreviations

Sector	Abbreviation
Basic materials	BasMat
Consumer cyclicals	ConCyc
Consumer Non-Cyclicals	ConNCy
Energy	Ene
Financials	Fin
Healthcare	Hea
Industrials	Ind
Real estate	ReaEst
Technology	Tec
Utilities	Uti

The target variable Y is $ESG.Score$ and its estimate $\widehat{Y}$ is the output of the RF algorithm used to estimate the model. Therefore, $\widehat{Y}$ is the RF estimator of $ESG.Score$. We use the RF algorithm implemented in the R package $randomForest$ developed by Liaw (2018).

The optimal parameter setting is found by performing the hyper-parameters tuning starting from a set of random seeds (100) for the pseudo-random generator and a number of trees which is appropriate for our dataset ($ntrees = 500$). We have to optimize the number of input variables that are selected at each splitting node, $mtry$, and the minimum size of a terminal node, $nodesize$. For example, the value $mtry = 3$ meaning that at each split, three variables would be randomly sampled as candidates and one of them is used to form the split. We select the combination of seed/$mtry$/$nodesize$ providing the lowest mean of squared residuals, $MSR = \frac{1}{J \cdot n_j} \sum_{j \in J} \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2$, with n_j the number of observations belonging to the region R_j, and the highest percentage of explained variance $RSS = \sum_{j \in J} \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2$.

We partition the dataset into training and test set according to the 80%-20% splitting rule. After the parameters tuning, the following parameters are set: $mtry$=10 and $nodesize$=1 for the $ESG.Score$[2]. The percentage of variance explained by the random forest algorithm, RSS, and the level of the resulting MSR are given in Table 1 for ESG score and its components.

Figure 9 shows MDG values for each model's feature, sorted decreasingly from top to bottom. From this plot, we observe that the most explicative features selected by RF algorithm are $Net.Sales$ and $EBIT$, followed by ROE. Note that the sector's name has been shortened and the shorter names are reported in Table 2.

---

[2] For the target variables $E.Score$, $S.Score$ and $G.Score$, the algorithm's parameters are set respectively as follows: $mtry$ = 9, 9, 10 and $nodesize$ = 1, 1, 1.
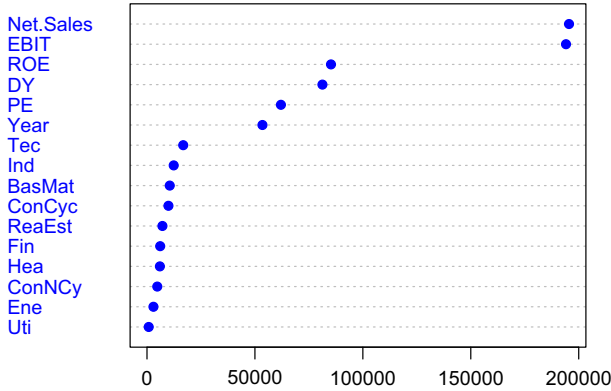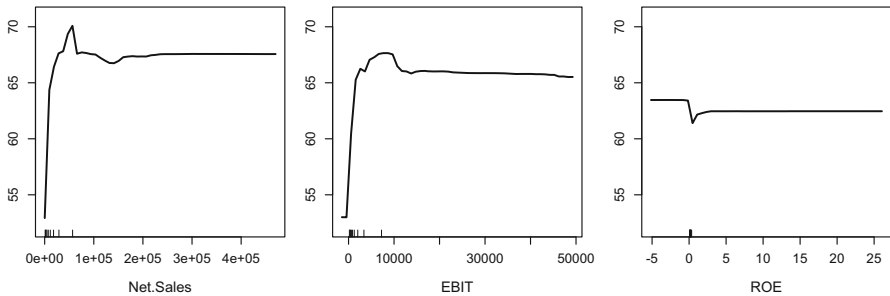
Fig. 9 Variable importance for ESG.Score

Fig. 10 Single variable PDP for $ESG.Score$. Predictors: $Net.Sales$, $EBIT$ and ROE

The partial dependence plots for the most important features are shown in Fig. 10. We observe that all the variables clearly show a non-linear pattern, which is suitable for machine learning. A simple linear regression model should be preferable only in case of linear variables.

According to the Principles for Responsible Investment (PRI), to predict revenues, investors generally look at how fast a company is growing and whether it will gain or lose market share. Companies with robust ESG practices have been found to have a lower cost of capital, lower volatility, and fewer instances of bribery, corruption and fraud over certain time periods. On the other hand companies with poorer ESG performance have had a higher cost of capital, higher volatility due to controversies and other incidences such as spills, labor strikes and fraud, and accounting and other governance irregularities. ESG risks have to be integrated in the company's performance, for instance company's sales growth rate may be affected by the level of ESG opportunity/risk. "For example, a food producer may stop selling a particular type of food due to environmental concerns, which is estimated to reduce sales by x% annually"(PRI 2016). So, for instance, we can expect sales to be linked to the ESG score. To deeply investigate the relationship between the sales amount and the $ESG.Score$, we display the values of these two variables in a scatterplot (Fig. 11). The points represent the observed values, the red line the locally estimated scatterplot smoothing
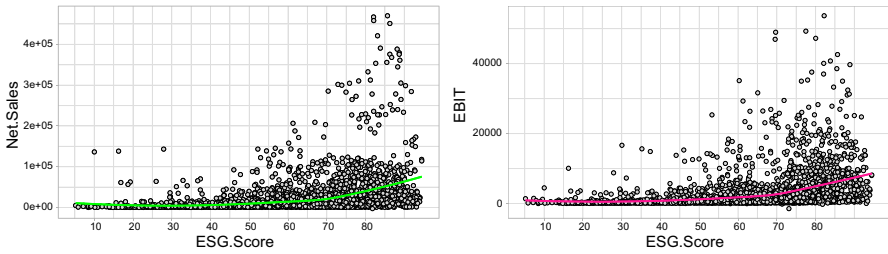
Fig. 11 Net.Sales (panel a) and EBIT (panel b) vs $ESG.Score$. Observed values (points) and LOESS (lines)

(LOESS), which is a local weighted (non-parametric) regression used to fit a smooth curve through the points. LOESS regression clearly highlights that $Sales$ increase as $ESG.Score$ grows.

The ESG score demonstrates a positive relationship with the sales. ESG scores has improved significantly as the firms make considerable efforts towards addressing their environmental and social risks. We could argue these companies are more willing to disclose information in order to enhance their brand image and good reputation. It also indicates strong, intuitive relationships between higher level of success in terms of higher sales for the company and particular attention to reputation. It is noteworthy that the investment decision process landscape is changing everywhere in the market, driven by both ESG and financial performance of individual companies. Particularly in actively-managed SRI and conventional funds, the funds' stock picking process reflecting the screening criteria higher portfolio weights to firms with higher ESG scores (Joliet and Titova (2018)).

4.3 RF predictive performance and comparison with GLM

We test the RF predictive performance compared with a traditional Generalized Linear Model (GLM). We measure the goodness of prediction by the Root Mean Square Error (RMSE) and Mean Absolute Percent Error (MAPE), which are defined as:

$$RMSE = \frac{\sum_i (y_i - \hat{y}_i)^2}{n} \qquad (8)$$

$$MAPE = \frac{100}{N} \sum_i \left| \frac{y_i - \hat{y}_i}{y_i} \right| \qquad (9)$$

In the GLM, the explanatory variables, $\mathbf{X} = (X_1, X_2, ..., X_p)$, are related to the response variable, Y, via a link function, $g()$, and each outcome of the response variable is assumed to be generated from a distribution belonging to the exponential family (e.g Gaussian, Binomial, Poisson). Denoting $\eta = g(E(Y))$ as the linear predictor, the following equation describes the dependency of the mean of the response variable from the linear predictors:
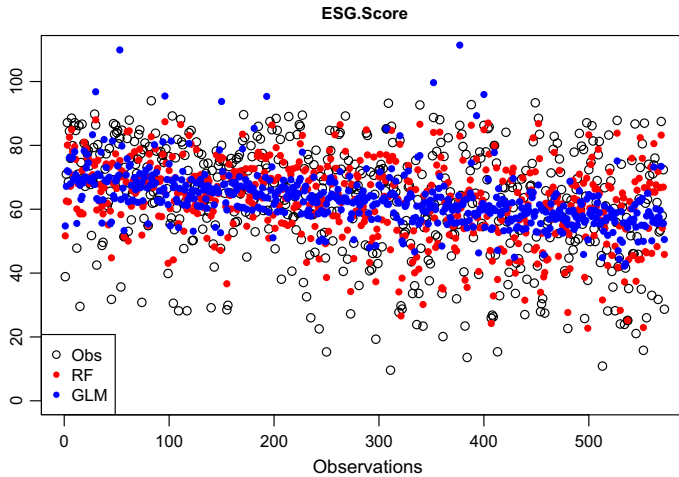
$$\eta = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p \qquad (10)$$

Fig. 12 $ESG.Score$: predicted (RF, GLM) versus observed values (Obs)

Where $\beta_1, ..., \beta_p$ are the regression coefficients to be estimated and β_0 is the intercept. We assume a Gaussian distribution for Y and an identity for the link function, so that: $\eta = E(Y)$.

To assess the importance of variables, it is often used in logistic regression the significance of the predictors, measured by the Wald test with null hypothesis: $H_0 : \beta = 0$.

The GLM performance and the estimate of the regression coefficients are reported in Table 3, where $z = \frac{\hat{\beta}}{SE(\hat{\beta})}$ is the value of the Wald test, $Pr(> |z|)$ is the corresponding p-value, and $SE(\hat{\beta})$ is the standard error of the model.

Apart from the intercept, the GLM assigns the greatest importance to predictors $Year$, $EBIT$ and $Net.Sales$ as well as Financials and Technology sectors. Such a result is, in part, similar to the RF output, which also assigns higher importance to $Net.Sales$ and $EBIT$.

Table 7 shows the values of RMSE and MAPE for RF and GLM, where the first two columns refer to the training set and the following two the test set. We can notice that RF obtains an improvement in the prediction with respect to GLM, reducing RMSE from 15.03 to 10.17, and MAPE from 26.62% to 16.77% in the test set.

Figure 12 illustrates the predicted $ESG.Score$ obtained by the random forest algorithm applied on the test set data compared to the values predicted by GLM. The RF algorithm obtains the best performance, showing higher flexibility and better adaptive capacity than the GLM.

The values of RMSE and MAPE in the training set for the E, S and G pillars are shown in Table 5. The corresponding values in the test set are reported in Table 6.

Table 3 GLM results for the *ESG.Score* regression model

Coefficient	$\hat{\beta}$	$SE(\hat{\beta})$	z value	$Pr(>|z|)$
(Intercept)	−2.645e+03	2.081e+02	−12.708	< 2e-16 ***
Year	1.344e+00	1.033e-01	13.012	< 2e-16 ***
I(Sector)Financials	−1.624e+01	1.789e+00	−9.075	< 2e-16 ***
EBIT	8.276e-04	1.060e-04	7.811	8.46e-15 ***
I(Sector)Technology	−9.299e+00	1.319e+00	−7.048	2.37e-12 ***
I(Sector)Industrials	−6.221e+00	1.116e+00	−5.573	2.78e-08 ***
Net.Sales	4.879e-05	1.165e-05	4.188	2.92e-05 ***
DY	3.748e+01	1.387e+01	2.702	0.00694 **
I(Sector)Consumer Cyclicals	−2.811e+00	1.148e+00	−2.448	0.01445 *
I(Sector)Healthcare	−3.132e+00	1.378e+00	−2.272	0.02316 *
I(Sector)Consumer Non-Cyclicals	−2.480e+00	1.256e+00	−1.974	0.04847 *
ROE	−4.159e-01	2.830e-01	−1.470	0.14179
I(Sector)Real Estate	−2.005e+00	1.604e+00	−1.250	0.21149
I(Sector)Utilities	1.184e+00	2.210e+00	0.536	0.59220
I(Sector)Energy	−9.271e-01	1.762e+00	−0.526	0.59885
PE	5.258e-04	4.395e-03	0.120	0.90478

Significance codes: $\cdot$ p <0.1, * p <0.05, ** p <0.01, *** p<0.001.

Table 4 RMSE and MAPE on $ESG.Score$ predicted values. RF and GLM

Measure	Training set		Test set	
	RF	GLM	RF	GLM
RMSE	11.15	15.65	10.17	15.03
MAPE	19.22%	28.57%	16.77%	26.62%

Table 5 RMSE and MAPE for the E,S,G components in the training set

Measure	E.Score		S.Score		G.Score	
	RF	GLM	RF	GLM	RF	GLM
RMSE	14.35	20.57	14.17	19.21	16.86	20.50
MAPE	64.73%	97.52%	30.17%	44.29%	36.70%	48.15%

Table 6 RMSE and MAPE for the E, S, G components in the test set

Measure	E.Score		S.Score		G.Score	
	RF	GLM	RF	GLM	RF	GLM
RMSE	12.93	19.93	12.84	18.50	17.27	20.84
MAPE	59.33%	125.24%	29.37%	39.37%	37.95%	48.55%

Table 7 RMSE and MAPE on $ESG.Score$ predicted values. RF, GLM and RFdyn

Measure	Training set			Test set		
	RF	GLM	**RFdyn**	RF	GLM	**RFdyn**
RMSE	11.15	15.65	**21.60**	10.17	15.03	**21.35**
MAPE	19.22%	28.57%	**37.35%**	16.77%	26.62%	**35.35%**

4.4 Dynamic RF predictive performance

Random forest regression for time series prediction. Rolling windows with same length (three years).

The following table shows RMSE and MAPE values for RF and GLM (the same as in Table 7), to which the result for the RFdyn are added.

The absolute differences between the predicted ESG.Score and the observed one are illustrated in Fig. 13 and in Fig. 14 for all the models.

5 Conclusions

In the last decade, growing attention by Institutional investors to adopt environmental, social and governance (ESG) investing fostered the growth of ESG assets under management. Many factors have contributed to this growth, we can mention certainly as principle drivers: sustainability challenges, shifts in investor preferences, and improve-
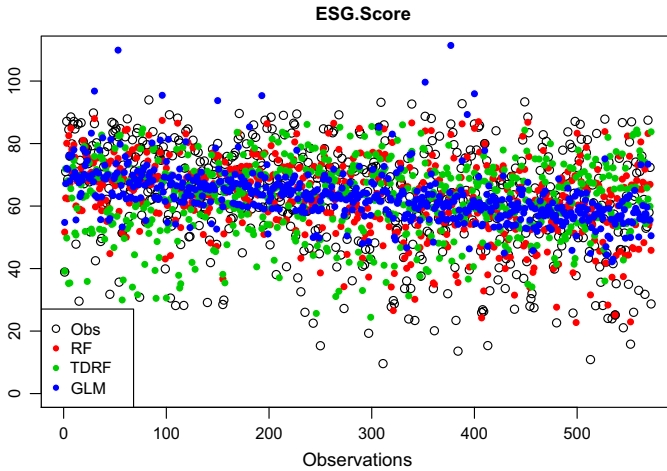
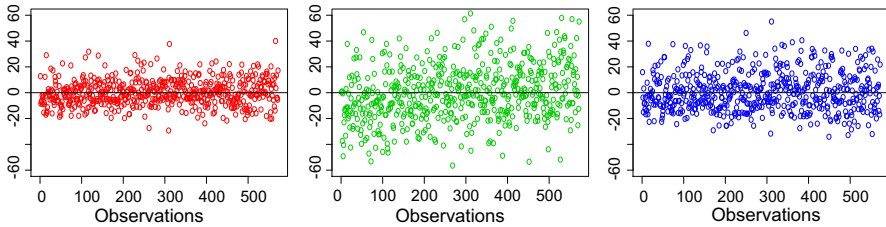**Fig. 13** $ESG.Score$: predicted (RF, $Dyn\ RF$, GLM) versus observed values (Obs)

Fig. 14 $ESG.Score$: predicted (RF, $Dyn\ RF$, GLM) versus observed values (Obs)

ment in data and analytics. The standardization of key metrics and broader transparency in financial markets simplifies the process of evaluating firms' ESG attributes by the rating agencies, avoiding the divergence between rating score which is merely noise, according to the recent literature (Berg et al. 2019).

In this paper, we identify the structural corporate variables, which affect the ESG score using a novel approach: the random forest algorithm. We use balance sheet data of a set of companies listed in the STOXX Europe 600 Index. The numerical results show that balance sheet items present a significant predictive power on ESG score. The RF algorithm achieves the best prediction performance compared to the classical regression approach based on GLM, demonstrating the ability to capture the non-linear pattern of the predictors. As regard to the importance of variables, the algorithm selects *Sales* as the most predictive variable. Further researches could be detect the different dependence structure between the determinants of the ESG score by developing, for instance, flexible r-vine copulas.

References

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

Adler P, Falk C, Friedler SA, Nix T, Rybeck G, Scheidegger C, Smith B, Venkatasubramanian S (2018) Auditing black-box models for indirect influence. Knowl Inf Syst 54(1):95–122

Alberg, J., Lipton, Z. (2017). Improving factor-based quantitative investing by forecasting company fundamentals. Working paper, Cornell University. Available at https://arxiv.org/abs/1711.04837

Angel J, Rivoli P (1997) Does ethical investing impose a cost upon the firm? a theoretical perspective. J Invest 6:57–61

Anderson TW (1984) An introduction to multivariate statistical analysis. Wiley, New York

Arribas I, Espinós-Vañó MD, García F, Oliver J (2019) Defining socially responsible companies according to retail investors preferences. Entrep Sustain Issues 7(2):1641–1653

Arribas I, Espinós-Vañó MD, Garcîa F, Morales-Bañuelos PB (2019) The inclusion of socially irresponsible companies in sustainable stock indices. Sustain 11:2047

Aydin AD, Caliskan Cavdar S (2015) Two different points of view through artificial intelligence and vector autoregressive models for Ex Post and Ex ante forecasting. Comput Intell (**Neuroscience**)

Batres-Estrada, G. (2015). Deep Learning for Multivariate Financial Time Series. Master's thesis, KTH Royal Institute of Technology. https://www.math.kth.se/matstat/seminarier/reports/M-exjobb15/150612a.pdf

Bhagat, S. Bolton, B. Corporate governance and firm performance. J Corp Financ. 14: 257–273

Bauer R, Koedijk K, Otten R (2005) International evidence on ethical mutual fund performance and investment style. J Bank Financ 29(7):1751–1767

Belghitar Y, Clark E, Deshmukh N (2014) Does it pay to be ethical? Evidence from the FTSE4Good. J Bank Financ 47:54–62

Bello ZY (2005) Socially responsible investing and portfolio diversification. J Financ Res 28(1):41–57

Benabou R, Tirole J (2010) Individual and corporate social responsibility. Economica. https://doi.org/10.1111/j.1468-0335.2009.00843.x

Berg F, Koelbel JF, Rigobon R (2019) Aggregate Confusion: The Divergence of ESG Ratings. MIT Sloan Research Paper No. 5822–19

Bernard S, Adam S, Heutte L (2012) Dynamic Random Forests. Pattern Recognit Lett 33(12)

Biau G, Scornet E (2016) A random forest guided tour. TEST 25:197–227

Boiral O, Paillé P (2012) Organizational citizenship behaviour for the environment: measurement and validation. J Bus Ethics 109(4):431–445

Bontempi G, Ben Taieb S, Le Borgne YA, (2013). Machine Learning Strategies for Time Series Forecasting. In: Aufaure MA., Zimányi E. (eds) Business Intelligence. eBISS, (2012) Lecture notes in business information processing, vol 138. Springer, Berlin, Heidelberg

Bowen HR (1953) Social responsibilities of the businessman. Ethics and economics of society, New York, Harper

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

Breiman L (1996) Bagging predictors. Mach Learn 24(2):123–140

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

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Brownlee J (2020) Introduction to Time Series Forecasting With Python, Mach Learn Mastery

Carannante M, D'Amato V (2021) Weather Index Insurance by satellite data: the basis risk measurement, presented at Energy Finance 2021

Cerin P, Dobers P (2001) What does the performance of the Dow Jones sustainability group index tell us? Eco-management and auditing. J Corp Environ Manag 8(3):123–133

Chatterji AK, Durand R, Levine DI, Touboul S (2016) Do ratings of firms converge? Implications for managers, investors and strategy researchers. Strategic Manag J 37(8):1597–1614

Chava, S., Environmental Externalities and Cost of Capital (June 15, 2011). Available at SSRN: https://ssrn.com/abstract=1677653 or http://dx.doi.org/10.2139/ssrn.1677653

Cohen B, Winn MI (2007) Market imperfections, opportunity and sustainable entrepreneurship. J Bus Ventur 22(1):29–49

Cremers M, Nair VN (2005) Governance mechanism and equity prices. J Financ 60(6):2859–2894

Cui Y, Geobey S, Weber O, Lin H (2018) The impact of green lending on credit risk in China. Sustain 10(6)

Derwall J, Guenster N, Bauer R, Koedijk K (2005) The eco-efficiency premium puzzle. Financ Anal J 61(2):51–63

Sustainable Investing: Establishing Long term Value and Performance. 2012

Dietterich TG (2002) Machine Learning for Sequential Data: A Review, : Caelli T., Amin A., Duin R.P.W., de Ridder D., Kamel M. (eds) Structural, Syntactic, and Statistical Pattern Recognition. SSPR, Lecture Notes in Computer Science, vol 2396. Springer, Berlin, Heidelberg

Drempetic S, Klein C, Zwergel B (2020) The influence of firm size on the ESG Score: corporate sustainability ratings under review. J Bus Ethics 167:333–360

EBA (2020). Discussion paper on management and supervision of ESG risks for credit institutions and investment firms. Consultation Paper (EBA/DP/2020/03)

EBA (2020). Risk Assessment of the European Banking System. Available at https://www.eba.europa.eu/sites/default/documents/files/document_library/Risk%20Analysis%20and%20Data/Risk%20Assessment%20Reports/2020/December%202020/961060/Risk%20Assessment_Report_December_2020.pdf

EBF (2021). EBF Response to the Discussion Paper on management and supervision of ESG risks for credit institutions and investment firms

European Investment Bank (2018). Sustainability Reporting Disclosures In accordance with the GRI Standards. https://www.eib.org/attachments/documents/gri_standards_2018_en.pdf. [Online; accessed 30-05-2018]

Friede G, Busch T, Bassen A (2015) ESG and financial performance: aggregated evidence from more than 2000 empirical studies. J Sustain Financ Invest 5(4):210–233

Garcia F, González-Bueno J, Guijarro F, Oliver J (2020) Forecasting the environmental, social, and governance rating of firms by using corporate financial performance variables: a rough set approach. Sustain 12:3324

Gu S, Kelly B, Xiu D (2020) Empirical asset pricing via machine learning. Rev Financ Stud 33:2223–2273

Gunning D, Stefik M, Choi J, Miller T, Stumpf S, Yang GZ (2019) XAI-Explainable artificial intelligence, Sci Robot

Hachenberg B, Schiereck D (2018) Are green bonds priced differently from conventional bonds? J Asset Manag 19(6):371–383

Hacker P, Krestel R, Grundmann S, Naumann F (2020) Explainable AI under contract and tort law: legal incentives and technical challenges. Artif Intell Law 28:415–439

Hamilton S, Jo H, Statman M (1993) Doing well while doing good? The investment performance of socially responsible mutual funds. Financ Anal J 49(6):62–66

Hartzmark SM, Sussman AB (2019) Do investors value sustainability? A natural experiment examining ranking and fund flows. J Financ 74(6):2789–2837

Harvard Law School Forum on Corporate Governance (2017). ESG Reports and Ratings: What They Are, Why They Matter. https://corpgov.law.harvard.edu/2017/07/27/esg-reports-and-ratings-what-they-are-why-they-matter/. [Online; accessed 27-07-2017]

James G, Witten D, Hastie T, Tibshirani R (2017) An Introduction to Statistical Learning: with Applications in R. Springer Texts in Statistics. ISBN 10: 1461471370

Jewell J, Livingston M (1998) Split ratings, bond yields, and underwriter spreads. J Financ Res 21(2):185–204

Joliet R, Titova Y (2018) Equity SRI funds vacillate between ethics and money: An analysis of the funds stock holding decisions. J Bank Financ 97:70–86

Kreander N, Gray R, Power D, Sinclair C (2005) Evaluating the performance of ethical and non-ethical funds: a matched pair analysis. J Bus Financ Account 32(7/8):1465–1493

Kumar E, Venkatasubramanian S, Scheidegger C, Friedler S (2020) Problems with Shapley-value-based explanations as feature importance measures. Computer Science-Artif Intell arXiv:2002.11097

In SY, Rook D, Monk A (2019) Integrating alternative data (Also Known as ESG Data) in investment decision making. Global Econ Rev 48(3):237–260

Lansilahti S (2012) Market reactions to Environmental, Social, and Governance (ESG)-news: Evidence from European Markets, WP AAlto University

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

Limkriangkrai M, Koh S, Durand RB (2017) Environmental, social, and governance (ESG) profiles, stock returns, and financial policy: Australian evidence. Int Rev Financ 17(3):461–471

Lin WL, Law SH, Ho JA, Sambasivan M (2019) The causality direction of the corporate social responsibility-Corporate financial performance Nexus: Application of Panel Vector Autoregression approach. North Am J Econ Financ 48:401–418

Lins KV, Servaes H, Tamayo A (2017) Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis. J Financ 72(4):1785–1824

Loh WY (2011) Classification and regression trees. Data Mining and Knowledge Discovery, Wiley Interdisciplinary Reviews

Mahjoub LB, Khamoussi H (2012) Environmental and Social Policy and Earning Persistence. Bus Strategy Environ, 22(3). https://doi.org/10.1002/bse.1739

Mahler D, Barker J, Belsand L, Schulz O (2009) Green winners: the performance of sustainability focused companies during the financial crisis. Technical report, A.T., Kearney

Makni R, Francoeur C, Bellavance F (2008) Causality between corporate social performance and financial performance: Evidence from Canadian firms. J Bus Ethics 89(3):409

Melas N, Padmakar K (2018) Factor investing and ESG integration. Expert Syst Appl. https://doi.org/10.1016/B978-1-78548-201-4.50015-5

Merrick L, Taly A (2019) The explanation game: Explaining machine learning models with cooperative game theory. arXiv preprint arXiv:1909.08128

Monk A, Prins M, Rook D (2019) Rethinking alternative data in institutional investment. J Financ Data Sci 1(1):14–31

Moritz B, and Zimmermann T (2016) Tree-based conditional portfolio sorts: The relation between past and future stock returns. Working Paper, Ludwig Maximilian University of Munich

Nakao Y, Amano A, Matsumura K, Genba K, Nakano M (2007) Relationship between environmental performance and financial performance: an empirical analysis of Japanese corporations. Bus Strategy Environ 16:106–118

Nami S, Shajari M (2018) Cost-sensitive payment card fraud detection based on dynamic random forest and k-nearest neighbors, Expert Syst Appl, 110: 381–392. ISSN 0957-4174

Oshiro M., Santoro Perez P., 2012, Baranauskas J. A., How Many Trees in a Random Forest?, Machine Learning and Data Mining in Pattern Recognition, Springer

Oztemel E (2003) Yapay Sinir Aglari (Artificial Neural Networks). Papatya Publisher, Istanbul

Petitjean M (2019) Eco-friendly policies and financial performance: Was the financial crisis a game changer for large us companies? Energ Econ 80:502–511

Principles for Responsible Investment (2016). A practical guide to ESG integration for equity investing. United Nations. Available on line at: https://www.unpri.org/listed-equity/a-practical-guide-to-esg-integration-for-equity-investing/10.article

Quinlan JR (1986) Induction of decision trees. Mach Learn 1:81–106

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

Renneboog L, Ter Horst J, Zhang C (2008) Socially responsible investments: institutional aspects, performance, and investor behavior. J Bank Financ 32(9):1723–1742

Searcy C, Elkhawas D (2012) Corporate sustainability ratings: an investigation into how corporations use the Dow Jones Sustainability Index. J Clean Prod 35:79–92

Schaltegger S, Figge F (2000) Environmental shareholder value: economic success with corporate environmental management. Eco-Manag Audit: J Corp Environ Manag 7(1):29–42

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

Simpson WG, Kohers T (2002) The link between corporate social and financial performance: Evidence from the banking industry. J Bus Ethics 35:97–109

Statman M (2000) Socially responsible mutual funds. Financ Anal J 56(3):30–39

Takeuchi L, Lee YYA (2013) Applying Deep Learning to Enhance Momentum Trading Strategies in Stocks. Working paper, Stanford University. http://cs229.stanford.edu/proj2013/TakeuchiLee-ApplyingDeepLearningToEnhanceMomentumTradingStrategiesInStocks.pdf

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

Trucost and Mercer (2010). Carbon Counts: Assessing the Carbon Exposure of Canadian Institutional Investment Portfolios. Posted: Sept 1, 2010

Tu JV (1996) Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. J Clin Epidemiol 49(11):1225–1231

Utz S, Wimmer M (2014) Are they any good at all? A financial and ethical analysis of socially responsible mutual funds. J Asset Manag 15(1):72–82

Van de Velde E, Vermeir W, Corten F (2005) Corporate social responsibility and financial performance. Corp Gov 5:129–138

Wang S, and Luo Y (2012) Signal Processing: The Rise of the Machines. Deutsche Bank Quantitative Strategy (5 June)

Wang S, Luo Y (2014) Signal Processing: The Rise of the Machines III. Deutsche Bank Quantitative Strategy

WBCSD (2019) ESG Disclosure Handbook. World Business Council for Sustainable Development, Available on line at (https://docs.wbcsd.org/2019/04/ESG_Disclosure_Handbook.pdf)

Weber O (2013) Measuring the Impact of Socially Responsible Investing. Working paper. Available at SSRN: https://ssrn.com/abstract=2217784 or http://dx.doi.org/10.2139/ssrn.2217784

Weber O (2017) Corporate sustainability and financial performance of Chinese banks. Sustain Acc, Manag Policy J, 8(3): 358-385. https://doi.org/10.1108/SAMPJ-09-2016-0066

Weber O, Koellner T, Habegger D, Steffensen H, Ohnemus P (2008) The relation between the GRI indicators and the financial performance of firms. Progress Ind Ecol, Int J 5(3):236–254

Weber O, Mansfeld M, Schirrmann E (2010) The Financial Performance of SRI Funds Between 2002 and 2009 (June 25, 2010). Available at SSRN: https://ssrn.com/abstract=1630502 or http://dx.doi.org/10.2139/ssrn.1630502

Windolph SE (2011) Assessing corporate sustainability through ratings: challenges and their causes. J Environ Sustain 1(1):61–80

Winn M, Pinkse J, Illge L (2012) Case studies on trade-offs in corporate sustainability. Corp Soc Responsib Environ MGMT 19:63–68. https://doi.org/10.1002/csr.293

Zerbib OD (2019) The effect of pro-environmental preferences on bond prices: Evidence from green bonds. J Bank Financ 98:39–60

Xie J, Nozawa W, Yagi M, Fujii H, Managi S (2019) Do environmental, social, and governance activities improve corporate financial performance? Bus Stratergy Environ 28:286–300

Xu X, Chen W (2017) Implementation and Performance Optimization of Dynamic Random Forest, International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), Nanjing, China, 2017: 283-289

Springer