Fundamental ratios as predictors of ESG scores: a machine learning approach

3 authors:

Valeria D'Amato
Università degli Studi di Salerno
76 PUBLICATIONS **467** CITATIONS

SEE PROFILE

Rita laura d'ecclesia
Sapienza University of Rome
61 PUBLICATIONS **506** CITATIONS

SEE PROFILE

Susanna Levantesi
Sapienza University of Rome
77 PUBLICATIONS **798** CITATIONS

SEE PROFILE

Fundamental ratios as predictors of ESG scores: a machine learning approach

Valeria D'Amato · **Rita D'Ecclesia** · **Susanna Levantesi** ·

Abstract Sustainable and responsible finance incorporates Environmental, Social, and Governance (ESG) principles into business decisions and investment strategies. In recent years investors have rushed to Sustainable and Responsible Investments in response to growing concerns about the risks of climate change. Asset managers look for some assessment of sustainability for guidance and benchmarking, for instance, $30 trillion of assets are invested using some ESG ratings. Several studies argue that good ESG ratings helped to prop up stock returns during the 2008 Global Financial Crisis (Lins et al., 2017). The ESG score represents a benchmark of disclosures on public and private firms, it is based on different characteristics which are not directly related to the financial performance (Harvard Law School Forum on Corporate Governance, 2017). The role of ESG ratings and their reliability have been widely discussed (Berg et al., 2019). Sustainable investment professionals are unsatisfied with publicly traded companies' climate-related disclosure. This negative sentiment is particularly strong in the US, and within asset managers who do not believe that markets are consistently and correctly pricing climate risks into company and sector valuations. We believe that ESG ratings, when available, still affect business and finance strategies and may represent a crucial element in the company's fundraising process and on shares returns. We aim to assess how structural data as balance sheet items and income statements items for traded companies affect ESG scores. Using the Bloomberg ESG scores, we investigate the role of structural variables adopting a machine learning approach, in particular, the Random Forest algorithm. We use balance sheet data for a sample of the constituents of the Euro Stoxx 600 index, referred to the last decade, and investigate how these explain the ESG Bloomberg ratings. We find that financial statements items represent a powerful tool to explain the ESG score.

Valeria D'Amato
University of Salerno

Rita D'Ecclesia
Sapienza University of Rome

Susanna Levantesi
Sapienza University of Rome
Viale Regina Elena 295-G, 00161 Roma, Italy
Tel.: +39-06-49255303
E-mail: susanna.levantesi@uniroma1.it

1 Introduction

The burgeoning interest in Environmental, Social, and Corporate governance investments, on short ESG, Socially Responsible Investments (SRIs), and Corporate Social Responsibility (CSR) by private and institutional investors, asset owners, society, central bankers, and researchers increasingly encourage firms to pursue non-monetary goals.
The confusion (Berg et al., 2019) about the topic of the ESG investments belongs to the lack of a standard terminology or taxonomy. The European Commission on June 2021 published the Taxonomy Rules on Reporting by Companies and Financial Institutions that are now Available. Companies ('non-financial undertakings') and financial institutions ('financial undertakings') need to do their Taxonomy-related reporting. This is a key change compared to previous versions and recommendations Both financial and non-financial undertakings start reporting on only a small subset of the requirements in 2022 and becoming totally effective in 2023. In this new context companies will report their own Taxonomy alignment, and the latest rules lay down the KPIs, content and presentation in detail. At the moment the ESG factors represent the fundamentals for informing Social Investor decisions "alongside financial criteria, as the basis for an investment" (Drempetic et al., 2020). Broadly speaking, Sustainable Investing, sometimes known as SRI involves both financial returns and moral values in investments decisions (S&P, 2020). According to this perspective, the financial returns downgrade to a secondary consideration in decision-making, after the investors' ethical values have been accounted for. The historical root of the SRI can be found in the conceptual paradigm of CSR, where corporate behavior begins to be influenced by social expectations. The role of executives and the discussion about the specific social responsibilities of companies began appearing in the literature (Carroll, 1999).

The regulations impose the mandatory disclosure of sustainable activity in some countries such as China, Denmark, Malaysia, and South Africa, however, the lack of uniform standards in the measurement of ESG efforts does not allow comparability of ESG scores, the disclosure being voluntary in other parts of the world. According to the European Sustainable Investment Forum (Eurosif), "Sustainable and responsible investment (SRI) is 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 to better capture long-term returns for investors, and to benefit society by influencing the behavior of companies." Sustainable investing assets in the world stood at $30.7 trillion at the start of 2018, showing an increase of 34% in only two years. In Europe according to Morningstar flows into sustainable companies more than doubled in 2019 reaching €120 bn. Europe has committed to huge investments in sustainability. In 2018, the European Investment Bank (2018) contributed for EUR 55 trillion in green projects launching the first Sustainability Awareness Bond, which is appealing to investors thanks to the same risk-return profiles of its conventionally counterpart (Hachenberg and Schiereck, 2018). The new green and sustainable debt market is principally composed of green bonds which offer a steady income and regulated contracts. Their popularity in the last years grew quickly accompanied by the emergence of new debt products as green loans and sustainability bond, just to mention the more popular.

Investing green signals the adoption of sustainable behaviors that reflect pro-environmental preferences, Zerbib (2019) shows the small negative premium difference between green

bonds and the conventional ones, which does not discourage investors to choose a sustainable product. It also signals a tendency to implement eco-friendly social policies that perform weakly better than non responsible ones even during period of low trust on markets (Petitjean, 2019). The proposed metrics to measure the environmental or social commitment is rarely comparable with financial accounting tools or well known impact analysis systems used in areas such as environmental, social or sustainability impact assessment. Yet developing a standard measure for social or environmental impact similar to standard financial tools would foster the development of SRI, giving investors and fund managers a tool to assess and compare the impact of their investments (Weber, 2013; Joliet and Titova, 2018; Hartzmark and Sussman, 2019). Asset managers look for some assessment of sustainability for guidance and benchmarking, for instance $30 trillion of assets are invested using some ESG ratings. Several studies argue that good ESG ratings helped to prop up stock returns during the 2008 global financial crisis (Lins et al. (2017)). Several institutions, including the well known rating agencies, have started to build sustainability ratings or corporate social responsibility ratings-also named ESG scores. 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 created ESG ratings, ESG indexes and ESG analytics. ESG Ratings aims to help investors to identify and quantify the ESG risks and opportunities, ESG Indexes provide institutional investors with indexes that can be used to manage and report on ESG mandates or as benchmarks to measure ESG investment performance. Bloomberg launched the Bloomberg ESG Data Service which 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 becoming a benchmark of disclosures on public and private firms, it is based on different characteristics which are not directly related to the financial performance (Harvard Law School Forum on Corporate Governance, 2017). It is obtained 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 CSR projects. Sustainable investment professionals and asset managers do not believe that markets are consistently and correctly pricing climate risks into company and sector valuations. 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. However, the accuracy of the existing ESG scores is widely questioned.

The fundamental issue relates to the ability of this tool to effectively discriminate between responsible and irresponsible firms. Searcy and Elkhawas (2012) explore the use of the Dow Jones Sustainability Index (DJSI), which evaluates the corporate sustainability performance of companies trading publicly, in the Canadian corporations. Antolín-López et al. (2019) systematically review the literature on Corporate Sustainability Performance Measurement (CSPM) to identify the most relevant instruments. They find both quantitative and qualitative differences in the CSPM instruments considered (Kinder, Lydenberg and Domini (KLD), DJSI, United Nations Global Compact (UNGC) and Global Reporting Initiative (GRI) among others).

Various firm-level attributes are likely to affect firm CSR participation, and understanding these effects is essential, for instance recent studies investigates how companies can more likely engage in CSR activities and analyze 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), Lin et al. (2019). These studies deal with heterogeneous data, as well as vague and uncertain (Garcia et al., 2020). The use of Rough Set Theory is proposed, by extracting the information from this context, which is not possible utilizing 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).

In this paper we want to relate ESG scores to structural information of the company. We choose to use the Bloomberg ESG scores to investigate the roles of structural variables as financial statements items on the ESG scores of a sample of the constituents of the STOXX Europe 600 index. We adopt a machine learning approach - the Random Forest algorithm - to detect which component explains the ESG Bloomberg ratings. We find that financial statements 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, Section 3 outlines the methodology we propose, describing the regression tree architecture, the random forest algorithm and the variable importance. In Section 4 we describe the empirical framework by illustrating the results and implications. Finally, Section 5 concludes.

2 Literature Review

In the past, the business community engaged in social issues, nevertheless the concept of CSR just recently emerged (Bowen, 1953). Most of the research on SRI studies the business case for sustainability rather than the sustainability case for business (Winn et al., 2012). 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). The sustainability risks is often measured using sustainability ratings of corporate securities or other investment opportunities. Therefore 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.

The bulk of research aimed to assess the impact of CSR investments on economic growth or on corporate financial performances is vast, however the lack of a precise set of universally recognized environmental and social variables makes the assessment of CSR activities quite complex. In the corporate reporting landscape, developments in reporting frameworks,

codes, rules and practices are committed to driving better alignment of sustainability. No universally agreed objectives, standards and thresholds for external ESG information disclosure determine significant differences in the quality of information, the adoption of guidelines depending on the specific choices of the companies. The current inadequacy of disclosures about ESG risks and opportunities outside the company's operational boundary has been stressed by World Business Council for Sustainable Development (WBCSD, 2019).

The role of ESG ratings and their reliability has been widely discussed (Berg et al., 2019). According to Berg et al. (2019) the ambiguity around ESG ratings is an impediment to prudent decision making that would contribute to an environmental, sustainable and socially just economy. Chatterji et al. (2016) find that ratings from different providers differ dramatically, showing that information received from rating agencies is quite noisy. According to Berg et al. (2019) the confusion generates three major consequences. 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. In this context investigating the relationship existing between structural data as financial statements items and the existing ESG scores may provide useful information to assess the accuracy of the score. We believe that without a reliable measure of "accurate ESG performance" to understand how structural indexes relate to the existing ESG ratings may provide important disclosure on the company's sustainable activity.

Application of machine learning techniques in finance has become quite common thanks to the large data set available nowadays. Many applications deal with the use of machine learning for stock selection. Wang et al. (2012) use the AdaBoost algorithm to forecast equity returns and Wang et al. (2014) show that using different training windows provide 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 model

Given a generic regression model for estimating the relationship between a target (or response) variable, Y, and a set of predictors (or features), $X_1, X_2, ..., X_p$:

$$Y = f(X_1, X_2, ..., X_p) + \varepsilon \tag{3.1}$$

where ε is the error term. The quantity $E(Y - \hat{Y})^2$ represents the expected squared prediction error, that can be written as

$$E[f(X_1, X_2, ..., X_p) - \hat{f}(X_1, X_2, ..., X_p)]^2 + E[\varepsilon]^2 \tag{3.2}$$

or the sum of the reducible error and irreducible error. Machine learning techniques aim at estimating f by minimizing the reducible error. Recent researches are using more and more models that can dynamically learn from past data. Simple regression techniques are not successful mostly because financial data is inherently noisy, in many cases the presence of multicollinearity affects the results, and relationships between factors and returns can be variable, non-linear and/or contextual. So estimating dynamic relationships between potential predictors and the target variable result quite complex. We believe machine learning algorithms (MLAs) can provide a better approach offering a natural way to combine many weak sources of information into a composite ESG score is stronger than any of its sources. Several machine learning algorithms have been developed, i.e. the gradient boosted regression trees, artificial neural networks, random forests, and support vector machines. They have proven to be able to uncover complex patterns and hidden relationships that are often difficult or impossible to detect with linear analysis, and in presence of multicollinearity they are more effective than linear regression. Machine learning approach has the ability to precisely classify observations. At the top of the classifier hierarchy is the random forest classifier that belongs to the family of ensemble methods. It is useful to get the error reduction pulling down the prediction variance, preserving the bias, i.e. the difference between the model's prediction and the true value of the target variable.

3.1 Regression tree architecture

The random forest algorithm is founded on the regression tree architecture. The regression trees allow to get the best function approximation $\hat{f}(X_1, X_2, ..., X_p)$ through a procedure consisting in the following steps (Loh (2011)):

- The predictor space (i.e. the set of possible values for $X_1, X_2, ..., X_p$) is divided into J distinct and non-overlapping regions, $R_1, R_2, ..., R_J$.
- For each observation that falls into the region R_j, the algorithm provides the same prediction, which is the mean of the response values for the training observations in R_j.

As described in James et al. (2017)), the basic idea is to divide the predictor space into high dimensional rectangles, finding the boxes $R_1, ..., R_J$ that minimize the Residual Sum of Squares (RSS):

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

Once the regions $R_1, ..., R_J$ have been created, the response is predicted for a given test observation using the mean of the training observations in the region to which that test observation belongs. The consideration of all the possible partitions of the feature is computationally infeasible, therefore we consider a top-down approach (Quinlan (1986)) by using a recursive binary splitting: the algorithm starts at the top of the tree, where all observations belong to a single region, and then successively splits the predictor space. The best split is identified according to the entropy or the index of Gini that is a purity measurement (homogeneity) for each node. The maximum purity is reached when only one class of Y is present in the node. Breiman (2001) has listed the most interesting properties of regression tree-based methods. They belong to non-parametric methods able to catch tricky relations between inputs and outputs, without involving any a-priori assumption. They handle heterogeneous data and intrinsically implement feature selection, making them robust to not significant or noisy variables. Finally, they are robust to outliers or missing values and are easily interpretable.

3.2 Random Forest

The fundamental concept behind random forest is the wisdom of crowds: a large number of relatively uncorrelated models (trees) operating as a committee will outperform any of the individual constituent models. The low correlation between models is the key to produce ensemble predictions that are more accurate than any of the individual predictions. The trees protect each other from their individual errors (as long as they don't constantly all err in the same direction. While some trees may be wrong, many other trees will be right, so as a group the trees can move in the correct direction. Random forest performs well when i) there are some actual signals in the features so that models built using those features do better than random guessing; ii) predictions (and therefore the errors) made by the individual tree must have low correlations with each other. The Random Forest (RF) technique basically consists of building an ensemble of decision trees grown from a randomized variant of the tree. Starting from a single learning set, the basic idea is to introduce a random perturbation into the learning procedure to introduce a differentiation among the trees and combine the predictions of all these trees using aggregation techniques. Breiman (1996) proposed a first aggregation method, the so-called bagging, in which different trees are built by using random bootstrap copies of the original data. Its natural evolution, the random forest, has been developed by the same author in 2001 Breiman (2001).

In the random forests, the bagging approach has been extended and combined with randomization of the input variables that are considered as candidate variables to split internal nodes t. In particular, instead of looking for the best split $s_t = s^*$ among all variables, the RF algorithm chooses a random subset of K variables for each node and then determines the best split using these variables. The RF estimator of the target variable $\hat{y}_{R_j}$ is function of the regression tree estimator, $\hat{f}^{tree}(\mathbf{X}) = \sum_{j \in J} \hat{y}_{R_j} \mathbf{1}_{\{\mathbf{X} \in R_j\}}$, where $\mathbf{X} = X_1, X_2, ..., X_p$ is the vector of the predictors, $\mathbf{1}_{\{.\}}$ represents the indicator function and $(R_j)_{j \in J}$ are the regions of the predictors space obtained by minimizing RSS. The regression tree estimator is identified by the average values of the variable belonging to the same region R_j. Therefore, denoting the number of bootstrap samples by B and the decision tree estimator developed on the sample $b \in B$ by $\hat{f}^{tree}(\mathbf{X}|b)$, the RF estimator is defined as follows:

$$\hat{f}^{RF}(\mathbf{x}) = \frac{1}{B} \sum_{b=1}^{B} \hat{f}^{tree}(\mathbf{X}|b) \tag{3.3}$$

The choice of the number of trees to include in the forest should be done carefully, in order to reach the highest percentage of explained variance and the lowest mean of squared residuals (MSR).

3.3 Variable importance

ML algorithms are usually viewed as a black box, as a large number of trees makes the understanding of the prediction rule hard. To get from the algorithm interpretable information on the contribution of different variables we follow the common approach consisting in the calculation of the variable importance measures.

Variable importance is determined according to the relative influence of each predictor, by measuring the number of times a predictor is selected for splitting during the tree building process, weighted by the squared error improvement to the model as a result of each split, and averaged over all trees. According to the definition provided by Breiman (2001), the RF

variable importance is a measure providing the importance of a variable in the RF prediction rule.

A weighted impurity measure has been proposed in Breiman (2001) for evaluating the importance of a variable X_m in predicting the target Y, for all nodes t averaged over all N_T trees in the forest. Among the variants of the variable importance measures, we consider to the Gini importance, obtained by assigning the Gini index to the impurity $i(t)$ index. This measure is often called Mean Decrease Gini, *MDG*:

$$MDG(X_m) = \frac{1}{N_T} \sum_{T} \sum_{t \in T: v(s_t) = X_m} p(t) \Delta i(s_t, t) \tag{3.4}$$

where $v(s_t)$ is the variable used in split 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:

$$\Delta i(s_t, t) = i(t) - \frac{N_{t_l}}{N_t} \cdot i(t_l) - \frac{N_{t_r}}{N_t} \cdot i(t_r) \tag{3.5}$$

where N is the sample size, $p(t) = \frac{N_t}{N}$ the proportion of samples reaching t, and $p(t_l) = \frac{N_{t_l}}{N}$ and $p(t_r) = \frac{N_{t_r}}{N}$ are the proportion of samples reaching the left node t_l and the right node t_r respectively. *MDG* presented in eq. 3.4 calculates the importance of each variable X_m as the sum over the number of splits that includes the variable, proportionally to the number of samples it splits. The analysis of the model's features importance can offer more intuition into the algorithm learning process.

4 Empirical Analysis

We first provide a description of the data used and their statistical features. We then proceed to identify the drivers of the Bloomberg ESG scores using a random forest approach. The performance of the used algorithm is assessed by comparing the results obtained using a classical Generalized Linear Model (GLM).

4.1 Outlook and motivation about the data choice

A large number of rating agencies composes the complex ESG landscape providing a wide array of data. Due to the fast growth of the ESG ratings marketplace and dynamism in driving merger and acquisition activity (recent examples include Moody's acquisition of a majority stake in Vigeo Eiris, as well as S&P Global's purchase of the ESG business of RobecoSAM, including its well-known Corporate Sustainability Assessment, connected to the Dow Jones Sustainability Index), a comprehensive overview of data providers as offered by some authors (i.e, for instance, Douglas et al. (2017)) become outdated quickly. Li and Polychronopoulos (2020) identified 70 different firms that provide some sort of ESG rating data by excluding investment banks, government organizations, and research organizations that conduct ESG-related research. Wong and Petroy (2020) estimate 600+ ESG ratings and rankings existing globally as of 2018, which is a number that continues to grow according to their report. As stressed by the European Banking Federation response (EBF, 2021), different ESG rating providers lead to a significant distortion in counterparties ESG risk assessment, being a priority for the providers and the whole market to come to an agreement on best practices and to become as transparent as possible about the reliability of

their data. To better understand the different types of ESG rating providers, we can refer to a three-tiered classification as proposed by Li and Polychronopoulos (2020): fundamental, comprehensive, and specialist. In the category of the fundamental are included ESG data providers that collect and aggregate publicly available data. Refinitiv (formerly, Thomson Reuters) and Bloomberg are examples of fundamental providers. In the comprehensive category, which corresponds to the majority of the providers, they "utilize a combination of objective and subjective data covering all ESG market segments. Typically these data providers will develop their own ratings methodology and combine publicly available data as well as data produced by their own analysts through company interviews/questionnaires and independent analysis" (Li and Polychronopoulos, 2020). The comprehensive providers' category can include MSCI, Sustainalytics, Vigeo Eiris, ISS, TruValue Labs, and RepRisk. The category called specialist consists of ESG data providers with specific expertise, that "specialize in a specific ESG issue, such as environmental/carbon scores, corporate governance, human rights, or gender diversity". In this category, for instance, we can insert TruCost (now owned by S&P Global), the nonprofit Carbon Disclosure Project (CDP), and Equileap (gender equality data). In our opinion, the advantage of using comprehensive environmental, social and governance (ESG) data consists in having exhaustive information. In the case of Bloomberg, the platform is easy to use and provides over 900 fields of information captured with up to 10 years of history and the high-quality of data coming from corporate responsibility reports, annual reports and specific ESG releases, proxy statements, and corporate governance reporting can add useful input to the ESG disclosure process. In addition Bloomberg emerges as mostly used by all the investors, as pointed out in the report "Rate Raters 2020" by SustainAbility. The report illustrates that Bloomberg is included in the group of providers that had the highest n-values for the investor survey.

4.2 The dataset

We use financial statements items and the Bloomberg ESG scores collected for the constituents of the STOXX Europe 600 Index which represents large, mid, and small-capitalization companies across 17 countries of the European region. Besides countries from the Eurozone, like France, Germany, Spain, the Netherlands, and Italy, you also have exposure to Great Britain, Switzerland, and Scandinavian countries. Due to its broad market exposure, the STOXX Europe 600 index is often quoted as the European equivalent of the U.S.-focused S&P 500 index. We selected a sample of 109 companies that have been included in the STOXX Europe 600 Index throughout the chosen period (2014-2018). This ensures that the sample universe remains unchanged over the reference time period. The 109 companies represent the 21% of the entire set of companies included in the index and belong to the following four industry sectors: Communications (1), Energy (2), Technology (3), Utilities (4), which proportions are illustrated in Fig. 4.1, panel a. These sectors have been quite active in terms of innovation in the last decade and result more exposed to choices in the sustainable investments. The Energy sector and the Utilities have long played a key role in the energy transition by contributing to a secure energy supply and decarbonization. Companies in the Energy and Utilities sectors are more and more forced to systematically pursue their transition to affordable, reliable, and more sustainable energy and to set ambitious goals in terms of reducing CO_2 emissions. At the same time, it is becoming increasingly important to integrate sustainability concepts within companies and governance structures and not only to understand the effects of one's own actions but also to measure them and ensure their implementation. The sectors of Technology and Communications provide an

example of sectors that could be on track to generate more carbon emissions than any other industry. The continuous search for new technologies and the advantages it brings to our daily lives often requires the consumption of a lot of energy; for instance, the technology as smartphones, IoT adoption, and big data led to massive growth in the need for data centers to store, manage and transfer all things digital, which also come with an environmental cost. According to Betti et al. (2018), Technology and Communications result the sectors with the fourth highest Average Sector Sustainable Development Goals Impact Index (ASSII)[1] after the Healthcare sector and the Energy and Utilities sectors. The 17 SDGs, ratified by the United Nations on 15 September 2015, have been described as "the closest thing the Earth has to a strategy", these goals set for 2030 are going achieved by the various sectors. Unfortunately, companies in the Healthcare sectors changed in the Euro STOXX index in the examined period so we were not able to include the Healthcare sector in our analysis.

Financial statements items which also include the attributed ESG scores for each company are obtained from Bloomberg. The ESG score is a proprietary Bloomberg score (methodologies are confidential) based on the extent of a company's ESG disclosure. The score is tailored to each industry sector, so that each company is evaluated only on the basis of data that are relevant for its industry sector. The sum is a weighted average of the E, S and G component scores, but the weights applied to each component vary according to the company's industry sector. The score ranges from 0.1 for companies disclosing a minimum amount of ESG data to 100 for those disclosing every ESG-related data point collected by Bloomberg. The companies in our sample have an ESG score, on average, of 43.29 in the period 2014-2018.

Unfortunately, there are no specific financial statements items on the individual components of the ESG score. However, Bloomberg also collects ESG data in a dedicated Terminal, which provides an ESG rank and its individual components: environmental rank, social rank and governance rank for each company. The ESG rank is calculated by equally weighing the ESG components (weights all equal to 1/3). It clearly differs from the ESG score reported in the companies' balance sheets, which weights depend on the company's industry sector. However, the E, S and G components, although not perfectly consistent with the ESG score from the balance sheet data, still provide additional information that can be useful in assessing as the balance sheet data affect the E, S and G scores.

The set of financial statements items that will be used to identify the ESG drivers and that constitute the explanatory variables in our model are listed in the following:

- *Year*: 2014-2018
- *Sector*: numerical variable indicating the company's industry sector (range: 1-4)
- *ESG.Score*: ESG score assigned by Bloomberg
- *Env.Rank*: Environmental score assigned by Bloomberg (ESG Terminal)
- *Soc.Rank*: Social score assigned by Bloomberg (ESG Terminal)
- *Gov.Rank*: Governance score assigned by Bloomberg (ESG Terminal)
- *Sales_to_Assets*: the ratio of Sales to Assets
- *EBIT_to_Sales*: the ratio of Earnings Before Interest and Taxes (EBIT) to Sales
- *DY*: Dividend Yield
- *NI_to_Sales*: the ratio of Net Income (NI) to Sales
- *Price_to_Earnings*: the ratio of Price to Earnings
- *Rating*: the Bloomberg Best Analyst Rating, which is a credit rating assigned by Bloomberg, calculated as a weighted average of opinions of various analysts. It indicates the analyst recommendation and their consensus on a single stock (range: 0-5, where 5 indicates the highest consensus)
- *LR*: liquidity ratio calculated as the ratio of current assets to current liabilities
- *SR*: solvency ratio calculated as he ratio of debt to total assets

---

[1] This index measures the ability of Sustainability Accounting Standards Board (SASB) sectors to impact Sustainable Development Goals (SDGs) in general.

We have considered the overall financial statements of companies by analyzing ratios representing profitability indexes, liquidity analysis, and solvency analysis (e.g., *Sales_to_Assets*, *EBIT_to_Sales* and *NI_to_Sales*). These ratios are more informative to absolute values and contribute to improving the characterization of companies, and to reflect a faithful and explicative understanding of the business activities and the financial performance of a company explaining the ESG scores. We also include a solvency ratio and a liquidity ratio to cover other important balance sheet features in relative terms. *LR* represents the current ratio (current assets/current liabilities), while *SR* the ratio of debt to total assets. Furthermore, the model works with static data including "year" as a separate variable while explicitly disregarding the year-on-year changes of any financial statement item.

Table 4.1 shows the main statistical features of the ESG score distribution and its components, E, S, and G by year, while Table 4.2 shows the descriptive statistics by industry sectors. Looking closer at Table 4.1, the E, S, and G components become more relevant over time showing lower standard deviation as time increases. For each calendar year, we can observe higher G scores that are better ranked than the S and E components. Furthermore, the E component affects the composite ESG score in terms of higher standard deviation. The strong effect of the G component still persists for each sector as illustrated in Table 4.2. The utilities sector shows the highest ESG score, and the lower standard deviation compared to the other sectors. The utilities confirm the best scores for each component, followed by the energy sector characterized by good results but more volatile. The technology sector reports the lowest average ESG score (and its components) compared to the other sectors (ESG=36) and at the same time shows very high standard deviation for both ESG score and its components. This is due to the fact that the technology sector is an energy intensive sector which may not easily become more environmentally friendly. The technology world in some cases generate more carbon emissions than any of the other industries. This explains the lower average ESG score and the more volatile.

Year	ESG score		E score		S score		G score	
	mean	std dev	mean	std dev	mean	std dev	mean	std dev
2014	40.52	17.56	40.29	26.11	58.22	23.03	64.41	16.10
2015	42.62	15.55	43.57	24.90	62.82	18.72	66.45	13.77
2016	43.46	15.22	44.78	23.74	64.20	17.38	67.35	13.92
2017	44.87	14.31	45.65	22.22	66.63	15.18	70.21	12.98
2018	45.00	14.05	45.87	23.32	67.01	14.75	69.42	12.98

Table 4.1: Main statistics of the ESG, E, S and G score distribution by year of the sample of 109 companies included in the STOXX Europe 600 Index: average value and standard deviation.

Sector	ESG score		E score		S score		G score	
	mean	std dev	mean	std dev	mean	std dev	mean	std dev
Communications (1)	39.58	12.23	40.97	17.18	64.91	17.35	69.33	12.54
Energy (2)	47.36	15.42	45.83	25.34	66.44	14.14	68.59	11.21
Technology (3)	36.00	16.27	44.99	30.57	59.37	24.03	63.98	19.63
Utilities (4)	53.46	11.71	45.76	22.79	65.22	13.47	68.46	9.20

Table 4.2: Main statistics of the ESG, E, S and G score distribution by industry sector of the sample of 109 companies included in the STOXX Europe 600 Index: average value and standard deviation.

Fig. 4.1 (panel a) shows the companies' percentage composition by industry sector for the four sectors considered in the analysis. About one-third of the sample is composed of companies in sector 1 (Communications). Sectors 3 and 4 (Technology and Utilities) follow each one constitutes about a quarter of the sample. The remaining percentage concerns sector 2 (Energy), which accounts for 17.4% of all companies included in the Euro STOXX index. Fig. 4.1 (panel b) shows the percentage composition of the credit rating assigned by Bloomberg (Bloomberg Best Analyst Rating), indicating the analyst recommendation and their consensus on a single stock. Score 5 indicates the highest consensus. We observe that scores 3 and 4 accounts for 85% of the total, score 2 for 12.5%, while the remaining scores (0-1 and 5) are residual.

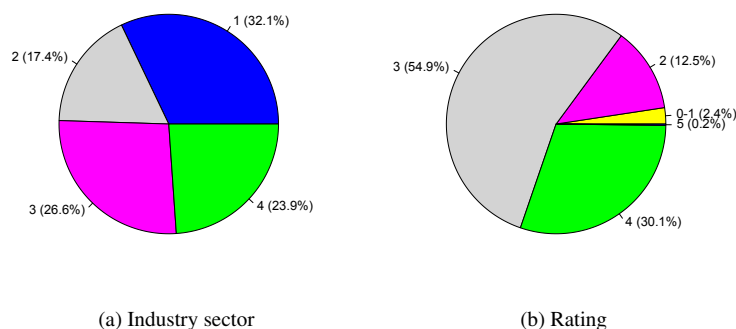(a) Industry sector   (b) Rating

Fig. 4.1: Industry sector and rating of the sample of 109 companies included in the STOXX Europe 600 Index.

All the following analyses, graphs, and tables only refer to the sample of 109 companies included in the STOXX Europe 600 Index and belonging to the industry sectors: communications, energy, technology, utilities. Focusing on the ESG data, we calculate the density functions of the *ESG.Score*, *Env.Rank*, *Soc.Rank* and *Gov.Rank* that are depicted in Fig. 4.2. In the first panel, the *ESG.Score* density is compared to the *ESG.Rank* one (dashed line), i.e. the ESG score obtained by equally weighing the ESG components. We observe that, in our sample, the Environmental and Social components seem to have a higher weight with respect to the Governance.
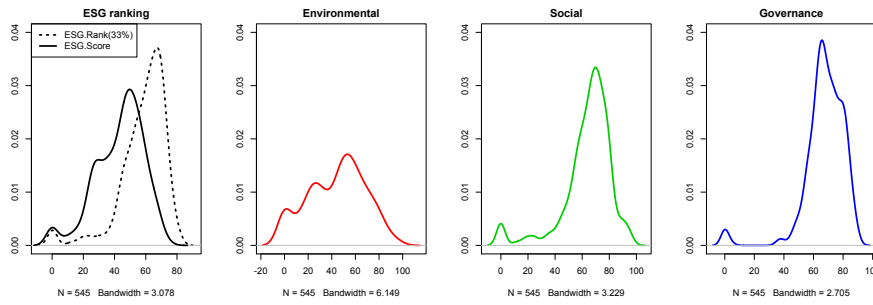
Fig. 4.2: Density functions of ESG variables.

We then estimate the pairwise correlations existing between the ESG score and the set of variables analyzed. We report the variables' correlation in the correlogram of Fig. 4.3. Positive correlations are shown in red while negative correlations in blue. The color intensity is proportional to the correlation coefficient. Noteworthy are the correlations between *LR* and *NI_to_Sales* (0.43), *SR* and *EBIT_to_Sales* (0.34) that are both positive, and those between *ESG score* and *NI_to_Sales* (-0.40), *ESG score* and *Asset_to_Sales* (-0.28) and *LR* and *SR* (-0.36) that are negative. The *NI_to_Sales* measures how much net income is generated as a percentage of sales. It helps investors assess if a company's management is generating enough profit from its sales, in this case we are not considering the impact of costs in this indicator. So an increase in *NI_to_Sales* will not translate in a direct improvement of the ESG score. It means that the increase of net income is not going to be used directly to improve the commitment in environmental or social goals in the sample we considered. Similar interpretations can be given for the *Sales_to_assets* which measures the amount of sales generated by the company's assets and *EBIT_to_Sales* which measures the company's operating profit as a percentage of its sales. It's interesting to see how the Solvency Ratio *SR* is the only variable to be almost non correlated with the ESG score while is positively correlated with the *EBIT_to_Sales* ratio. Total-debt-to-total-assets is a leverage ratio that defines the total amount of debt relative to assets owned by a company. This information can reflect how financially stable a company is. The higher the ratio, the higher the degree of leverage (DoL) and, consequently, the higher the risk of investing in that company. So in presence of increasing *EBIT_to_Sales* an increase in the SR may occur. While the increase of the first negatively affect the ESG scores the increase of the latter will not affect the ESG score. The correlogram detects the linear dependence description among variables which in this case do not provide a correct understanding of the various dynamics while the use of machine learning allowing to capture hidden correlations and non-linear patterns.
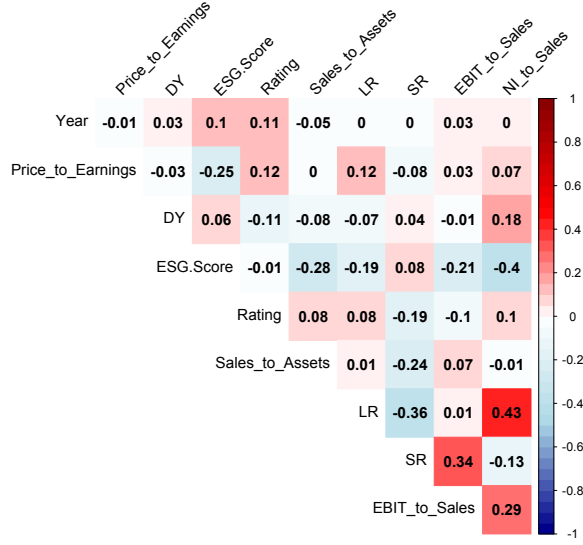
Fig. 4.3: Correlogram of the data set.

4.3 RF estimation of the ESG score

We apply the RF algorithm (Liaw (2018)) using the following model to measure the explanatory capacity of each predictor in explaining the target variable Y.

$$
\begin{aligned}
Y \sim & Year + Sector + Price\_to\_Earnings + DY + Rating + Sales\_to\_Assets \\
& + LR + SR + EBIT\_to\_Sales + NI\_to\_Sales
\end{aligned}
\tag{4.1}
$$

We use first the global ESG score as target variable, Y_{ESG}, and then the single components, Y_E, Y_S and Y_G. We denote as $\widehat{Y}$ the random forest estimator of the target variable. In order to find the optimal parameter setting, we perform the hyper-parameter tuning by initially considering a set of random seeds (100) for the pseudo-random generator and a reasonable number of trees ($ntrees = 500$). The hyper-parameters to be optimized are: $mtry$ representing the number of input variables that are selected at each splitting node and $nodesize$ the minimum size of a terminal node. 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 choose the combination of seed/$mtry$/$nodesize$ producing 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: $mtry$=5 and $nodesize$=1 for

the target variable $Y_{ESG}{}^2$.

The percentage of variance explained by the random forest algorithm, *RSS*, and the level of the resulting *MSR* are given in Table 4.3 for the ESG score and its components.

Indicator	ESG.Score	Env.Rank	Soc.Rank	Gov.Rank
RSS	66.96%	50.60%	45.50%	46.76%
MSR	80.14	286.60	195.70	112.35

Table 4.3: RSS and MSR obtained by the RF algorithm.

The MDG values assigned by the random forest to the predictors are illustrated in Fig. 4.4, sorted decreasingly from top to bottom. This plot allows to identify the most important variables in predicting the ESG score. The most explicative variables selected by the RF algorithm are *NI_to_Sales*, *SR* and *Sales_to_Assets*. It is noteworthy that the variable *Year* shows very low variable importance.
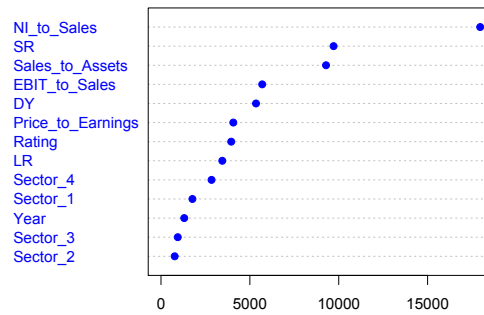

Fig. 4.4: Variable importance for ESG.Score.

Fig. 4.5 shows the partial dependence plots, i.e. the marginal effect of the (most important) predictors on the the target variable averaged over the joint values of the other predictors. 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.

---

[2] The algorithm's parameters are set as follows: *mtry*=5,5,5 and *nodesize*=1,1,1 for Y_E, Y_S and Y_G respectively.
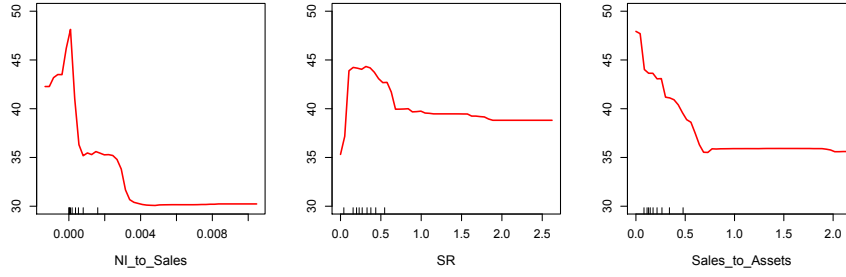
Fig. 4.5: Single variable partial dependence plot for Y_{ESG}. Predictors: *NI_to_Sales*, *SR*, *Sales_to_Assets*.

It is interesting to see how the most important variable results *NI_to_Sales* which according to the linear model had a negative correlation with the ESG score. If we look at the PDP, we observe how an increase in *NI_to_Sales* causes a reduction of the ESG score. This implies that the net profit generated by a unit of sales is not going to generate a better assessment in terms of ESG commitments in the set of companies we examined. The second important variable in explaining ESG dynamics is the *SR*. We may think that companies that have a high leverage ratio may put more effort into moving toward a more sustainable framework. If we look at the single variable PDP we observe how an initial increase of SR (from 0 to 0.35) translates in an improvement of the ESG score; when the debt gets close to 50% of total assets the ESG score worsens to be not affected any more when the SR becomes higher than 1. In this case, companies will have to worry about their financial health more than the sustainability commitment. Finally, the variable *Sales_to_Assets* has the same level of importance as the SR ratio. If we look at the PDP plot we can see how an increase in sales per unit of assets generates a worsening of the ESG score. The formula for the asset turnover ratio evaluates how well a company is utilizing its assets to produce revenue. The numerator of the asset turnover ratio formula shows revenues that are found on a company's income statement and the denominator shows total assets which are found on a company's balance sheet. It should be noted that the asset turnover ratio formula does not look at how well a company is earning profits relative to assets. The asset turnover ratio formula only looks at revenues and not profits. ESG factors can be integrated into the revenue forecasts by increasing or decreasing the company's sales growth rate by an amount that reflects the level of ESG opportunity/risk. "For example, a carmaker may stop selling a particular type of car in a particular country due to environmental concerns, which is estimated to reduce sales by x% annually"(PRI (2016)). Therefore, it is straightforward to expect that the sales amount will be linked to the ESG score.

To deeply investigate the relationship between the *NI_to_Sales* and the Y_{ESG}, we display the values of these two variables in a scatterplot (Fig. 4.6). The points represent the observed values, the red line the locally estimated scatterplot smoothing (LOESS) that is a local weighted (non-parametric) regression used to fit a smooth curve through the points. LOESS regression clearly highlights that *NI_to_Sales* decreases as Y_{ESG} grows, then showing the existence of a negative relationship between them.
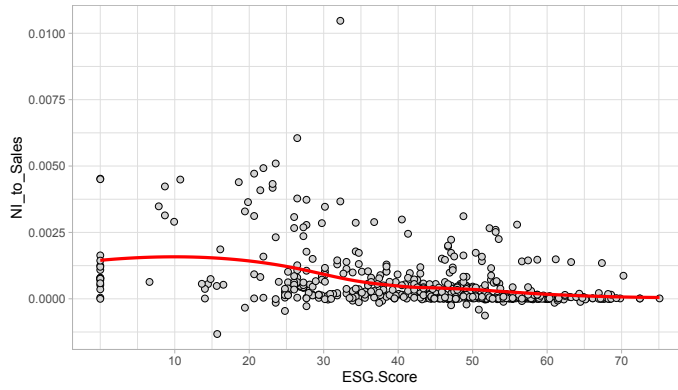
Fig. 4.6: *NI_to_Sales* vs *ESG.Score*. Observed values (points) and LOESS (red line).

4.4 RF predictive performance and comparison with GLM

In this sub-section we compare the predictive performance of the RF algorithm to those obtained by a classical Generalized Linear Model (GLM).

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:

$$\eta = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p \tag{4.2}$$

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, identity for the link function, so that: $\eta = E(Y)$.

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

The GLM performance and the estimate of the regression coefficients are reported in table 4.4, 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.

Coefficient	$\hat{\beta}$	$SE(\hat{\beta})$	z value	$Pr(>|z|)$
(Intercept)	-2.354e+03	8.469e+02	-2.780	0.00568 **
Year	1.189e+00	4.203e-01	2.830	0.00488 **
Price_to_Earnings	-8.138e-02	2.954e-02	-2.755	0.00612 **
DY	1.048e+00	2.578e-01	4.067	5.68e-05 ***
Rating	1.397e+00	8.742e-01	1.598	0.11070
Sales_to_Assets	-1.383e+01	2.494e+00	-5.546	5.16e-08 ***
LR	3.367e-01	8.164e-01	0.412	0.68024
SR	-5.985e+00	3.411e+00	-1.755	0.08004 .
EBIT_to_Sales	-5.533e+00	4.640e+00	-1.192	0.23377
NI_to_Sales	-5.165e+03	6.832e+02	-7.561	2.49e-13 ***
as.factor(Sector)2	6.408e+00	1.870e+00	3.427	0.00067 ***
as.factor(Sector)3	9.276e-02	1.787e+00	0.052	0.95862
as.factor(Sector)4	1.209e+01	1.669e+00	7.242	2.10e-12 ***

Significance codes: ·p<0.1, *p<0.05, **p<0.01, ***p<0.001.

Table 4.4: GLM results for the *ESG.Score* regression model.

The GLM assigns the greatest importance to the predictors *NI_to_Sales*, *Sales_to_Assets*, *DY* and Sectors 2 and 4. This result is partially in line with the RF output, which ascribes the greatest importance to *NI_to_Sales* and *Sales_to_Assets*, but also to *SR*, which is not included in the most significant features by the GLM.

The goodness of prediction is measured by the Root Mean Square Error (RMSE) and Mean Absolute Percent Error (MAPE), respectively defined as:

$$RMSE = \frac{\sum_i (y_i - \hat{y}_i)^2}{n} \tag{4.3}$$

$$MAPE = \frac{100}{N} \sum_i \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{4.4}$$

The resulting values of RMSE and MAPE are shown in Table 4.5 for RF and GLM both in the training and test set: the improvement in the prediction obtained by applying RF with respect to the traditional GLM is strong, reducing MAPE from 4.54 to 3.74 and RMSE from 10.18 to 7.99 in the test set.

Measure	Training set		Test set	
	RF	GLM	RF	GLM
RMSE	8.02	11.02	7.99	10.18
MAPE	2.11	2.50	3.74	4.54

Table 4.5: RMSE and MAPE on *ESG.Score* predicted values.

Fig. 4.7 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.
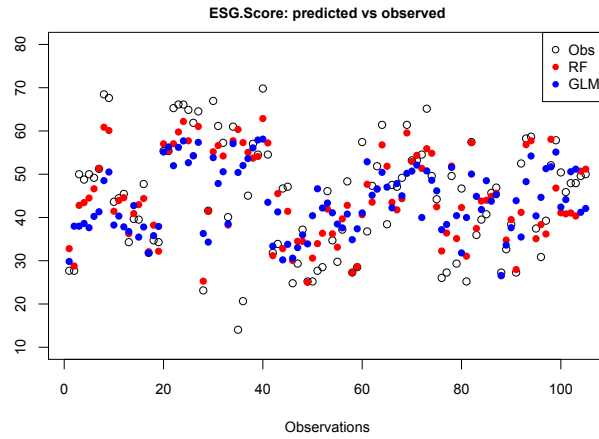
Fig. 4.7: *ESG.Score*: predicted (RF, GLM) versus observed values (Obs).

In addition to Fig. 4.7, we report in Fig. 4.8 the scatterplots between the observed and predicted values of *ESG.Score* data points by GLM (left panel) and RF (right panel) models and the regression error (R^2) values for the two regression models. We can visually appreciate the higher predictive performance of RF compared to the GLM.
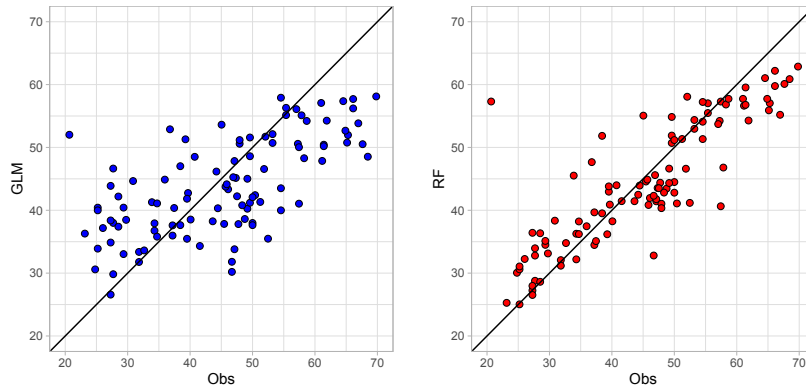

Fig. 4.8: Scatterplot between the observed and predicted values of *ESG.Score*: GLM in the left panel, RF in the right panel. Regression error: R^2=0.4356 (GLM), R^2=0.6181 (RF).

Finally, we provide the results obtained by using every single component of the ESG score as a target variable (Y_E, Y_S and Y_G) in the model in Eq. 4.1. We remind that the Environmental, Social, and Governance components are not equally weighted by Bloomberg in building the ESG score, where weights depend on the company's industry sector. Therefore, it is appropriate to consider a regression model for each component. The single components show proper dynamics (see main statistics in Section 4.2) which are different from that of

the ESG.score and may provide useful information in evaluating as the balance sheet data influence the E, S, and G scores. The results reported in Tables 4.6-4.8 show the values of RMSE and MAPE for the *Env.Rank*, *Soc.Rank* and *Gov.Rank*, respectively. We observe that the relationship between each ESG score component and the balance sheet items appears to be weaker than that which considers the global ESG score as the target variable. The RF model is in any way superior in predicting the score in all the cases.

Measure	*Training set*		*Test set*	
	RF	GLM	RF	GLM
RMSE	14.80	19.01	12.01	18.54
MAPE	2.84	3.44	5.25	6.63

Table 4.6: RMSE and MAPE on *Env.Rank* predicted values.

Measure	*Training set*		*Test set*	
	RF	GLM	RF	GLM
RMSE	11.29	14.02	11.63	12.19
MAPE	1.98	2.21	4.18	4.35

Table 4.7: RMSE and MAPE on *Soc.Rank* predicted values.

Measure	*Training set*		*Test set*	
	RF	GLM	RF	GLM
RMSE	9.04	10.69	7.87	10.34
MAPE	2.06	2.17	2.91	3.43

Table 4.8: RMSE and MAPE on *Gov.Rank* predicted values.

5 Conclusions

The sustainability disclosure regulations have been characterized by a proliferation of reporting rules, frameworks, and metrics aiming to incentivize companies to adopt disclosure practices. The standardization of key metrics and broader transparency in markets could facilitate the process of evaluating firms' ESG attributes by the rating agencies, avoiding the divergence between rating scores which is merely noise, according to the recent literature (Berg et al., 2019).

In this paper, the novelty of our contribution consists in explaining the determinants of the ESG score by performing the random forest algorithm. We use overall financial statements information such as profitability indexes, liquidity and solvency ratios of a subset of companies listed in the STOXX Europe 600 Index to predict ESG score. In particular, in order to improve the characterization of the companies, we set up a ratio analysis which is traditionally a useful management tool enriching the understanding of financial results and trends

over time, and provide key indicators of organizational performance. We have included in the analysis all the companies belonging to the industry sectors communications, energy, technology, and utilities. The numerical results show that financial statements items present a significant predictive power on the ESG score provided by Bloomberg. The RF algorithm achieves the best prediction performance with respect to the classical regression approach based on GLM, demonstrating the ability to capture the non-linear pattern of the predictors. As regards the importance of variables, the algorithm selects *NI_to_Sales* as the most predictive variable. We observe how an increase in *NI_to_Sales* causes a reduction of the ESG score. This implies that the net profit generated by a unit of sales is not going to generate a better assessment in terms of ESG commitments in the set of companies we examined.

According to the paradigm of the "quality analysis", the determinants of the ESG score could be interpreted as responsible for the quality premium (Novy-Marx, 2013) by suggesting a tool for overcoming the lack of consistency when implementing quality. On the basis of the firm fundamentals analysis, our findings show that higher-rated ESG companies are not always higher-quality companies (as stressed in Chen and Deleon (2020)). Also Pedersen et al. (2020) study how ESG investments are connected to the future firm fundamentals. ESG factors can be integrated into the revenue forecasts by increasing or decreasing the company's sales growth rate by an amount that reflects the level of ESG opportunity/risk. Therefore, it is straightforward to expect that the sales amount will be linked to the ESG score.

Further researches could detect the different dependence structures between the determinants of the ESG score by developing flexible r-vine copulas.

References

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. The Journal of Investing, 6: 57–61.

Antolín-López, R., Delgado-Ceballos, J., Montiel, I. (2016). Deconstructing corporate sustainability: a comparison of different stakeholder metrics. Journal of Cleaner Production, 136: 5-17.

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.`

Bauer, R., Koedijk, K., Otten, R. (2005). International evidence on ethical mutual fund performance and investment style. Journal of Banking and Finance, 29(7):1751–1767.

Belghitar Y., Clark E., Deshmukh, N. (2014). Does it pay to be ethical? Evidence from the FTSE4Good. Journal of Banking and Finance, 47: 54–62.

Bello, Z.Y. (2005). Socially responsible investing and portfolio diversification. Journal of Financial Research, 28(1): 41–57.

Benabou R., Tirole J. (2010). Individual and Corporate Social Responsibility. Economica, 77. doi: 10.1111/j.1468-0335.2009.00843.x.

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

Betti, G., Consolandi, C., Eccles, R.G. (2018). The Relationship between Investor Material-
ity nd the Sustainable Development Goals: A Methodological Framework. Sustainability,
10, 2248. doi:10.3390/su10072248

Boiral O., Paillé P. (2012). Organizational Citizenship Behaviour for the Environment: Mea-
surement and Validation. Journal of Business Ethics,.109 (4): 431-445.

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

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

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

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

Carroll, A. B. (1999). Corporate social responsibility. Business & Society, 38(3), 268–295.

Chatterji, A.K., Durand, R., Levine, D.I., Touboul, S. (2016). Do ratings of firms con-
verge? Implications for managers, investors and strategy researchers. Strategic Manage-
ment Journal, 37(8): 1597-1614.

Chen Y., Deleon A. (2020). Financial Quality Metrics and ESG Factor Interactions in Eq-
uity Markets. The Journal of Impact and ESG Investing, Winter 2020, 1(2): 7-15. DOI:
https://doi.org/10.3905/jesg.2020.1.005

Cohen, B., Winn, M.I. (2007). Market imperfections, opportunity and sustainable en-
trepeneurship. Journal of Business Ventuaring, 22: 29-49.

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

Douglas, E., Van Holt, T., Whelan, T. (2017). Responsible Investing: Guide to ESG Data
Providers and Relevant Trends, Journal of Environmental Investing, 8 (1): 92-114.

Drempetic, S., Klein, C., Zwergel, B. (2020). The Influence of Firm Size on the ESG Score:
Corporate Sustainability Ratings Under Review. Journal of Business Ethics, 167: 333-
360.

European Banking Federation (EBF) (2021). REBF 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 accor-
dance with the GRI Standards. https://www.eib.org/attachments/documents/
gri_standards_2018_en.pdf. [Online; accessed 30-05-2018].

Garcia, F., González-Bueno, J., Guijarro, F., Oliver, J. (2020). Forecasting the Environmen-
tal, Social, and Governance Rating of Firms by Using Corporate Financial Performance
Variables: A Rough Set Approach. Sustainability, 12, 3324.

Gu, S., Kelly, B., Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. The Re-
view of Financial Studies, 33: 2223-2273.

Hachenberg, B., Schiereck, D. (2018). Are Green Bonds Priced Differently from Conven-
tional Bonds? Journal of Asset Management, 19(6): 371-383.

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

Hartzmark, S. M. and Sussman, A. B. (2019). Do investors value sustainability? A natural
experiment examining ranking and fund flows. The Journal of Finance, 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; ac-
cessed 27-07-2017].

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

Jewell, J., Livingston M. (1998), Split ratings, bond yields, and underwriter spreads. Journal of Financial Research, 21(2):185–204

Joliet, R. and Titova, Y. (2018). Equity SRI funds vacillate between ethics and money: An analysis of the funds stock holding decisions. Journal of Banking & Finance, 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. Journal of Business Finance and Accounting, 32(7/8): 1465–1493.

Li, F., Polychronopoulos, A. (2020). What a Difference an ESG Ratings Provider Makes!, Research Affiliates. https://www.researchaffiliates.com/en_us/publications/articles/what-a-difference-an-esg-ratings-provider-makes.html

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

Limkriangkrai, M., Koh, S., Durand, R.B. (2017). Environmental, Social, and Governance (ESG) Profiles, Stock Returns, and Financial Policy: Australian Evidence. International Review of Finance, 17(3): 461–471

Lin, W.L., Law, S.H., Ho, J.A., Sambasivan, M. (2019). The causality direction of the corporate social responsibility - Corporate financial performance Nexus: Application of Panel Vector Autoregression approach. North American Journal of Economics and Finance, 48: 401-418.

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

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

Mahjoub, L.B., Khamoussi, H. (2012). Environmental and Social Policy and Earning Persistence. Business Strategy and the Environment, 22(3). DOI: 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. Journal of Business Ethics, 89(3), 409.

Moritz, B., and T. Zimmermann. (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. Business Strategy and the Environment, 16: 106-118.

Novy-Marx, R. (2013). The Other Side of Value: The Gross Profitability Premium. Journal of Financial Economics, 108: 1-28.

Pedersen H.L., Fitzgibbons S., Pomorski L. (2020). Responsible investing: The ESG-efficient frontier. Journal of Financial Economics. DOI: https://doi.org/10.1016/j.jfineco.2020.11.001.

Petitjean, M. (2019). Eco-friendly policies and financial performance: Was the financial crisis a game changer for large us companies? Energy Economics, 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, J.R. (1986). Induction of decision trees. Machine Learning, 1: 81–106.

Renneboog, L., Ter Horst, J., Zhang, C. (2008). Socially Responsible Investments: Institutional Aspects, Performance, and Investor Behavior. Journal of Banking and Finance 32(9): 1723–1742.

Searcy, C and Elkhawas, D. (2012). Corporate sustainability ratings: an investigation into how corporations use the Dow Jones Sustainability Index. Journal of Cleaner Production, 35: 79-92.

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.

S&P Global (2020). `https://www.spglobal.com/en/research-insights/articles/ what-is-the-difference-between-esg-investing-and-socially-responsible-investing`, February 25, 2020.

Statman, M. (2000). Socially responsible mutual funds. Financial Analysts Journal, 56(3): 30–39

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

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

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

Van de Velde E., Vermeir W., Corten F. (2005). Corporate Social Responsibility and Financial Performance. Corporate Governance, 5: 129–138.

Wang, S., and Y. Luo. (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., 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

Winn, M., Pinkse, J., Illge, L. (2012). Case Studies on Trade-Offs in Corporate Sustainability. Corp. Soc. Responsib. Environ. Mgmt., 19: 63-68. doi:10.1002/csr.293

Wong, C., Petroy, E. (2020). Rate the Raters 2020: Investor Survey and Interview Results, StustainAbility.

Zerbib, O.D. (2019). The effect of pro-environmental preferences on bond prices: Evidence from green bonds. Journal of Banking & Finance, 98: 39-60.