No 507

DECEMBER 2022

CefES
Center for European Studies
PAPER SERIES

ESG Factors and Firms' Credit Risk

Laura Bonacorsi, Vittoria Cerasi, Paola Galfrascoli, and Matteo Manera

DIPARTIMENTO DI ECONOMIA, METODI QUANTITATIVI E STRATEGIA DI IMPRESA

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

ESG Factors and Firms' Credit Risk

Laura Bonacorsi[1], Vittoria Cerasi[2], Paola Galfrascoli[3], and Matteo Manera[4]

[1]*Department of Social and Political Sciences, Bocconi University; Milano, Italy*
[2]*Italian Court of Auditors (Corte dei Conti) & CefES & O-Fire, University of Milano-Bicocca, Italy*
[4]*Department of Economics, Management and Statistics (DEMS), University of Milano-Bicocca, Italy, and Fondazione Eni Enrico Mattei (FEEM), Milano, Italy*
[3]*Department of Economics, Management and Statistics (DEMS), CefES & O-Fire, University of Milano-Bicocca, Italy*

December 19, 2022

Abstract

We study the relationship between the risk of default and Environmental, Social and Governance (ESG) factors using Machine Learning (ML) techniques on a cross-section of European listed companies. Our proxy for credit risk is the z-score originally proposed by Altman (1968).We consider an extensive number of ESG raw factors sourced from the rating provider MSCI as potential explanatory variables. In a first stage we show, using different SML methods such as LASSO and Random Forest, that a selection of ESG factors, in addition to the usual accounting ratios, helps explaining a firm's probability of default. In a second stage, we measure the impact of the selected variables on the risk of default. Our approach provides a novel perspective to understand which environmental, social responsibility and governance characteristics may reinforce the credit score of individual companies.

Keywords: credit risk, z-scores, ESG factors, Machine learning.

JEL Classification: C5, D4, G3.

We are particularly grateful to Gianna Monti for helpful suggestions on statistical methodological issues. We also appreciated comments from Marzio Galeotti, Alessandro Lanza and Sergio Vergalli and from seminar participants at DEMS ReLunch Seminars (Bicocca University), FEEM Seminars, GRETA Conference 2021 (Venice) and Computational and Financial Econometrics 2021 (London, UK). We also appreciated the excellent research assistance by Martina Bianchi. We acknowledge financial support from the European Investment Bank for the EIBURS project *ESG-Credit.eu - ESG Factors and Climate Change for Credit Analysis and Rating* and from Fondi MUR Dipartimento di Eccellenza 2018-2022.

1 Introduction

In recent years, Environmental, Social responsibility and Governance (ESG) characteristics of companies have become important indicators for evaluating companies and criteria to be carefully considered by investors. For this reason, they have as well become the object of specific regulatory actions. Many empirical studies investigate whether companies performing better from an ESG perspective are indeed performing better financially (Friede et al., 2015) and whether they are also more resilient during the initial outbreak of the Covid-19 pandemic (Albuquerque et al., 2020). There is still little evidence on whether companies in the highest ranking from the ESG perspective are safer borrowers as well, i.e. whether they exhibit a lower probability of bankruptcy. Understanding such a relationship could be particularly relevant for all the players involved in the credit markets: companies would have an incentive to improve their ESG ranking as it benefits also their creditworthiness and to disclose their ESG type as a signal in credit markets; financial intermediaries could improve their traditional models to measure credit risk of potential borrowers by adding variables related to the ESG dimensions; regulators might know which ESG variables are more informative and deserve mandatory disclosure for reducing asymmetric information in credit markets and preventing rationing (Stiglitz and Weiss, 1981).

This paper contributes to the literature by providing evidence on the link between ESG dimensions and companies' creditworthiness. In particular, we investigate whether and how ESG *factors*, representing a firm's concern towards sustainability and ethical behaviour, affect the creditworthiness of a company. The ESG *factors* we refer to is the ESG "raw" information used by the ESG rating agencies for constructing the widely used ESG scores. The rationale for the use of such variables is motivated by recent contributions by Berg et al. (2022) and Berg et al. (2020), showing that one should be cautious when relying on ESG *scores*: the same company belongs to different quartiles of the score distribution according to different providers; moreover, the scoring methods themselves exhibit changes over time, resulting in overwriting and shuffling of past rankings. For this reason, we decided to focus on ESG *factors* as opposed to *scores*, i.e. we take into account the original information used by rating companies for the construction of ESG scores. The advantage of this procedure is twofold: on the one hand, our results are independent from the ESG rating provider chosen, hence from different rating schemes or weights as well as from potential revisions; on the other hand, our results can be applied to non-rated corporations and used by lenders in their screening process, reducing adverse selection problems.

Our aim is contributing to design a credit rating framework augmented by ESG dimensions. Our results consist in the selection, among many possible candidates, of the ESG dimensions that are meaningful for credit risk analysis and, in a second stage, on the interpretation of their relationship with creditworthiness. The paper will proceed as follows. Section 2 will describe the current European regulatory framework for ESG disclosure. Section 3 will summarize the main strands of economic literature analyzing ESG topics. Section 4 will present the data used in the empirical analysis, which will be described in Section 5.

Section 6 will present the selected variables and Section 7 will analyze their relationship with creditworthiness. Section 8 will conclude.

2 Regulatory framework and rating agencies

The COP26[1] held in Glasgow in 2021 expressed alarm that human activities have caused around 1.1 °C of warming to date. Climate and weather extremes and their adverse impacts on people and nature will continue to increase with every additional increment of rising temperatures (COP, 2021). It is well recognized that climate change is a common and serious concern of humankind and emphasis is put on the urgency of taking action to address its causes. Such an increasing global awareness on climate change and environmental issues, together with concerns on social and governance matters, has led the European Union (EU) to introduce directives and define uniform standards and guidelines for the disclosure of non-financial information.

This process was initiated after the Paris Agreement on climate change (2015) and the adoption, in the same year, of the UN 2030 Agenda for Sustainable Development, which encouraged the EU to enhance sustainable growth and develop a renewed sustainable finance strategy in the framework of the European Green Deal. The latter was proposed in 2019, aiming at "*transforming the EU into a modern, resource-efficient and competitive economy*"[2] and "*making Europe the first climate neutral continent in the world*"[3]. All 27 EU Member States endorsed the European Climate Law[4] in 2021 and committed to reduce emissions by at least 55% by 2030 (compared to 1990 levels), to address energy poverty and to reduce external energy dependence. Ambitious targets include making transport sustainable, creating markets for clean technologies and products, adopting renewable energy and achieving greater energy efficiency. In the same direction, the European Commission delivered in 2018 the first concrete directives to enable the EU financial sector to lead the way to a greener and cleaner economy through the Action Plan on Sustainable Finance[5]. Part of this plan is the introduction of the EU taxonomy[6], introduced as a classification system containing the technical criteria necessary to identify environmentally sustainable economic activities that make a substantial contribution (without causing significant damage) to 6 main objectives: climate change mitigation, climate change adaptation, sustainable use and protection of water and marine resources, transition to a circular economy, pollution prevention and control, protection and restoration of biodiversity and ecosystems.

[1]Conference of the Parties

[2]https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_en

[3]https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal/delivering-european-green-deal_en

[4]https://ec.europa.eu/clima/eu-action/european-green-deal/european-climate-law_en

[5]https://finance.ec.europa.eu/publications/renewed-sustainable-finance-strategy-and-implementation-action-plan-financing-sustainable-growth_en

[6]Regulation (EU) 2020/852 (Taxonomy) on the establishment of a framework to facilitate sustainable investment

In March 2021, the EU adopted the Sustainable Finance Disclosure Regulation, which provides sustainability disclosure obligations for financial institutions and financial advisers toward end-investors. In April 2021 the European Commission adopted a proposal for a Corporate Sustainability Reporting Directive to amend the Non-Financial Reporting Directive. The latter requires large public-interest companies (with more than 500 employees) to disclose information about non-financial aspects, such as the way firms manage social and environmental challenges and the level of diversity on their boards. The proposed policy will extend a more detailed version of the reporting requirements to all large companies and all listed companies.

The importance of the role of regulatory authorities in contributing to the green transition has been confirmed by the former Chair of th European Securities and Markets Authority, Steven Maijoor, who declared: "*The financial sector plays a pivotal role as a key transmission channel of the much-needed transformation towards a more sustainable economic system. [...] Key pillars supporting the shift towards a more sustainable financial system are the measurement, verification and disclosure of ESG factors. It is important that public authorities step in and establish robust ESG standards and supervise the relevant actors and products to prevent the risk of greenwashing*". Beyond this, the Taxonomy Regulation requires all large and all listed companies (excluding micro-enterprises) to publish their Taxonomy Eligibility Report starting from 2022, namely reporting information on the alignment of their business with the taxonomy rules, and further disclosure will be requested from 2023.

The interest on part of institutions towards sustainability topics mirrors the one exhibited in recent years by players in the financial markets. Assets managers have been assigning a growing importance to ESG dimensions during the selection of assets in their portfolio. As a result, sustainable and responsible investments have grown in terms of assets by more than 38% since 2016 (Forum for Sustainable and Responsible Investment). This growth is expected to continue in the future (Global Insights Report, February 2019), revealing the importance attributed by investors to corporate social performance. Moreover, in the last decade we also witnessed the emergence of many rating agencies (such as Vigeo Eiris, MSCI, Sustainalitics, Refinitiv, RepRisk) providing ESG ratings to corporations. These agencies collect firm-level and sector-level information on several dimensions, such as emissions, corporate responsibility issues and governance practices, and employ such ESG "raw" data for constructing sub-indexes at different levels of aggregation, following proprietary pyramid-schemes and weighting systems, for eventually assigning a grade (A, A+, B...) to each company. These ratings are then used as criteria for constructing ESG portfolios, providing investors with a selection of ESG companies to invest in.

3 Related Literature

Several empirical studies have recently used ESG ratings as characteristics for explaining corporate performance at the firm-level. For instance, Cornett et al. (2016) find that banks' financial performance is positively and significantly related to the scores reflecting Corporate Social Responsibility. Friede et al. (2015) provide a review on the link between ESG and corporate performance. Albuquerque et al. (2020) find that companies with higher ES ratings exhibit higher returns, lower volatility, and higher trading volumes compared to other stocks during the recent Covid-19 pandemic. Some studies find that ESG portfolios provide a positive yield to investors (Derwall et al. (2005); Kempf and Osthoff (2007); Statman and Glushkov (2009); Belu and Manescu (2013); Eccles et al. (2014); Nofsinger and Varma (2014); Auer and Schuhmacher (2016); Henke (2016)). Hvidkjær (2017) reveal a strong outperformance of stocks with high ESG ratings when compared to alternative benchmarks over the last 30 years. Halbritter and Dorfleitner (2015) show that the extra performance of ESG portfolios strongly depends on the rating provider. On the same line, some recent contributions have reported how ESG scores can be misleading: different agencies propose very different scores for the same company and the criteria underlying score formation change over time, leading to re-classifications of companies altering empirical analysis (see Berg et al. (2022) and Berg et al. (2020)).

ESG characteristics might be linked not only to firm performance but also to creditworthiness, even though little empirical work is available on this issue. Scholz et al. (1995) note that a substantial percentage of bank credit losses in Germany could be attributed to environmental risks, while Weber et al. (2010) show that sustainability criteria can be used to predict the financial performance of a borrower and improve the predictive validity of the credit rating process. They also prove that banks using environmental and social criteria during the loan assessment process are more accurate in classifying the risks of loans in their portfolios: this indicates that factors affecting sustainability of a firm influences its creditworthiness. Chava (2014) finds that lenders charge a significantly higher interest rate on the bank loans issued to firms with environmental concerns; moreover, they show that these firms exhibit a smaller institutional ownership and fewer banks in their loan syndicate with respect to firms without such environmental issues. Devalle et al. (2017) investigate the impact of ESG performance on credit ratings on a sample of 56 Italian and Spanish public firms: their results confirm that social and governance variables affect credit ratings contrary to environmental factors. They suggest integrating the creditworthiness evaluation of firms with ESG factors as they affect borrowers' cash flows and the likelihood of default on their debt obligations. The first large-sample empirical study of the link between ESG performance and credit risk is provided by Henisz and McGlinch (2019). They find that companies with stronger ESG performances have lower credit yield spreads than other companies. They also show that better ESG performers have fewer revenue-reducing negative events and fewer lawsuits. Wood et al. (2015) state that the integration of ESG issues into the credit-research process provides a more comprehensive and forward-looking evaluation of a borrower's ability to repay a debt.

Credit rating agencies have proven to be aware of the link existing between ESG performance and creditworthiness. According to the Financial Times[7] (2019), "*credit rating agencies increasingly view risks through an ESG lens when they assess if corporate bond issuers will be able to pay back their obligations and stay in business*". A recent report by Moody's stresses how social issues can be material to credit quality, affecting it through four potential channels: reputational, operational, litigation and regulatory risks. Such an awareness of market participants can be found in the UNPRI (United Nations Principles for Responsible Investment) Statement on ESG in credit risk and ratings, which has currently been signed by more than 170 investors, with more than US 36 trillion dollars in collective Asset Under Management, and 26 Credit Rating Agencies (CRAs), including Fitch, Moody's and S&P. The agreement recognises that *"ESG factors can affect borrowers' cash flows and the likelihood that they will default on their debt obligations"* and their importance in the assessment of the borrowers' creditworthiness. CRAs are committed to evaluate the credit-relevance of ESG factors for different issuers and identify and understand how ESG risks affect creditworthiness. For this reason, Fitch published in January 2019 the ESG Relevance Scores (ESG.RS): these scores vary from 1 (No Impact) to 5 (High Impact) and display both the relevance and materiality of each ESG risk elements to the rating decision. These scores are analyst observations on how ESG factors impact the final credit rating decision and can be used by market participants to understand how a specific sub-category of ESG risks has impacted credit rating decisions and how an ESG issue directly affects a company's current credit rating.

Despite this evidence, there is still lack of a credit risk model that explicitly includes ESG dimensions among the determinants of the risk of default. Attig et al. (2013) show that the sustainability criteria can help predicting traditional credit ratings. Some studies look at the relationship between ESG and Corporate Social Responsibility factors on the cost of capital, however their results are mixed (see El Ghoul et al. (2011), Menz (2010), Goss and Roberts (2011), Sharfman and Fernando (2008)). Our project aims at filling this gap in the literature: we use a cross-section for 2019 of European listed companies for which we were able to obtain ESG raw factors and, at the same time, build a measure of creditworthiness; using Machine Learning techniques, we provide a list of the ESG characteristics relevant for predicting the risk of default; we analyze their relationship with creditworthiness.

4 Data

Our dataset is obtained by merging different data sources. In this section, we provide a description of the different datasets used as well as some summary statistics on the main variables of interest.

[7]https://www.ft.com/content/c1f29e0c-6012-3ac5-9a05-13444b89c5ec

Our starting point is the ESG "raw" information we sourced from MSCI, our choice among the ESG rating providers as it stores this information for a sufficiently large number of companies. This database includes about seven hundred ESG variables, ranging from carbon emissions to worker fatalities per company as well as governance information (for instance, on board diversity and composition). We include in our analysis individual company's characteristics -such as age and export status- as potential firm-level explanatory variables for creditworthiness.

4.1 Credit risk

Credit risk is usually measured through scores, constructed by rating agencies such as Moody's, S&P, Fitch. Alternatively, CDS prices can be employed for the same purpose, since they can be used to compute the implied probability of default. A third possible strategy implies computing the Altman (1968) z-score, constructed using accounting ratios. While credit ratings and CDS are only available for listed companies, a measure of credit worthiness can be computed for non-listed firms, allowing for potential analysis of smaller size companies. For this reason, we decided to follow the method suggested by Altman (1968), based on balance sheet ratios.

Altman (1968) developed a model to predict a risk that a company goes bankrupt: companies can be classified into different groups (distressed/uncertain/safe) according to a score computed as a linear combination of accounting ratios. The weights were obtained by the author using MDA (Multiple Discriminant Analysis) on a sample of firms that already filed for bankruptcy. For non-US companies, such as the ones in our paper, the formula turns out to be the following:

$$Z = 6.56\ X_1 + 3.26\ X_2 + 6.72\ X_3 + 1.05\ X_4 + 3.25$$

where:
X_1 = Working Capital / Total Assets
X_2 = Retained Earnings / Total Assets
X_3 = Earnings Before Interest and Taxes (EBIT)/ Total Assets
X_4 = Book Value of Equity / Book Value of Total Liabilities

Working capital is a proxy for liquidity, whereas retained earnings and the EBIT are measures of a firm's profitability. All this figures are normalized by total assets to account for a company's size. The latter indicator, X_4, represents the debt-to-equity ratio at market prices, that is a measure of leverage.

The Altman z-score classifies companies in three categories: for scores below 1.1, companies are considered to be distressed; if the score is above 2.6, a company is considered safe; all companies whose score is in between these thresholds are instead considered to be in a "grey" area, i.e. we cannot robustly classify them based only on balance sheet information. The majority of companies in our sample (48 %) are safe, whereas 23% are distressed and 29% are in the grey area. Figure 3 presents the distribution of the Altman z-score in our sample. Figure 4 and 5 show the mean and the median values for the Altman z-score in our sample, by country and by sector. We can see that Financial and Insurance companies (Sector K) show higher values for the Altman z-score: for this reason, we will also be performing our analysis separately on all non-financial companies.

4.2 Information about ESG factors

We have access to ESG raw information for a cross-section of European companies in 2019 from MSCI ESG Manager. The dataset includes the information that MSCI elaborates using a proprietary algorithm for rating companies: for the year 2019, it includes up to 656 ESG factors on more than 12200 European firms. For 5330 of them, we were able to compute the z-score, by retrieving the necessary balance-sheet figures from Orbis and FactSet. After removing all firms without ESG coverage and all ESG factors without available information, we obtained a dataset of 537 ESG factors for 1251 European companies (see Table 3 and Figure 6, 7 and 8 for firms characteristics). Figure 1 provides a break-down of the companies included in our sample by country and Figure 2 shows how many firms belong to each sector included in our sample. Table 1 provides the list of sectors considered, while Table 2 provides the numbers of companies per country and sector are included in the dataset.

The information about ESG factors is quite heterogeneous, both in terms of coverage as well as of the quality of the information included. Figure 9 reports the number of variables considered within each subcategory. The *Environmental* factors (E) are related to the environmental impact of each company. Among them, we find records on CO2 emissions according to Scope 1, Scope 2 and Scope 3, representing respectively direct emissions, indirect emissions via energy consumption and emissions that indirectly impact the company's value chain. Moreover, there is information on the emissions of other pollutants, on the intensity of water usage, on the use of renewable energy sources, as well as on a company's impact on biodiversity. The *Social responsibility* variables (S) include very different information on how a company deals with social and ethical aspects. For instance, we can find the percentage of revenues linked to products containing substances of chemical safety concern, whether there are ethical policies in place on the company, whether the company invests in community development projects as well as records on workers fatalities. The *Governance* set of factors (G) includes instead detailed information on board members such as gender, years of tenure, wages, as well as detailed information on board votes.

Even after removing firms and factors with no available information, we still face coverage issues. As a result, we perform a thorough analysis of the distribution of missing values in our data. First, we evaluate the number of missing ESG values per firm, either considering the full set of ESG factors or partitioning them according to the three main pillars to which they belong (E for Environmental, S for Social Responsibility and G for Governance). As shown in Figure 10, 50% of firms in the dataset present at least 200 missing ESG factors.

Moreover, a substantial group of companies have barely 150 ESG entries out of 537 potential variables. Looking at the different levels of disclosure between E, S and G factors (Figure 11), there is on average a greater coverage on the E and G dimensions rather than on the S one, even though G factors present a greater variability, with higher numbers of missing values per firm. More in detail, Figure 12 shows that more than 60% of companies have at least 1/3 of the E factors unavailable, while in Figure 13 we clearly see a generalized lower level of disclosure for S variables. Last, the majority of firms have at least 60% of available values for G factors, even though a group of around 50 companies have hardly any information at all on this pillar (Figure 14).

To check for a potential and concerning distortion in our sample, we examine whether "safer" firms provide more information than distressed ones. Figure 15 proves that our sample is balanced in terms of degree of disclosure across the 3 type of companies. We exclude that differences in coverage among companies are linked to their creditworthiness measured by the Altman's z-score. We also check whether coverage may depend on industry characteristics, in particular by partitioning our sample between non-financial firms and manufacturing. As shown in Figure 16, our data do not seem to show this problematic issue.

Another way of analysing the distribution of missing values is by counting the number of missing entries for each factor. As before, we do this either by considering separately each E, S and G factor or studying them all together. In Figure 17, we observe that more than half of the ESG variables have less than 250 missing values (out of 1251 entries), and this figure is confirmed for all categories (see Figures 18, 19, 20). Nonetheless, there is a group of E and S factors with poor coverage, especially the S variables (at least 40% have less than 20% of available data). This issue calls for an additional cleaning of the data. To this aim, we proceed through the following steps. After removing the variables with very poor coverage, we employ the solutions proposed by Kuhn et al. (2003) for handling data pre-processing. The variables that we consider are numerous, but they often report detailed information on a specific topic (such as environmental concerns). For this reason, we observe in our sample several groups of highly correlated factors showing similar information. On the other hand, we have also many non-informative variables, namely those with low or zero variance. Aiming at removing these predictors lacking of information content, we adopt the following approach:

- we delete factors with no variance whatsoever (only one value observed) or factors that have both of the following characteristics: they have few single values relative to the number of samples (i.e. the percentage of distinct values out of the number of total samples, cutoff: 10%), and the ratio of the frequency of the most common value to the frequency of the second most common value is large (cutoff: 19).

- we keep only one variable within each group of highly correlated variables: if two variables have a high correlation (cutoff: 0.75), we remove the one with the largest mean absolute correlation.

We apply this process of data cleaning on the entire dataset as well as separately on two subsamples: the first excludes non-financial firms, while the second concentrates on manufacturing firms. The reason of this choice is the presence of some factors related to specific characteristics of a sector or a industry, thus non-informative for firms belonging to other sectors.

The resulting dataset comprises 1076 European companies, 464 of which are manufacturing firms, while 43 are financial (so that the subset excluding them includes 1033 firms). There are 102 ESG factors left in the "cleaned" sample, divided into 51 E factor, 43 S factors and 8 G factors (this last equal in both sub-samples). Similar results are obtained when excluding non-financial firms (51 E factor, 41 S factors) and when looking at manufacturing companies only (47 E factors, 38 S factors). We consider again the characteristics of firms in such a reduced dataset and notice that there are no significant differences between this treated version of the data and the original data (see Tables 3 and 4). We also observe a definite shift in financial firms' distributions of their balance sheet information that reflects the specificity of this sector compared to the rest (Figures 21, 22 and 23). This feature avails our decision of separating the financial firms from the remaining companies in our analysis. On these datasets, merged with variables containing balance sheet information, we apply the empirical strategy described in the following sections.

5 Variable selection strategy

Our objective is to link the z-score to ESG variables. Due to the huge number of variables involved, we employ techniques of Supervised Machine Learning (SML) and Unsupervised Machine Learning (UML). This section provides a basic description of the techniques used.

Our preferred method for the analysis that we perform is the Least Absolute Shrinkage and Selection Operator (LASSO), firstly proposed by Tibshirani (1994), belongs to the family of SML techniques, where a dependent variable has to be identified. It solves the optimization problem

$$\hat{\beta} = argmin\Big\{\frac{1}{2n}\sum_{i=1}^{n}(y_i - x_i\beta')^2 + \lambda\sum_{j=1}^{p}\omega_j|\beta_j|\Big\} \tag{1}$$

where:

- $\lambda > 0$ is the lasso penalty parameter (if $\lambda = 0$, LASSO reduces to OLS)
- y_i is the dependent variable (in our case, the z-score), where $i = 1, ..., n$ indicates the number of observations

- x_i contains the n potential explanatory variables
- β is the n-th dimensional vector of coefficients on x_i
- ω_j are parameter-level weights known as penalty loadings

If λ increases, the penalty term increases and we see a shrinkage of the coefficients towards zero. Selecting a model (i.e. selecting a set of explanatory variables) corresponds to selecting a value for the penalty parameter λ using the Cross Validation (CV) method.

Being LASSO a method for selecting variables based on prediction, classical inference procedures no longer have the properties established by classical theory and this invalidates parametric inference. A method often used as an alternative to parametric inference is bootstrapping. The basic idea of bootstrapping is that inference about a population from sample data can be modeled by performing inference about the sample from re-sampled data. In our case, we want to assign confidence intervals to the parameters estimated by LASSO so to improve the robustness of the selection process. The Bootstrap-LASSO algorithm runs as follows:

- Draw a random sample (with replacement) of the same size of the original data.
- On this bootstrap sample apply LASSO and store the estimated parameters.
- Repeat the previous steps for b=1 to B, where B is the number of bootstrap samples (here we use B=1000).
- For each coefficient, evaluate the confidence interval based on the empirical distribution function.

In order to select a manageable number of variables, we employ the results of combining the bootstrap with LASSO: we choose a variable only if its estimated coefficient is significantly different from 0 at least in 60% of the bootstrap replications (for non-financial firms it is possible to increase the threshold to 90%).

As a way of improving the robustness of our results, we employ other techniques for selecting variables. Random Forests (RF) is a SML technique based on a combination of tree predictors such that a large collection of low-correlated trees is built and their average is used to form the ultimate predictor (Breiman, 2001). Each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest, so that the variance reduction derived from the bagging (i.e. bootstrap aggregating) procedure is combined with a reduction of correlation between the trees. This injection of randomness in the algorithm makes this method an accurate classifier and regressor. The Bootstrap-RF algorithm operates as follows:

1. Draw a bootstrap sample of size n from the training data on the vector x of p original variables.

2. Grow a random-forest tree T_b to the bootstrapped data, by recursively repeating the following steps for each node of the tree, until the minimum node size is reached:
   (a) select m variables at random from the p variables.
   (b) choose the best variable among them and split the node into two daughter nodes at the best split-point.
3. Repeat this procedure for b=1 to B to obtain the ensemble of trees.
4. The random forest predictor is $\hat{f}_B(x^*) = \sum_{b=1}^{B} T_b(x^*)$.

Random Forest is also used to construct a variable importance measure that evaluates the prediction strength of each variable. This measure is computed as follows. After each tree is grown, its prediction accuracy is recorded. Then the values of a variable are randomly permuted and the accuracy is again evaluated. The decrease in accuracy due to this permuting is averaged over all trees and is used as a measure of the importance of this variable in the random forests predictor. This procedure can be used for each variable in the dataset, so that we can identify the most powerful factors in predicting the z-score.

Another SML technique, based on a linear regression model like LASSO, is the selection by the AIC criterion. Its formula is $AIC = -2ln(\hat{L}) + 2p$, where p is the number of estimated parameters in the model and $\hat{L}$ is the maximum value of the likelihood function. Thus, the AIC evaluates the goodness of fit of a model, including a penalty factor that is an increasing function of the number of parameters, so that the trade-off between overfitting and simplicity is taken into account. The selection is achieved through an algorithm that can be run forward, backward, or stepwise. In the first case, the algorithm starts by estimating the regression with no covariates and at each step it adds a variable. Each time the AIC criterion is recorded and the algorithm stops when the AIC stops to significantly improve. The covariates included at this step of the algorithm are those which best fit the model without compromising its simplicity. The backward procedure works in the opposite direction, namely starting from the full model with all possible covariates and removing one variable at a time. The stepwise method is a combination of both forward and backward, since at each step of the algorithm a covariate is added or removed, depending on its power in improving the AIC.

As a last approach, we use the Principal Components Analysis (PCA). This Unsupervised ML technique has been developed with the aim of reducing the dimensions of a dataset, but it does not consider the predictive power of each variable. In other words, we run PCA on the ESG factors *without* considering their possible relation with creditworthiness, but only with the aim of selecting the factors that best summarize the initial set of variables according to the following definition. PCA is defined as an orthogonal linear transformation of the data by a scalar projection such that the greatest variance of the data comes to lie on the first coordinate (called the first principal component), the second coordinate explains most of the variance of what is left once the effect of the first is removed, and so on until the whole variance is explained. Given a $n \times p$ data matrix X, the transformation is defined by a set w of k (with $k < p$) p-dimensional vectors $w_{(j)}$, $j = 1, ...k$ such that the principal components are computed as $t = Xw$. The first weight vector, $w_{(1)}$, must satisfy the following maximization problem:

$$arg \max_{||w||=1} \Big\{ ||Xw||^2 \Big\}. \tag{2}$$

Then, the first principal component of a set of variables is the derived vector formed as a linear combination of the original variables that explains most of the variance. This new variable often preserves up to 80/90% of the whole variance of the data. When this is the case, it can be used as covariate itself or to select the most important variables in the dataset. Since the first alternative causes many difficulties in the interpretation of the estimated parameters of the model, we proceed by following the second approach. The importance of each variable is assessed by its weight (w_j) in the principal component, that is the magnitude (in absolute value) of the corresponding coefficient in the scalar projection. The larger its coefficient in the principal component, the greater the variable importance in explaining the variance of the data. In our study, we apply the PCA on three different sub-set of variables, namely the E, S and G factors separately. Our intention is to identify for each factor its main variables.

6 Selected variables

In this section, we present the ESG factors selected by each of the methods mentioned in the previous section, distinguishing between supervised and unsupervised techniques, and provide a brief description of their content (see Table 9 for the complete list). Due to the specific characteristics of financial firms (see Figure 5), we perform the analysis both on the whole sample and on the sample obtained by removing financial companies. Another sub-sample considered for variable selection concentrates on manufacturing firms, since they represent a substantial portion of the companies in our sample which may show a specific behaviour in terms of ESG practices.

We decide to keep for the analysis described in Section 7 only those variables that are selected by the majority (at least 3 out of 5) of the supervised ML techniques (we consider forward, backward and stepwise AIC algorithms as separate methods). As a result of the common regression framework, LASSO and the AIC algorithms provide similar results, while RF perform differently (in fact, it is not constrained to be a linear model). In Tables 5, 6, and 7 we report the selected ESG factors according to the different supervised methods on the three samples. Summary statistics and distributions are reported in Tables 10, 11, 12 and 13.

From PCA we obtain a significantly different set of variables. First, notice that we managed to isolate the first principal component as the one explaining most of the variance of the data only for the E factor. Indeed, S and G factors often present a dichotomic behaviour or do not have regular distributions, resulting in the PCA incapability to provide a useful outcome. There is not such a problem for many E variables, nonetheless the list of factors selected by the PCA is very different from the one derived from the supervised techniques presented above. The most likely explanation for this result stands in the unsupervised nature of the method: PCA does not consider the variables' importance in terms of their explanatory power of the z-score, but only in terms of their weight in representing the variance of the data. Related to this feature, we may assume that variables with higher variance and standard distributions are better candidates than variables with few categories or values and irregular distributions. Thus, the most important factors explaining nearly all the variance of the E pillar are Carbon Emissions Scope 1 and 2, as we can see in Table 8 (their weights definitely have a higher magnitude than those of the other factors proposed below). See Table 14 for summary statistics.

Notice that, before operating the selection procedure, we added to the ESG information the ordinary balance sheet variables, since they could still be pivotal in explaining the creditworthiness of firms. In fact, as we expected, the selection methods highlighted the following variables as the most important to this aim. Solvency Ratio, Current Ratio, ROA and firm's Age were selected in our analysis of the complete sample as well as in the non-financial subsample. For manufacturing firms, instead, Solvency Ratio, Current Ratio, ROE and export status were selected for the prediction of the z-score (summary statistics in tables 15 and 16).

7 The determinants of credit risk

After the selection process, we used the common subset of variables selected by the different SML methods as regressors in a OLS framework in order to study their relationship with the z-score. Consistent with our previous analysis, we estimate OLS models on the three different sub-set of the original data: full sample, non-financial firms and manufacturing companies (see Tables 17, 18 and 19 in the Appendix). We also keep these results separated from the ones obtained using the variables selectied by the unsupervised technique (Tables 20, 21, and 22 respectively).

Our regression results present first the models estimated without the ESG explanatory variable (Columns 1 and 2), then the integrated models. The increase in predictive power due to the introduction of the ESG factors among the covariates in the full sample is greater by 5-6% in terms of adjusted R^2, while excluding financial firms substantially improves the overall explanatory power up to an adjusted R^2 of 0.812, with an increases of nearly 2% (4% without fixed-effects). In the manufacturing sub-sample the effect is weaker but still present.

When including environmental covariates, the following evidence emerges. Carbon intensive lines of business negatively affect creditworthiness, which is instead enhanced by having activities in countries subject to adequate carbon regulations, especially for manufacturing firms. Companies involved in financing projects with an environmental impact exhibit higher creditworthiness only if they belong to the manufacturing sector. Effective reduction target for carbon emissions do not improve the z-score, this may be due to a rising impact on costs necessary to achieve the substantial reduction in carbon emissions. The same rationale applies to companies that are compelled to meet high energy requirements and green regulations with respect to the buildings in which they operate. Holding assets in regions that are typically highly water intensive has also a detrimental, though moderate, effect on creditworthiness perception.

Regarding social responsibility, we observe a consistent result through all sub-samples in two aspects, health safety and privacy security. Participation in activities with low worker injury rates has a harmful effect on corporations, while operating in regions with high injury rates seems to be beneficial for creditworthiness. We speculate that this is also due to a cost channel: companies allocating lower resources for safety protection of their workers have higher liquidity and higher z-scores. Companies with large amounts of revenues from countries where unsustainable lending practices are present face an adverse effect on their z-score. The opposite if revenues originate from lines of business that are reliant on highly-skilled or better-educated workers. Poor safety levels and quality induce a loss in creditworthiness. To a similar consequence leads the involvement in management of investment assets (both for the firm itself and on behalf of customers).

We now provide a few comments on the relation between the ESG score and balance sheet variables. As expected, a greater solvency capability, as well as a higher level of liquidity to meet short-term obligations and better profitability impact favourably on the z-score. The same is true when comparing long-established companies to the younger ones. For manufacturing firms, being an exporter (reaching destinations outside Europe) appears to be rewarding, even though those effects are incorporated into country fixed-effects, providing evidence of a country-specific characteristic influencing the z-score.

PCA results provide quite a different picture, probably because PCA, as a unsupervised approach, does not select the variables with the objective to predict creditworthiness but chooses the ones that better summarize the category to which they are belonging. The results show that large amounts of Scope 2 greenhouse gas emissions have detrimental implications on creditworthiness of firms, pointing at the crucial role of emissions caused by the generation of electricity purchased by companies rather than those from sources directly owned or controlled by them. This effect is not remarkable when emissions are normalized by a measure of firms' enterprise value. Regarding toxic emissions, lines of business intensively producing greenhouse gas seem positively associated with the z-score, while assets devoted to highly water intensive activities have a damaging effect. Probably due to the unsupervised nature of the PCA, however, we must notice that adding the E factors selected by the PCA as covariates in the regression does not substantially improve the goodness of fit.

8 Conclusions

In this paper, we study the association between ESG factors and credit risk, measured by the Altman z-score, at individual firm level for a sample of European companies with the aim to build an augmented model for credit risk where ESG factors help in predicting the company's creditworthiness.

The large number of potential explanatory variables for each company in our dataset has required the application of Machine Learning techniques, both supervised and unsupervised. We experimented several approaches to select the final ESG factors that survived among the different Machine Learning techniques.

The empirical evidence in this paper shows that ESG factors help explaining the probability of default: when including the ESG factors in a OLS model in order to explain credit risk, together with the traditional accounting variables, they contribute to improve the fit of the model by reducing the mean squared errors.

Our main findings suggest that companies with a moderate, rather than large, proportion of revenues related to carbon emissions or to green building have a higher credit risk, implying that an effort in reducing pollution or energy requirements is costly. On the contrary, hiring more skilled workers reduces credit risk, as it is associated to a greater company's productivity. Interestingly, we provide evidence of a positive externality from environmental friendly locations, since companies located in regions where carbon regulation is stricter exhibit a lower credit risk. Also companies located in regions with better data protection show a lower credit risk.

Our study suffers from few limitations, mostly related to the nature of the data. In particular, since the variability of the ESG factors across time is very limited, we cannot exploit the time dimension within a panel data framework. Hence we cannot infer any causal relation between ESG factors and credit risk. More importantly with a time dimension we would be able to capture the effect that a costly investment, for instance to improve on a ESG dimension today, has on the ability to pledge a higher revenue to investors tomorrow.

Finally, the selection of variables resulting from the application of Machine Learning techinques may have suffered from the lack of information on some ESG factors: for instance we observed a limited selection of S and G factors, mostly due to the absence of disclosure on these factors by many companies, while accounting records are available for most of them.

9 Tables and Figures

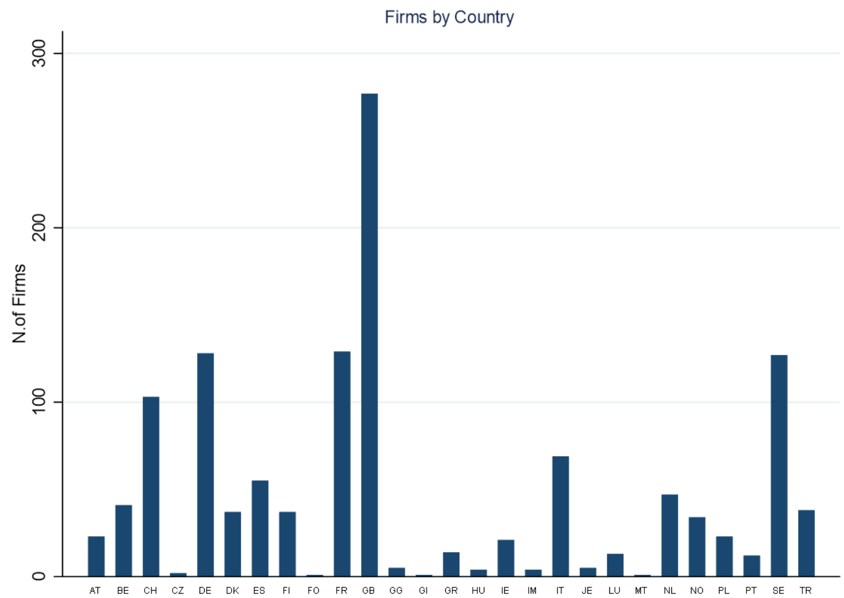

Figure 1: This figure shows the companies included in our sample by country.

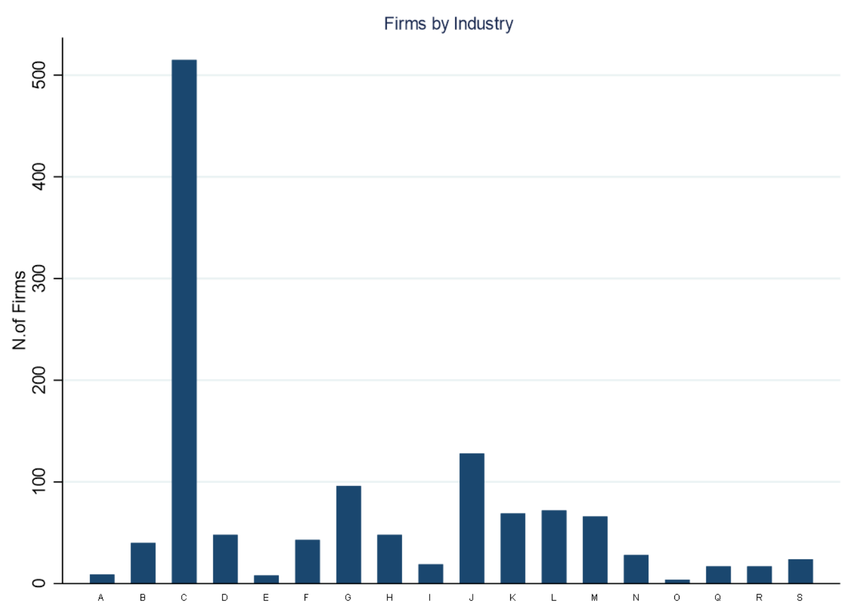

Figure 2: This figure shows the companies included in our sample by sector (see Table 1 for sector codes and description).

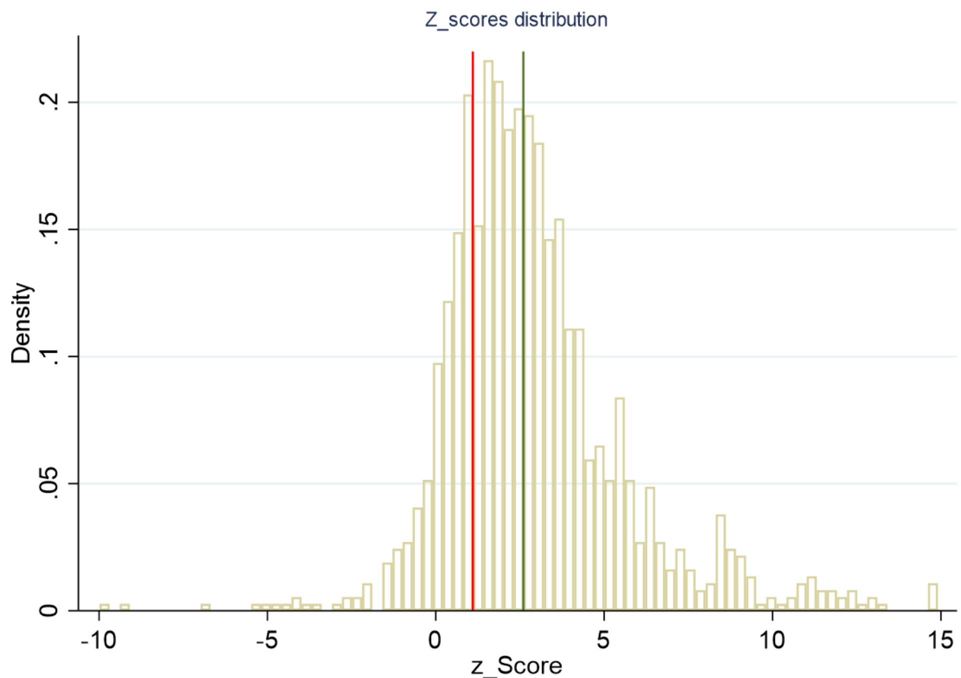

Figure 3: This figure presents the distribution of the Altman z-score in our sample.

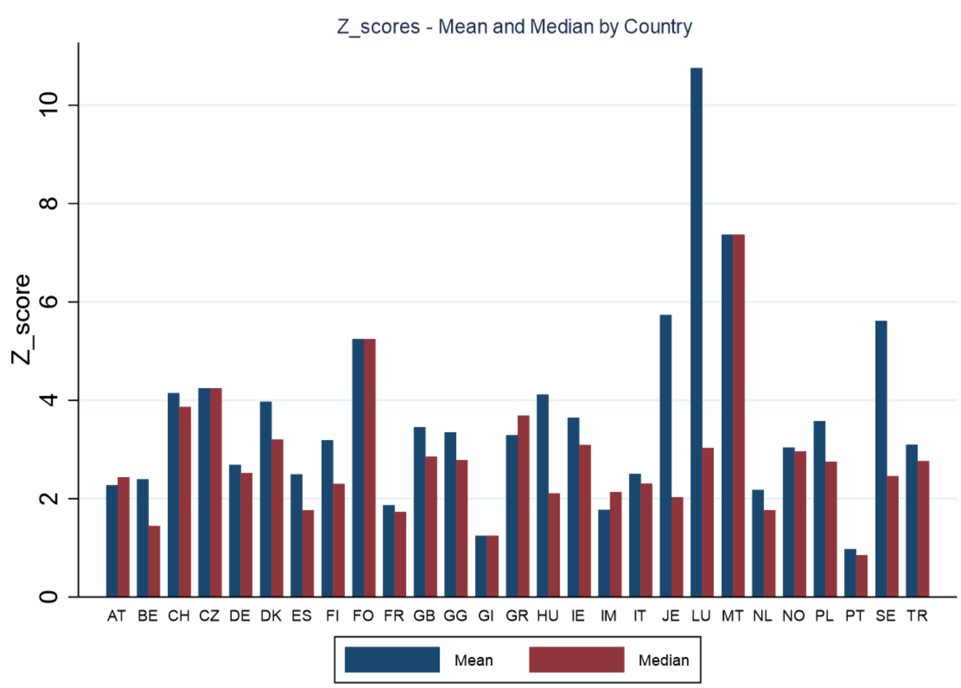

Figure 4: This figure shows the mean and median values by country for the Altman z-score in our sample.

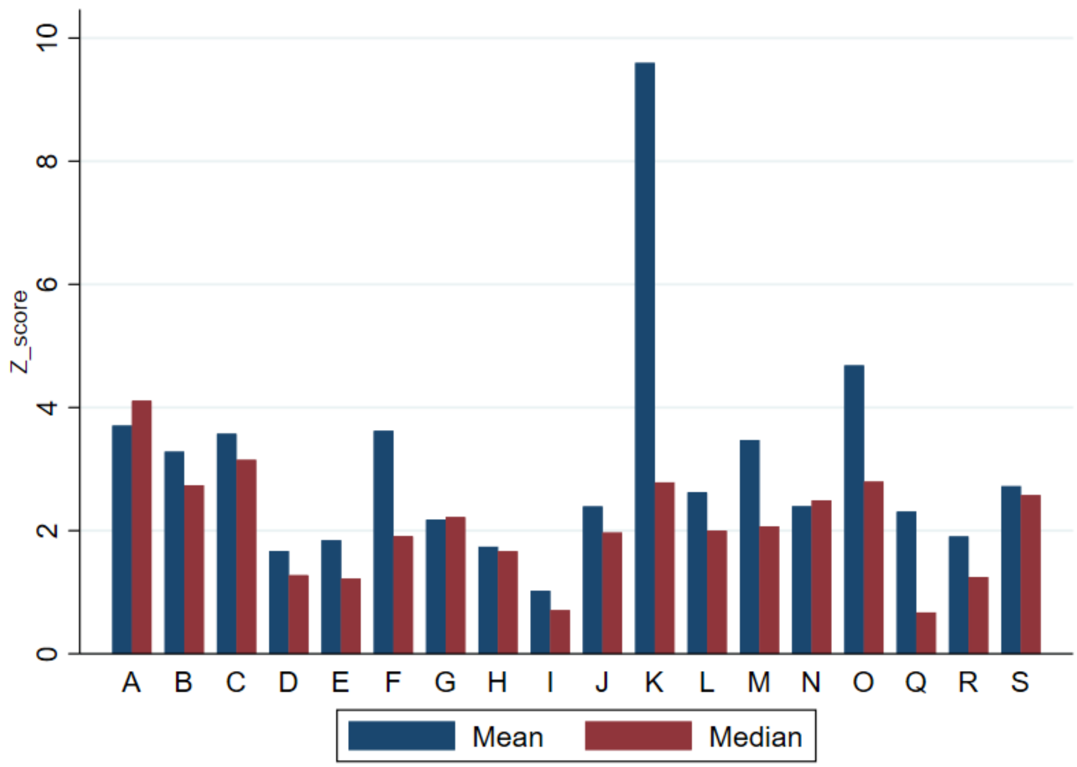

Figure 5: This figure shows the mean and median values by sector for the Altman z-score in our sample. See Table 1 for sector codes and description.

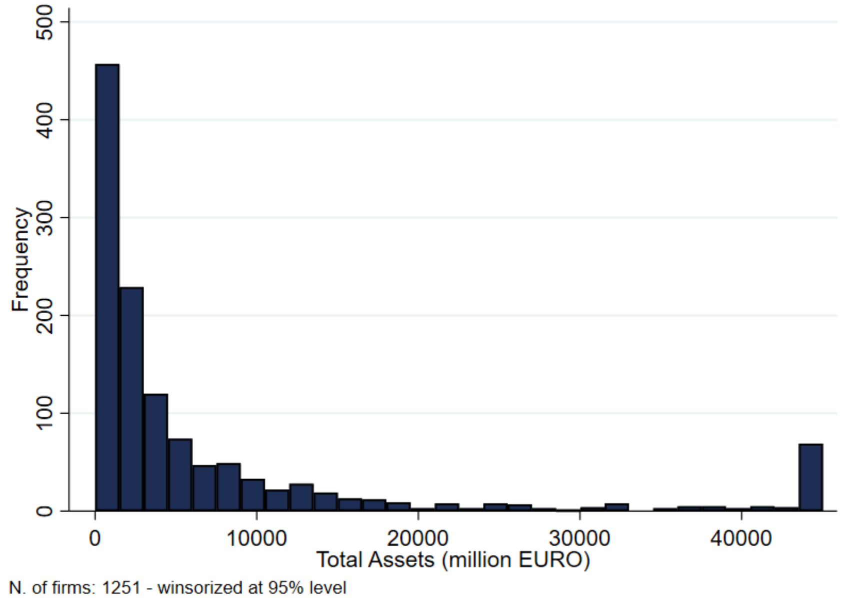

Figure 6: This figure presents the distribution of Total Assets for the 1251 firms in our sample. Extreme values have been winsorized at 95% level.

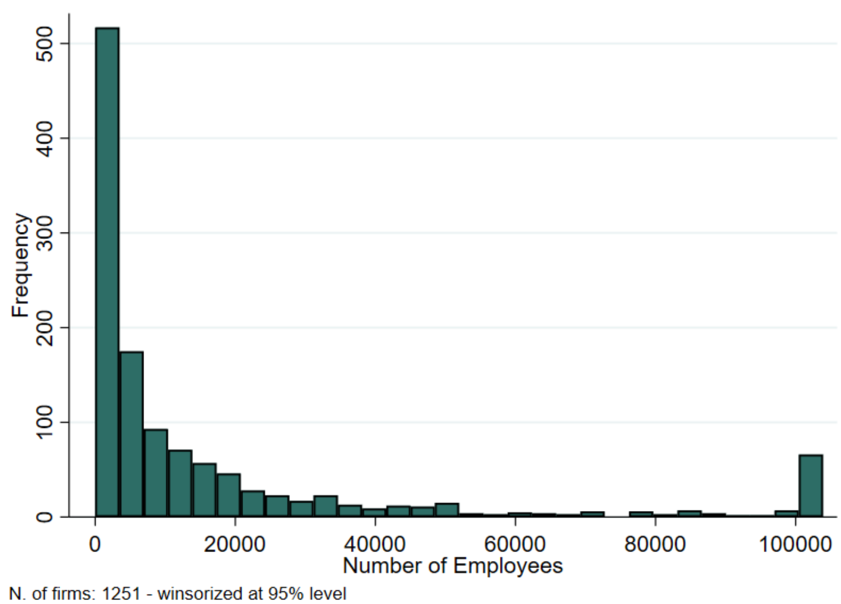

Figure 7: This figure presents the distribution of the Number of Employees for the 1251 firms in our sample. Extreme values have been winsorized at 95% level.

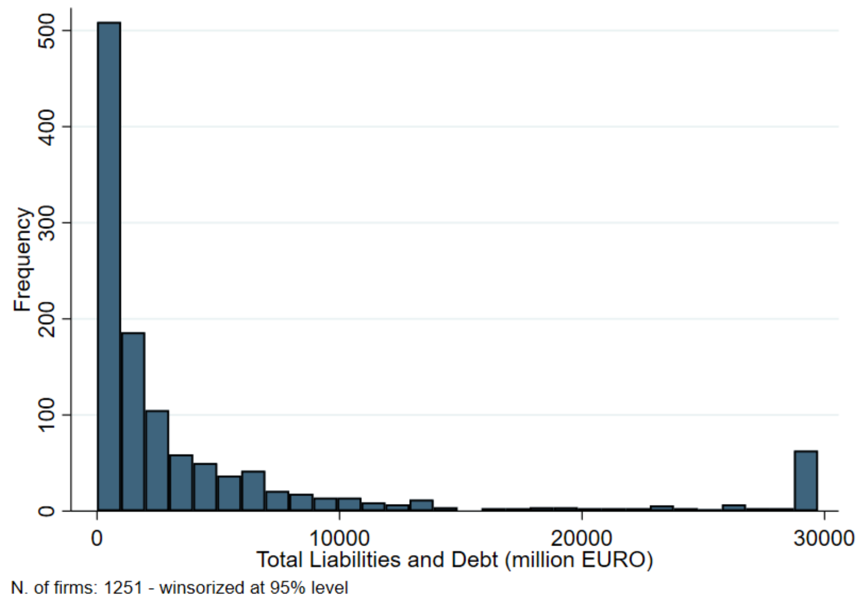

Figure 8: This figure presents the distribution of Total Liabilities and Debt for the 1251 firms in our sample. Extreme values have been winsorized at 95% level.

ENVIRONMENTAL		SOCIAL		GOVERNANCE	
Biodiversity & Land Use	9	Access to Communications	3	**Corporate Behaviour:**	
Carbon Emissions	39	Access to Finance	6	Business Ethics	22
Electronic Waste	6	Access to Healthcare	6	Legacy Data	33
Energy Efficiency	3	Chemical Safety	9	Tax Transparency	6
Financing Environmental Impact	4	Community Relations	21	**Corporate Governance:**	
Opportunities in Clean Tech	3	Consumer Financial Protection	6	Accounting	6
Opportunities in Green Building	6	Controversial Sourcing	3	Board	28
Opportunities in Renewable Energy	7	Health & Safety	28	Director Data:	
Packaging Material & Waste	9	Human Capital Development	6	- Board Committees	1
Product Carbon Footprint	3	Insuring Health & Demographic Risk	6	- Board Seats	26
Raw Material Sourcing	3	Labor Management	15	- Director Election Vote Results	19
Toxic Emissions and Waste	32	Opportunities in Nutrition and Health	9	- Individual Data	6
Water Stress	23	Privacy & Data Security	19	- Pay	9
Total	**148**	Product Safety and Quality	6	Management and Shareholders Proposals	24
		Responsible Investment	6	Ownership and Control	29
		Supply Chain Labor Standards	12	Pay	13
		Total	**161**	**Total**	**222**

Figure 9: This table reports the number of ESG variables included in the dataset by each subcategory.

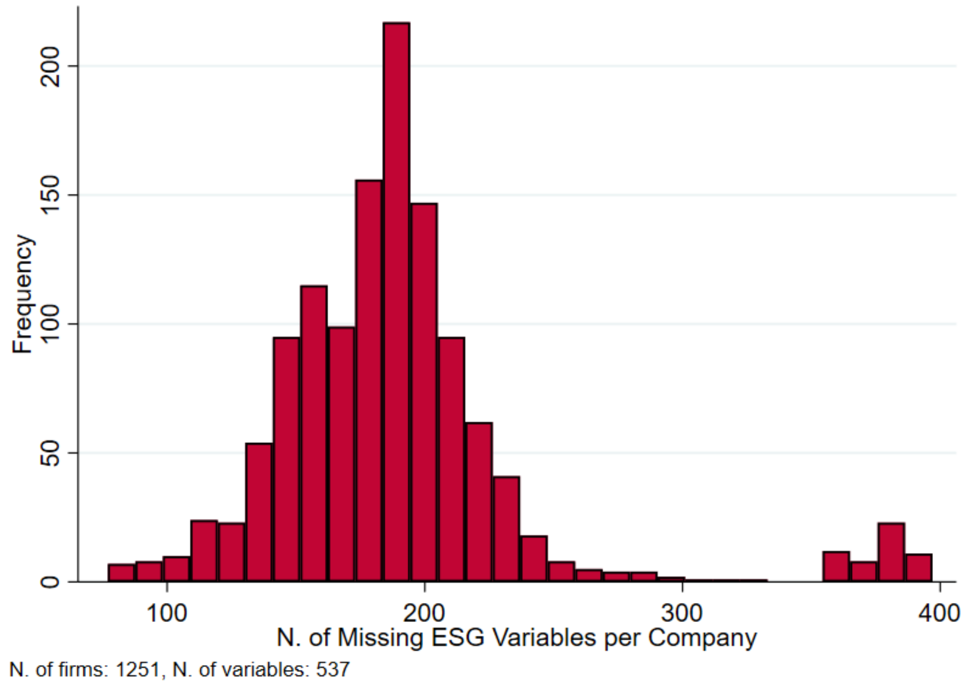

Figure 10: This figure shows the distribution of missing values per firm for the ESG variables.

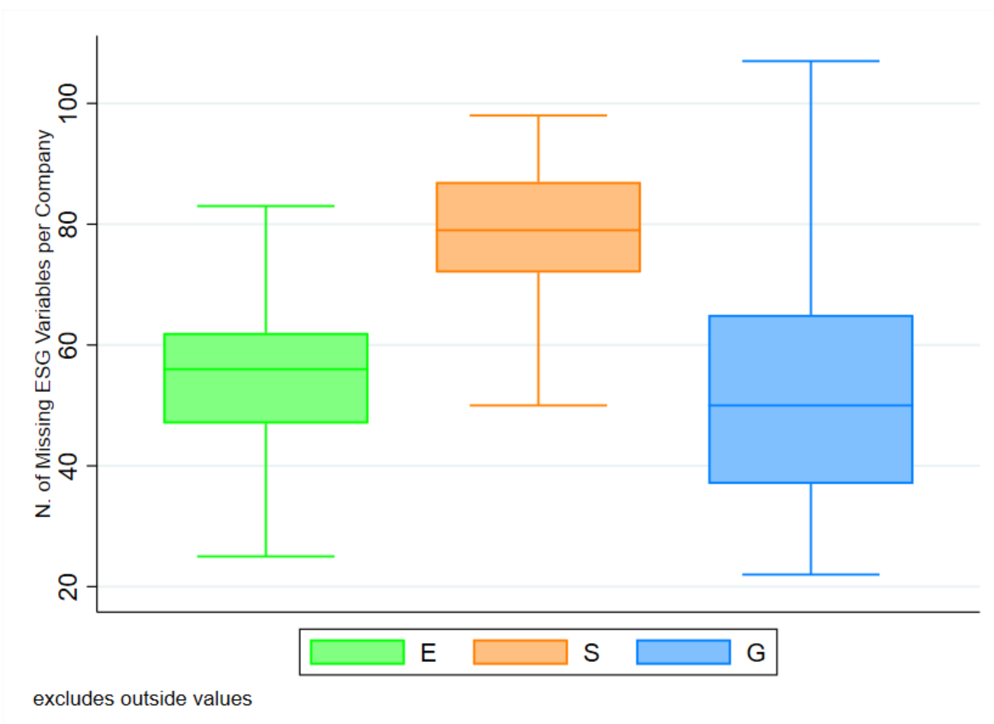

Figure 11: This figure shows the distributions of missing values per firm by E,S and G factor.

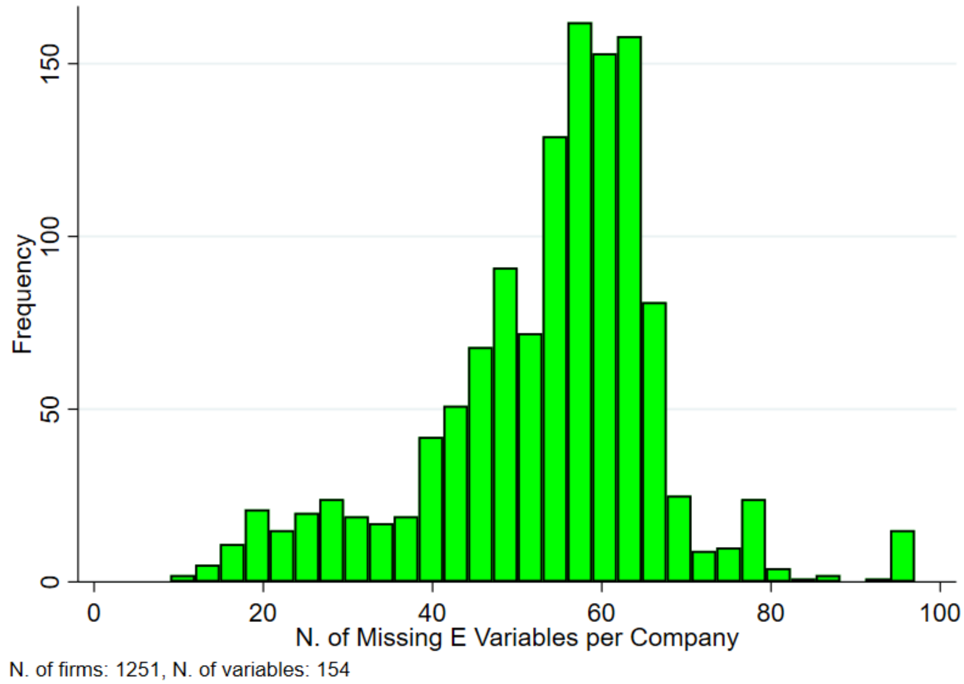

Figure 12: This figure shows the distribution of missing values per firm for the E variables.

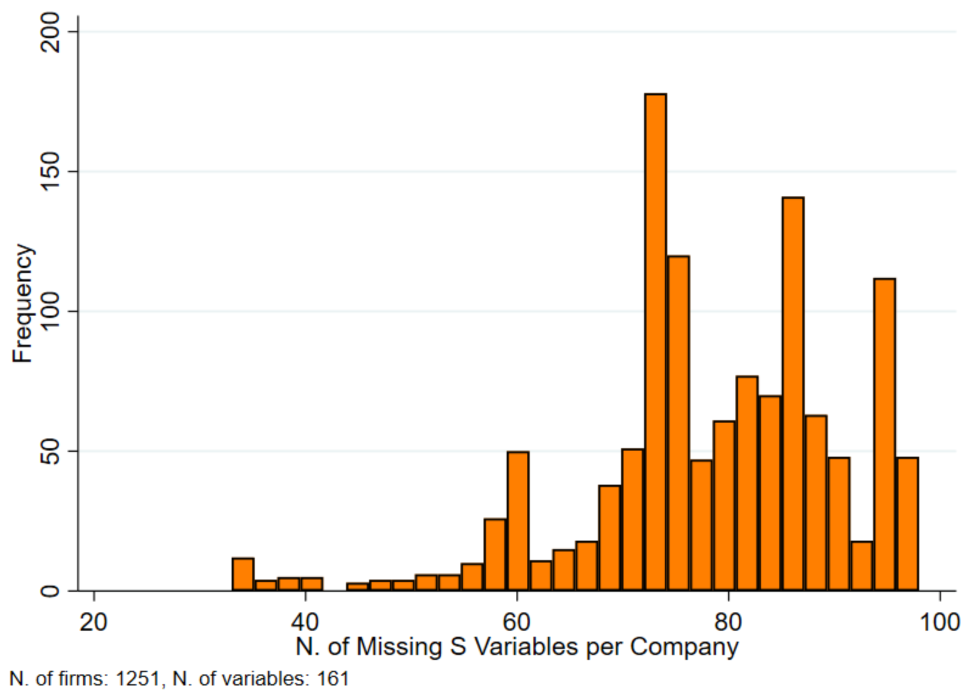

Figure 13: This figure shows the distribution of missing values per firm for the S variables.

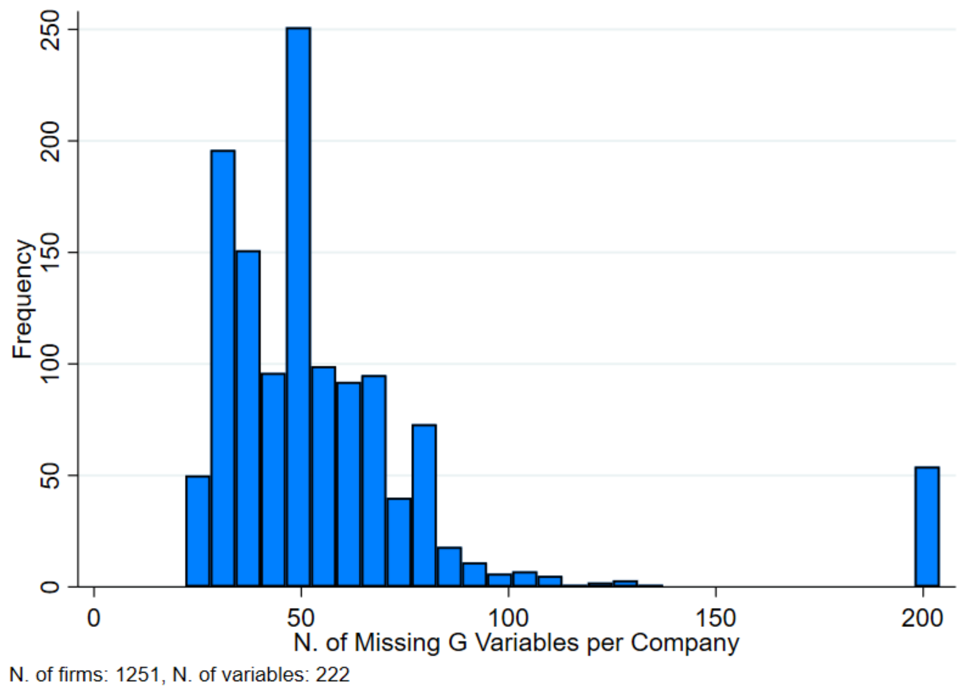

Figure 14: This figure shows the distribution of missing values per firm for the G variables.

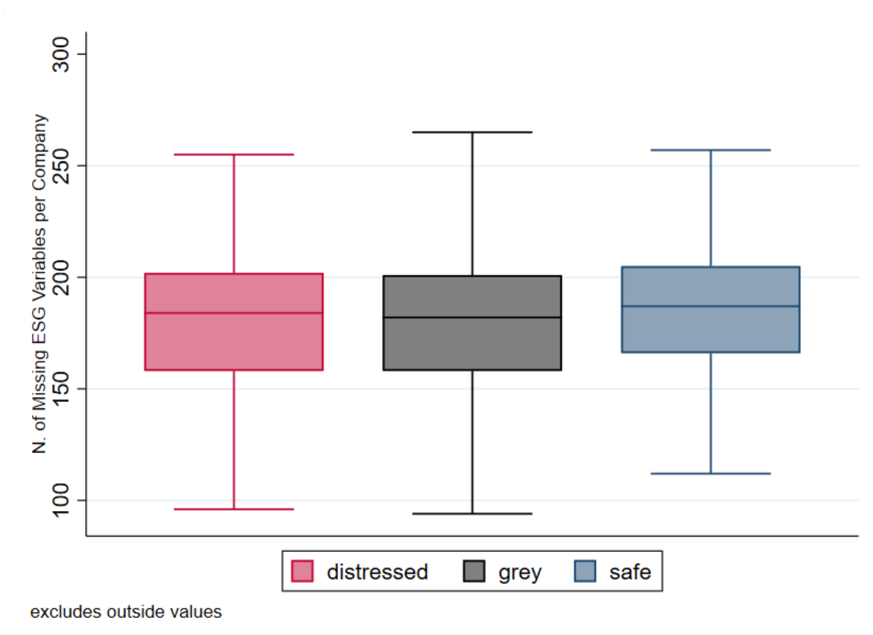

Figure 15: This figure shows the distributions of missing values per firm of the ESG factors by the z-score categories: safe, grey, and distressed companies.

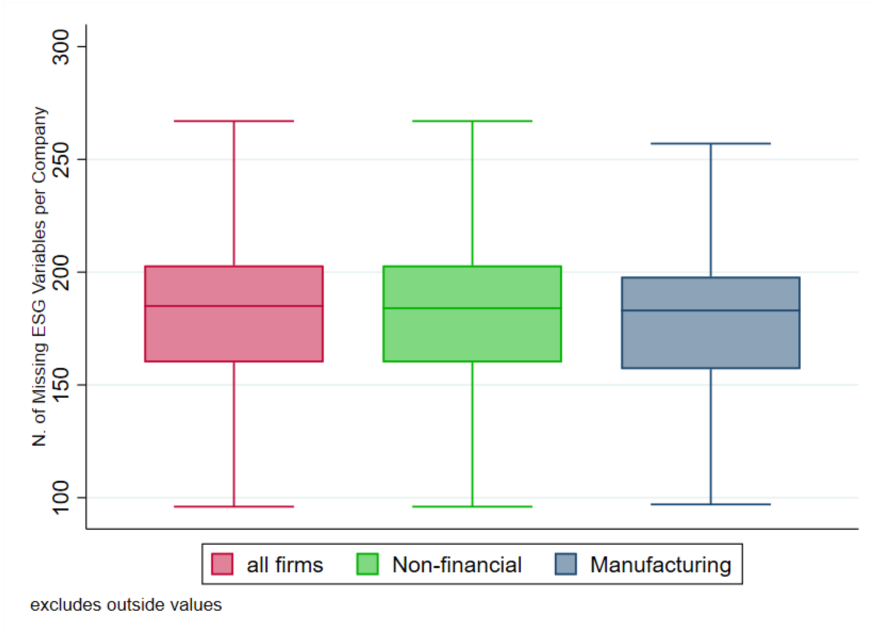

Figure 16: This figure shows the distributions of missing values per firm of the ESG factors by the three sub-samples: all companies, non-financial and manufacturing.

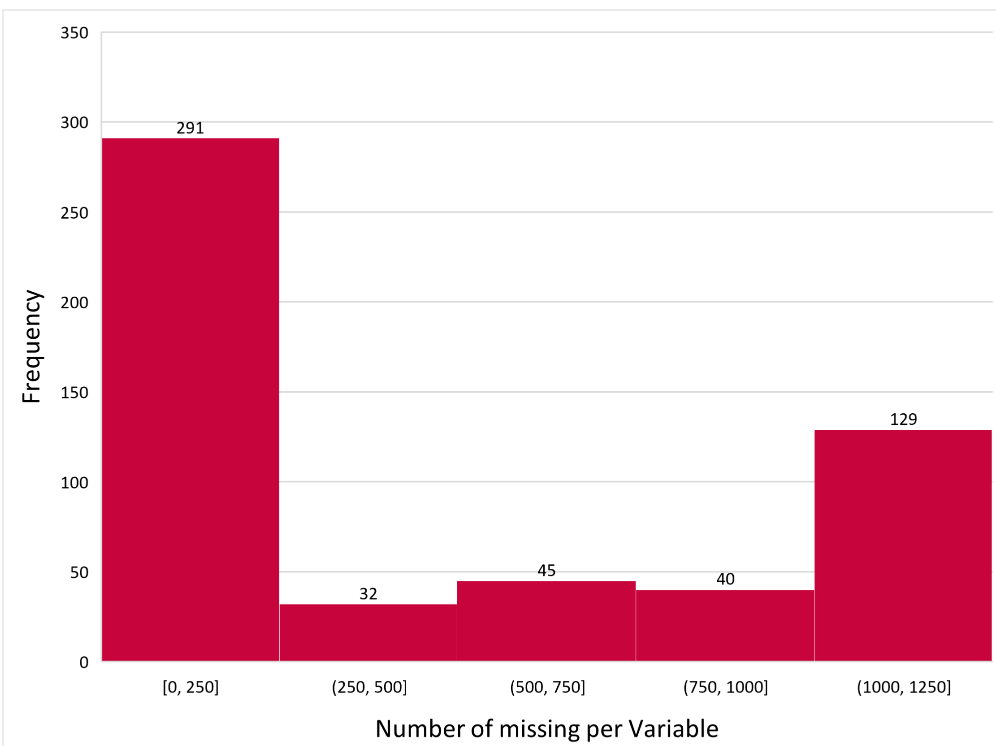

Figure 17: This figure presents the distribution of the number of missing values per variable over all the ESG factors.

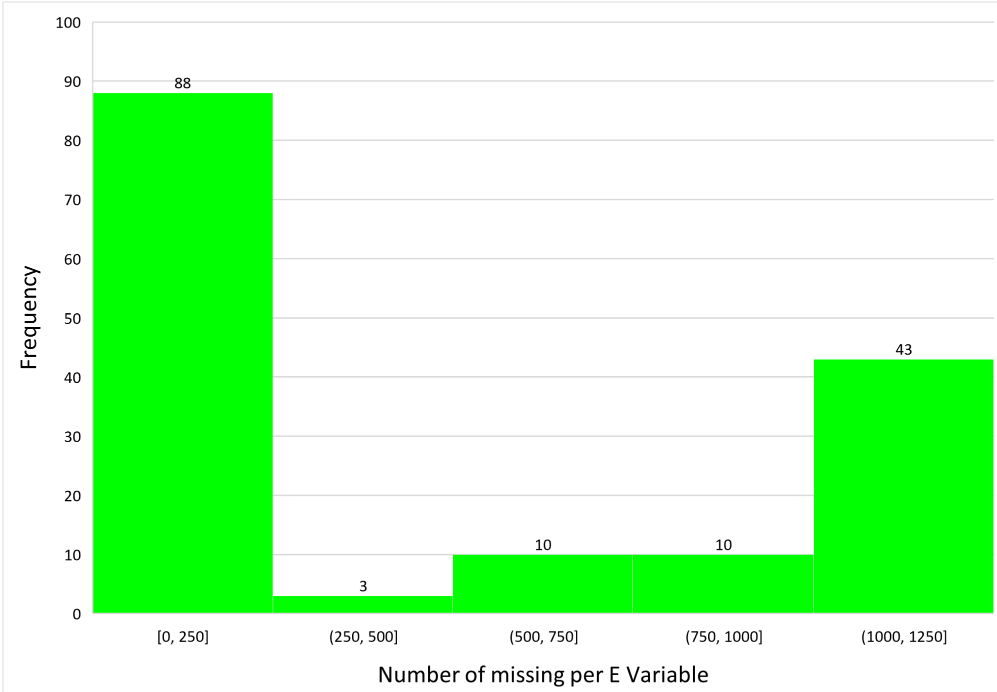

Figure 18: This figure presents the distribution of the number of missing values per variable over the E factors.

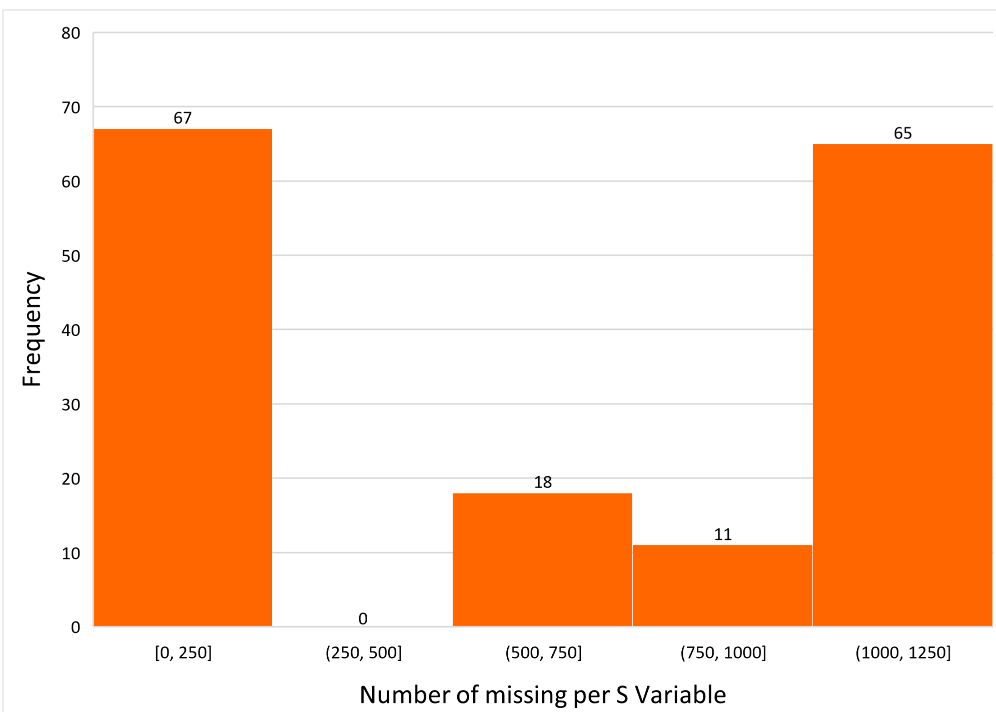

Figure 19: This figure presents the distribution of the number of missing values per variable over the S factors.

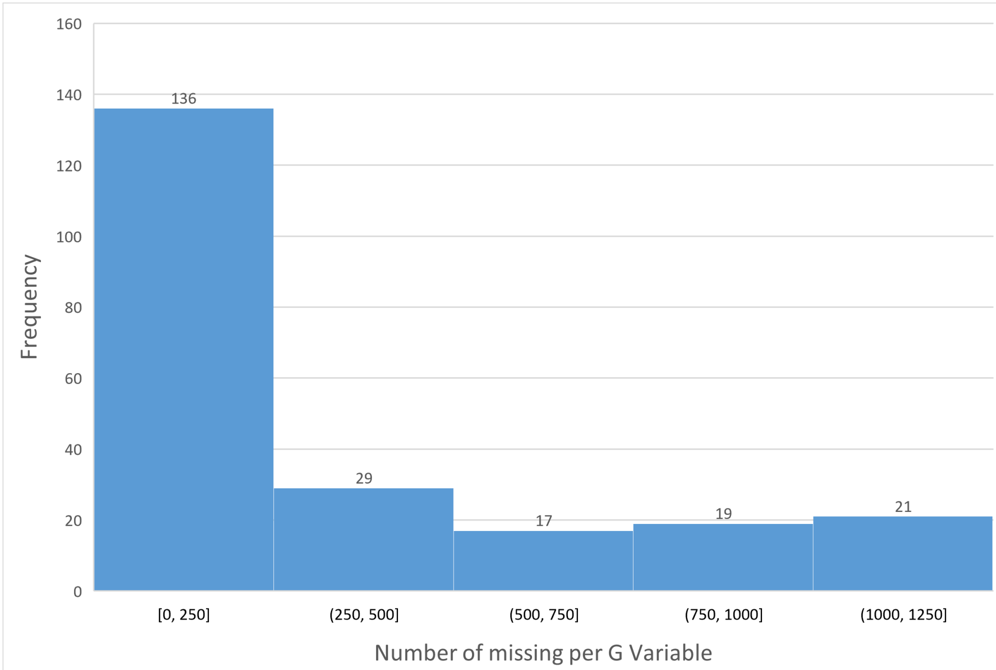

Figure 20: This figure presents the distribution of the number of missing values per variable over the G factors.

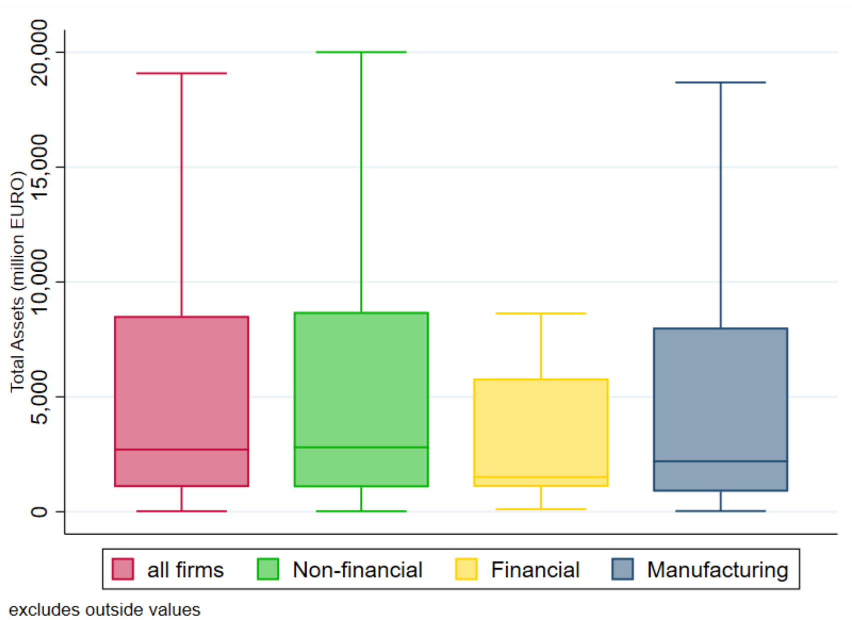

Figure 21: This figure presents the distributions of Total Assets by sub-samples: all 1076 firms, excluding financial firms (1033 obs.), financial firms (43 obs.) and manufacturing firms only (464 obs.).

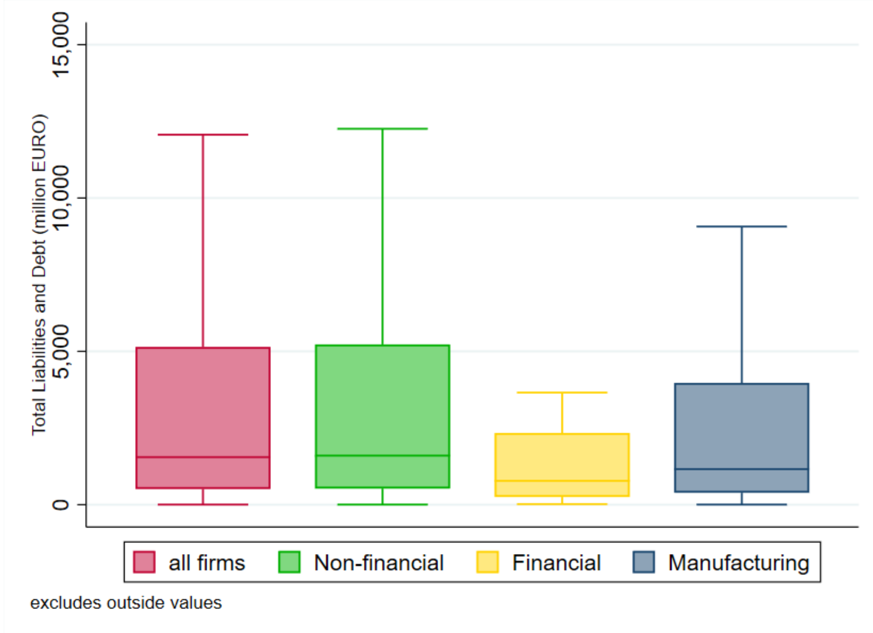

Figure 22: This figure presents the distributions of Total Liabilities and Debt by sub-samples: all 1076 fimrs, excluding financial firms (1033 obs.), financial firms (43 obs.) and manufacturing firms only (464 obs.).

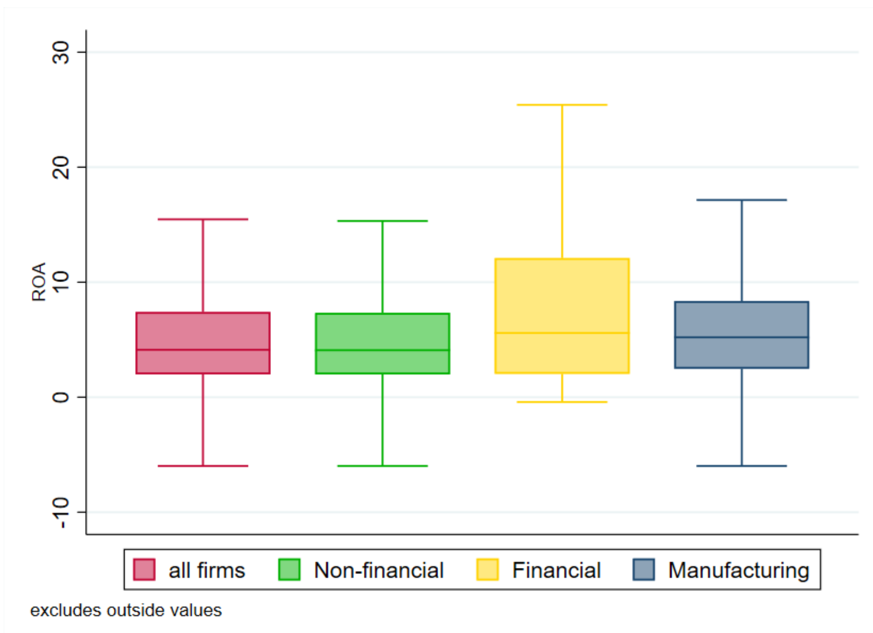

Figure 23: This figure presents the ROA distributions by sub-samples: all 1076 fimrs, excluding financial firms (1033 obs.), financial firms (43 obs.) and manufacturing firms only (464 obs.).

Table 1: This table reports the list of sectors in our data.

Code	Description
A	Agriculture, forestry and fishing
B	Mining and quarrying
C	Manufacturing
D	Electricity, gas, steam and AC supply
E	Water supply; sewerage, waste management and remediation act.
F	Construction
G	Wholesale and retail trade; repair of motor vehicles and motorcycles
H	Transportation and storage
I	Accommodation and food service act.
J	Information and communication
K	Financial and insurance activities
L	Real estate activities
M	Professional, scientific and technical activities
N	Administrative and support service activities
O	Public administration and defence; compulsory social security
Q	Human health and social work activities
R	Arts, entertainment and recreation
S	Other service activities

Table 2: This table reports how many companies per country and sector are included in our dataset. See Table 1 for sector codes and description.

	AT	BE	CH	CZ	DE	DK	ES	FI	FO	FR	GB	GG	GI	GR	HU	IE	IM	IT	JE	LU	MT	NL	NO	PL	PT	SE	TR	
A	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	4	0	0	1	0	9
B	1	1	2	0	1	1	2	0	0	4	13	0	0	0	1	0	0	2	3	0	0	1	5	1	1	1	0	40
C	13	14	64	1	58	20	17	19	1	51	88	0	0	3	2	10	1	33	1	4	0	22	11	5	3	53	21	515
D	2	1	2	1	5	1	6	1	0	3	5	0	0	3	0	0	0	8	0	0	0	0	3	4	2	1	0	48
E	0	0	0	0	0	0	1	0	0	2	4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	8
F	1	2	1	0	2	1	7	1	0	4	9	0	0	0	0	1	0	1	0	0	0	2	2	2	1	6	0	43
G	0	3	8	0	14	1	1	4	0	8	22	0	0	1	0	4	0	2	0	1	0	2	2	6	2	10	5	96
H	1	2	2	0	4	3	1	2	0	5	14	0	0	2	0	0	0	4	0	0	0	2	1	0	1	2	2	48
I	1	0	0	0	0	0	2	0	0	3	9	0	0	0	0	1	0	1	0	0	0	0	0	0	0	2	0	19
J	1	5	5	0	16	4	7	7	0	18	33	0	0	1	1	0	1	6	0	1	0	5	3	2	1	10	1	128
K	0	1	5	0	4	1	0	0	0	3	20	1	0	0	0	2	0	5	0	1	0	3	0	2	1	14	6	69
L	3	5	6	0	5	0	4	2	0	7	17	0	0	1	0	0	0	1	0	2	0	3	1	0	0	15	0	72
M	0	3	5	0	8	2	2	0	0	12	18	0	0	0	0	0	0	2	1	0	0	4	2	0	0	6	1	66
N	0	0	1	0	3	2	2	1	0	2	13	0	0	1	0	0	0	0	0	1	0	1	0	0	0	1	0	28
O	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	4
Q	0	2	2	0	1	0	0	0	0	3	2	0	0	0	0	0	0	0	0	1	0	0	0	1	0	4	1	17
R	0	0	0	0	3	0	0	0	0	2	2	0	1	1	0	1	1	2	0	0	1	1	0	0	0	1	1	17
S	0	2	0	0	3	0	2	0	0	0	8	4	0	0	0	1	1	1	0	2	0	0	0	0	0	0	0	24
Tot	23	41	103	2	128	37	55	37	1	129	277	5	1	14	4	21	4	69	5	13	1	47	34	23	12	127	38	1251

Table 3: This table reports the main balance sheet characteristics of the 1251 European companies in our sample.

	Mean	Median	Std.Dev.	Min	Max	Obs.
z_Score	3.3	2.5	6.2	-18.4	101.9	1251
Total Assets (million EUR)	11459.8	2492.5	39716.8	5.5	938677.9	1251
Number of Employees	22840.3	5235.0	57579.5	0.0	687280.0	1232
Total Liabilities and Debt (million EUR)	7690.6	1415.2	34309.7	0.6	934238.3	1197
Long-term Debt (million EUR)	2598.6	512.2	7935.1	0.0	113556.0	1251
ROA	4.4	4.0	11.6	-108.1	213.9	1251

Table 4: This table reports the main balance sheet characteristics of the 1076 European companies resulting from the process of data cleaning.

	Mean	Median	Std.Dev.	Min	Max	Obs.
z_Score	3.2	2.5	3.5	-9.2	70.4	1076
Total Assets (million EUR)	12088.9	2709.5	41473.6	24.9	938677.9	1076
Number of Employees	24925.5	6397.5	60694.1	2.0	687280.0	1076
Total Liabilities and Debt (million EUR)	7987.0	1551.0	35485.2	3.0	934238.3	1063
Long-term Debt (million EUR)	2704.8	547.9	8200.2	0.0	113556.0	1076
ROA	5.1	4.1	10.0	-54.2	213.9	1076

Table 5: ESG Factors selected through LASSO, AIC Criterion and Random Forests on the full sample of firms: 1076 obs., 102 ESG factors.

Factor	**Variable Name**	**Lasso**	**AIC selection**	**Random Forest**
E	Energy Management and Efficiency	×	×	
	Moderate Carbon Emission Business	×	×	
	Financing Environmental projects		×	
S	Responsible Investment (High Risk)	×	×	×
	Privacy Security (Medium-risk area)	×	×	
	Health Safety (Low Risk Business)	×	×	
	Access to Finance (High level)		×	
	Unsustainable Lending (High Risk)		×	

Table 6: ESG Factors selected through LASSO, AIC Criterion and Random Forests on non-financial firms: 1033 obs., 100 ESG factors.

Factor	**Variable Name**	**Lasso**	**AIC selection**	**Random Forest**
E	Moderate Carbon Emission Business	×	×	
	Green Building (High level)	×	×	
	Moderate Carbon Emission Area		×	
	Water stress (High risk area)		×	
	Carbon Emissions Disclosure		×	
	Biodiv./Land Use (Medium Intensity)		×	
	Carbon Emissions Reduction Target		×	
S	Unsust. Lending (Medium-risk area)	×	×	×
	Humanical Capital (Medium level)	×	×	
	Health Safety (Low Risk Business)	×	×	
	Responsible Investment (High Risk)		×	
	Product Safety/Quality (High Risk)		×	
	Privacy Security (Medium-risk area)		×	

Table 7: ESG Factors selected through AIC Criterion and Random Forests on manufacturing firms: 464 obs., 93 ESG factors. The reduced sample does not permit the use of the bootstrap procedure on LASSO.

Factor	**Variable Name**	**AIC selection**	**Random Forest**
E	Carbon Emissions Reduction Target	×	
	Financing Environmental projects	×	
	Green Building (Low level)	×	
	Moderate Carbon Emission Area	×	×
G	Number of Executives	×	
S	Chemical Safety (Low Risk)	×	
	Health Safety (High Risk Area)	×	
	Privacy Security (High Risk Area)	×	
	Privacy Security (Medium-risk area)	×	
	Responsible Investment (High Risk)	×	

Table 8: This table reports the most important variables in explaining the variance of all E factors. The PCA is run on three different sub-samples: the full sample (FS), excluding non-financial firms (NF), and manufacturing firms only (M). Many factors are selected in all of these cases (all).

Score	Description	Sample
Carbon Emissions Scope 1	Scope 1 greenhouse gas emissions: those from sources owned or controlled by the company, typically direct combustion of fuel as in a furnace or vehicle (metric tons).	all
Carbon Emissions Scope 2	Scope 2 greenhouse gas emissions: those caused by the generation of electricity purchased by the company (metric tons).	all
Carbon Emission Intensity Scope 1	Scope 1 greenhouse gas emissions normalized by enterprise value including cash (t/USD million).	all
Carbon Emission Intensity Scope 2	Scope 2 greenhouse gas emissions normalized by enterprise value including cash (t/USD million).	all
Intensive Carbon Emission Business	Revenues (%) from lines of business that are highly carbon intensive.	all
Energy Consumption (High Intensity)	Revenues (%) from operations whose energy intensity is typically high.	all
Toxic Emissions (High Intensity)	Revenues (%) from lines of business that generate large amounts of toxic emissions.	all
Water Intensive Business (High risk)	Portion of the company's assets devoted to lines of business that are highly water intensive.	NF, FS
Biodiv./Land Use (Medium Intensity)	Revenues (%) from activities that involve moderate disturbance to land or marine areas.	all
Product Carbon Footprint (High level)	Revenues (%) from highly carbon intensive products.	M

Table 9: This table describes the content of each raw factor selected by the supervised techniques used in this paper as reported by the MSCI provider

Factor		Description
Biodiv./Land Use (Medium Intensity)	E	Revenues (%) from activities that involve moderate disturbance to land or marine areas.
Carbon Emissions Disclosure	E	Does the company report its carbon emissions to the CDP? Yes/No
Energy Management and Efficiency	E	How aggressively the company seek to mitigate its carbon emissions by managing energy consumption and improving the energy efficiency of its operations.
Moderate Carbon Emission Business	E	Revenues (%) from lines of business that have a moderate level of carbon intensity.

Moderate Carbon Emission Area	E	Revenues (%) from countries or regions where carbon regulations are of moderate strength.
Carbon Emissions Reduction Target	E	Effectiveness of the carbon emissions reduction target.
Financing Environmental projects	E	Revenues (%) from business lines that are heavily involved in financing projects or corporate entities.
Green Building (High level)	E	Portion of the company's portfolio comprised of building types with high energy requirements and subject to green building regulations.
Green Building (Low level)	E	Portion of the company's portfolio comprised of building types that do not have high energy requirements and are not subject to green building regulations.
Water stress (High risk area)	E	Portion of the company's assets in regions that are typically highly water intensive.
Number of Executives	G	Total Number of Executives
Access to Finance (High level)	S	Revenues (%) from financial business segments that have substantial retail operations.
Chemical Safety (Low Risk)	S	Revenues (%) from products that typically contain few substances of concern.
Unsustainable Lending (High Risk)	S	Revenues (%) from segments where retail services to borrowers, individuals, and households are a core activity.
Unsust. Lending (Medium-risk area)	S	Revenues (%) from countries where the regulatory environment is such that unsustainable lending is present but not widespread.
Health Safety (High Risk Area)	S	Revenues (%) from countries or regions with high rates of worker injuries.
Health Safety (Low Risk Business)	S	Revenues (%) from lines of business that have low worker injury rates.
Human Capital (Medium level)	S	Revenues (%) from lines of business that are moderately reliant on highly-skilled or -educated workers.
Privacy Security (High Risk Area)	S	Revenues (%) from countries where privacy and data security regulations are weak or nonexistent and are not expected to become stronger in the near to medium term.
Privacy Security (Medium-risk area)	S	Revenues (%) from countries where privacy and data security regulations are of moderate strength.
Privacy Security (Moderate Risk)	S	Revenues (%) from business activities that involve the handling of sensitive personally identifiable information as well as selling advertising, marketing to children, selling access to data, and/or that are moderately frequent subject to data breaches.
Product Safety/Quality (High Risk)	S	Revenues (%) from business segments most likely to experience negative material effects from product safety or quality problems.
Responsible Investment (High Risk)	S	Revenues (%) from activities that involve both management of investment assets for itself and on behalf of clients.

Table 10: This table reports Summary statistics of the selected variables in Tables 5, 6 and 7. Summary statistics are reported separately for the different sub-samples in case variables are selected over more than one of them. The full-sample comprises 1076 firms, while the two sub-samples of non-financial and manufacturing firms comprise respectively 1033 and 464. The list of variables is divided into the three factors E, S and G.

	Mean	Median	Std.Dev.	Min	Max	Obs.
Moderate Carbon Emission Business	24.7	0.0	39.4	0	100	1076
	25.2	0.0	39.7	0	100	1033
Financing Environmental projects	4.0	0.0	18.2	0	100	1076
	0.7	0.0	7.2	0	100	464
Green Building (High level)	2.9	0.0	16.1	0	100	1033
Moderate Carbon Emission Area	42.0	36.4	35.8	0	100	1033
	47.5	44.2	33.5	0	100	464
Water stress (High risk area)	7.8	0.0	21.7	0	100	1033
Biodiv./Land Use (Medium Intensity)	15.2	0.0	31.3	0	100	1033
Green Building (Low level)	99.5	100.0	5.7	0	100	464
Responsible Investment (High Risk)	1.6	0.0	12.2	0	100	1076
	0.5	0.0	6.4	0	100	1033
	0.0	0.0	0.1	0	2.5	464
Privacy Security (Medium-risk area)	10.9	0.0	18.7	0	100	1076
	11.2	0.0	18.9	0	100	1033
	15.1	8.4	19.8	0	100	464
Health Safety (Low Risk Business)	55.1	78.1	46.4	0	100	1076
	53.8	71.4	46.4	0	100	1033
Access to Finance (High level)	1.6	0.0	11.8	0	100	1076
Unsustainable Lending (High Risk)	0.8	0.0	8.1	0	100	1076
Unsust. Lending (Medium-risk area)	67.9	76.1	33.1	0	100	1033
Humanical Capital (Medium level)	50.8	51.7	46.1	0	100	1033
Product Safety/Quality (High Risk)	18.9	0.0	36.6	0	100	1033
Chemical Safety (Low Risk)	17.1	0.0	32.4	0	100	464
Health Safety (High Risk Area)	5.6	0.0	10.9	0	83	464
Privacy Security (High Risk Area)	84.9	91.6	19.8	0	100	464
Number of Executives	23.8	21.0	12.5	0	76	464

Table 11: Distribution of **Energy Management and Efficiency** evaluated on the whole sample: this variable indicates MSCI assessment of how aggressively the company has sought to mitigate its carbon emissions by managing energy consumption and improving the energy efficiency of its operations. Possible values are: 3 'Aggressive efforts'; 2 'Some efforts'; 1 'Limited efforts/information'; 0 'No efforts'.

	Freq.	Percent
0	83	7.7
1	868	80.7
2	108	10.0
3	17	1.6
Total	1076	100

Table 12: Distribution of **Carbon Emissions Disclosure** evaluated excluding financial firms: this variable indicates whether the company reports its carbon emissions to the CDP. Possible values: 'Yes' or 'No'

	Freq.	Percent
No	501	48.5
Yes	532	51.5
Total	1033	100

Table 13: Distribution of **Carbon Emissions Reduction Target** evaluated on the sub-samples of non-financial firms and manufacturing firms. If a company has a carbon emissions reduction target, this variable indicates MSCI assessment of how aggressive that target is. The highest scores go to companies aggressively seeking to reduce emissions from a level that is already relatively low. Apart from companies with no target, the lowest scores go to those with high emissions levels that are seeking to make only minor reductions. For smaller companies, among which carbon reduction targets are relatively uncommon, a moderately high score is given for any type of carbon reduction target.

	Non-financial firms		Manufacturing firms	
	Freq.	Percent	Freq.	Percent
0	456	44.1	184	39.7
1	446	43.2	208	44.8
2	26	2.5	10	2.2
3	16	1.5	5	1.1
4	35	3.4	16	3.4
5	35	3.4	29	6.3
6	19	1.8	12	2.6
Total	1033	100	464	100

Table 14: This table reports Summary statistics of the selected variables in Tables 8. Summary statistics are reported separately for the different sub-samples in case variables are selected over more than one of them. The full-sample comprises 1076 firms, while the two sub-samples of non-financial and manufacturing firms comprise respectively 1033 and 464.

	Mean	Median	Std.Dev.	Min	Max	Obs.
Carbon Emissions Scope 1 (th t)	1438.8	18.2	6617.3	0	91700	1076
	1482.7	20.6	6741.7	0	91700	1033
	1030.5	31.9	4938.2	0	73080	464
Carbon Emissions Scope 2 (th t)	283.4	25.8	1108.8	0	21012	1076
	292.5	28.1	1130.0	0	21012	1033
	356.9	47.1	1346.2	0	21012	464
Carbon Emission Intensity Scope 1	119.7	4.1	499.5	0	8194	1076
	119.9	4.4	496.9	0	8194	1033
	100.8	6.3	375.2	0	3494	464
Carbon Emission Intensity Scope 2	27.1	5.5	70.3	0	1175	1076
	27.6	5.8	70.6	0	1175	1033
	34.1	9.9	59.7	0	443	464
Intensive Carbon Emission Business	6.1	0.0	20.8	0	100	1076
	6.3	0.0	21.1	0	100	1033
	7.6	0.0	22.9	0	100	464
Energy Consumption (High Intensity)	11.5	0.0	29.0	0	100	1076
	11.8	0.0	29.4	0	100	1033
	12.6	0.0	30.2	0	100	464
Toxic Emissions (High Intensity)	16.3	0.0	33.1	0	100	1076
	16.7	0.0	33.5	0	100	1033
	26.6	0.0	39.7	0	100	464
Water Intensive Business (High risk)	15.7	0.0	33.4	0	100	1076
	16.2	0.0	33.9	0	100	1033
Biodiv./Land Use (Medium Intensity)	14.9	0.0	31.1	0	100	1076
	15.2	0.0	31.3	0	100	1033
	16.8	0.0	32.0	0	100	464
Product Carbon Footprint (High level)	27.0	0.0	39.8	0	100	464

Table 15: This table reports Summary statistics of the selected variables referring to balance sheet information. Summary statistics are reported separately for the different sub-samples: the full-sample comprises 1076 firms, while the two sub-samples of non-financial and manufacturing firms comprise respectively 1033 and 464.

	Mean	Median	Std.Dev.	Min	Max	Obs.
Solvency Ratio	43.6	41.8	18.3	-1	98	1076
	43.4	41.6	17.8	-1	93	1033
	47.2	45.4	16.6	5	91	464
Current Ratio	1.8	1.3	1.8	0	22	1076
	1.7	1.3	1.7	0	22	1033
	2.0	1.6	1.7	0	22	464
ROA	5.1	4.1	10.0	-54	214	1076
	5.0	4.1	10.1	-54	214	1033
Age	53.7	34.0	48.8	0	365	1076
	54.7	35.0	49.2	1	365	1033
ROE	11.4	11.1	15.8	-94	77	464

Table 16: Distribution of the dummy **Exporter out of Europe** evaluated on the sub-sample of 464 manufacturing firms.

	Freq.	Percent
no	25	5.4
yes	439	94.6
Total	464	100

Table 17: This table reports the results of the OLS estimation of the regression where the dependent variable is the Z-score and the covariates are the E and S variables selected by the SML techniques on the full sample (see Section 6, table 5). Balance sheet information such as Solvency Ratio, Current Ratio, ROA and firm's Age are also included as covariates. The sample comprises 1076 European firms for the year 2019. Equations are estimated both with and without country and sector fixed-effects. R^2, adjusted R^2, F-test for joint significance of coefficients, and Root Mean Square Error are reported. p-values in parentheses.

	(1) z_Score	(2) z_Score	(3) z_Score	(4) z_Score
X Solvency Ratio	0.0835***	0.0854***	0.0889***	0.0876***
	(0.000)	(0.000)	(0.000)	(0.000)
X Current Ratio	0.685***	0.659***	0.545***	0.550***
	(0.000)	(0.000)	(0.000)	(0.000)
X ROA	0.122***	0.117***	0.122***	0.119***
	(0.000)	(0.000)	(0.000)	(0.000)
X Age	0.00437***	0.00339***	0.00304***	0.00252**
	(0.000)	(0.000)	(0.000)	(0.002)
E Energy Management and Efficiency			-0.402	-0.365

			(0.142)	(0.205)
E Moderate Carbon Emission Business			-0.00734***	-0.00606***
			(0.000)	(0.000)
E Financing Environmental projects			-0.00557*	-0.00185
			(0.041)	(0.448)
S Access to Finance (High level)			-0.0770	-0.0815
			(0.268)	(0.248)
S Unsustainable Lending (High Risk)			0.0838	0.0807
			(0.235)	(0.234)
S Health Safety (Low Risk Business)			-0.00552***	-0.00531***
			(0.000)	(0.000)
S Privacy Security (Medium-risk area)			0.00962***	0.0154***
			(0.000)	(0.000)
S Responsible Investment (High Risk)			0.0776	0.0725
			(0.254)	(0.267)
Constant	-2.524***	-2.303***	-1.677***	-1.642***
	(0.000)	(0.000)	(0.000)	(0.000)
Country Fixed-effects	No	Yes	No	Yes
Sector Fixed-effects	No	Yes	No	Yes
R^2	0.625	0.650	0.671	0.687
Adjusted R^2	0.624	0.637	0.667	0.672
F	82.68	57.58	124.7	54.63
rmse	2.175	2.138	2.046	2.032
Observations	1076	1076	1076	1076

p-values in parentheses

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 18: This table reports the results of the OLS estimation of the regression where the dependent variable is the Z-score and the covariates are the E and S variables selected by the SML techniques on the sub-sample of non financial firms (see Section 6, table 6). Balance sheet information such as Solvency Ratio, Current Ratio, ROA and firm's Age are also included as covariates. The sample comprises 1033 non-financial European firms for the year 2019. Equations are estimated both with and without country and sector fixed-effects. R^2, adjusted R^2, F-test for joint significance of coefficients, and Root Mean Square Error are reported. p-values in parentheses.

	(1)	(2)	(3)	(4)
	z_Score	z_Score	z_Score	z_Score
X Solvency Ratio	0.0800***	0.0816***	0.0839***	0.0837***
	(0.000)	(0.000)	(0.000)	(0.000)
X Current Ratio	0.581***	0.554***	0.566***	0.556***
	(0.000)	(0.000)	(0.000)	(0.000)
X ROA	0.109***	0.105***	0.107***	0.105***
	(0.000)	(0.000)	(0.000)	(0.000)
X Age	0.00479***	0.00297***	0.00333***	0.00235**
	(0.000)	(0.000)	(0.000)	(0.001)
E Biodiv./Land Use (Medium Intensity)			-0.000144	-0.000626
			(0.902)	(0.610)
E Moderate Carbon Emission Business			-0.00498***	-0.00409***
			(0.000)	(0.000)
E Moderate Carbon Emission Area			0.00325**	0.00338*
			(0.009)	(0.032)
E Carbon Emissions Reduction Target			-0.0922***	-0.0581*
			(0.000)	(0.035)
E Green Building (High level)			-0.00810**	-0.00647*
			(0.001)	(0.024)
E Water stress (High risk area)			-0.00628**	-0.00483

			(0.005)	(0.147)
S Unsust. Lending (Medium-risk area)			-0.00452**	-0.00316+
			(0.001)	(0.096)
S Health Safety (Low Risk Business)			-0.00405***	-0.00384**
			(0.000)	(0.001)
S Human Capital (Medium level)			0.00350***	0.00272**
			(0.000)	(0.002)
S Privacy Security (Medium-risk area)			0.0108***	0.0128***
			(0.000)	(0.000)
S Product Safety/Quality (High Risk)			-0.00255+	-0.00211
			(0.052)	(0.137)
S Responsible Investment (High Risk)			-0.00875	-0.00711
			(0.692)	(0.733)
E Carbon Emissions Disclosure			0.194*	0.160+
			(0.021)	(0.070)
Constant	-2.209***	-2.206***	-1.933***	-2.077***
	(0.000)	(0.000)	(0.000)	(0.000)
Country Fixed-effects	No	Yes	No	Yes
Sector Fixed-effects	No	Yes	No	Yes
R^2	0.776	0.805	0.807	0.822
Adjusted R^2	0.775	0.797	0.804	0.812
F	355.8	57.80	113.1	53.64
rmse	1.334	1.267	1.246	1.220
Observations	1033	1033	1033	1033

p-values in parentheses
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 19: This table reports the results of the OLS estimation of the regression where the dependent variable is the Z-score and the covariates are the E, S and G variables selected by the SML techniques on the sub-sample of manufacturing firms (see Section 6, table 7). Balance sheet information such as Solvency Ratio, Current Ratio, ROE and export status are also included as covariates.The sample comprises 464 manufacturing European firms for the year 2019. Equations are estimated both with and without country fixed-effects. R^2, adjusted R^2, F-test for joint significance of coefficients, and Root Mean Square Error are reported. p-values in parentheses.

	(1)	(2)	(3)	(4)
	z_Score	z_Score	z_Score	z_Score
X Solvency Ratio	0.105***	0.103***	0.108***	0.106***
	(0.000)	(0.000)	(0.000)	(0.000)
X Current Ratio	0.427***	0.430***	0.408***	0.412***
	(0.000)	(0.000)	(0.000)	(0.000)
X ROE	0.0589***	0.0598***	0.0591***	0.0606***
	(0.000)	(0.000)	(0.000)	(0.000)
X Exporter out of Europe	0.731+	0.621	0.714+	0.624
	(0.074)	(0.123)	(0.080)	(0.125)
E Moderate Carbon Emission Area			0.00585**	0.00731**
			(0.005)	(0.003)
E Carbon Emissions Reduction Target			-0.0390	-0.0337
			(0.274)	(0.341)
E Financing Environmental projects			0.0274***	0.0276***
			(0.000)	(0.000)
E Green Building (Low level)			-0.0276**	-0.0284***
			(0.001)	(0.000)
S Chemical Safety (Low Risk)			0.00131	0.00112
			(0.479)	(0.535)
S Health Safety (High Risk Area)			0.0132*	0.00874

			(0.033)	(0.176)
S Privacy Security (High Risk Area)			1.207+	1.263
			(0.080)	(0.159)
S Privacy Security (Moderate Risk)			-0.00283	-0.00247
			(0.111)	(0.186)
S Responsible Investment (High Risk)			-1.723***	-1.539***
			(0.000)	(0.000)
G Number of Executives			-0.00285	-0.000305
			(0.576)	(0.950)
Constant	-3.477***	-3.431***	-1.028	-0.928
	(0.000)	(0.000)	(0.217)	(0.223)
Country Fixed-effects	No	Yes	No	Yes
R^2	0.751	0.773	0.769	0.789
Adjusted R^2	0.749	0.761	0.762	0.773
F	140.4	40.53	47.62	36.13
rmse	1.354	1.321	1.319	1.289
Observations	464	464	464	464

p-values in parentheses
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 20: This table reports the results of the OLS estimation of the regression where the dependent variable is the Z-score and the covariates are the E variables selected by the PCA on the full sample (see Section 6, table 8). Balance sheet information such as Solvency Ratio, Current Ratio, ROA and firm's Age are also included as covariates. The sample comprises 1076 European firms for the year 2019. Equations are estimated both with and without country and sector fixed-effects. R^2, adjusted R^2, F-test for joint significance of coefficients, and Root Mean Square Error are reported. p-values in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)
	z_Score	z_Score	z_Score	z_Score	z_Score	z_Score
X Solvency Ratio	0.0835***	0.0854***	0.0835***	0.0853***	0.0838***	0.0856***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
X Current Ratio	0.685***	0.659***	0.684***	0.658***	0.683***	0.657***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
X ROA	0.122***	0.117***	0.122***	0.117***	0.122***	0.117***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
X Age	0.00437***	0.00339***	0.00434***	0.00341***	0.00429***	0.00344***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
E Carbon Emissions Scope 1			5.57e-09	5.82e-09	4.02e-09	4.58e-09
			(0.244)	(0.278)	(0.383)	(0.397)
E Carbon Emissions Scope 2			-6.91e-08**	-7.64e-08**	-6.62e-08*	-7.91e-08**
			(0.004)	(0.002)	(0.011)	(0.003)
E Carbon Emission Intensity Scope 1			-0.000145+	-0.000120	-0.000162+	-0.000133
			(0.097)	(0.260)	(0.077)	(0.231)
E Carbon Emission Intensity Scope 2			0.00114+	0.000936*	0.00113+	0.000910+
			(0.053)	(0.040)	(0.080)	(0.080)
E Intensive Carbon Emission Business					0.00171	0.00153
					(0.381)	(0.506)
E Energy Consumption (High Intensity)					-0.000622	-0.0000515
					(0.655)	(0.975)
E Toxic Emissions (High Intensity)					0.00268*	0.00302*
					(0.020)	(0.028)
E Water Intensive Business (High risk)					-0.00253+	-0.00170
					(0.051)	(0.233)
E Biodiv./Land Use (Medium Intensity)					0.00123	-0.000921
					(0.321)	(0.488)
E Product Carbon Footprint (High level)					-0.00126	-0.00170

					(0.366)	(0.204)
Constant	-2.524***	-2.303***	-2.519***	-2.297***	-2.532***	-2.294***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Country Fixed-effects	No	Yes	No	Yes	No	Yes
Sector Fixed-effects	No	Yes	No	Yes	No	Yes
R^2	0.625	0.650	0.626	0.651	0.627	0.651
Adjusted R^2	0.624	0.637	0.623	0.636	0.622	0.634
F	82.68	57.58	51.64	54.26	41.08	48.81
rmse	2.175	2.138	2.178	2.141	2.181	2.144
Observations	1076	1076	1076	1076	1076	1076

p-values in parentheses
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 21: This table reports the results of the OLS estimation of the regression where the dependent variable is the Z-score and the covariates are the E variables selected by the PCA on the sub-sample of non financial firms (see Section 6, table 8). Balance sheet information such as Solvency Ratio, Current Ratio, ROA and firm's Age are also included as covariates. The sample comprises 1033 non-financial European firms for the year 2019. Equations are estimated both with and without country fixed-effects. R^2, adjusted R^2, F-test for joint significance of coefficients, and Root Mean Square Error are reported. *p*-values in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)
	z_Score	z_Score	z_Score	z_Score	z_Score	z_Score
X Solvency Ratio	0.0800***	0.0816***	0.0799***	0.0814***	0.0801***	0.0817***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
X Current Ratio	0.581***	0.554***	0.580***	0.553***	0.579***	0.551***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
X ROA	0.109***	0.105***	0.109***	0.105***	0.109***	0.105***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
X Age	0.00479***	0.00297***	0.00475***	0.00301***	0.00457***	0.00302***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
E Carbon Emissions Scope 1			2.23e-09	2.81e-09	1.01e-09	1.58e-09
			(0.586)	(0.518)	(0.802)	(0.724)
E Carbon Emissions Scope 2			-6.70e-08**	-8.27e-08***	-7.00e-08**	-8.41e-08***
			(0.006)	(0.000)	(0.007)	(0.001)
E Carbon Emission Intensity Scope 1			-0.000111	-0.0000723	-0.000131+	-0.0000895
			(0.138)	(0.361)	(0.099)	(0.296)
E Carbon Emission Intensity Scope 2			0.00146*	0.000897+	0.00126+	0.000934+
			(0.011)	(0.059)	(0.078)	(0.088)
E Intensive Carbon Emission Business					0.00155	0.00195
					(0.454)	(0.389)
E Energy Consumption (High Intensity)					-0.000834	-0.000211
					(0.549)	(0.894)
E Toxic Emissions (High Intensity)					0.00307**	0.00286*
					(0.007)	(0.030)
E Water Intensive Business (High risk)					-0.00224+	-0.00229+
					(0.061)	(0.082)
E Biodiv./Land Use (Medium Intensity)					0.00159	-0.000824
					(0.193)	(0.535)
E Product Carbon Footprint (High level)					-0.000172	-0.00117
					(0.882)	(0.333)
Constant	-2.209***	-2.206***	-2.210***	-2.197***	-2.236***	-2.194***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Country Fixed-effects	No	Yes	No	Yes	No	Yes
Sector Fixed-effects	No	Yes	No	Yes	No	Yes
R^2	0.776	0.805	0.777	0.806	0.779	0.807
Adjusted R^2	0.775	0.797	0.776	0.798	0.776	0.798

F	355.8	57.80	184.4	54.00	108.9	48.50
rmse	1.334	1.267	1.333	1.266	1.332	1.266
Observations	1033	1033	1033	1033	1033	1033

p-values in parentheses
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 22: This table reports the results of the OLS estimation of the regression where the dependent variable is the Z-score and the covariates are the E variables selected by the PCA on the sub-sample of manufacturing firms (see Section 6, table 8). Balance sheet information such as Solvency Ratio, Current Ratio, ROE and export status are also included as covariates. The sample comprises 464 manufacturing European firms for the year 2019. Equations are estimated both with and without country fixed-effects. R^2, adjusted R^2, F-test for joint significance of coefficients, and Root Mean Square Error are reported. p-values in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)
	z_Score	z_Score	z_Score	z_Score	z_Score	z_Score
X Solvency Ratio	0.105***	0.103***	0.106***	0.103***	0.108***	0.105***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
X Current Ratio	0.427***	0.430***	0.423***	0.428***	0.412***	0.415***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
X ROE	0.0589***	0.0598***	0.0592***	0.0601***	0.0601***	0.0610***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
E Carbon Emissions Scope 1			5.82e-09	4.47e-09	3.89e-09	-4.50e-10
			(0.570)	(0.694)	(0.656)	(0.961)
E Carbon Emissions Scope 2			-0.000000108***	-9.74e-08***	-0.000000103**	-9.31e-08**
			(0.000)	(0.000)	(0.002)	(0.003)
E Carbon Emission Intensity Scope 1			-0.0000710	0.0000117	-0.0000622	0.0000267
			(0.737)	(0.954)	(0.791)	(0.905)
E Carbon Emission Intensity Scope 2			0.00113	0.000778	0.00163	0.000862
			(0.292)	(0.469)	(0.181)	(0.494)
E Intensive Carbon Emission Business					0.00361	0.00389
					(0.385)	(0.331)
E Energy Consumption (High Intensity)					-0.00396	-0.00370
					(0.148)	(0.168)
E Toxic Emissions (High Intensity)					0.00405*	0.00503**
					(0.024)	(0.007)
E Water Intensive Business (High risk)					-0.00386*	-0.00374*
					(0.014)	(0.020)
E Biodiv./Land Use (Medium Intensity)					0.000372	0.00153
					(0.871)	(0.504)
E Product Carbon Footprint (High level)					-0.00168	-0.00172
					(0.255)	(0.247)
Constant	-3.477***	-3.431***	-3.483***	-3.454***	-3.478***	-3.506***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Country Fixed-effects	No	Yes	No	Yes	No	Yes
R^2	0.751	0.773	0.753	0.775	0.760	0.782
Adjusted R^2	0.749	0.761	0.749	0.761	0.752	0.765
F	140.4	40.53	79.58	37.57	49.69	41.07
rmse	1.354	1.321	1.354	1.322	1.345	1.311
Observations	464	464	464	464	464	464

p-values in parentheses
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

References

Albuquerque, Rui, Yrjö Koskinen, Shuai Yang, and Chendi Zhang, "Love in the Time of COVID-19: The Resiliency of Environmental and Social Stocks," CEPR Discussion Papers 14661, C.E.P.R. Discussion Papers 2020.

Altman, Edward I., "Financial Ratios, Discriminant Analysis and The Prediction of Corporate Bankruptcy," *The Journal of Finance*, 1968, *23* (4), 589–609.

Attig, Najah, Sadok El Ghoul, Omrane Guedhami, and Jungwon Suh, "Corporate social responsibility and credit ratings," *Journal of business ethics*, 2013, *117* (4), 679–694.

Auer, Benjamin R and Frank Schuhmacher, "Do socially (ir) responsible investments pay? New evidence from international ESG data," *The Quarterly Review of Economics and Finance*, 2016, *59*, 51–62.

Belu, Constantin and Cristiana Manescu, "Strategic corporate social responsibility and economic performance," *Applied Economics*, 2013, *45* (19), 2751–2764.

Berg, Florian, Julian F Kölbel, and Roberto Rigobon, "Aggregate Confusion: The Divergence of ESG Ratings*," *Review of Finance*, 05 2022. rfac033.

___ **, Kornelia Fabisik, and Zacharias Sautner**, "Rewriting History II: The (Un)Predictable Past of ESG Ratings," *SSRN Electronic Journal*, 11 2020.

Breiman, Leo, "Random forests," *Machine learning*, 2001, *45* (1), 5–32.

Chava, Sudheer, "Environmental externalities and cost of capital," *Management science*, 2014, *60* (9), 2223–2247.

COP26, United Nations, *Glasgow Climate Pact,* 2021. Report of the Conference of the Parties serving as the meeting of the Parties to the Paris Agreement on its third session.

Cornett, Marcia Millon, Otgontsetseg Erhemjamts, and Hassan Tehranian, "Greed or good deeds: An examination of the relation between corporate social responsibility and the financial performance of U.S. commercial banks around the financial crisis," *Journal of Banking & Finance*, 2016, *70*, 137–159.

Derwall, Jeroen, Nadja Guenster, Rob Bauer, and Kees Koedijk, "The eco-efficiency premium puzzle," *Financial Analysts Journal*, 2005, *61* (2), 51–63.

Devalle, Alain, Simona Fiandrino, and Valter Cantino, "The linkage between ESG performance and credit ratings: a firm-level perspective analysis," 2017.

Eccles, Robert G, Ioannis Ioannou, and George Serafeim, "The impact of corporate sustainability on organizational processes and performance," *Management Science*, 2014, *60* (11), 2835–2857.

Friede, Gunnar, Timo Busch, and Alexander Bassen, "ESG and financial performance: aggregated evidence from more than 2000 empirical studies," *Journal of Sustainable Finance & Investment*, 2015, *5* (4), 210–233.

Ghoul, Sadok El, Omrane Guedhami, Chuck CY Kwok, and Dev R Mishra, "Does corporate social responsibility affect the cost of capital?," *Journal of Banking & Finance*, 2011, *35* (9), 2388–2406.

Goss, Allen and Gordon S Roberts, "The impact of corporate social responsibility on the cost of bank loans," *Journal of Banking & Finance*, 2011, *35* (7), 1794–1810.

Halbritter, Gerhard and Gregor Dorfleitner, "The wages of social responsibility—where are they? A critical review of ESG investing," *Review of Financial Economics*, 2015, *26*, 25–35.

Henisz, Witold J and James McGlinch, "ESG, material credit events, and credit risk," *Journal of Applied Corporate Finance*, 2019, *31* (2), 105–117.

Henke, Hans-Martin, "The effect of social screening on bond mutual fund performance," *Journal of Banking & Finance*, 2016, *67*, 69–84.

Hvidkjær, Søren, "ESG investing: a literature review," *Report prepared for Dansif*, 2017.

Kempf, Alexander and Peer Osthoff, "The effect of socially responsible investing on portfolio performance," *European Financial Management*, 2007, *13* (5), 908–922.

Kuhn, Max, Kjell Johnson et al., *Applied predictive modeling*, Vol. 26, Springer, 2003.

Menz, Klaus-Michael, "Corporate social responsibility: Is it rewarded by the corporate bond market? A critical note," *Journal of Business Ethics*, 2010, *96* (1), 117–134.

Nofsinger, John and Abhishek Varma, "Socially responsible funds and market crises," *Journal of Banking & Finance*, 2014, *48*, 180–193.

Scholz, RW, O Weber, J Stünzi, W Ohlenroth, and A Reuter, "Umweltrisiken systematisch erfassen [Systematically inventorying environmental risks]," *Schweizer Bank: Monatsmagazin für Führungskräfte aus Bank und Finanz [Swiss Bank: Magazine for Managers in Banks and Finance]*, 1995, *10* (4), 45–47.

Sharfman, Mark P and Chitru S Fernando, "Environmental risk management and the cost of capital," *Strategic management journal*, 2008, *29* (6), 569–592.

Statman, Meir and Denys Glushkov, "The wages of social responsibility," *Financial Analysts Journal*, 2009, *65* (4), 33–46.

Stiglitz, Joseph E and Andrew Weiss, "Credit rationing in markets with imperfect information," *The American economic review*, 1981, *71* (3), 393–410.

Tibshirani, Robert, "Regression Shrinkage and Selection Via the Lasso," *Journal of the Royal Statistical Society, Series B*, 1994, *58*, 267–288.

Weber, Olaf, Roland W Scholz, and Georg Michalik, "Incorporating sustainability criteria into credit risk management," *Business strategy and the environment*, 2010, *19* (1), 39–50.

Wood, Richard, Konstantin Stadler, Tatyana Bulavskaya, Stephan Lutter, Stefan Giljum, Arjan De Koning, Jeroen Kuenen, Helmut Schütz, José Acosta-Fernández, Arkaitz Usubiaga et al., "Global sustainability accounting—Developing EXIOBASE for multi-regional footprint analysis," *Sustainability*, 2015, *7* (1), 138–163.