Working Paper presented at the

Peer-to-Peer Financial Systems 2023 Workshop

2023

Environmental Data And Scores: Lost In Translation

Enrico Bernardini
Bank of Italy

Marco Fanari
Bank of Italy

Enrico Foscolo
Bank of Italy

Francesco Ruggiero
Bank of Italy

Powered by

P2P Financial Systems

Exponential Science

Environmental data and scores: Lost in translation

Enrico Bernardini[‡], Marco Fanari[†], Enrico Foscolo[†], and Francesco Ruggiero[†]

[‡]Bank of Italy, Climate change and Sustainability Hub
[†]Bank of Italy, Financial Risk Management Directorate

Abstract

This paper aims to address methodological issues and limited coverage of providers' environmental scores, which are increasingly employed by investors, financial institutions and policy makers for the corporate environmental assessment. The contribution of the paper is twofold. First, regression analysis shows a substantial heterogeneity among the environmental scores of seven providers in the reliance on raw data, although some variables are material across providers. We find that the unexplained component of the regression is heterogeneous across providers and can be attributed to a judgmental factor related to the providers' different focus on financial risk or environmental impact. Second, we propose a classification system based on corporate disclosure that allows the extension of the environmental assessment to companies not rated by providers. This system has been calibrated to implement two common investment strategies, i.e. *best-in-class* and *exclusion*. The resulting portfolios exhibit both environmental and financial profiles similar to portfolios based on providers' scores. The paper confirms that it is of utmost importance to improve the disclosure of corporate data and providers' methodologies to enhance the environmental assessment, thus fostering the development of sustainable finance.

1 Introduction

Climate and environmental risks are becoming more apparent through damages by extreme events of physical risks and biodiversity loss, while global warming has exceeded 1.1°C above pre-industrial levels (IPCC, 2023). Furthermore, such risks have become more prominent among regulators and supervisors, as testified by the regulatory initiatives in the EU (e.g. EU green taxonomy) and at the international level (Task Force on Nature-related Disclosure) as well as supervisory expectations drawn in several jurisdictions (ECB, 2020 and NGFS, 2020). In addition, climate change risks have attracted a growing interest of policymakers, public institutions and investors. Meaningful initiatives have been taken, for instance, by several central banks (Visco, 2021, Signorini, 2020, Dunz et al., 2021, Bolton et al., 2020 and Bailey, 2021) and the United Nations race to zero campaign (UNEP, 2021).

The views expressed herein are those of the authors and do not necessarily represent the views of the Bank of Italy. Beyond data sources directly available, the research project benefited from data set kindly provided by Prof. Monica Billio and the team of University of Venice - Ca' Foscari within its project on ESG data analysis. We thank Ivan Faiella, Stefano Siviero, Antonio Scalia, Patrizio Pagano, Franco Panfili, Tommaso Perez, Marco Taboga, Alessandro Mistretta, Cristina Angelico, Francesco Columba and Pier Luigi Migliorati for their helpful comments. All remaining errors are our own.

The scores of specialised providers are widely used to assess environmental and climate risks, although such scores usually refer to broader sustainability issues, combining environmental, social and governance profiles (the so-called ESG scores). The development of ESG scores has boosted the growth of sustainable investments, which have doubled during the last four years (2018-2022) reaching 35 trillion USD globally, almost one-third of the assets under management (GSIA, 2020). In general, sustainability scores represent key information to channel capital towards transition investment in the long term. Despite their pivotal role, the ESG scores are currently far from being transparent (Angelini, 2022). In the absence of clear and shared methodologies for sustainability assessment, there is a significant risk of greenwashing, which can translate into regulatory arbitrage or misrepresentation to the investors. Moreover, decision-makers on sustainable investment face challenges to reduce the reliance on third-party ESG scores, since most of the investors struggle to gather and assess sustainability metrics or do not have the analytical skills to make their own ESG assessment yet (Schumacher, 2021). The weak scrutiny of ESG scores in the financial markets may be indeed due to a lack of incentives among passive managers and index providers within the long ESG-data value chain (Pagano et al., 2018).

A better understanding of the analytical tools to assess environmental and climate risks and opportunities is nowadays crucial. Differently from most of the studies that investigate the contribution of ESG raw data to overall scores (Billio et al., 2021 and Lee et al., 2020), this paper aims at assessing the link between environmental (hereafter E) raw indicators and scores. Indeed, the environmental and climate-related measures among ESG profiles have become the most prominent given the rising awareness of the urgency of tackling climate-change threats to the economic and financial system. Moreover, the environmental profile offers wider coverage of firms and a larger array of data points in comparison with the Social and Governance pillars.

To our knowledge, the paper offers the uniqueness of a rich data set of environmental scores and raw data from seven specialized providers over a period of seven years (i.e., 2015-2021).

The paper specifically deals with two main research questions. The first one is related to what extent the E-scores stem from data or rely on a qualitative judgement of the ESG providers. To gauge the contribution of these two components, this work applies both traditional econometric methods (i.e. fixed effects quantile panel regression) and machine learning techniques (i.e. Lasso regression). According to both approaches a limited set of variables play a major role, such as the presence of reduction targets for emissions and resource use as well as environmental and renewable energy policies. Moreover, the outcomes hint at the possibility that a qualitative component plays a significant role. We explore this hidden component through a Kalman filter finding a 'judgemental' latent variable more material for some providers than others. This evidence highlights the differences among providers concerning the methodologies and the focus on either environmental impact or the relevant financial risk or both. Nonetheless, the paper does not investigate how the judgmental component can help to predict future changes of the E-score.

The second research question attains the design of an environmental assessment relying only on raw data through a classification rule. The classification system allocates each company among three environmental performance classes consistently with the actual grading of providers. The design of the three classes is aimed to replicate two common sustainable investment strategies. We train the system for each provider through two techniques: the Linear Discriminant Analysis (LDA) and the K-Nearest Neighbours algorithm (KNN). We test the classification based on LDA and KNN in the period 2017-2021 by simulating portfolios consistent with *best-in-class* (or positive screening) and an *exclusion* (or negative screening) investment strategy. The portfolios based on classification rules present risk/return and environmental profiles similar to portfolios built by using the providers' scores. The classification system could be easily extended to companies that disclose environmental variables, although not rated with E-scores. There is a potential application connected with the EU Corporate Sustainability Reporting Directive, which will broaden the voluntary disclosure also of unlisted Small and Medium-Sized Enterprises (SMEs) starting from the 2026 financial year.

The paper is structured as follows. In Section 2 we review the relevant theoretical and empirical literature. Section 3 presents the data set and the descriptive analysis of environmental variables and the E-scores. Section 4 focuses on the results of the regression approach used to detect the role of raw data in defining the E-score. We then discuss the significance and the pattern of the latent qualitative components. Section 5 shows the classification system and its results in terms of accuracy also employing portfolio simulations. Finally, Section 6 concludes.

2 Literature on ESG scores and environmental indicators

There is a broad and still growing literature on dissecting the effects of ESG profiles on business performance and corporate evaluation. The evidence among studies is mixed, also due to heterogeneity over ESG terminology and diverging imputations methods in ESG scoring (SSGA, 2017 and Ehlers et al., 2023). One of the most recent meta-study by Atz et al. (2023) on over 1.100 studies between 2015 and 2020 finds that ESG investing can provide superior financial performance in one out of three studies. This holds especially during a social or economic crisis while being indistinguishable in the rest of the studies.[1]

Despite this evidence, the divergence among ESG scores of different providers may lead to controversial investment decisions and are subject to rising criticism. Berg et al. (2019) find an average correlation among global ESG scores of around 60%, compared to the nearly over 90% correlation among credit ratings.[2] The divergence of ESG scores can derive from differences in data sources and selection of indicators as well as assigned materiality. Heterogeneity in definitions and assessment methodologies can prompt such divergences, which can be misunderstood by investors (Billio et al., 2021). In addition, this heterogeneity produces weak signals in asset pricing and may jeopardise the efficient allocation of capital to companies committed to the transition.

To face the lack of sustainability information, either regulatory or voluntary disclosure may play a significant role in increasing the firm valuation (Ioannou and Serafeim, 2017). Jebe (2019) proposes to overcome the disconnect between financial and ESG information by clarifying the definition of financial materiality in sustainability reporting. Currently, several standard-setting bodies (SASB, IFRS, EFRAG, ISSB) are working to provide key guidance towards this avenue. However, we are still far from setting robust governance of sustainability information (Aramonte and Packer, 2022).

Among ESG profiles, the environmental and climate-related metrics have become the most prominent given the rising awareness of the urgency of tackling climate-change threats to the

[1] A previous meta-study on more than 2,000 research papers by Friede et al. (2015) finds that almost 90 per cent of papers evidenced that companies more careful of environmental, social and governance issues show a positive (or a non-negative) relation with higher financial and market performance. Besides, decarbonisation strategies can potentially capture a climate risk premium. Looking at the economic roots of this phenomenon, Clark et al. (2015) state that efforts to enhance environmental, social and governance profiles lead corporates to higher operational efficiency, innovation of products, and manufacturing processes and they are also perceived as less risky by investors that demand a lower risk premium for investing equity and credit in such companies. The combination of these factors brings such a superior performance of ESG-leading firms. Ehlers et al. (2022) underline that investors can improve the sustainability of their investments through ESG investing without jeopardising the risk-return profile of their portfolios.

[2] Lanza et al. (2020) compute a slightly lower correlation between 0.4 and 0.6 of ESG scores for euro-area equities.

economic and financial system. Furthermore, the need for a deeper understanding of environmental data is also due to the challenges in measuring environmental and climate risks. These assessments need to consider forward-looking factors and their associated high levels of uncertainty, tipping points and complex compounding effects (NGFS, 2019, 2022). In addition, the importance of measuring and managing these sources of risk is underlined by initiatives from regulators (e.g. the EU Taxonomy and the Corporate Sustainability Reporting Directive in Europe, and the SEC Climate Disclosure rule in the US), and supervisors (e.g. NGFS, 2020).

The lack of quality, consistency and availability of environmental raw data hampers the soundness and transparency of E-scores. A disconnect between E-scores and the relevant environmental and climate indicators is found by OECD (2022), which identifies areas to improve the alignment of E-pillar scores to low-carbon objectives and to give higher weight to forward-looking climate measures. It also highlights the importance of effective processes to track and verify data to ensure credibility among market participants. Considering separately each ESG pillar, Lee et al. (2020) show that some E- indicators (e.g. carbon emissions) have long-lasting effects because they tend to accumulate, while the G indicators are more influential in the short term. Papadopoulos (2022) documents discrepancies in greenhouse gas emission data among providers across time and sectors. Such discrepancies increase as one moves from direct (Scope 1) emissions to indirect emissions (Scope 2 and 3). They can translate into diverging carbon performance assessments and might affect E-pillar scores and the ESG ratings as a whole. Similar to these works, our research focuses on the E-pillar. Compared to OECD (2022), which covers a large sample of almost 2,500 companies globally and four providers for one year, we consider environmental data from seven providers for a set of European companies over a several year time-span.

3 Sources and data description

Our analysis starts from the European listed equities belonging to the Euro Stoxx index[3], whereby all sectors are represented. The initial sample is based on the 343 constituents of the Euro Stoxx index, combining the composition at the end of 2011 (288 stocks) with that at the end of 2021 (309 stocks). We exclude 65 financial stocks (banks, insurance companies and diversified financials, while real estate companies are included) due to their intermediation function in the economy. In other words, their environmental profile depends greatly on that of non-financial companies represented in the banking and investment portfolio holdings. Moreover, we filter out stocks for which neither E-scores nor raw environmental data are available, without altering sector composition. The final sample includes 211 equities, which represent 86% of the current market capitalization (see Table 1).

The environmental scores and raw data for the sample are sourced from seven data providers: MSCI ESG, Bloomberg, RobecoSAM/S&P Global, ISS-Institutional Shareholder Services, Carbon Disclosure Project, Sustainalytics and Datastream-Reuters-Asset4. Such providers offer either the E-scores only or both scores and raw data. In particular, raw data are taken from Bloomberg and Datastream-Asset4 and CDP. They provide 209 environmental raw indicators, whose coverage is heterogeneous across companies: for only one-third of the variables, the coverage is higher than 70% (see Figure 1).[4] This evidence highlights the relevance of data gaps, hampering the environmental

[3]The Euro Stoxx Index is a broad yet liquid subset of the Stoxx Europe 600 Index. The index includes large, mid and small capitalisation companies of 11 euro area countries: Austria, Belgium, Finland, France, Germany, Ireland, Italy, Luxembourg, the Netherlands, Portugal and Spain.

[4]Among the starting data set of 209 variables, 125 are taken from Datastream, 49 from Bloomberg and 35 from

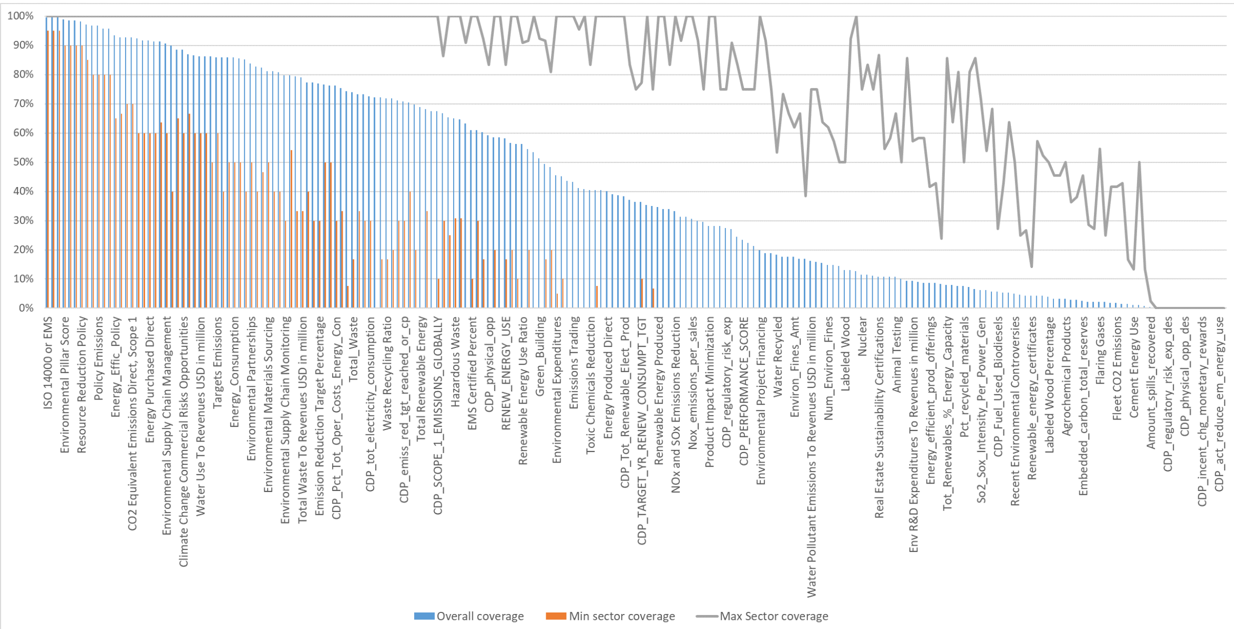
Table 1. Corporates sample – sector and market cap weights.

Sectors	No. of Corporates	Weight
Consumer Goods	17	24%
Discretionary Goods	28	23%
Energy	10	3%
Industrials	57	15%
Information Technology	10	8%
Real Estate	11	1%
Materials	33	6%
Health Care	13	7%
Communications	11	5%
Public Utility	21	8%
Total	211	100%

assessment of corporate sustainability from specialised evaluators and investors.

From the initial 209 environmental raw variables we select those with higher coverage among firms, considering a time window spanning from 2015 to 2021.

Figure 1. Coverage of environmental variables.



Note: The figure shows for each indicator the overall coverage computed on the whole corporate sample and the minimum and maximum levels of coverage among sectors.

Besides the coverage, we exclude variables following two additional criteria. First, we exclude the highly collinear variables as detected by their cross-correlation, gauged by the Variance Inflation Factor (VIF). Second, the variables less correlated with the E-scores are excluded to improve the explanatory efficiency.[5] Based on these criteria, we finally select 62 variables (8 sourced by Bloomberg and the remaining by Datastream), which can be classified into six groups

CDP. The time series of data from CDP is shorter compared to the other providers.

[5]The description of the selected variables, sources and the VIF are reported in Table 1 in the Appendix.

Table 2. Grouping of key environmental issues and contents.

Key issues	Contents
carbon emissions	Emission reduction target, Carbon pricing, Revenues considered in the emission reporting
climate and environmental risk and opportunities	Environmental quality management and environmental management system, Worksites certified by the environmental management standards, Exposure to climate change regulatory risk and opportunity, Exposure to physical risk and opportunity, Environmental fines
energy risks and opportunities	Energy intensities, Renewable Electricity production and capacity, Electricity production and consumption, Fuel consumed for energy purposes
green and clean opportunities	Environmental investments or expenditures, Design of products for reuse, Recycling or abating environmental impacts, Products improving the energy efficiency of buildings
waste and pollution risks and opportunities	Discards, Volatile organic compounds and particulate matter, Recycled and reused waste produced, Take-back procedures and recycling programs, Environmental criteria for sourcing or eliminating materials
water and biodiversity risk and opportunities	Water withdrawal, Discharged and treated, Water pollutant emissions, Animal testing, Impact on biodiversity risk, Native ecosystems and species, and protected and sensitive areas

of environmental key issues covering climate change risks and opportunities (carbon emissions and opportunities), resources use and efficiency (energy, waste and water), and clean technologies for manufacturing products (see Table 2). Most importantly, a relevant number of data points refers to binary variables (e.g. *yes* or *no* for the adoption of environmental policies or carbon reduction commitments, etc.). Among the 62 selected, only nine variables are continuous.

Finally, the E-scores of each provider have been normalized on a scale between 0 and 100 using a min-max scaler, while the environmental variables regarding absolute figures (e.g. carbon emissions, water and energy consumption) have been normalised according to corporate revenues.

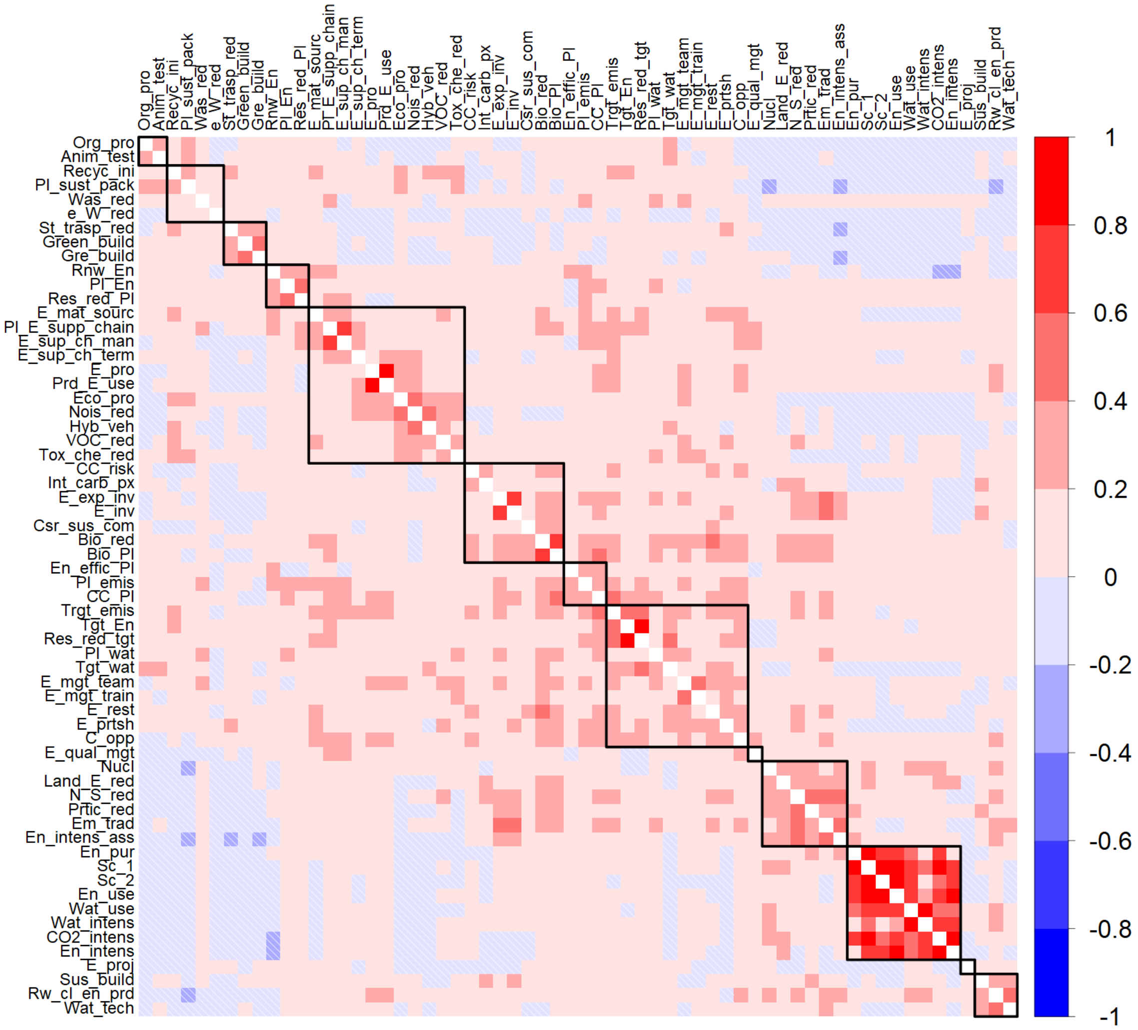
3.1 Descriptive statistics of the environmental variables and scores

A snapshot of the median cross-correlation among the 62 environmental raw variables for the 211 firms is shown in Figure 2. We spot a cluster of higher correlations among variables relating to carbon intensity, energy and water use.

On the other side, the E-scores of the seven data providers display a linear correlation in a range between 0.14 and 0.46 over the period 2015-2021, with an average of 0.27 (see Figure 3). Notably, the correlation is lower than that among credit ratings of the same corporate sample as of June 2022, which falls in the range 0.7–0.9 (Table 3).

A higher pairwise correlation is found between Sustainalytics and Robeco (0.46), Robeco and Datastream (0.45), and MSCI and Robeco (0.38). The average pairwise correlation for each

Figure 2. Cross-correlation of environmental variables.



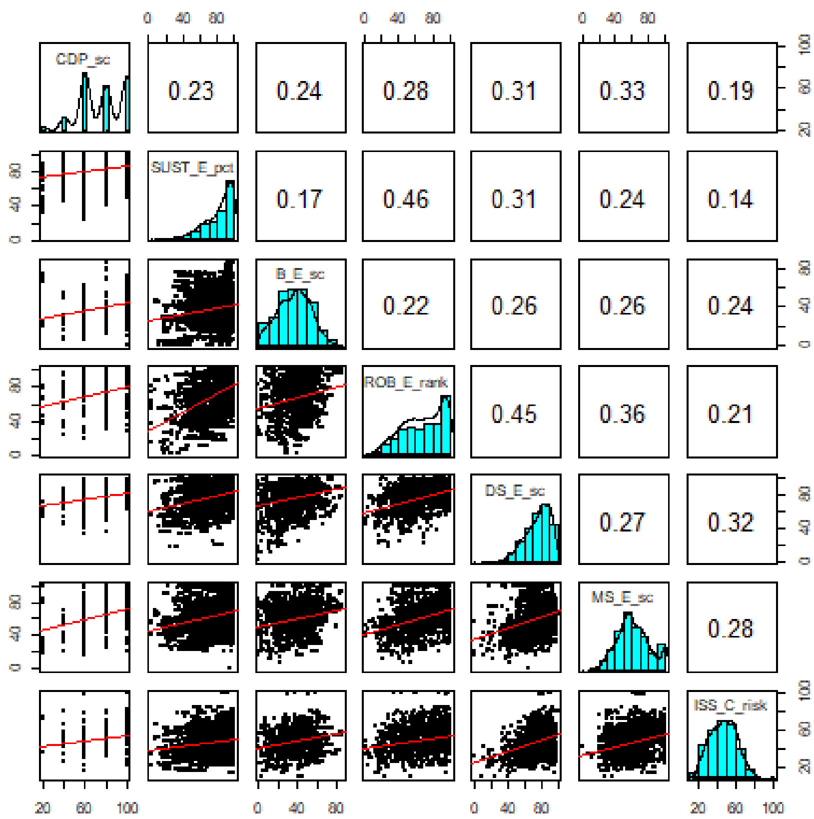
Note: The figure shows median annual values of Pearson correlations among each pair of variables. Variables have been ordered according to hierarchical clustering. Hidden patterns are highlighted by means of black boxes. The description of the raw data is reported in Table 1 in the Appendix A.

Table 3. Credit rating correlation.

	S&P	Moody's	Fitch
S&P			
Moody's	0.9		
Fitch	0.8	0.7	

Note: Own elaboration on Bloomberg data.

Figure 3. E-scores distributions, scatter plots and correlations across providers (2015-2021).



Legend: CDP = Carbon Disclosure Project; SUST = Sustainalytics; B = Bloomberg; ROB = Robeco; DS = Datastream/Asset4; MS = Morgan Stanley ESG; ISS_sc= ISS carbon risk rating.

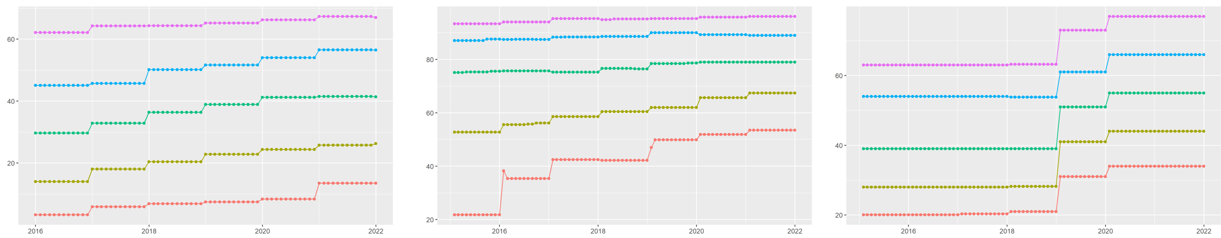
provider is particularly low for Bloomberg and ISS, while it is higher for Robeco and MSCI (see Table 4). We identify different patterns in the distribution of E-scores. The distribution of the scores from Bloomberg, MSCI ESG and ISS is close to a normal , while those from Robeco, Sustainalytics, Datastream and CDP are skewed towards lower values (Figure 3). Moreover, the pairwise correlation patterns seem dependent on the shape of E-score distributions, where more skewed ones show a higher correlation. This heterogeneity in the distribution shape may refer to the different underlying provider's assessment methodologies. Brandon et al. (2021) underline that Sustainalytics and MSCI aim to provide ratings on ESG performance, while some providers like Bloomberg are more geared towards capturing specific profiles such as ESG disclosure quality. In addition, the CDP score reflects a particular focus on the corporate commitment to emission reduction, rather than a broad environmental assessment.

Table 4. Average pairwise correlation per provider.

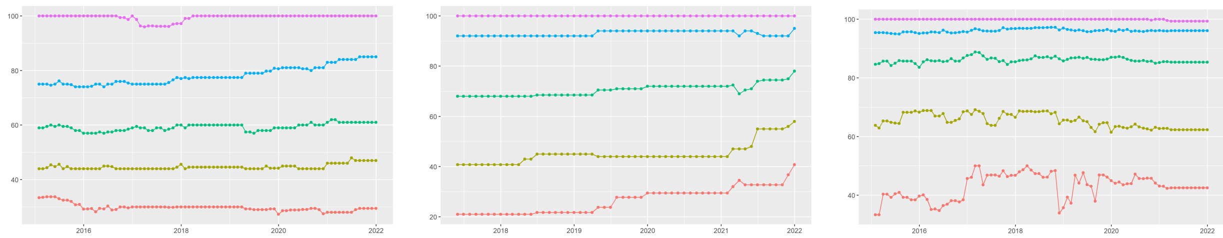
Provider	Mean
CDP	0.29
Sustainalytics	0.23
Bloomberg	0.22
Robeco	0.33
Datastream	0.28
MSCI	0.30
ISS	0.22

Analysing the changes over time (see Figure 4), we record an increase of the average E-scores across the selected quantiles (from 5% to 95%), especially for lower quantiles among most of the providers (Bloomberg, Datastream, ISS, Robeco). This pattern is consistent across the whole sample, including the Covid period. It may be due to several reasons, such as an increase in coverage (e.g. Sustainalytics), an improvement of the environmental assessment of the worst-rated companies (Datastream), a review of the assessment methodology (ISS), a re-calibration of the scores (Bloomberg), or an increased transparency in the firms' reports.

Figure 4. Quantile average scores per provider.



(a) Bloomberg **(b)** Datastream **(c)** ISS



(d) MSCI **(e)** Robeco **(f)** Sustainalytics

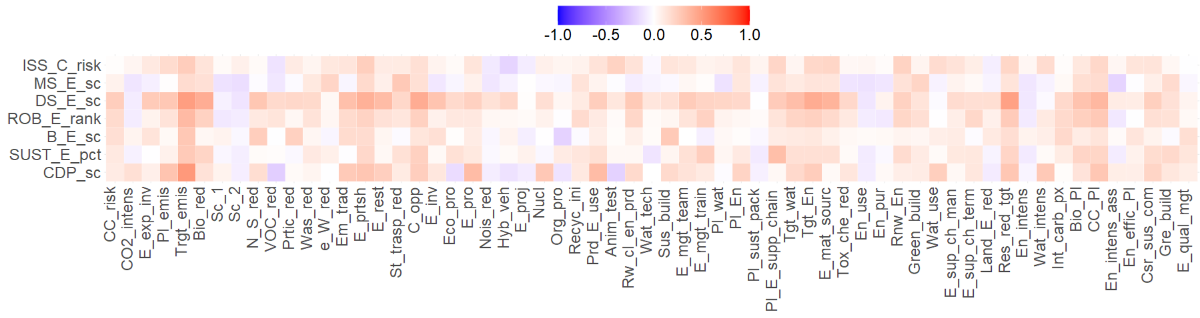
Note: In each panel, lines from bottom to top stand for the 5%, 20%, 50%, 80%, and 95% quantiles, respectively.

Looking at the correlation between E-scores and raw data as an early signal of their relationship, we find that some providers (i.e. Datastream, CDP, Robeco and MSCI) are more reliant on raw environmental data (Figure 5).

Looking at the sectoral breakdown of the correlation between E-scores and raw data (Figure 6), some raw variables appear more relevant for the E-scores in specific sectors. For instance, Scope 1 and 2 emissions (`Sc_1` and `Sc_2` in the plot) exhibit a negative correlation with E-scores in most of the sectors, while their explanatory power softens in the average correlation.

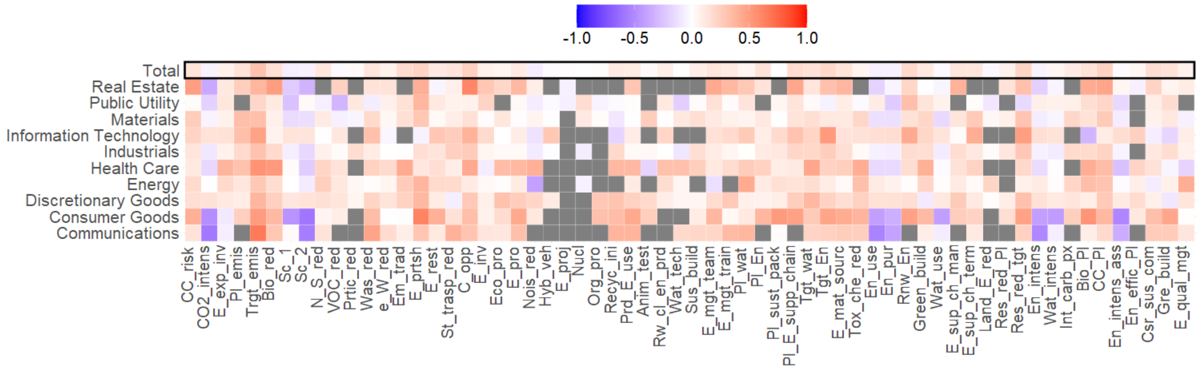
To summarize, the descriptive analysis points out that the low correlation among E-scores may be partially due to different distribution patterns across providers and different reliance on

Figure 5. Cross-correlation between environmental variables and E-scores.



Note: The levels reported are the median of annual correlation figures. The description of the raw data is shown in Table 1 in Appendix A.

Figure 6. Correlation by sector between environmental raw data and E-scores.



Note: The grey boxes mean missing raw data. The description of the raw data is shown in Table 1 in Appendix A.

raw data. We explore such a nexus in the following Section. To avoid any possible qualitative assessment of the providers based on the paper results, we anonymise their names (from A to G) from this point onwards.

4 Dissecting E-scores with raw data and regression techniques

We apply Lasso[6] and quantile[7] regression to analyse the contribution of environmental data to E-scores, finding more promising results for the Lasso technique in terms of explanatory power and variables identified. The advantage of using the Lasso estimation is twofold. On the one hand, we obtain a signal on relevant variables that are crucial to create an implicit rule to detect and

[6]De Lucia et al. (2020) combine statistical inference and machine learning techniques such as the K-Nearest Neighbours algorithm to explore the causal relation of ESG data points with corporate financial indicators. Bouyé and Menville (2020) apply five types of regressions to ESG scores for sovereigns, spanning from step-wise, principal component analysis, ridge, Lasso and elastic net regressions. Although all five techniques (see Hastie et al., 2016) are roughly equally robust, they find it preferable to apply the Lasso model for its virtue of limiting the number of explanatory parameters.

[7]Teng et al. (2021) apply a quantile regression (shortly, QR) to assess the relevance of environmental variables among ESG risk scores. QR technique may be more efficient to integrate information of the tails of the distribution than the standard ordinary least squares (hereafter, OLS) as QR allows to estimate the nexus between the dependent variable and its explanatory variables at any specific quantile, while OLS focus on the mean.

distinguish green issuers from less green ones. On the other hand, the regularisation imposed by Lasso reduces the risk of overfitting. Therefore, we explore the possibility to obtain meaningful variable selection using a penalised regression methodology. The norm of the penalisation could vary and should be chosen according to the peculiarity of the analysis. The regression equation is written in the form of a simple linear model with Gaussian errors:

$$E_{it} = X_{it}\beta + \varepsilon_{it}$$

Where E_{it} is the E-score, X_{it} is a matrix of explanatory variables and ε_{it} is the Gaussian error term for issuer i in year t. The matrix of coefficients is obtained by minimising the following objective function which includes the Lasso regularisation (i.e. penalty) term:

$$\beta^{\mathcal{L}} = \sum_i (E_{it} - X'_{it}\beta)^2 + \lambda \sum_j \mid \beta_j \mid . \tag{1}$$

The Lasso term is the absolute sum of all the j estimated coefficients multiplied by the factor λ. Where $\lambda \in \mathbb{R}^+$ is the tuning parameter which determines the amount of penalisation imposed and is calibrated using cross-validation.

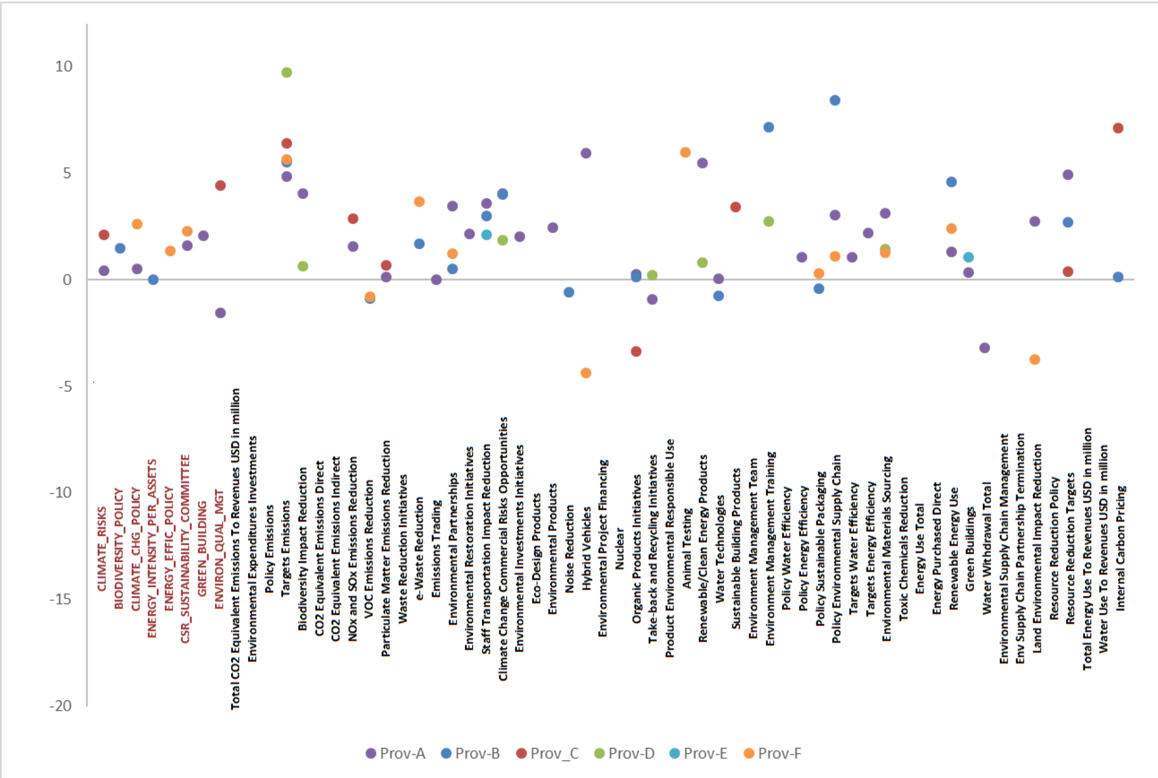
Separately for each provider, we run a Lasso regression for each E-score on the set of 62 raw variables over the period 2015-2021. Despite the different relevance of some raw data at the sector level as shown in the previous Section, we are not able to perform sector-specific regression analysis due to the scarcity of data for certain sectors (see Figure 6). Table 5 summarises the main results. Notwithstanding the ex-ante exclusion of potentially redundant variables based on the VIF, we believe that not all of the 62 remaining variables in the data set provide significant explanatory power for measuring environmental performance. Thus, some collinearity may remain. By estimating equation 1, we obtain a sparse matrix of coefficients only for the significant variables and for six out of seven E-scores. We leave Provider-G out as we cannot find significant variables for this provider due to the scarcity of data. The zero coefficients filter out the redundant variables. The highest explanatory power is recorded for Provider-A (62%) with the largest number of 31 significant variables, while for Provider-F and Provider-B the R^2 is around 20% and 15 relevant variables on average. Provider-E exhibits very few variables reaching a very limited explanatory power for this technique.

Table 5. No. of significant variables, R^2 in Lasso regression, and correlation between providers and Lasso-estimated E-scores.

Provider	No. Variables	R^2	Correlation
Provider-A	31	62%	0.76
Provider-F	14	21%	0.48
Provider-B	17	20%	0.49
Provider-C	9	17%	0.46
Provider-D	7	13%	0.44
Provider-E	2	3%	0.29

The analysis of the coefficients (see Figure 7) underlines that some of the most meaningful variables are common across E-scores, such as the presence of reduction targets for emissions and resource use as well as environmental and renewable energy policies. However, the coefficients of several raw variables vary remarkably across E-scores, such as the presence of emission targets,

Figure 7. Lasso coefficients for different E-scores.



Note: The first eight variables are sourced by Bloomberg, the others from Datastream.

environmental policies for the supply chain and internal carbon pricing. This evidence hints at a different materiality assessment among providers.

The quantile regression identifies a similar number of significant variables for five out of seven providers. [8] In particular, Provider-A seems to rely on the largest number of raw variables. For Provider-C, Provider-F and Provider-B, the number of identified variables decreases. Moreover, the quantile regression analysis reveals a higher explanatory power in the tails of E-scores distribution, in particular for the companies performing *worst* on environmental issues. In other words, raw data seems to better explain the worst companies in comparison with those that exhibit the best environmental practices. This outcome is particularly interesting for investors seeking to use E-scores for *best-in-class* and *exclusion* strategies.

Moving forward in this direction, in the following Section we investigate whether Lasso regression residuals could be interpreted as an unobserved component of providers' E-score, not explained by raw data. This component could reflect non-linear relationships or qualitative judgements.

4.1 Latent variable analysis

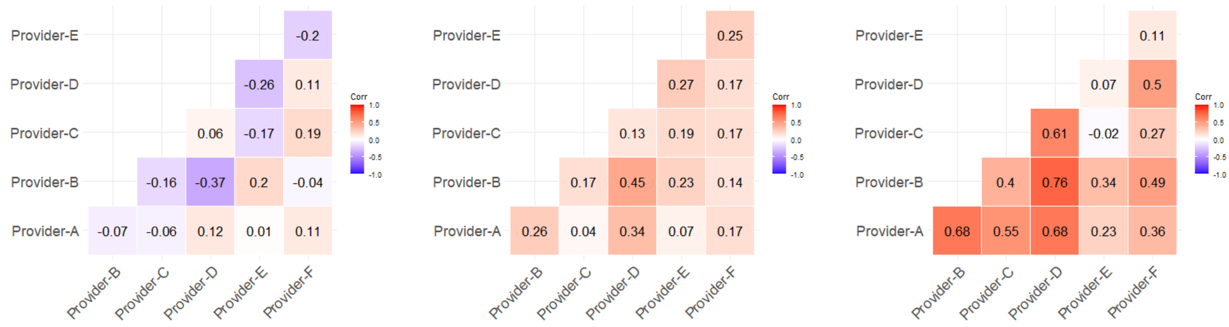
To further investigate the unexplained variance in our regression analysis, we filter the data by applying the Kalman filter (see details in Section B.3 of the Appendix). By using the estimated Lasso regression residuals, we build a latent variable which can be interpreted as a judgemental or non-linear component of E-scores. The results can help to identify a possible common element across providers which is not related to raw data. The analysis shows a very low correlation of

[8] for extensive results of the quantile regression see Section B.1 in the Appendix

the estimated latent variables among providers (with an average correlation at -0.03, see Figure 8), which is below the correlation of the corresponding actual E-scores (on average at 0.20), while the correlation between Lasso-estimated E-scores is higher (on average at 0.40, see Figure 8).

This evidence suggests that the environmental scoring based only on raw data can lead to more similar E-scores, while the remaining component, arguably a judgemental or non-linear one, brings more heterogeneity across providers' scores. This heterogeneity may be due to different sustainability definitions and perspectives taken from providers in their environmental assessments as highlighted by Billio et al. (2021). Indeed, some raters focus their analysis on how environmental issues affect corporate financial conditions, while others consider how corporate conduct can affect environmental conditions, others consider environmental performance, and others consider both perspectives ('double materiality'). Moreover, some providers consider prominently the environmental risk implications for firms, while others consider also the related opportunities as observed by Larcker et al. (2022).

Figure 8. Correlation among estimated Latent variables, actual, and Lasso-ruled E-scores.



(a) Latent variables **(b)** Actual E-scores **(c)** Lasso-estimated E-scores

By summarising, we conclude that the unexplained components remain relevant across providers, while appear not related to each other. This suggests raters employ different assessment approaches when treating raw data, which could be influenced by non-linear relationships or judgmental considerations

5 Classification systems for sustainable investment strategies

In this Section, pursuing the second research question, we explore how the environmental assessment can be translated from a continuous variable (i.e. score) into a discrete grading trained on the provider's E-scores. To this aim, we apply classification techniques based on raw variables to rate each company in a grade consistent with the rank resulting from the actual E-scores.[9] The classification system allows the extension of the environmental assessment to companies not rated by providers using only raw data. Nevertheless, the grading is calibrated to implement sustainable strategies, i.e. *best-in-class* or *exclusion*, which are among the most adopted investment strategies

[9]The set of raw data is the same for all the providers.

at the global level (GSIA, 2020). Therefore, we finally test the classification rules by means of portfolio simulations.

5.1 Classification rules

For each provider, we apply two well-known classification techniques, the LDA and the KNN, to predict the probability of belonging to one of three environmental performance classes based on actual E-scores; i.e., worst, intermediate, and best.[10] Companies are thus assigned to three environmental rating grades identified by thresholds corresponding to the first and the fourth quintile of the distribution of each E-score (i.e., 20 and 80%). We split the whole data set into training samples and test samples to prevent overfitting issues and to produce more accurate estimates. For both techniques, the training set consist of 2-year rolling samples, while each test set is based on the observations in the year following the training sample. We therefore train the LDA and the KNN from 2017 to 2020,[11] and we test our system over the period 2019–2021. Slightly different from the regression analysis, we omit raw variables exhibiting no variability over time. In particular, we exclude 15 variables,[12] thus reducing the number of raw variables from 62 to 47.

5.1.1 Linear Discriminant Analysis

The LDA aims to find combinations of raw variables that maximise the separation between classes similar to those obtained with each provider's E-scores. Overall, the in-sample accuracy ratio of the LDA rule across providers is high, ranging within the interval of 67% - 80%. The higher accuracy ratio is found for the rule mimicking E-scores of Provider-A and Provider-C (80% on average), holding through the considered time windows (see Figure 9).

The general picture is confirmed by the out-of-sample accuracy, showing consistent results across time windows for Provider-A, Provider-E and Provider-C (on average at 74%, 71% and 67% respectively), whilst the LDA shows lower accuracy ratios for Provider-F (53%), Provider-B (55%) and Provider-D (60%).

The promising results in accuracy underline that classification techniques can help investors to make better decisions on how to classify companies (as "best" or "worst"), without the need to use or replicate (e.g., through regression) the providers' E-scores. The correlation between the class allocation based on actual E-scores and the one estimated using the LDA ranges between 0.20 and 0.67 with an average of 0.42.

Finally, the estimated coefficients of raw data for the LDA rule across data providers,[13] are

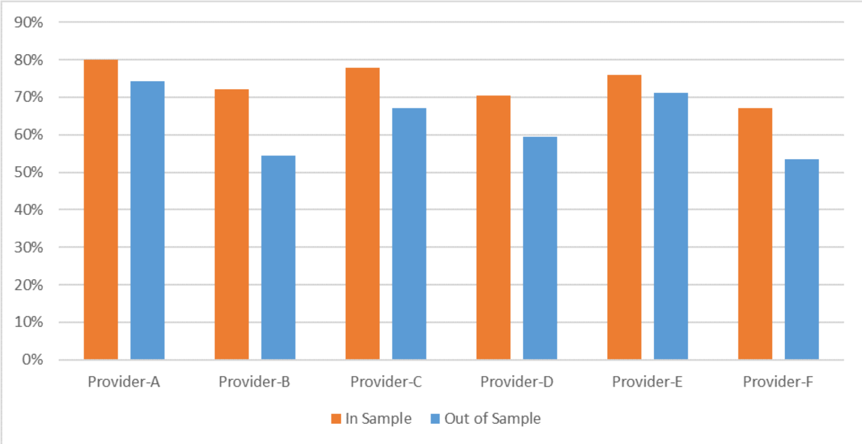
[10] We also tested multi logit analysis that returns worse results both in-sample and out-of-sample with respect to the LDA and the KNN.

[11] The rolling time windows are 2017-2018, 2018-2019 and 2019-2020, given that for years 2015 and 2016 the repeated data for Provider-F prevents from running the LDA and the KNN estimation.

[12] The list of disregarded variables contains climate risks, emissions reduction, particulate matter emissions reduction, eco-design products, noise reduction, animal testing, environment management team, environment management training, policy sustainable packaging, land environmental impact reduction, resource reduction policy, green buildings, environmental quality management, environmental project financing, and energy efficacy policy.

[13] These coefficients cannot be interpreted as in the standard regression, rather they project all the information of the 46 variables into the one-dimensional space of the E-score. Nevertheless, thanks to the standardization of the variables, absolute estimated coefficients greater than 1 (taken as a threshold) can help to identify the most material predictors to discriminate firms among different environmental rating classes.

Figure 9. In-sample and Out-of-sample accuracy per data provider-implied LDA rule.



Note: The accuracy ratios are averages of the three time windows and classes for each provider.

Table 6. Correlation between classes obtained through classification based on the empirical distribution of providers' E-scores and the LDA rule, respectively.

Variable	Correlation
Provider-A	0.67
Provider-B	0.28
Provider-C	0.49
Provider-D	0.40
Provider-E	0.50
Provider-F	0.20

Note: The table displays the value of the pairwise correlations between firms allocated in grades according to E-score and the relevant LDA allocation rule.

consistent with those of the regression analysis. Most material variables are related to carbon and energy intensities.[14]

5.1.2 K-Nearest Neighbours

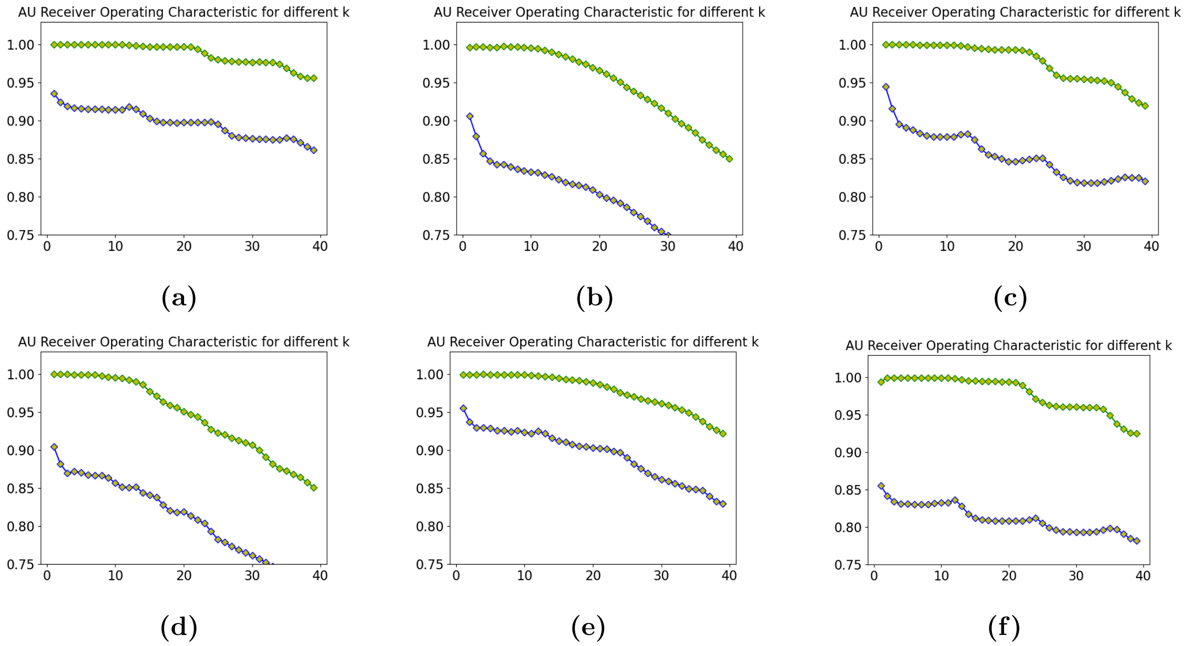
The second classification rule is developed by applying the KNN algorithm. First, we determine the appropriate number of neighbour points, denoted as k, which is used to classify observations into different classes based on information derived from the Area Under the Receiver Operating Characteristic (AUROC) for each provider (see Figure 10). Common choices in the literature range between 3 and 20, and we select an interval within this range. Our choice is also consistent with the rule of thumb suggesting to set $k = \sqrt{N}/2)$. We carefully select the appropriate k to strike a balance between comparison needs, concise description, and satisfactory accuracy.[15] Second, as for the LDA, the KNN assigns each issuer to one of three grades (i.e. worst, best, and intermediate)

[14]Such as total CO2 equivalent emissions to revenues in USD million, CO2 equivalent direct emissions (Scope 1 and 2), total energy use, total amount of water withdrawn from a surface water or groundwater source, total energy use to revenues, and water use to revenues, respectively

[15]In the case of an even k value, we add an additional element to break any potential ties, following the practice found in the literature.

based on the similarities detected among the features of the k nearest peers.

Figure 10. AUROC for different k by provider.



Note: Panels from (a) to (f) show the AUROC plotted for different values of k of our training samples across the time period (2015-2020). The yellow diamonds show the AUROC calculated for the train sample while the blue diamonds are AUROC calculated for the test sample.

The KNN achieves higher accuracy ratios, both in-sample and out-of-sample, compared to the LDA rule. Notably, the highest accuracy is observed for the classification of Provider-E, Provider-A, and Provider-C (see Table 7).

Table 7. KNN accuracy.

Provider	Selected k	AUROC in-sample	AUROC out-of-sample
Provider-A	3	0.999	0.919
Provider-B	17	0.977	0.815
Provider-C	15	0.996	0.875
Provider-D	3	0.999	0.870
Provider-E	3	0.999	0.930
Provider-F	13	0.999	0.837

Note: The k is selected by considering the needs of comparison, parsimonious description, and satisfactory accuracy.

Moreover, the correlation between the allocation in classes based on E-scores and relevant allocation estimated by using the KNN rule ranges from 0.45 to 0.70, with higher values for Provider-A, Provider-E, and Provider-D (see Table 8). The average correlation for the KNN rule is 0.55, higher than the 0.42 resulting from the LDA rule. The comparison between the LDA and the KNN reveals that the KNN consistently outperforms the LDA in classifying firms. The higher accuracy achieved by the KNN may be due to the model-free property of this classification technique, as compared to the LDA which assumes linearity and homoscedasticity. Compared to the Lasso method, both LDA and KNN show a stronger correlation with classes based on the actual E-scores of providers.

Table 8. Correlation between classes obtained through classification based on the empirical distribution of providers' E-scores and the KNN rule.

Provider	Correlation
Provider-A	0.68
Provider-B	0.45
Provider-C	0.46
Provider-D	0.54
Provider-E	0.68
Provider-F	0.48

Note: The table displays the value of the pairwise correlations between allocations implied by each E-score and the respective KNN implicit rule allocation.

Therefore, converting the environmental assessment outcomes from a continuous variable (E-score) to a discrete grading simplifies the estimation, improves predictability, and provides a solid ground to set 'green' strategies for equity investment. In the next Section, we test two common strategies through portfolio simulations.

5.2 Portfolio simulations

We assess the classification techniques from a financial viewpoint simulating portfolios based on two well-established sustainable investment strategies, i.e. the *best-in-class* and the *exclusion* strategies. We compare the environmental and financial features of portfolios built through the LDA and the KNN with those of portfolios built from actual providers' grades. Simulations are repeated for each provider. Therefore, the aim of this exercise is not to compare the financial performance of sustainable strategies with the whole market portfolio, but to test the accuracy of the rating systems based on the classification techniques as an alternative to the E-scores of professional providers.

The first investment strategy known as *exclusion* aims at tilting the standard market portfolio by filtering out the lower 20th percentile of the E-score distribution, i.e. the group of companies rated as *worst*.[16] The E-grade of companies is obtained through the two described classification rules. The proceeds from selling the excluded companies are reinvested into the remaining 80% of companies (*best* and *intermediate*) according to their respective market capitalisation weights in each sector (to safeguard sector neutrality). The strategy is compared with the same strategy tilted with the E-scores of the six data providers .

The second strategy known as *best-in-class* aims at investing in the top 20th percentile of the E-scored companies, according to their market weights. This strategy seeks a higher sustainability profile for the portfolio and is more suitable for active managers as it is more distant from market neutrality. Similarly to the previous investment exercise, the financial and environmental features of the portfolios built on the LDA and the KNN rules of classification are compared with the portfolios built by using the E-scores of the six providers.

[16] The percentile threshold reflects the common investment practice and is consistent with the results of the classification system.

Portfolio simulations are based on the same time-splitting rule adopted to train and test the LDA and the KNN rules. Thus, the portfolio is rebalanced yearly with a two-year rolling training sample, e.g. the rebalancing in 2019 refers to the period 2017-2018 used as training sample. The financial results are measured vis-a-vis the provider's portfolio according to the cumulative returns and two risk-adjusted measures, i.e. the tracking error volatility and the Sharpe ratio. The environmental profile of each simulated portfolio (LDA, KNN and market) is gauged by the weighted average E-score of each provider. The outcomes reported in tables are averages of annualised data, while figures show cumulative returns over the same period.

Regarding the results of the *exclusion* strategy for the overall period (from 2019 to 2021), classification rules deliver portfolios with E-scores slightly lower than those built upon the providers' scores, with a standard deviation equal to 16% for both techniques (see Table 9). This result may hint at a systematic prediction bias of the automatic rule compared with the provider's scores.

Table 9. E-scores of portfolios with *exclusion* strategy.

	LDA	KNN	provider
Provider-A	80.2	79.9	81.9
Provider-B	73.8	71.9	74.0
Provider-C	32.5	30.2	32.7
Provider-D	73.2	72.4	75.3
Provider-E	65.4	66.8	68.4
Provider-F	56.5	56.7	58.5

Concerning the financial performance, the investment portfolios built on the KNN technique achieve risk-adjusted returns (measured by Sharpe ratio) similar to those built on the providers' scores, with a limited tracking error volatility (see Table 10).

Table 10. Financials of the portfolios with *exclusion* strategy.

	Provider-A			Provider-B			Provider-C		
	LDA	KNN	prov.	LDA	KNN	prov.	LDA	KNN	prov.
TEV vs. prov.	1.1	1.5		1.5	1.0		0.9	1.9	
Sharpe R.	101.6	99.4	96.0	101.2	110.1	108.8	101.5	103.3	99.7

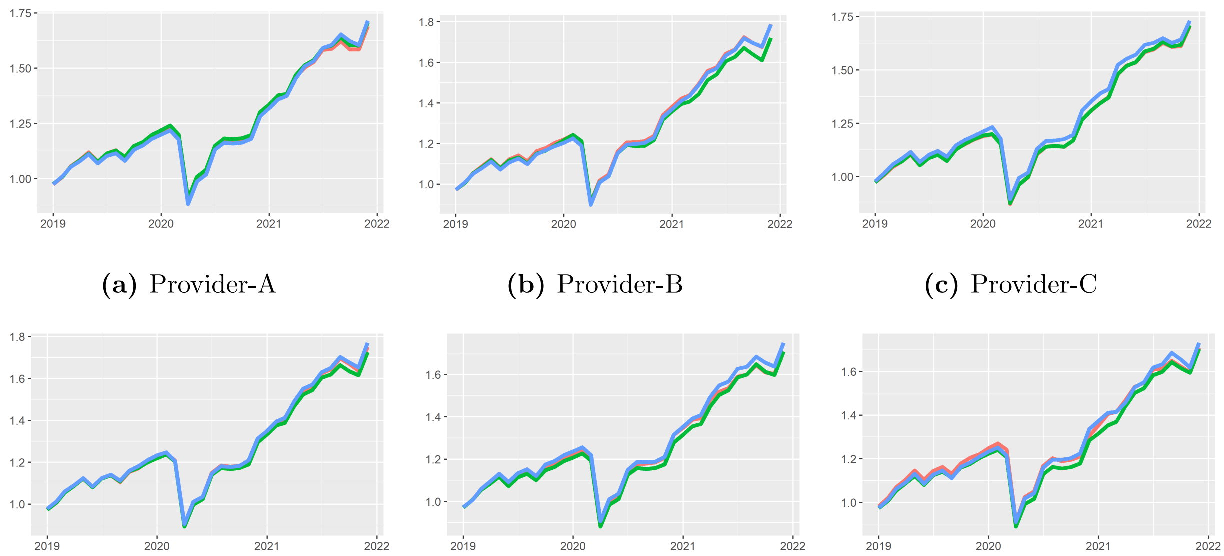
	Provider-D			Provider-E			Provider-F		
	LDA	KNN	prov.	LDA	KNN	prov.	LDA	KNN	prov.
TEV vs. prov.	0.9	1.2		1.4	1.2		2.1	2.2	
Sharpe R.	98.4	105.8	100.7	95.5	101.9	96.6	95.3	99.8	94.5

The cumulative returns of three simulated portfolios are aligned (see Figure 11). Nevertheless, the KNN confirms a slightly higher performance than the LDA (except for Provider-A), as confirmed also by the Sharpe ratio reported in Table 10. It is worth noting that the portfolios built according to different E-scores show quite diverse cumulative returns.

Regarding the *best-in-class* strategy, we find that portfolios based on the classification rules have a weighted average E-score similar to the portfolios built on providers' scores, with a small bias due to misclassification (see Table 11).

As far as the financial indicators are concerned, portfolios built on the KNN and the LDA achieve risk-adjusted returns that are comparable to those based on E-scores. However, they show tracking error levels higher than the exclusion strategies, given the less diversified portfolios

Figure 11. Portfolio cumulative returns with *exclusion* strategy.



(a) Provider-A **(b)** Provider-B **(c)** Provider-C

(d) Provider-D **(e)** Provider-E **(f)** Provider-F

Note: The red line stands for data provider portfolios, while green and blue lines stand for LDA- and KNN-based portfolios, respectively.

Table 11. E-scores of portfolios with *best-in-class* strategy.

Environmental scores	LDA	KNN	provider
Provider-A	86.9	88.9	92.9
Provider-B	68.9	86.4	97.1
Provider-C	51.4	50.4	63.7
Provider-D	86.7	85.4	95.2
Provider-E	84.5	80.9	93.1
Provider-F	61.5	63.4	71.7

(concentrated on high-grade firms) which amplify the classification differences (see Table 12).

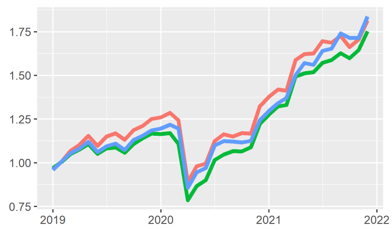
Table 12. Financials of the portfolios with *best-in-class* strategy.

	Provider-A			Provider-B			Provider-C		
	LDA	KNN	prov.	LDA	KNN	prov.	LDA	KNN	prov.
TEV vs. prov.	4.4	5.1		7.7	5.0		7.1	7.0	
Sharpe R.	87.8	98.8	92.9	80.1	64.8	47.7	70.5	106.2	102.0
	Provider-D			**Provider-E**			**Provider-F**		
	LDA	KNN	prov.	LDA	KNN	prov.	LDA	KNN	prov.
TEV vs. prov.	6.3	4.1		5.9	5.9		6.5	3.9	
Sharpe R.	42.0	81.4	64.0	69.6	141.4	86.8	121.5	102.9	108.5

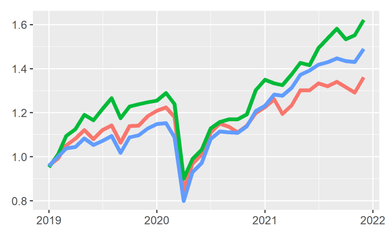
The pattern of cumulative returns among the three simulated portfolios is more differentiated with respect to the *exclusion strategies* (see Figure 12).

Once more, the financial results obtained with both sustainable investment strategies are quite

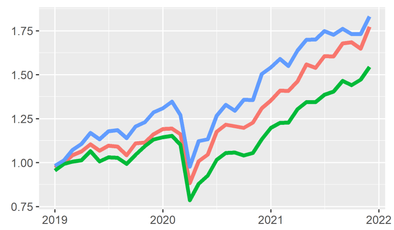
Figure 12. Portfolio cumulative returns with *best-in-class* strategy.



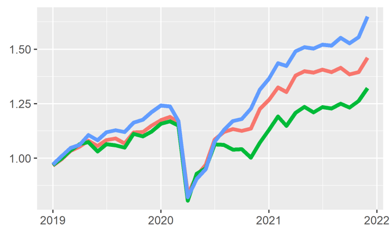
(a) Provider-A



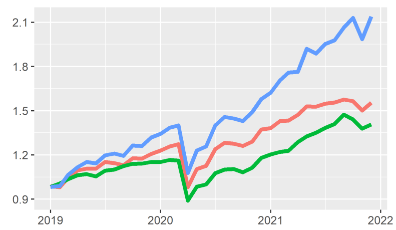
(b) Provider-B



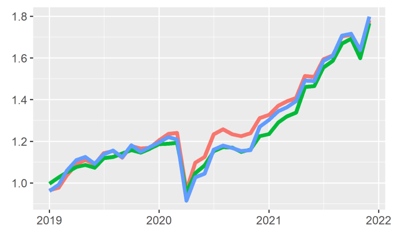
(c) Provider-C



(d) Provider-D



(e) Provider-E



(f) Provider-F

Note: The red line stands for data provider portfolio, while green and blue lines stand for LDA- and KNN-based portfolios, respectively.

diverse for different providers. This confirms that the financial materiality of the environmental assessment by the providers can vary significantly and eventually, it leads to different market performance of sustainable investment strategies.

6 Conclusions

The paper studies the relationship between E-scores and raw data with the aim of providing insights into the extent to which raw data contribute to the environmental scores of providers.

In particular, we find that environmental data have meaningful, although limited, explanatory power for the E-scores. The quantile and Lasso regression analysis offer consistent evidence that the scores of some providers are more related to raw data, while for others a missing component, possibly due to a judgemental assessment, plays a major role. We identify some variables as significant and common across several providers, such as forward-looking measures like the presence of reduction targets for emissions and resource use as well as environmental and renewable energy policies. To further investigate the structure of regression residuals, we consider the hypothesis of a judgmental component by testing for the presence of a hidden latent variable. We find the latent component to be heterogeneous across providers and this evidence may be due to different materiality in the providers' assessments. Indeed, some providers focus their analysis on how the corporate financial conditions are affected by environmental issues, while others consider how corporate conduct can affect environmental conditions and others consider both perspectives (”double materiality”).

The paper also proposes two classification rules, aimed at allocating each company among three environmental performance classes consistent with the actual grading of providers. Such classification rules rely only on raw data, therefore they allow the extension of the environmental assessment to companies not rated by providers. For instance, a potential application could be offered by the broader scope of sustainability disclosure envisaged by the EU Corporate Sustainability Reporting Directive. That is expected to involve up to 50,000 companies, including unlisted SMEs on a voluntary basis and with a simplified disclosure framework starting from the 2026 financial year. According to our simulations, the proposed rating rules deliver portfolios with environmental profiles and return-risk characteristics similar to those obtained using the providers' E-scores. This result holds under two well-established investment strategies. These results emphasize that the challenges associated with environmental scores should not hamper the integration of sustainability considerations into investment decisions.

From a policy perspective, our findings highlight the importance of enhancing the disclosure, comparability, and quality of raw data related to environmental profiles, as promoted by TNFD (2023). The issues raised by the study regarding E-scores call for further research aimed at enhancing the providers' transparency and robustness of environmental scores. Such efforts are crucial for promoting the development of sustainable finance. The insights gained from our analysis are also valuable for companies committed to improve the corporate environmental disclosure and for financial authorities, including central banks and supervisors, to detect the key variables for their environmental assessment. Expanding this line of research would be beneficial for financial authorities to assess the reliability and consistency of the scores and data used by supervised intermediaries or counterparts in their strategies. This has implications for the overall stability and the resilience of the financial system as a whole.

References

Andriana, A. E. and Anisykurlillah, I. (2019), 'The effects of environmental performance, profit margin, firm size, and environmental disclosure on economic performance', *Accounting Analysis Journal* **8**(2), 143–150.

Angelini, P. (2022), 'Long-term investing and sustainable finance: challenges and perspectives'. **URL:** *https://www.bancaditalia.it/pubblicazioni/interventi-direttorio/int-dir-2022/en-Angelini-11-luglio-2022.pdf?language_id=1*

Aramonte, S. and Packer, F. (2022), Information governance in sustainable finance, BIS Working Papers 132, Bank for International Settlements.

Atz, U., Van Holt, T., Liu, Z. Z. and Bruno, C. C. (2023), 'Does sustainability generate better financial performance? review, meta-analysis, and propositions', *Journal of Sustainable Finance & Investment* **13**(1), 802–825.

Bailey, A. (2021), Tackling climate for real: progress and next steps, *in* 'Speech given at BIS-BDF-IMF-NGFS Green Swan 2021 Global Conference'.

Berg, F., Koelbel, J. F. and Rigobon, R. (2019), 'Aggregate confusion: The divergence of esg ratings', *Forthcoming Review of Finance* .

Billio, M., Costola, M., Hristova, I., Latino, C. and Pelizzon, L. (2021), 'Inside the esg ratings:(dis) agreement and performance', *Corporate Social Responsibility and Environmental Management* **28**(5), 1426–1445.

Bolton, P., Despres, M., Pereira da Silva, L. A., Samama, F. and Svartzman, R. (2020), *The green swan*, Bank for International Settlements.
URL: *https://EconPapers.repec.org/RePEc:bis:bisbks:31*

Bouyé, E. and Menville, D. (2020), 'The convergence of sovereign environmental, social and governance ratings', *Social and Governance Ratings (December 21, 2020)* .

Brandon, R. G., Krueger, P. and Schmidt, P. S. (2021), 'Esg rating disagreement and stock returns', *Financial Analysts Journal* **77**(4), 104–127.
URL: *https://doi.org/10.1080/0015198X.2021.1963186*

Clark, G. L., Feiner, A. and Viehs, M. (2015), 'From the stockholder to the stakeholder: How sustainability can drive financial outperformance', *Available at SSRN 2508281* .

De Lucia, C., Pazienza, P. and Bartlett, M. (2020), 'Does good esg lead to better financial performances by firms? machine learning and logistic regression models of public enterprises in europe', *Sustainability* **12**(13), 5317.

Dunz, N., Emambakhsh, T., Hennig, T., Kaijser, M., Kouratzoglou, C. and Salleo, C. (2021), 'Ecb's economy-wide climate stress test', *ECB Occasional Paper* (2021/281).

ECB (2020), 'Guide on climate-related and environmental risks. supervisory expectations relating to risk management and disclosure'.

Ehlers, T., Elsenhuber, U., Jegarasasingam, A. and Jondeau, E. (2022), Deconstructing ESG scores: how to invest with your own criteria, BIS Working Papers 1008, Bank for International Settlements.
URL: *https://ideas.repec.org/p/bis/biswps/1008.html*

Ehlers, T., Elsenhuber, U., Jegarasasingam, A. and Jondeau, E. (2023), 'Deconstructing esg scores: How to invest with your own criteria'.

Friede, G., Busch, T. and Bassen, A. (2015), 'Esg and financial performance: aggregated evidence from more than 2000 empirical studies', *Journal of sustainable finance & investment* **5**(4), 210–233.

GSIA (2020), 'Global sustainable investment review'.

Hastie, T., Tibshirani, R., Friedman, J. H. and Friedman, J. H. (2016), *The elements of statistical learning: data mining, inference, and prediction. Second Edition*, Springer Series in Statistics.

Ioannou, I. and Serafeim, G. (2017), 'The consequences of mandatory corporate sustainability reporting', *Harvard Business School research working paper* (11-100).

IPCC (2023), Urgent climate action can secure a liveable future for all, Technical report.
URL: *https://www.ipcc.ch/2023/03/20/press-release-ar6-synthesis-report/*

Jebe, R. (2019), 'The convergence of financial and esg materiality: taking sustainability mainstream', *American Business Law Journal* **56**(3), 645–702.

Lanza, A., Bernardini, E. and Faiella, I. (2020), 'Mind the gap! machine learning, esg metrics and sustainable investment', *Machine Learning, ESG Metrics and Sustainable Investment (June 26, 2020). Bank of Italy Occasional Paper* (561).

Larcker, D. F., Pomorski, L., Tayan, B. and Watts, E. M. (2022), 'Esg ratings: A compass without direction', *Rock Center for Corporate Governance at Stanford University Working Paper Forthcoming* .

Lee, L.-E., Giese, G. and Nagy, Z. (2020), 'Combining e, s, and g scores: An exploration of alternative weighting schemes', *The Journal of Impact and ESG Investing* **1**(1), 94–103.

NGFS (2019), First comprehensive report: A call for action, Technical report.

NGFS (2020), 'Guide for supervisors. integrating climate-related and environmental risks into prudential supervision'.

NGFS (2022), Final report on bridging data gaps, Technical report.

OECD (2022), 'Esg ratings and climate transition. an assessment of the alignment of e pillar scores and metrics', *OECD Business and Finance Policy Papers* (06).
URL: *https://www.oecd-ilibrary.org/content/paper/2fa21143-en*

Pagano, M. S., Sinclair, G. and Yang, T. (2018), Understanding esg ratings and esg indexes, *in* 'Research handbook of finance and sustainability', Edward Elgar Publishing.

Papadopoulos, G. (2022), Discrepancies in corporate ghg emissions data and their impact on firm performance assessment, Technical report, Joint Research Centre, European Commission.

Schumacher, K. (2021), 'We should not equate awareness or passion with subject matter expertise', *Sustainability & ESG News Central Europe* .
URL: *https://sustainabilitynews.eu/dr-kim-schumacher-on-esg-competence-greenwashing-we-should-not-equate-awareness-or-passion-with-subject-matter-expertise/*

Signorini, L. F. (2020), Sustainable investment in uncertain times: The future of public sector asset management, *in* 'OMFIF Roundtable for Public Sector Asset Managers London, 6 February 2020'.

SSGA (2017), Technical report.

Teng, X., Wang, Y., Wang, A., Chang, B.-G. and Wu, K.-S. (2021), 'Environmental, social, governance risk and corporate sustainable growth nexus: Quantile regression approach', *International Journal of Environmental Research and Public Health* **18**(20), 10865.

TNFD (2023), The tnfd nature-related risk and opportunity management and disclosure framework final draft – beta v0.4, Technical report, Taskforce on Nature-related Financial disclosures.
URL: *https://framework.tnfd.global/publications/*

UNEP (2021), Race to net zero campaign, Technical report.

Visco, I. (2021), 'The g20 presidency programme on sustainable finance'.
URL: *https://www.bancaditalia.it/pubblicazioni/interventi-governatore/integov2021/Visco_30092021.pdf*

Zhang, D. and Xie, Y. (2022), 'Customer environmental concerns and profit margin: Evidence from manufacturing firms', *Journal of Economics and Business* **120**, 106057.
URL: *https://www.sciencedirect.com/science/article/pii/S0148619522000133*

A Environmental raw variables

Table 1. Variables definitions, the VIF, and sources.

Variables	Symbol	VIF factor	Source
CO2 Equivalent Emissions Indirect, Scope 2	Sc_1	26561,0	Datastream
CO2 Equivalent Emissions Direct, Scope 1	Sc_2	6406,5	Datastream
Total CO2 Equivalent Emissions To Revenues	CO2_intens	3157,5	Datastream
Policy Emissions	Pl_emis	38,9	Datastream
Targets Emissions	Trgt_emis	9,6	Datastream
Emissions Trading	Em_trad	3,6	Datastream
NOx and SOx Emissions Reduction	N_S_red	2,8	Datastream
Particulate Matter Emissions Reduction	Prtic_red	2,0	Datastream
Internal Carbon Pricing	Int_carb_px	1,2	Datastream
Resource Reduction Policy	Res_red_Pl	232,0	Datastream
Environmental Supply Chain Management	E_sup_ch_man	33,7	Datastream
Policy Environmental Supply Chain	Pl_E_supp_chain	27,6	Datastream
Resource Reduction Targets	Res_red_tgt	23,1	Datastream
ENVIRON_QUAL_MGT	E_qual_mgt	17,1	Bloomberg
Environment Management Training	E_mgt_train	12,0	Datastream
CLIMATE_CHG_POLICY	CC_Pl	11,9	Bloomberg
Environment Management Team	E_mgt_team	10,4	Datastream
Climate Change Commercial Risks Opportunities	C_opp	9,9	Datastream
Environmental Partnerships	E_prtsh	7,8	Datastream
Staff Transportation Impact Reduction	St_trasp_red	2,9	Datastream
Env Supply Chain Partnership Termination	E_sup_ch_term	2,7	Datastream
Policy Sustainable Packaging	Pl_sust_pack	2,7	Datastream
Green Buildings	Green_build	2,6	Datastream
Environmental Restoration Initiatives	E_rest	2,6	Datastream
CLIMATE_RISKS	CC_risk	2,0	Bloomberg
CSR_SUSTAINABILITY_COMMITTEE	Csr_sus_com	2,0	Bloomberg
Land Environmental Impact Reduction	Land_E_red	1,6	Datastream
Environmental Project Financing	E_proj	1,1	Datastream
Energy Use Total	En_use	59246,0	Datastream
Total Energy Use To Revenues USD in million	En_intens	5587,3	Datastream
Energy Purchased Direct	En_pur	182,5	Datastream
ENERGY_EFFICIENCY_POLICY	En_effic_Pl	158,9	Bloomberg
Policy Energy Efficiency	Pl_En	94,1	Datastream
Targets Energy Efficiency	Tgt_En	15,0	Datastream
Renewable Energy Use	Rnw_En	13,0	Datastream
ENERGY_INTENSITY_PER_ASSETS	En_intens_ass	2,5	Bloomberg
Nuclear	Nucl	2,0	Datastream
Product Environmental Responsible Use	Prd_E_use	23,1	Datastream
Environmental Products	E_pro	16,4	Datastream
Environmental Expenditures Investments	E_exp_inv	7,0	Datastream
Environmental Investments Initiatives	E_inv	4,4	Datastream
Renewable/Clean Energy Products	Rw_cl_en_prd	2,9	Datastream
Eco-Design Products	Eco_pro	2,8	Datastream
Noise Reduction	Nois_red	2,5	Datastream
Hybrid Vehicles	Hyb_veh	2,2	Datastream
GREEN_BUILDING	Gre_build	2,2	Bloomberg
Sustainable Building Products	Sus_build	1,9	Datastream
Water Technologies	Wat_tech	1,8	Datastream
Organic Products Initiatives	Org_pro	1,5	Datastream
Waste Reduction Initiatives	Was_red	28,5	Datastream
Environmental Materials Sourcing	E_mat_sourc	7,1	Datastream
Toxic Chemicals Reduction	Tox_che_red	2,7	Datastream
Take-back and Recycling Initiatives	Recyc_ini	2,5	Datastream
VOC Emissions Reduction	VOC_red	2,4	Datastream
e-Waste Reduction	e_W_red	1,7	Datastream
Water Withdrawal Total	Wat_use	624,8	Datastream
Water Use To Revenues USD in million	Wat_intens	289,2	Datastream
Policy Water Efficiency	Pl_wat	9,7	Datastream
BIODIVERSITY_POLICY	Bio_Pl	7,2	Bloomberg
Biodiversity Impact Reduction	Bio_red	6,7	Datastream
Targets Water Efficiency	Tgt_wat	3,5	Datastream
Animal Testing	Anim_test	1,7	Datastream

B Methodologies applied for dissecting E-scores

B.1 Results of quantile regression

Quantile regression has been performed over the period 2015-2021 for seven quantiles simultaneously: 5, 20, 25, 50, 75, 80 and 95% (see Table 1). Except for Provider-G, the number of significant variables across quantiles is on average in the range of 10 to 19 across the providers. The larger the number of variables, the wider the dispersion across quantiles. Notably more variables are identified as significant for the lower quantile. For instance, for Provider-A and Provider-E we record an increase of more than a half in the number of variables for the 5-th lower quantile compared to the 95-th quantile. This result suggests that to discriminate the environmentally worst-performing firms, a larger number of data points are needed. When we consider the 80-20 quantile - which is often used by investors for *best-in-class* and *exclusion* strategies - this consideration holds across four data providers (Provider-A, Provider-E, Provider-F and Provider-C).

Table 1. No. of significant variables in quantile regressions.

τ	Provider-A	Provider-B	Provider-C	Provider-D	Provider-E	Provider-F	Provider-G
5%	22	10	13	16	18	12	3
20%	20	10	16	12	19	16	3
25%	22	10	15	14	19	16	3
50%	22	9	13	14	16	15	3
75%	19	10	15	14	14	15	3
80%	17	10	13	15	16	14	3
95%	14	9	14	15	12	15	3
Average	19	10	14	14	16	15	3

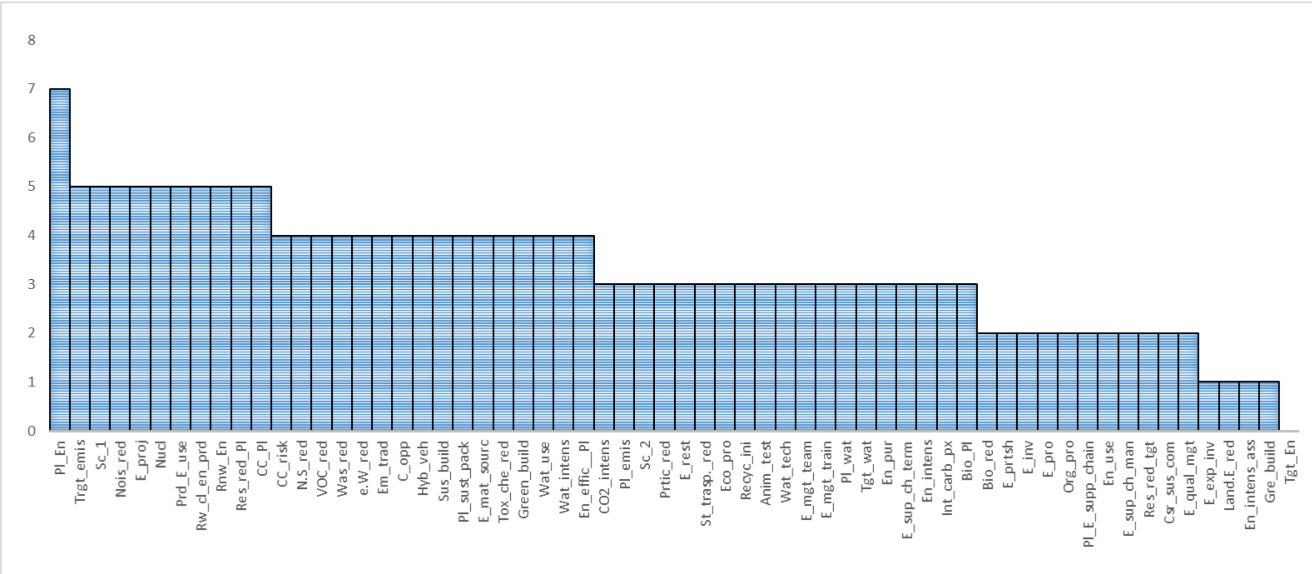
The pseudo R^2 in Table 2 computed for the pooled quantile regressions point at the higher explanatory power for the E-scores of those providers where more variables were found significant (Provider-A, Provider-C and Provider-E).

Table 2. Pseudo R^2 for quantile regressions.

τ	Provider-A	Provider-B	Provider-C	Provider-D	Provider-E	Provider-F	Provider-G
5%	52%	33%	36%	29%	27%	25%	60%
20%	48%	26%	30%	24%	21%	23%	50%
25%	47%	26%	30%	24%	20%	22%	49%
50%	42%	22%	26%	24%	20%	20%	52%
75%	38%	16%	24%	21%	24%	16%	51%
80%	36%	15%	24%	20%	25%	16%	48%
95%	28%	8%	34%	13%	30%	23%	44%
Average	42%	21%	29%	22%	24%	21%	51%

Furthermore, some variables are found meaningful across different E-scores, as shown in Figure 2. Noticeably, both current and forward-looking indicators, such as those related to environmental targets (for emissions or water use), policies for reducing resource use and waste, energy intensity and efficiency, are identified among the most relevant variables for several E-scores. Arguably, the coefficients of sectoral dummies are not very meaningful since the carbon emission variables can explain part of the sector-specific variance. Regarding non-environmental variables, we exclude from the investigation financial indicators commonly used for fundamental analysis (e.g. EBITDA margin, financial leverage, ROE and ROI) because they do not seem directly related to the environmental performance, while rather an inverse relation is found by Zhang and Xie (2022) and Andriana and Anisykurlillah (2019). Besides, we do not include any autoregressive component to investigate the persistence of the E-score, neither we investigate how the judgmental component could predict future changes of raw data. This investigation could warrant a follow-up research.

Figure 2. Quantile regression - variables frequently found relevant across E-scores.



Note: the figure sort the variables according to their significance across the seven estimated regression equations. The leftmost variables are those found significant in most regressions.

Finally, estimated coefficients are relatively similar for most of the common variables, such as the presence of emission targets, climate change policies and climate risks disclosure, and renewable energy use. Some remarkable differences arise for variables such as sustainable buildings, environmental projects, policies for energy efficiency and resource reduction and staff transportation impact reduction (Figure 3).

B.2 Joint analysis of results from quantile and Lasso regression

Looking jointly at the results of both regression techniques, the number of regressors found per each provider varies significantly between the quantile regression and Lasso regression (see Figure 4). For five out of seven providers, the two techniques identify a similar number of significant variables while for Provider-E there is a remarkable difference (16 versus 3) and R^2 (24% versus 3%). Both techniques give consistently the most promising results for Provider-A, with the largest number of variables and (by construction) the highest explanatory power. The R^2 for both techniques is on average around 20% for Provider-C, Provider-F and Provider-B. When the number of indicators decreases, the explained part of E-scores diminishes accordingly (with R^2 between 29% with 14 variables for Provider-C and 21% with ten variables for Provider-B). Lower performance is found for Provider-D showing an R^2 of 13% with 8 variables with Lasso regression. As a general remark, since the identified indicators cannot largely explain the environmental score, we can assume that the environmental assessment of the providers substantially depends on unknown factors, arguably related to qualitative judgment, as discussed in Section 4.1.

The most ambiguous result is found for Provider-G (with a high R^2 upfront just 3 significant variables in estimated quantile regression). This evidence and the very low correlation with

Figure 3. Quantile regression - average variables' coefficients per E-score.

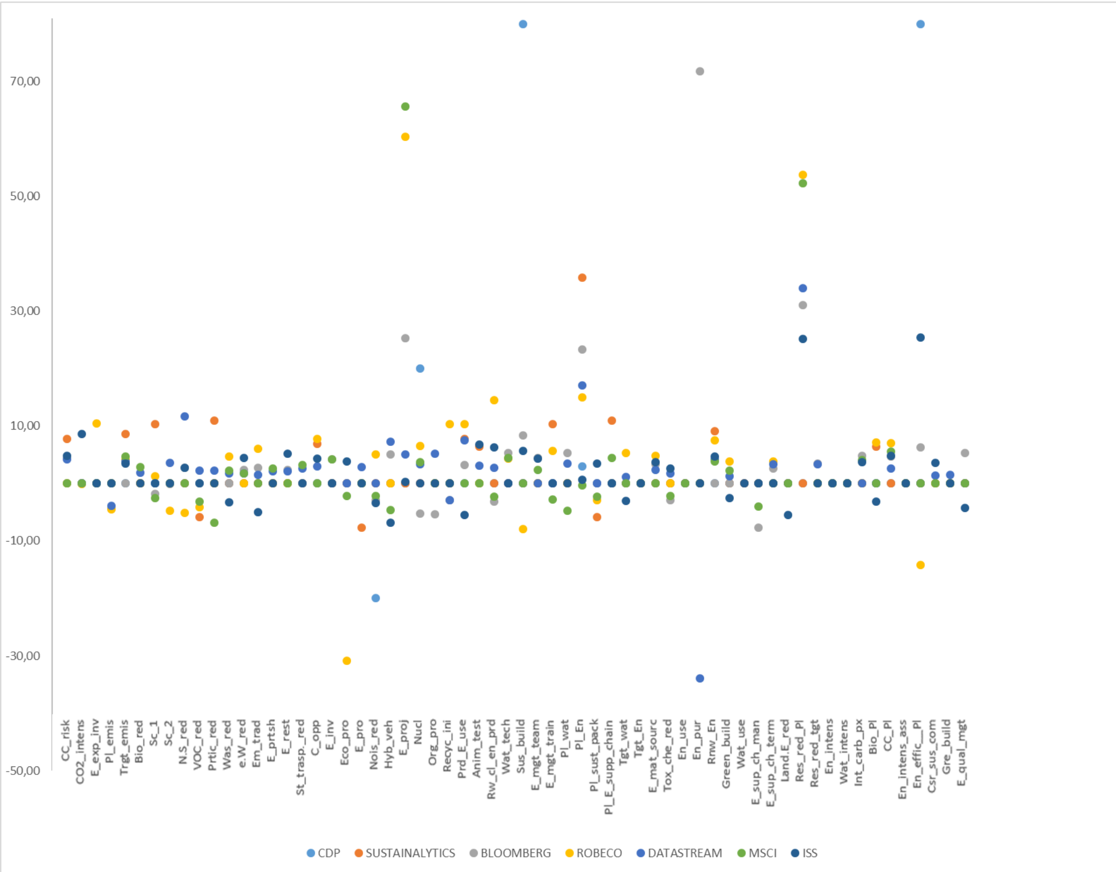
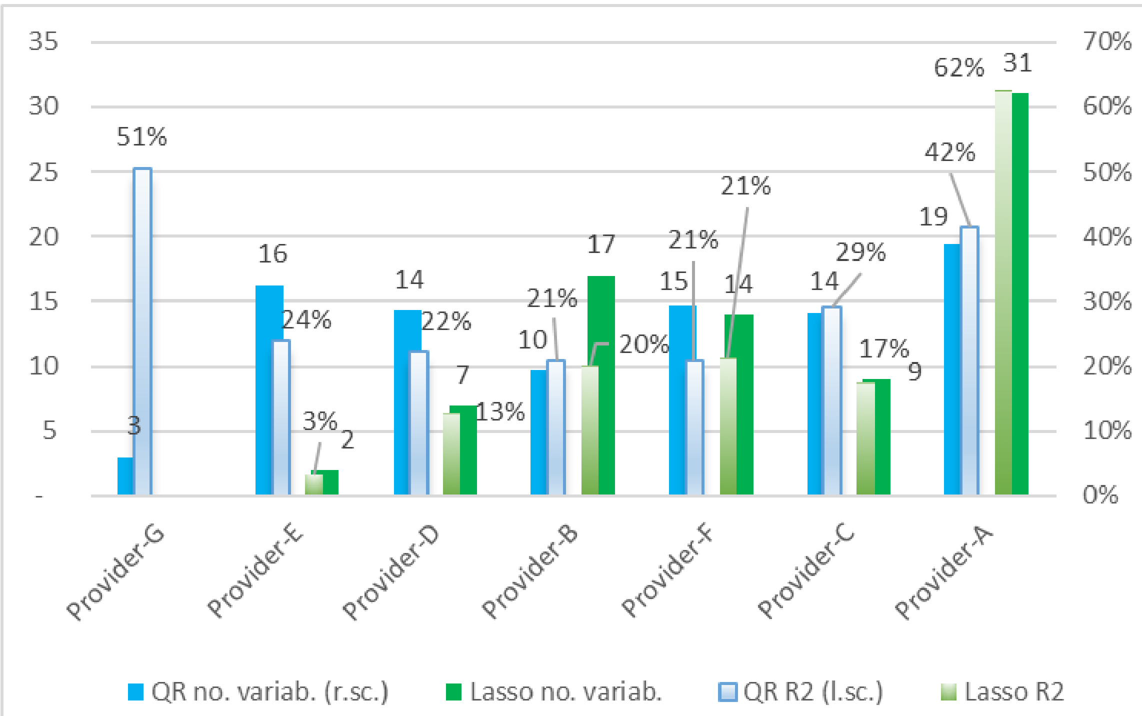


Figure 4. No. of regressors and R^2 per E-score.



other providers, could be due to provider's specific focus on climate rather than on a broader environmental assessment. These considerations suggest excluding Provider-G from the analysis.

B.3 Latent variable estimation

The relatively low R^2 observed in the quantile and Lasso regression estimations calls for further investigation relative to the unexplained components of the regression. Indeed, we conjecture that a relevant variable, which has been omitted in the regression equation because of its hidden nature, should be included in the analysis to have a complete picture of the relationship between E-scores and their underlying elements. We may interpret this hidden latent variable as providers' judgemental component which is not measured by quantitative raw data although it affects significantly the E-scores. We use the Kalman filter to estimate the state-space model for each provider as follows:

$$X_t = fX_{t-1} + w_t \tag{2}$$

$$Y_t = Z_t D + hX_t + v_t \tag{3}$$

where X_t is the recursive hidden variable for each provider (judgemental component) at time t, Y_t is the E-score of each provider at time t and Z_t is the vector of 62 explanatory variables at time t employed in the regression approaches and common for all providers. The vector D contains the Lasso-estimated coefficients for each provider. Therefore, $Y_t - DZ_t$ are the Lasso-estimated residuals which we seek to explain with the latent component. Finally, we assume $w_t \sim \mathcal{N}(0, Q_t)$ and $v_t \sim \mathcal{N}(0, R_t)$. After having set all standard parameter restrictions on the autoregressive coefficients and the variances, we estimate the model with the Kalman filter to identify the judgemental component variable.