Statistical learning models to measure the impact of ESG factors on credit ratings

Arianna Agosto  
University of Pavia

Paola Cerchiello  
University of Pavia

Paolo Giudici (✉ paolo.giudici@unipv.it)  
University of Pavia

Research Article

Keywords: Sustainability, Explainability, ESG scores, Credit ratings, Machine learning, Bayesian models

Posted Date: April 5th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2756660/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.
Statistical learning models to measure the impact of ESG factors on credit ratings

Arianna Agosto\textsuperscript{1}, Paola Cerchiello\textsuperscript{1} and Paolo Giudici\textsuperscript{1*}

\textsuperscript{1*}Department of Economics and Management, University of Pavia, Italy.

*Corresponding author(s). E-mail(s): paolo.giudici@unipv.it; Contributing authors: arianna.agosto@unipv.it; paola.cerchiello@unipv.it;

Abstract

Artificial Intelligence methods, based on either statistical or machine learning models, are rapidly changing financial services, credit lending in particular, complementing traditional bank lending with platform lending. While financial technologies improve user experience, and possibly lower costs, they may increase risks and, in particular, the model risks that derive from inaccurate credit rating assessments. In this paper we will show how to reduce model risks, using a S.A.F.E. statistical learning model, which means improving: Sustainability, by taking environmental, social and governance factors into account; Accuracy, by building a model for which ESG factors do predict credit ratings; Fairness, by dealing with data inconsistencies; Explainability, by merging ESG scores into one measure.

Keywords: Sustainability, Explainability, ESG scores, Credit ratings, Machine learning, Bayesian models

1 Background

Artificial Intelligence methods, based on machine learning applied to the data, are rapidly changing financial services, in all areas, such as lending, asset management and payment services, transforming “finance” into “financial technologies”.
Learning the impact of ESG factors

While financial technologies, and peer to peer lending in particular, improve user experience, and possibly lower costs, they may increase risks. Among them, the risk of inaccurate estimates in credit scoring, i.e. non-proper measures of creditworthiness of the borrowers (“model risk”). The occurrence of this risk may lead to important credit losses, especially when credit is given to large companies. Indeed, incentives are rather different: while in classic bank lending the costs of wrong credit rating assessments are paid by banks themselves, in peer to peer lending they are paid by the borrowers.

These considerations suggest, given the increased economic importance of platform lending, that regulators and supervisors should carefully supervise the model risks that arise from credit ratings and their use by lending platforms.

A first important model risk arises when sustainability factors, represented by Environmental, Social and Governance (ESG) scores, are not taken into account. The problem is quite challenging. First of all, it is not clear whether ESG factors do impact on credit ratings, particularly as they refer to a long term horizon, differently from credit ratings. The most important problem is however the lack of standardisation of ESG scores. ESG scores are currently made available by various specialised companies, including rating agencies. The presence of ESG scores in the market can push companies to improve their Corporate Social Performance (CSP) or ESG behaviour ([1]), but it also presents possible drawbacks. Multiple ESG ratings for a given company can differ and create opaqueness in the company’s actual ESG standing or greenwashing misbehaviour [2]. A recent survey by KPMG [3] showed the existence of more than 160 ESG ratings and data providers, with multiple agencies (eg. Bloomberg, Thomson Reuters, S&P, etc.) whose ESG ratings may however differ. [4] showed little convergence between different ESG ratings. More recently, [5] provided evidence of the low correlation between ESG ratings issued by different providers. The lack of standardisation of ESG metrics is a problem for both investors and borrowers. From the investors’ point of view, it could be challenging to understand and choose among the ESG ratings to select the best investment opportunities. Similarly, it would be difficult for borrower companies to establish financing plans in a correct way.

We believe that taking into account ESG factors is a necessary step for a sustainable finance and, for this reason, we will consider the issue of Sustainable credit scores, through the investigation of the impact of ESG scores on credit ratings, as the main focus of our paper.

A second important model risk concerns lack of predictive accuracy. Credit scorings in peer to peer lending has been studied in a few recent papers, that propose network models to take into account platform risk arising from the connectivity between companies. In these papers, financial network models allow to improve the predictive accuracy of the individual probability of default by considering similarities or linkages among borrowers. This becomes crucial for peer-to-peer lending platforms, in which individuals are able to directly provide small and, in most cases, unsecured loans to small and medium enterprises, without the availability of financial and behavioral information typically
leveraged by banks. A network based scoring model built upon balance sheet similarities between P2P borrowing companies was applied by [6], while [7] improved P2P credit scoring models by clustering SMEs based on latent risk factors, deduced from financial ratios. In [6], a network is instead built upon trade flows between the companies joining the platform, proxied by input-output data at the sector level. While network models, and similarly complex machine learning models, may seem appealing, capturing non-linearities and, thereby, improving predictive accuracy, in some cases they can be limited by their “black-box” nature, which makes it difficult to interpret the results.

We believe that complex machine learning models may be useful when they improve model accuracy in a manner that overcompensates their lack of explainability, making the further computational burden of making them explainable affordable. This may not be the case when data is of limited quality.

Indeed, following what we already discussed, a third important model risk that may arise in machine learning credit scoring is that of data quality, whose lack may lead to unfair results, as stated, for example, in the recent European Artificial Intelligence Act [8]. The problem of data quality arises in credit scoring when some necessary information is missing or contradictory. This is the case of sustainability factors, encoded in ESG measures: they are not yet standardised, with different data providers assigning a different ESG value to the same company, and with a relatively short time series available. This lack of standardisation may lead to unfair credit ratings, which creates a distorted credit allocation.

We believe that lack of data quality is a real concern which prevents from a correct understanding of the impact of ESG factors on credit ratings. However, in line with our focus, we will employ the data available so far, trying to leverage not only the disadvantages but also the advantages of inconsistent ESG databases.

A fourth important model risk is lack of explainability of the credit scores. This is a very relevant problem for many stakeholders: for investors, who cannot rationalise their investment decisions, not knowing why some companies have a higher score than others; for borrowers, who cannot improve their scores, without knowing the drivers of their values; for regulators and supervisors, which cannot evaluate the impacts of the proposed models, particularly under stress scenarios and, therefore, may not validate them. Complex machine learning models, may be highly accurate, as they can capture non-linearities and interdependences, but are typically “black-box”: they assign predictive scores without explaining their determinants, in terms of the most correlated explanatory variables, as “classic” regression models do, leading to a lack of model explainability. The recent machine learning literature has proposed methods to explain black box models, by means of further processing of the predictive output: see e.g. [9], [10], [11].

We believe that Explainable AI methods are useful, but their extra computational burden is not justified when the available data is of limited quality and/or size. In this case it would be better to build a model that is, while
Learning the impact of ESG factors

complex, and capable to capture non-linearities, “explainable by design”, as a simple regression model.

To fill the gaps identified above, we have been working in a close collaboration between academics and policy makers, within the Milano Hub of the Bank of Italy, aimed at developing Sustainable, Accurate, Fair and Explainable (S.A.F.E.) AI methods, in line with the European AI Act.

The result of the collaboration shown in the present paper is a credit scoring model for companies that, given the available data, is: Sustainable, as credit scores take ESG factors into account; Accurate, as it indicates that ESG factors do indeed predict credit ratings, even when controlling for balance sheet information; Fair, as it “compensates” different data providers into one combined measurement; Explainable, as based on a mixture model whose weights indicate the importance of each ESG score in determining the credit scores.

From a methodological viewpoint, the main contribution of the paper is a data-driven model that describes how ESG scores affect credit ratings, by means of a statistical learning model that is explainable by design, as the final ESG score is a linear combination of the ESG sources, with weights that are proportional to their predictive accuracies. Indeed, the very aim of the proposed methodology is to measure the quality of the ESG scores, in terms of their accuracy in predicting credit ratings. The combined ESG score will be strongly impacted by good scores and less impacted by bad scores.

To our knowledge, this is the first data-driven model based on the relationship between credit ratings and ESG scores, by means of a statistical learning model that is explainable by design, as the final ESG score is a linear combination of the ESG sources, with weights that are proportional to their predictive accuracies.

The remainder of this paper is organized as follows: Section 2 presents a discussion on the main focus of our paper: the relationship between ESG factors and credit ratings; Section 3 introduces the proposed modelling approach; Section 4 presents an application of the methodology to a sample of European companies and, finally, Section 5 concludes.

2 ESG scores and credit rating

Corporate Social Performance (CSP) is aimed at evaluating the degree to which companies are sustainable, that is, how they perform their business activities in relation to the external stakeholders and taking into account the economic, environmental, social, and time factors [12–14]. Environmental, Social and Governance (ESG) factors are often taken as a proxy for the sustainable behaviour of companies.

Environmental factors (E) relate to the impact on the environment deriving from the production of goods or services and include carbon emissions, preservation of the natural environment, biodiversity protection, and waste and water management [15–17]. A company that operates with less harm to the environment might reduce the probability of future scandals, legal actions,
losses related to legal claims etc. and benefit from a better reputation and lower risks [18].

Social factors (S) refer to the impacts of companies on society, including issues of employee satisfaction, diversity, inequality, gender gap, protection of young and children, investment in human capital and communities, and human rights [15, 19].

Governance factors (F) measure the quality of corporate governance. Shortcomings in governance have been in the past the cause of major scandals and crises, such as the Enron crisis in the US, Volkswagen in Germany, Parmalat in Italy, and the banking crisis of 2007-2008 [20, 21]. Improved governance settings can contribute to a more sustainable and balanced firms’ growth, therefore contributing to a more sustainable economic development [22, 23].

The above factors are the basis for investment decisions and drive the choice of investors in terms of which companies to finance through equity or debt. To improve the interpretability of ESG, specialised companies (including rating agencies) have started to provide measures and proxies for ESG behaviour, publishing ESG ratings or ESG scores that convey the level of sustainability of companies and the degree of accountability of these companies on ESG aspects [24, 25].

Each rating provider collects information from different sources (company reports, news, stock exchange information, etc.) and applies proprietary methodologies to combine information and produce a summary measure of ESG behaviour. Different methodologies yield different measures, that often produce divergent results [4, 5, 26, 27], and this induces lack of standardisation.

The importance of ESG metrics is bound to grow in the future, with ESG ratings likely to affect investors’ decisions, firms’ ability to finance their investments and pursue a sustainable business model. It follows that understanding whether and how ESG ratings affect creditworthiness is a very important managerial and policy challenge.

To our knowledge, this is the first work to: 1. analyse the relationship between ESG scores and credit ratings through a data-driven model that predicts the company’s credit rating class based on the ESG rating; 2. use the ESG scores assigned by different providers to create a combined metric where each ESG score is weighted based on its predictive accuracy.

In the next section we describe our proposed methodology, which is applied to real data in Section 4. Section 5 concludes the paper with a final discussion.

3 Proposal

This paper provides a summary indicator for the ESG performance of listed companies by integrating the ESG scores assigned by different providers. The indicator is obtained by attributing to each available ESG score a weight that is a function of the likelihood of the observed counts of companies belonging to the different credit rating classes, under the alternative partitions generated by
the ESG scores. The likelihood weights are obtained through the application of Bayes’ theorem.

To investigate the validity of our combined score from the investor’s perspective, we leverage the relationship between ESG scores and credit rating. Indeed, we investigate the possibility of improving credit risk prediction through the efficient use of different ESG data sources. To this end, we extend the methodology proposed by [28], who considered the case of estimating a company’s probability of default using a set of explanatory financial variables and whose proposal was extended to the multinomial case by [29].

In [28], based on the mixture of Dirichlet processes model proposed by [30], it is assumed that the partition $g_k$ generated by the $k$-th among $K$ covariates is made up of $j = 1, ..., J_k$ levels and that the probability of default of company $i$ ($\text{Prob}(Y_i = 1)$, where $Y_i$ is a binary variable equal to 1 if company $i$ defaults, 0 otherwise) is constant within the same $j$ level of the covariate and equal to $\theta_j$.

Here we extend their work assuming that the partition $g_k$ is generated by the values of the ESG scores assigned by the $k$-th data provider, and that $Y_i$ is a binary variable which indicates whether a company rating is speculative (equal to 1) or investment grade (equal to 0). These assumptions do not imply a loss of generality: different partitions can be assumed, for example corresponding to a combination of ESG scores, and a different binarisation of the rating can be considered to obtain $Y$.

Letting $Y_i$ be a Bernoulli($\theta_j$) variable and the $\theta_j$’s Beta random variables with parameters $\alpha$ and $\beta$, which implies that, a priori, $E(\theta_j) = \frac{\alpha}{\alpha + \beta}$, the marginal likelihood contribution of level $j$ can be obtained as:

$$
p(y||j) = \int_0^1 p(y||\theta_j)p(\theta_j)d\theta_j
$$

$$
= \int_0^1 \theta_j^{d_j} (1 - \theta_j)^{n_j - d_j} \frac{1}{B(\alpha, \beta)} \theta_j^{\alpha - 1} (1 - \theta_j)^{\beta - 1} d\theta_j
$$

$$
= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} \frac{\Gamma(\alpha + d_j) \Gamma(\beta + n_j - d_j)}{\Gamma(\alpha + \beta + n_j)} (1)
$$

where $p(\theta_j)$ is the prior distribution of $\theta_j$, $d_j$ is the number of defaulted companies and $n_j$ is the total number of companies sharing level $j$ of the $k$ covariate. Furthermore, $B$ is the Beta function, defined by:

$$
B(z_1, z_2) = \frac{\Gamma(z_1) \Gamma(z_2)}{\Gamma(z_1 + z_2)},
$$

where, for each positive integer $n$:

$$
\Gamma(n) = (n - 1)!
$$
Under the assumption that the \( \theta_j \)'s are independent random variables, the marginal likelihood of the partition \( g_k \) is:

\[
p(y \mid g_k) = \prod_{j=1}^{J_k} p(y \mid j),
\]

which determines the posterior probability of the partition:

\[
p(g_k \mid Y) \propto p(y \mid g_k)p(g_k),
\]

where \( p(g_k) \) can be set a priori, for example according to the uniform distribution: \( p(g_k) \propto 1/M \) where \( M \) is a constant.

The expected probability of default of company \( i \), conditional on the available set of covariates \( X \), can then be obtained as follows:

\[
E(\theta_i \mid X, Y) = \sum_{k=1}^{K} E(\theta_j \mid g_k, Y)p(g_k \mid Y),
\]

with \( E(\theta_j \mid g_k, Y) = \frac{\alpha + d_j}{\alpha + \beta + n_j} \),

in which the posterior probability \( p(g_k \mid Y) \) acts as \( k \)-th covariate weight in determining the expected probability of the default event.

Equations (3) and (4) summarise the essence of our proposed machine learning model. It is a Sustainable model, as it assumes that ESG factors can affect credit ratings; it is a Fair model, as it averages the contribution of different ESG providers, compensating their differences; it is an Explainable model, as it is a linear combination of weights with posterior probabilities, which, although calculated in a non-linear way, have a clear meaning. In the next section, we will verify whether the model is also accurate, that is whether ESG factors have predictive relevance for credit ratings.

4 Application

4.1 Data

In the present section we apply our proposed methodology to a sample of 1382 European companies for which we retrieve:

- the MSCI ESG Score: a continuous variable ranging from 0 (lowest sustainability) to 10 (highest sustainability);
- the Refinitiv ESG Score: a continuous variable ranging from 0 to 100. As for the MSCI ESG score, higher values indicate better sustainability profiles;
Learning the impact of ESG factors

- the Standard and Poor’s (S&P) Global ESG Rank: a discrete variable defined as the total sustainability percentile rank, ranging from 0 (lowest sustainability) to 100 (highest sustainability);
- the risk class assigned to the company based on the Bloomberg Issuer Default Risk model generated probability of default over the next one year: an ordinal variable whose categories in the sample range from IG1 (highest credit worthiness) to D4 (lowest credit worthiness). Specifically, classes from IG1 to IG10 identify Investment Grade bond issuers, while classes from HY to H6 and from D1 to D4 identify High Yield and Distressed bond issuers respectively;
- a set of 13 financial ratios\(^1\) which should reflect company profitability, growth and liquidity, together with the value of market capitalization, which serves as a dimensional indicator.

To allow the comparability of the scores, the MSCI ESG score has been rescaled in the 0-100 range.

Data is the last available as of August 3, 2022 and is retrieved from various sources: MSCI ESG Research (for the MSCI ESG scores), Refinitiv LSEG business (for the Refinitiv ESG scores), Bloomberg (for the S&P Global ESG rank and the credit ratings). All Data is pre-processed so that no missing values are present in our sample. In our setting, among the European companies having an ESG rating, we only select those (1382) for which all three ESG scores are available at the considered date. The data has got a cross-sectional structure, all being referred to a single date, the 3rd of August, 2022.

The distribution of sample companies among the credit rating classes is shown in Figure 1.

\(^1\)Our balance sheet dataset includes the following indicators: Return on Equity, Return on Asset, Return on Investment, Short Term Debt on 1-year Growth, Total Debt on 1-year Growth, Free Cash Flow on 1-year Growth, Free Cash Flow on 5-year Growth, EBITDA to Interest Expenses, Long Term Debt to Total Equity, Quick Ratio, Capital Expenditure Ratio, Financial Leverage, Asset Turnover
Figure 1 shows that, for the considered companies, the distribution of ratings is quite skewed to the right, and that there is a large group of companies with very high ratings (IG1). Both aspects will make it more challenging to attain a good level of predictive accuracy.

As it can be seen from Figure 2, the distribution of the three ESG scores in the analysed sample is instead left-skewed, meaning that a few number of companies have a much worst ESG evaluation than the mean one.

---

2 Reproduced by permission of MSCI ESG Research LLC, copyright 2023 MSCI ESG Research LLC, All rights reserved.
Concerning the concordance between the ESG scores, it can then be noticed from Tables 1 to 4 that correlation between the Refinitiv and the S&P ESG scores is relatively high according to the Pearson and Spearman measures, but decreases to nearly 50% when moving to rank-based concordance measures. Correlation between the MSCI ESG scores and the other two indicators is instead low, never reaching 40%. This increases the interest in reaching a sustainability metric that combines alternative ESG scores based on their capability to order the observed companies by their creditworthiness.

Table 1 Pearson correlation between the ESG scores. Source: own elaborations based on MSCI ESG Research, Refinitiv (LSEG business), S&P Global and Bloomberg data.

<table>
<thead>
<tr>
<th></th>
<th>S&amp;P</th>
<th>Refinitiv</th>
<th>MSCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>S&amp;P</td>
<td>1</td>
<td>0.692</td>
<td>0.373</td>
</tr>
<tr>
<td>Refinitiv</td>
<td>0.692</td>
<td>1</td>
<td>0.383</td>
</tr>
<tr>
<td>MSCI</td>
<td>0.372</td>
<td>0.383</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2 Spearman correlation between the ESG scores. Source: own elaborations based on MSCI ESG Research, Refinitiv (LSEG business), S&P Global and Bloomberg data.

<table>
<thead>
<tr>
<th></th>
<th>S&amp;P</th>
<th>Refinitiv</th>
<th>MSCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>S&amp;P</td>
<td>1</td>
<td>0.689</td>
<td>0.356</td>
</tr>
<tr>
<td>Refinitiv</td>
<td>0.689</td>
<td>1</td>
<td>0.376</td>
</tr>
<tr>
<td>MSCI</td>
<td>0.356</td>
<td>0.376</td>
<td>1</td>
</tr>
</tbody>
</table>
Learning the impact of ESG factors

Table 3 Kendall’s tau correlation between the ESG scores. Source: own elaborations based on MSCI ESG Research, Refinitiv (LSEG business), S&P Global and Bloomberg data.

<table>
<thead>
<tr>
<th></th>
<th>S&amp;P</th>
<th>Refinitiv</th>
<th>MSCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>S&amp;P</td>
<td>1</td>
<td>0.501</td>
<td>0.246</td>
</tr>
<tr>
<td>Refinitiv</td>
<td>0.501</td>
<td>1</td>
<td>0.259</td>
</tr>
<tr>
<td>MSCI</td>
<td>0.246</td>
<td>0.259</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4 Somers D correlation between the ESG scores. Source: own elaborations based on MSCI ESG Research, Refinitiv (LSEG business), S&P Global and Bloomberg data.

<table>
<thead>
<tr>
<th></th>
<th>S&amp;P</th>
<th>Refinitiv</th>
<th>MSCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>S&amp;P</td>
<td>1</td>
<td>0.498</td>
<td>0.247</td>
</tr>
<tr>
<td>Refinitiv</td>
<td>0.498</td>
<td>1</td>
<td>0.262</td>
</tr>
<tr>
<td>MSCI</td>
<td>0.247</td>
<td>0.262</td>
<td>1</td>
</tr>
</tbody>
</table>

4.2 Results
4.2.1 In-sample analysis

The first step in our empirical analysis consists of the calculation of the posterior probability-based weights according to the methodology described in Section 3. Having no a priori reasons to assign different weights to the scores, we set the $M$ constant in (3) equal to 3, which means that the three scores are a priori equally weighted.

The posterior weights associated to the scores are estimated on a random training sample of 829 companies (60% of the available observations) and are shown in the second column of Table 5. The third column of Table 5 reports instead the weights obtained by applying the same methodology to the residuals of stepwise linear regression models where the dependent variable is a given ESG score (MSCI, Refinitiv or S&P) and the regressors are the company’s balance sheet variables and market capitalisation. This allows indeed to consider the extent to which the financial information - on which both the ESG scores and the credit ratings are supposed to be related to - influences the capability of ESG scores to predict the credit ratings. Coefficient estimates for the estimated linear regression models are shown in table A1-2-3.

Table 5 Weights derived from the posterior probabilities associated to the ESG scores, before and after controlling for financial ratios. Source: own elaborations based on MSCI ESG Research, Refinitiv (LSEG business), S&P Global and Bloomberg data.

<table>
<thead>
<tr>
<th>ESG Score</th>
<th>Before control</th>
<th>After control</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSCI</td>
<td>0.36</td>
<td>0.34</td>
</tr>
<tr>
<td>Refinitiv</td>
<td>0.37</td>
<td>0.32</td>
</tr>
<tr>
<td>S&amp;P</td>
<td>0.27</td>
<td>0.34</td>
</tr>
</tbody>
</table>
Table 5 shows that model weights are somewhat different, before controlling for financial ratios. But also that such difference nearly disappears, once controlling for the same ratios. This may be the effect of different attention given by the providers to the financial ratios. Once they are taken into account, however, the ESG scores have a similar importance, in determining credit worthiness. This shows that our proposed model is able to improve fairness, reducing inconsistencies among the data providers. And, by taking an equally weighted average of the ESG scores, it does not generate any bias deriving from using one rather than the other.

We also remark that the weights in Table 5 are the main output of our proposed model: a set of weights which is easy to interpret and implement in the monitoring of credit risk.

In other words, with the in-sample analysis we have shown that our proposed model is fair and explainable.

### 4.2.2 Out-of-sample analysis and Robustness check

We now provide a predictive analysis where the probability that a company belongs to a certain rating class - conditional on the ESG score - is estimated based on the methodology described in Section 3.

Specifically, we use the weights associated to the ESG scores estimated on the training sample (see Section 4.2.1) to predict the credit rating in the validation sample (40% of the available observations). According to the proposed merged scoring methodology, the weights are then used to determine, for each company, and for each provider domain (Refinitiv, Standard and Poors, MSCI) the probability associated to each of the two considered rating categories: Investment Grade or Speculative (High Yield or Distressed) class.

Figure 3 shows the posterior probabilities associated to the different classes of the ESG score distribution, for each of the three scores considered. These probabilities are used to determine the probabilities assigned by the merged score. Indeed, for each company, the probability of belonging to a speculative rating class is calculated as the weighted mean of the probabilities assigned by the three scores, using the Bayesian likelihood-based weights.

![Fig. 3 Estimated probability of belonging to a High Yield or Distressed credit rating class by ESG score class, before (left) and after (right) controlling for financial and dimensional indicators. Source: own elaborations based on MSCI ESG Research, Refinitiv (LSEG business), S&amp;P Global and Bloomberg data.](image-url)
Figure 4 shows the ROC curves of the credit rating prediction based in the ESG scores, obtained by applying the Bayesian model.

From Figure 4 note that there is no absolute dominance of one specific ROC curve. The relationship depends strictly on the quantiles of reference. More in detail, if we compare the related AUROC measures, the two leading models are the Merged score and the MSCI ones. The Merged score model is, furthermore, more robust (more sustainable in the statistical sense) as it does better in modelling the tails of the distribution, where the more extreme financial profiles lie: companies that are either very bad or very good.

The results are confirmed after controlling for the financial ratios. We can conclude, from Figure 4, that the proposed model is a sustainable credit rating model, as it shows that ESG factors are important to predict credit ratings, even when financial variables are inserted into the model. The proposed model also improves predictive accuracy, with respect to what the separate ESG scores would do.

A question that may arise, especially for the sake of comparison, is whether a different (non-Bayesian) machine learning model would improve predictive accuracy, although being not explainable. If it were so, computationally expensive explainable AI methods, such as Shapley values (Lundberg and Lee, 2016; Bussman et al., 2021) could be applied as an “add-on” to the model.

To this end, we additionally fit a competing model, which is typical expression of Machine Learning approaches.

Among the several proposals, we choose the XGBOOST [31], for its well-known capability of modelling non-linearity in a very efficient way, without imposing any distributional assumption. We fit it by means of the package ’xgboost’ of R software and set three tuning parameters: a parameter d, which determines the depth of each boosted tree; a learning parameter η, which determines the updating rate, and a parameter B, which determines the number of boosted trees. We select the values of the tuning parameters after a cross-validation exercise that maximizes the AUROC: specifically, we take $d = 1; \eta = 0.001; B = 5000$. The features employed are the three ESG scores, exactly as for the Bayesian model. We ended up with a boosting model whose predictive performance is reported in Figure 5.
From Figure 5 we observe that the AUC, calculated on the test set, is very close to the Bayesian model one. This represents an important result: XGBOOST is not able to outperform the Bayesian approach in the context of our problem and of the available data.

Indeed, the proposed Bayesian model does not offer an exceptional performance, especially because the effect of ESG factors on credit ratings is probably limited, but it has a clear and unavoidable advantage: it is explainable by design and it offers a system of weights that can be used in further analysis. On the other hand, the XGBOOST model, which is not explainable by design, does not lead to a gain in predictive accuracy that can justify the use of a computationally expensive AI method, such as Shapley values.

Although a computationally expensive explainable AI method may not be justified in our context, we have tried to interpret the predictions obtained from XGBoost with a graphical method, comparing the plots of the estimated probabilities by the three ESG scores, similar to what obtained in Figure 3 for the Bayesian model, reported in Figure 6 below.
Comparing Figure 6 with Figure 3 (left), note that the behaviours of the estimated probabilities are rather similar. In both cases, there is an overall negative dependence between the ESG score class and the probability of default; moreover, the range of variation of Refinitiv is the smallest. This implies a higher weight in the Bayesian model for Refinitiv than for more discriminant scores, such as S&P. Figure 6 shows that the probabilities estimated by the XGBOOST have generally a lower variability, with respect to those from the Bayesian model. This is in line with the smoothing effect carried out by the (non-linear) XGBOOST model.

5 Conclusions

In the paper we have shown how credit worthiness could be measured by means of a S.A.F.E. machine learning model which reduces model risks, in line with the emerging regulations of Artificial Intelligence, which aim to measure the risks to promote its usage.

The model is Sustainable, as credit ratings take Environmental, Social and Governance factors into account. The model is Accurate as it indeed shows that ESG scores have an effect in the prediction of credit ratings. The model is Fair as it can level out differences between different ESG data providers, taking an averaged score. The model is Explainable as it can be easily interpreted by means of a set of normalised weights assigned to the different ESG providers.

The paper is the first of this kind, and it may generate debate and impact, in the AI and in the financial community altogether.

This, in particular, because it can improve ESG standardisation, providing a solution to the problem of multiple ESG ratings. The increased attention to sustainability issues has yielded the proliferation of rating agencies and ESG scores, with multiple ESG scores on the market that are often divergent and provide different types of information. In the paper we show how to combine different ESG scores into a single ESG one that combines the information from different providers. A combined score which brings evidence of the capability of the combined ESG score to predict credit rating classes.

Our findings have many implications for the application of statistical learning and Artificial intelligence methods in the financial sector. What presented can be useful for investors in financial markets, who can exploit the information provided by different ESG scores in a comprehensive setting, reducing information asymmetries on ESG company performance. It can also be useful for lenders in credit markets, as they can make a better informed use of ESG factors in determining credit worthiness, to the benefit of the best-performing companies in terms of sustainable behaviour. It can be of interest also for insurance companies, helping to assess pricing of climate and ESG related events.

Our research is also of interest for regulators and supervisors in the financial sector, as it provides a standardised metric to measure the impact of different ESG scores, along with a combined score, thereby improving the assessment
of the sustainability of the company which receives ESG ratings. And, finally, it is important for ESG data providers, as they can receive feedback on the relative quality of their metrics, and, possibly, improve them.

We also remark that the scope of this paper is to provide indications to financial institutions on the relative quality of different ESG providers (in terms of their predictive accuracy).

Moreover the shown results, obtained on the available data, clean of missing values, can be extended, without loss of generality, to a larger database.

Future research should extend our work to cover also companies for which ESG scores are missing for some providers. Our approach can be easily generalized to this context assigning companies with missing scores to a distinct new category that contains all companies with missing information.

Future research should also concern the implementation of the proposed methodology to other regulated industries, such as the health care and the automotive sectors, and, possibly, to other high risk Artificial Intelligence applications.

Funding, Authors contribution, Conflict of interest, Data Availability Statement.

Funding. The authors acknowledge the European HORIZON 2020 PERISCOPE project (contract number 101016233) and the Italian MUR PRIN project FIN4GREEN for financial support. They also acknowledge continuous support from the Milano Hub of the Bank of Italy and, in particular, Juri Marcucci for the coordination activity; Marco Fanari, Johnny Di Giampaolo and Enrico Foscolo for useful discussion and the provision of data.

Authors contribution. While Paolo Giudici supervised the work, and wrote Sections 1 and 5 of the paper, Arianna Agosto and Paola Cerchiello developed the methodology and the data analysis, and wrote Sections 2 (PC), 3 and 4 (AA).

Conflict of interests. The authors declare they have no conflicts of interest.

Data availability statement. Concerning the data employed in the paper, the Authors remark that they have used what available within the academic Bloomberg license available for the University of Pavia researchers. Additionally, the MSCI ESG data was reproduced by permission of MSCI ESG research LLC, copyright 20[] MSCI ESG Research LLC, All rights reserved. The ESG data contained herein is the property of MSCI ESG Research LLC (ESG). ESG, its affiliates and information providers make no warranties with respect to any such data. The ESG data contained herein is used under license and may not be further distributed or disseminated without the express written consent of ESG.
Declarations

- Funding: This research received support from the European HORIZON 2020 PERISCOPE project (contract number 101016233) and the Italian MUR PRIN project FIN4GREEN.
- Conflict of interest/Competing interests: No conflict of interest/competing interest to declare.
- Ethics approval: Not applicable.
- Consent to participate: Not applicable.
- Consent for publication: Not applicable.
- Availability of data and materials: Data cannot be distributed.
- Code availability: Code available upon reasonable request.
- Authors’ contributions: While Paolo Giudici supervised the work, and wrote Sections 1 and 5 of the paper, Arianna Agosto and Paola Cerchiello developed the methodology and the data analysis, and wrote Sections 2 (PC), 3 and 4 (AA).

Editorial Policies for:

Springer journals and proceedings:  
https://www.springer.com/gp/editorial-policies

Nature Portfolio journals:  
https://www.nature.com/nature-research/editorial-policies

Scientific Reports:  
https://www.nature.com/srep/journal-policies/editorial-policies

BMC journals:  
https://www.biomedcentral.com/getpublished/editorial-policies
Learning the impact of ESG factors

Appendix A

The following tables report the coefficient estimates\(^3\) for the linear regression models of the ESG scores on balance sheet ratios and dimensional indicators.

**Table A1** Results of stepwise linear regression model of the MSCI ESG score on balance sheet ratios and dimensional indicators. Source: own elaborations based on MSCI ESG Research and Bloomberg data.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>68.9</td>
<td>&lt;2e-16***</td>
</tr>
<tr>
<td>EBITDA to Revenue</td>
<td>6.55e-03</td>
<td>0.029**</td>
</tr>
<tr>
<td>Free Cash Flow to Total Equity</td>
<td>-2.40e-03</td>
<td>0.007***</td>
</tr>
<tr>
<td>Quick Ratio</td>
<td>-1.21</td>
<td>0.002***</td>
</tr>
<tr>
<td>Market capitalisation</td>
<td>9.22e-06</td>
<td>0.041**</td>
</tr>
</tbody>
</table>

**Table A2** Results of stepwise linear regression model of the Refinitiv ESG score on balance sheet ratios and dimensional indicators. Source: own elaborations based on Refinitiv (LSEG business) and Bloomberg data.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>63.8</td>
<td>&lt;2e-16***</td>
</tr>
<tr>
<td>EBITDA to Revenue</td>
<td>7.87e-03</td>
<td>0.003***</td>
</tr>
<tr>
<td>Short Term Debt on 1-year Growth</td>
<td>-1.33e-03</td>
<td>0.108*</td>
</tr>
<tr>
<td>EBITDA on Interest Expenses</td>
<td>-7.13e-04</td>
<td>0.138</td>
</tr>
<tr>
<td>Quick Ratio</td>
<td>-2.182</td>
<td>5.94-10***</td>
</tr>
<tr>
<td>Financial Leverage</td>
<td>0.148</td>
<td>0.009***</td>
</tr>
<tr>
<td>Market capitalisation</td>
<td>1.83e-05</td>
<td>6.66e-06***</td>
</tr>
</tbody>
</table>

\(^3\)In all tables, ***, **, * denote statistical significance at the 1%, 5% and 10% respectively.
Table A3 Results of stepwise linear regression model of the S&P ESG score on balance sheet ratios and dimensional indicators. Source: own elaborations based on S&P Global and Bloomberg data.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>64.6</td>
<td>&lt;2e-16***</td>
</tr>
<tr>
<td>EBITDA to Revenue</td>
<td>5.83e-03</td>
<td>0.121</td>
</tr>
<tr>
<td>Long Term Debt to Total Equity</td>
<td>-1.60e-03</td>
<td>0.081*</td>
</tr>
<tr>
<td>Quick Ratio</td>
<td>-3.18</td>
<td>1.59e-10***</td>
</tr>
<tr>
<td>Financial Leverage</td>
<td>0.210</td>
<td>0.014**</td>
</tr>
<tr>
<td>Asset Turnover</td>
<td>-3.382</td>
<td>0.003*</td>
</tr>
<tr>
<td>Market capitalisation</td>
<td>2.30e-05</td>
<td>5.52e-05***</td>
</tr>
</tbody>
</table>

References


Learning the impact of ESG factors


[19] Van Duuren, E., Plantinga, A., Scholtens, B.: Esg integration and
Learning the impact of ESG factors


Learning the impact of ESG factors

Discovery and Data Mining, pp. 785–794 (2016)
Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- E