BANCA D'ITALIA
EUROSISTEMA

Mercati, infrastrutture, sistemi di pagamento

(Markets, Infrastructures, Payment Systems)

Environmental data and scores: lost in translation

by Enrico Bernardini, Marco Fanari, Enrico Foscolo, and Francesco Ruggiero

October 2024

Number 51

BANCA D'ITALIA
EUROSISTEMA

Mercati, infrastrutture, sistemi di pagamento

(Markets, Infrastructures, Payment Systems)

Environmental data and scores: lost in translation

by Enrico Bernardini, Marco Fanari, Enrico Foscolo, and Francesco Ruggiero

Number 51 – October 2024

The papers published in the 'Markets, Infrastructures, Payment Systems' series provide information and analysis on aspects regarding the institutional duties of the Bank of Italy in relation to the monitoring of financial markets and payment systems and the development and management of the corresponding infrastructures in order to foster a better understanding of these issues and stimulate discussion among institutions, economic actors and citizens.

The views expressed in the papers are those of the authors and do not necessarily reflect those of the Bank of Italy.

The series is available online at www.bancaditalia.it.

Printed copies can be requested from the Paolo Baffi Library: richieste.pubblicazioni@bancaditalia.it.

ISSN 2724-6418 (online)
ISSN 2724-640X (print)

Banca d'Italia
Via Nazionale, 91 - 00184 Rome - Italy
+39 06 47921

Designed and printing by the Printing and Publishing Division of the Bank of Italy

Environmental data and scores: lost in translation

by Enrico Bernardini,* Marco Fanari, Enrico Foscolo,** and Francesco Ruggiero****

Abstract

This paper addresses the methodological issues and limited coverage of environmental scores, which are widely used by financial institutions and policymakers. Its contribution is twofold. First, regression analysis highlights significant deviations among the environmental scores of seven providers compared to the granular data published by companies, despite some environmental indicators play a significant role in the scores of multiple providers. The residual component of the regression analysis varies across providers, presumably because some pay more attention to the environmental impact itself, while others place greater emphasis on the related financial risks. Second, we propose a classification system based on granular data, which can also be applied to assess unrated companies and to implement investment strategies, such as best-in-class and exclusion. The resulting portfolios have similar environmental and financial profiles to those based on providers' scores. The paper underscores the importance of improving corporate disclosure on granular data and transparency on providers' methodologies to foster sustainable finance development.

JEL Classification: G32, C55, G11, Q56.

Keywords: Climate risk, Environmental Scores, Machine Learning, Portfolio analysis, Sustainable investment.

Sintesi

Il lavoro affronta i problemi metodologici e la scarsa copertura dei punteggi ambientali delle società non finanziarie, ampiamente utilizzati da intermediari finanziari e istituzioni. Il contributo è duplice. In primo luogo, un'analisi di regressione mette in luce ampi scostamenti dei punteggi di sette fornitori professionali rispetto ai dati granulari pubblicati dalle imprese, sebbene alcuni indicatori ambientali giochino un ruolo rilevante nel determinare i punteggi di più fornitori. La componente residuale dell'analisi di regressione varia tra i diversi fornitori, presumibilmente in conseguenza del fatto che alcuni prestano maggiore attenzione all'impatto ambientale in quanto tale mentre altri pongono maggior enfasi sui connessi rischi finanziari. In secondo luogo, si propone un sistema di classificazione basato su dati granulari, applicabile anche a società non valutate dai fornitori e utilizzabile per realizzare strategie di investimento, quali *best-in-class* e *exclusion*. I portafogli generati mostrano caratteristiche ambientali e finanziarie in linea con quelle dei portafogli costruiti con i punteggi dei fornitori. Il lavoro sottolinea che, al fine di favorire lo sviluppo della finanza sostenibile, è importante migliorare sia la divulgazione dei dati granulari delle imprese sia la trasparenza sulle metodologie adottate dai fornitori per la propria valutazione ambientale.

---

* Bank of Italy, Climate Change and Sustainability Hub.

** Bank of Italy, Financial Risk Management Directorate.

CONTENTS

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, regulators and supervisors have become increasingly aware of such risks, as testified by the initiatives in the EU (e.g. the EU green taxonomy) and at the international level (e.g. the 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 growing interest from policymakers, public institutions and investors. Meaningful initiatives have been taken, for instance, by some 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 scores of specialised providers are widely used to assess environmental and climate risks. However, such scores usually refer to broader sustainability issues, combining environmental, social and governance profiles (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 total assets under management (GSIA, 2022). In general, sustainability scores represent essential information to channel capital towards transition investment in the long term. Despite their pivotal role, the ESG scores are far from transparent (Angelini, 2022). Without clear and shared methodologies for sustainability assessment, there is a significant risk of greenwashing, which can turn into regulatory arbitrage or misrepresentation to investors. Moreover, decision-makers on sustainable investment need help reducing reliance on third-party ESG scores since most investors need help gathering and assessing sustainability metrics or need more analytical skills to make their own ESG assessment (Schumacher, 2021). The weak scrutiny of ESG scores in the financial markets may be due to scant incentives within the long value chain of ESG services, arising from the limited room of manoeuvre for passive managers as well as the high competition among index providers (Pagano et al., 2018).

Nowadays, a better understanding of the analytical tools used to assess environmental and climate risks and opportunities is 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 focuses on 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 pillar offers broader coverage of firms and more extensive data points than the Social and Governance pillars within the ESG fields. The paper offers a rich data set - unique to our knowledge - of environmental scores and raw data from seven specialized providers over seven years.

This study deals with two main research questions. The first concern is the extent to which the E-scores rely either on quantitative data or qualitative judgement of the ESG providers. To gauge the contribution of the two components, we apply traditional econometric methods (i.e. fixed effects quantile panel regression) and machine learning techniques (i.e. lasso regression). According to both approaches, the number of meaningful variables is limited, such as reduction targets for emissions, waste, and resource use, as well as partnerships for reducing the environmental

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 a data set kindly provided by Prof. Monica Billio and the team of the University of Venice - Ca' Foscari within its project on ESG data analysis. We thank Stefano Siviero, Ivan Faiella, 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.

impacts and policies for renewable energy and green buildings. Moreover, the outcomes hint at the possibility that a qualitative component plays a significant role. We explore this hidden component through a Kalman filter and a particle filter (through a Gibbs sampler as a Monte Carlo Markov Chain algorithm), finding a "judgemental" latent variable that is more material for some providers than others. This evidence highlights the divergences among providers concerning methodologies and the different focus on environmental impacts, financial risks, or both.

The second research question aims to design an environmental assessment relying only on raw data. We train a classification system that allocates each company among three environmental performance classes consistently with the actual grading of providers. The classification rule enables users to extend the environmental assessment to those companies that disclose raw environmental data even though providers do not rate them. Such potential application is connected with the EU Corporate Sustainability Reporting Directive, which will likely broaden the voluntary disclosure of unlisted small and medium-sized enterprises (SMEs) since the 2026 financial year. Moreover, the classification technique allows for circumventing the potentially distorting effects of the unexplained component highlighted in the regression analysis. The three-classes framework also enables the implementation of two common sustainable investment strategies. Specifically, we train the classification system through two techniques applied to each provider's scores, i.e. the Linear Discriminant Analysis (LDA) and the K-Nearest Neighbours algorithm (KNN). We then test the financial results and the environmental profile obtained with the classification based on LDA and KNN in 2017-2021 by simulating portfolios consistent with the so-called *best-in-class* (or positive screening) and *exclusion* (or negative screening) investment strategies. We find that portfolios based on classification rules present risk/return and environmental profiles similar to those built using the providers' scores.

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

2 Literature on ESG scores and environmental indicators

A growing body of literature dissects 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, 2016 and Ehlers et al., 2023). One of the most recent meta-studies 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-third of the studies. It holds especially during a social or economic crisis, while results are indistinguishable in the rest of the studies.[1]

[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. Decarbonisation strategies can 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 corporations to higher operational efficiency, product innovation, and manufacturing processes. Investors who demand a lower risk premium for investing equity and credit in such companies also perceive them as less risky. The combination of these factors brings such superior performance to ESG-leading firms. Ehlers et al. (2023) underline that investors can improve the sustainability of their investments through ESG investing without jeopardising the risk-return profile of their portfolios.

Against this evidence, the divergence among ESG scores of different providers may lead to controversial investment decisions and make them subject to rising criticism. Berg et al. (2022) find an average correlation among global ESG scores of around 60%, compared to the nearly 90% correlation among credit ratings.[2] The divergence of ESG scores can be derived from differences in data sources, the selection of indicators, and assigned materiality. Heterogeneity in definitions and assessment methodologies can prompt such divergences, which investors can misunderstand (Billio et al., 2021). In addition, this heterogeneity produces weak signals in asset pricing and may jeopardise the efficient capital allocation to companies committed to the transition.[3]

Berg et al. (2021) analyzed the impact of ESG performance on firms' stock returns, considering ratings from various data providers and addressing the noise and confusion inherent in ESG ratings. Their findings reveal a positive correlation between a firm's sustainable performance and stock returns. It underlines the importance of integrating multiple ESG ratings in investment decisions despite variations in noise levels. It highlights that, even amidst this variability, the informational value of ESG ratings remains significant. Bams and van der Kroft (2022) delve into the issue of ESG information asymmetry and rating inflation. They emphasize the reliance of sustainable investors on third-party ESG assessments when making investment choices. This reliance stems from the substantial information asymmetry resulting from the lack of standardized information in ESG disclosures. According to this study, firms have incentives to boost their sustainability ratings to lower their cost of capital.

Regulatory or voluntary disclosure may address the lack of sustainability information and significantly increase the firm's 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. Several standard-setting bodies (SASB, IFRS, EFRAG, ISSB) are working to provide critical guidance towards this avenue. However, we must still set 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 economic and financial system. Likewise, the challenges in measuring environmental and climate risks are compelling to understand the environmental data further. In particular, forward-looking data are featured by high levels of uncertainty due to tipping points and complex compounding effects (NGFS, 2019, 2022). In addition, regulators' initiatives underline the importance of measuring and managing these sources of risk (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. OECD (2022) found a disconnect between E-scores and the relevant environmental and climate indicators. It identified areas to improve E-scores' alignment with low-carbon objectives and forward-looking climate measures. It also highlights the importance of effective processes to track and verify data to ensure credibility among market participants. Considering each ESG pillar separately, Lee et al. (2020) show that some E-indicators (e.g. carbon emissions) have long-lasting effects because they tend to accumulate over time, while the G-indicators are more influential in the short term. Papadopoulos (2022) reports discrepancies in greenhouse gas emission data among providers across time and sectors. Such discrepancies in-

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

[3]Heterogeneity in raw data is also apparent within the green bond market. Post-issuance disclosures in this market vary regarding the granularity of reported information. There is a meaningful and positive size bias, as large-cap issuers tend to disclose more detailed information.

crease from direct emissions (Scope 1) to indirect emissions (Scope 2 and 3). They can translate into diverging assessments of carbon performance, affecting E-pillar scores and ESG ratings. A growing body of academic research explores the intersection of machine learning methods and ESG ratings. For instance, D'Amato et al. (2022) use deep learning algorithms to forecast companies' ESG scores using balance sheet data. In contrast, De Lucia et al. (2020) investigate the opposite, predicting companies' Return on Equity (ROE) and Return on Assets (ROA) from ESG indicators. Likewise, Zanin (2022) examines the impact of firms' ESG scores on credit ratings, employing both statistical and machine learning methods.

Like our work, Del Vitto et al. (2023) develop machine learning algorithms capable of replicating and predicting ESG scores and their components. They discover that E-scores appear to be strongly influenced by a few common variables, such as resource reduction policies and levels of CO2 equivalent emissions. However, they also find evidence of persistent unlearnable noise that even more complex models can not eliminate. Our study delves into the Environmental pillar of ESG assessments and examines environmental data sourced from seven providers across several years for a data set of European companies. In contrast, the OECD (2022) investigates a sample of nearly 2,500 companies worldwide, using data from four providers over one year. Del Vitto et al. (2023) also analyse data and scores from a single provider, encompassing approximately 6,400 companies within one year.

3 Sources and data description

Our analysis embraces the European listed equities belonging to the Euro Stoxx index[4], 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 mainly 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, representing 86% of the current market capitalization as of December 31, 2021 (see Table 1).

The environmental scores and raw data are sourced from seven data providers: MSCI ESG, Bloomberg, RobecoSAM/S&P Global, ISS-Institutional Shareholder Services, Carbon Disclosure Project (CDP), Sustainalytics and Datastream-Reuters-Asset4. While for CDP we observe the lowest total coverage percentage of its E-score (27.9%), for Sustainalytics, Bloomberg, and Robeco the percentages of the relevant E-scores exceed 60% (66.6%, 63.8%, and 61.8%, respectively). Only MSCI, Datastream-Reuters and ISS reach 90% (93.5%, 96.1%, and 91.5%, respectively). The picture is similar at the sector level, where the lowest coverage percentages are found for the first four mentioned providers (CDP, Sustainalytics, Bloomberg, and Robeco; see Figure 2, Appendix B). Companies in the healthcare and information technology sectors are covered less by Bloomberg and CDP, respectively.

Some providers only offer E-scores, while others offer scores and raw data. Our analysis takes raw data from Bloomberg, Datastream-Asset4 and CDP. They provide 209 environmental raw indicators, whose coverage is heterogeneous across companies: for only one-third of the variables,

[4]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 in 11 euro area countries: Austria, Belgium, Finland, France, Germany, Ireland, Italy, Luxembourg, the Netherlands, Portugal, and Spain.

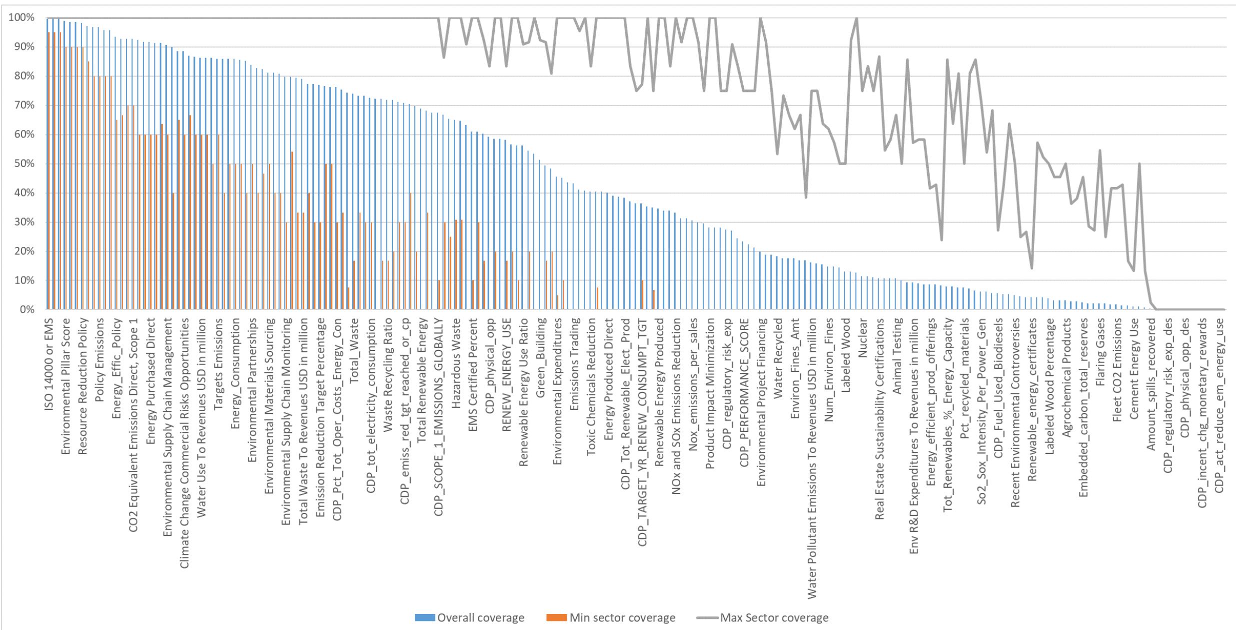
Table 1. Corporates sample – sector and market cap weights.

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

the coverage is higher than 70% (see Figure 1).[5] This evidence highlights the criticality of data gaps, which hampers the environmental assessment of corporate sustainability from specialised evaluators and investors.

From the initial 209 environmental raw variables, we select those with average coverage percentages among sectors greater than 80% (see Figure 2, Appendix B), considering a time window from 2015 to 2021.

Figure 1. Coverage of environmental variables.



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

Besides the coverage, we exclude other variables using two additional criteria. First, we exclude the highly collinear variables detected by their cross-correlation to reduce information redundancy,

[5] Among the starting data set of 209 variables, 125 are taken from Datastream, 49 from Bloomberg and 35 from CDP. The time series of CDP data is shorter than that of the other providers.

Table 2. Grouping of 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

gauged by the Variance Inflation Factor (VIF). Second, the variables that show a low correlation with the E-scores are excluded to improve sample efficiency.[6] Based on these criteria, we finally select 62 variables (8 sourced by Bloomberg, the remaining by Datastream), which can be classified into six groups of environmental critical issues (summarised in Table 2). Most importantly, a relevant number of data points refers to binary variables (e.g. *yes* or *no* for adopting environmental policies or carbon reduction commitments). Among 62 selected variables, only nine are continuous.

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

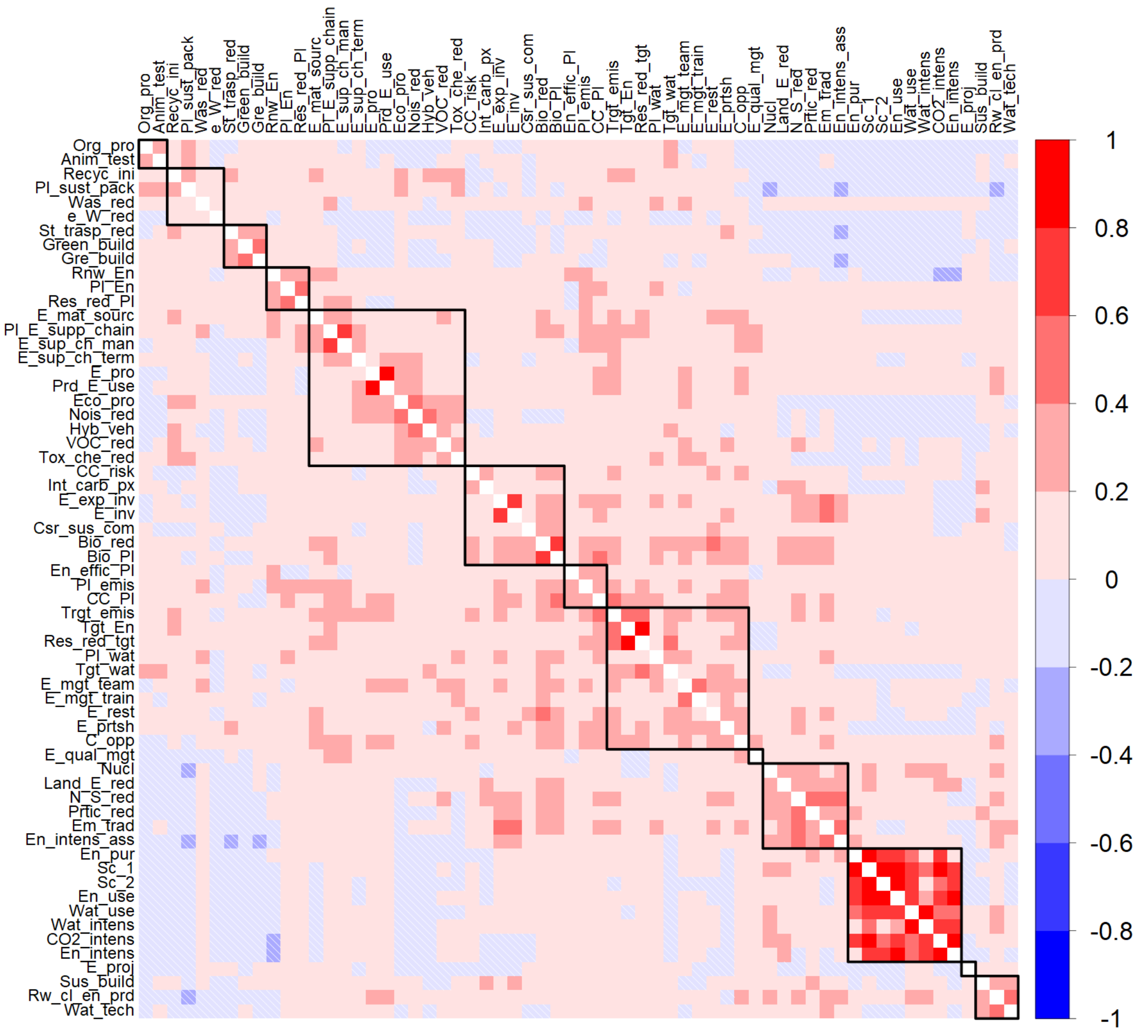
3.1 Descriptive statistics of the environmental variables and scores

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 the variables related to carbon intensity, energy and water use.

Conversely, the E-scores of the seven data providers display a linear correlation between 0.14 and 0.46 over 2015-2021, with an average of 0.28 (see Figure 3). Notably, the correlation is lower

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

Figure 2. Cross-correlation of environmental variables.



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

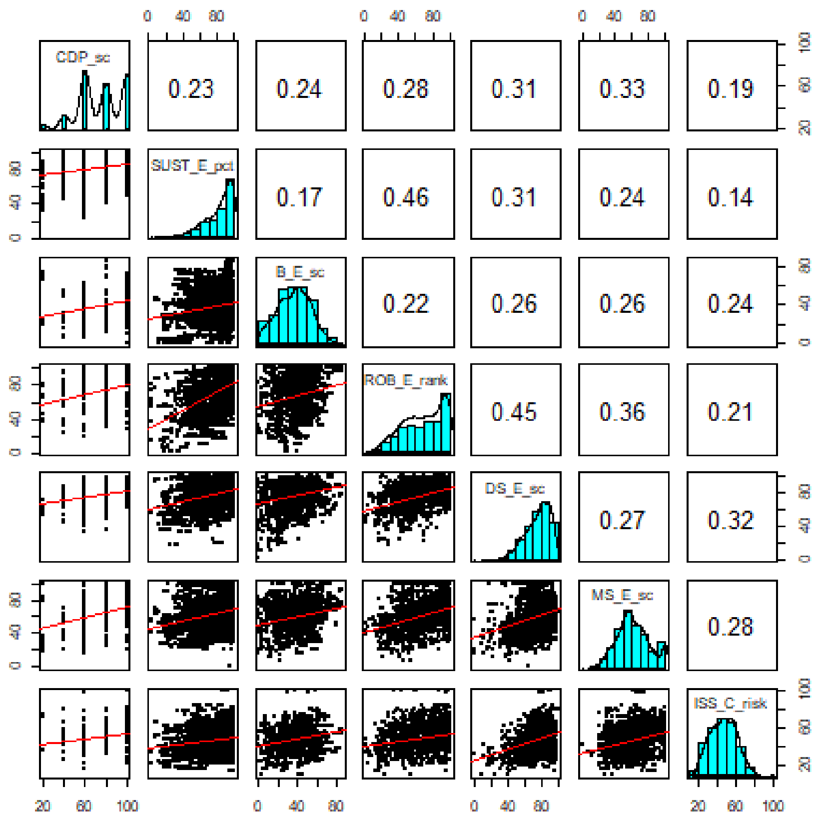
than that among credit ratings of the same corporate sample as of June 2022, which falls in the range of 0.7–0.9 (Table 3).

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 = MSCI ESG; ISS sc = ISS carbon risk rating.

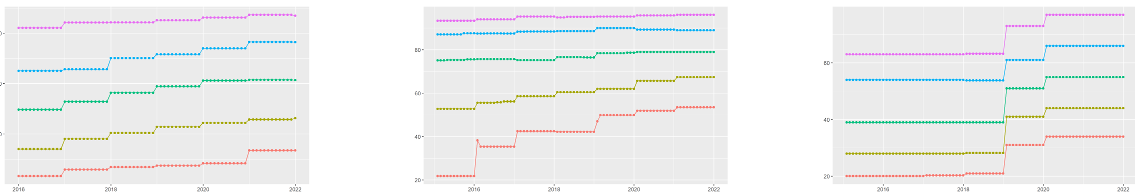
A higher pairwise correlation is found between Sustainalytics and Robeco (0.46), Robeco and Datastream (0.45), and MSCI and Robeco (0.38). The average pairwise correlation for each provider is particularly low for Bloomberg and ISS, while it is higher for Robeco and MSCI (see Table 4). Specifically, we identify different patterns in the distribution of E- scores. The distribution of the scores from Bloomberg, MSCI ESG and ISS is nearly normal-shaped, 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 depend on the underlying provider's assessment methodologies. Brandon et al. (2021) underline that Sustainalytics and MSCI aim to provide ratings on ESG performance. Meanwhile, other 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.

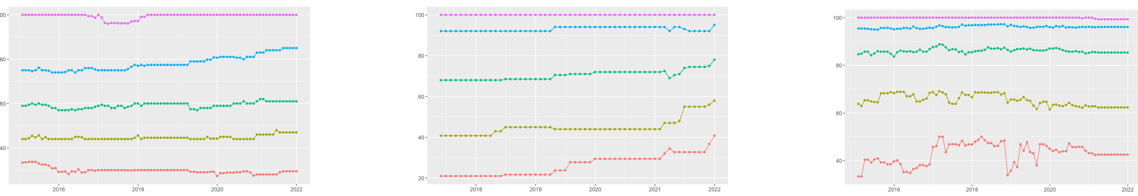
Provider	Mean
Robeco	0.33
MSCI	0.30
CDP	0.29
Datastream	0.28
Sustainalytics	0.23
Bloomberg	0.22
ISS	0.22

Analysing the changes over time (see Figure 4), we record an average E-score increase across the selected quantiles (from 5% to 95%), especially for lower quantiles among most providers (Bloomberg, Datastream, ISS, Robeco). This pattern is consistent across the whole sample, including the Covid period. It may be arguably 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. Nonetheless, we do not have evidence to exclude that the increase in the average E-scores observed over time could partly reflect more benevolent conduct on the providers' side, potentially dictated by growing competition in the industry.

Figure 4. Quantile average scores per provider.



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



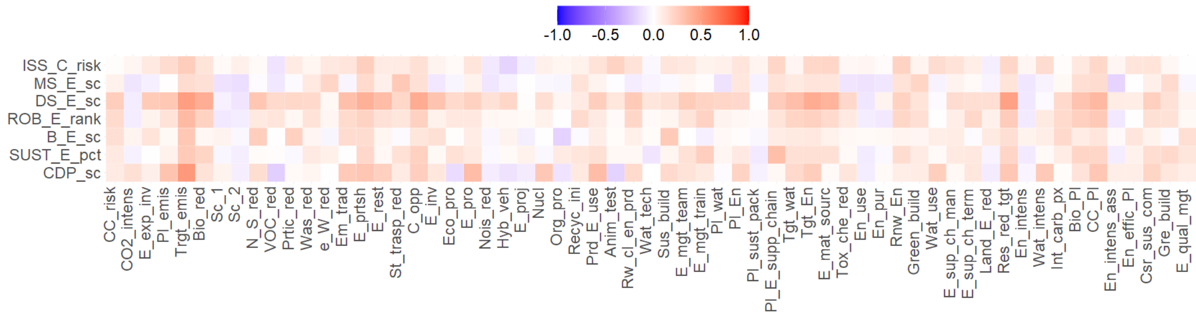
(d) MSCI **(e)** Robeco **(f)** Sustainalytics

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 a first signal of their relationship, we find that some providers (i.e. Datastream, CDP, Robeco and MSCI) rely more on raw environmental data (Figure 5).

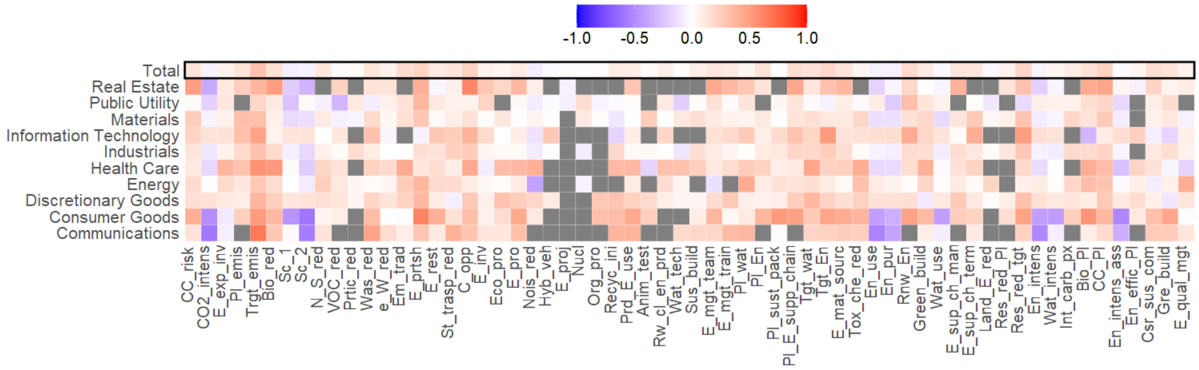
Looking at the sectoral breakdown of the correlation between E-scores and raw data (Figure 6), we point out that some raw variables appear more relevant for the E-scores in specific sectors. For

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



Note: The coloured boxes represent median annual pairwise Pearson correlation coefficients. The description of the raw data is shown in Table 1, Appendix A.

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



Note: The coloured boxes represent median annual pairwise Pearson correlation coefficients. The grey boxes stand for missing outcomes. The description of the raw data is shown in Table 1, Appendix A. Evidence on the overall sectoral coverages are reported in Figure 2, Appendix B.

instance, Scope 1 and 2 emissions (`Sc_1` and `Sc_2` in the plot) negatively correlate with E-scores in most of the sectors. At the same time, their explanatory power softens in the average correlation. Regardless of the provider or sector, the variables that exhibit higher average correlation with E-scores are *Target Emissions*, *Environmental Partnerships*, *Resource Reduction Targtes*, and *Biodiversity Impact Reduction*.

To summarise, 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 raw data reflecting, to some extent, some differences in the specialisation of the providers. We explore such a nexus in the following Section. To avoid any possible quality 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[7] and quantile[8] 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, it signals the relevant variables, which is crucial to endow ourselves with an implicit rule to detect and distinguish green issuers from less green ones. On the other hand, the regularisation imposed by lasso reduces the risk of overfitting. Therefore, we aim to obtain meaningful variable selection using such 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 at time 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 sum of all the j coefficients in absolute value 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.

For each provider, we run a standalone 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 can not perform sector-specific regression analysis due to the scarcity of data for specific 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 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 providers. The zero coefficients filter out the redundant variables. We also conduct a sensitivity analysis of the tuning parameter, and details are provided in Section C.2 of the Appendix. We leave Provider-G out as we can not find significant variables for this provider due to the scarcity of data. The highest explanatory power is recorded for Provider-A (63%), with the largest number of significant coefficients (i.e. 33). In comparison, for Provider-D, Provider-B, and Provider-F, the R^2 equals 30%, 26%, and 23%, while the relevant coefficients cut down to 28, 20, and 16, respectively. We find the lowest number of significant coefficients for Provider-E with only seven significant coefficients and R^2 equal to 8%.

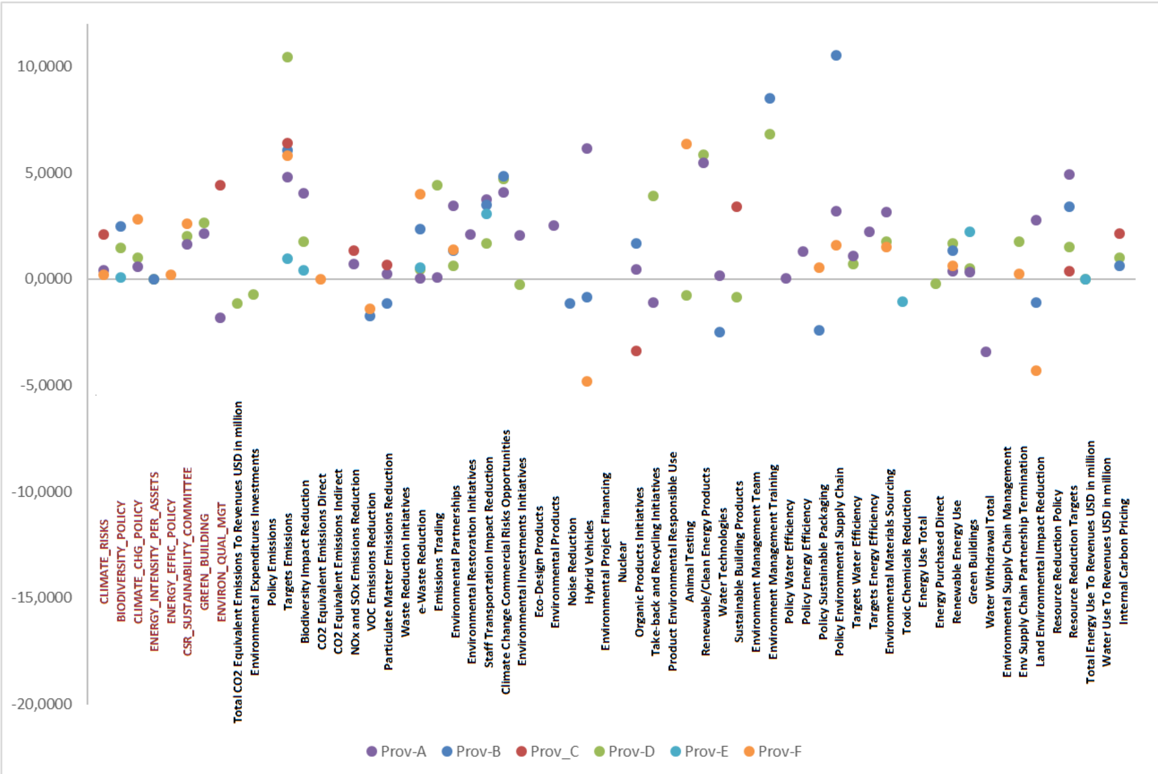
[7]De Lucia et al. (2020) combine statistical inference and machine learning techniques, such as the K-Nearest Neighbours algorithm, to explore the causal relation of ESG data points with corporate financial indicators. Bouyé and Menville (2020) apply five types of regressions to ESG scores for sovereigns: spanning stepwise, 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.

[8]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 in integrating information of the tails of the distribution than the standard ordinary least squares (hereafter, OLS) as QR allows the estimate of the nexus between the dependent variable and its explanatory variables at any specific quantile, while OLS focus on the mean.

Table 5. Number of significant variables by provider, R^2 in lasso regression, and correlation between actual and lasso-estimated E-scores.

Provider	No. Variables	R^2	Correlation
Provider-A	33	63 %	0.80
Provider-B	20	26 %	0.53
Provider-C	9	17 %	0.46
Provider-D	28	30 %	0.57
Provider-E	7	8%	0.39
Provider-F	16	23 %	0.50

Figure 7. Lasso coefficients for different E-scores.



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

The analysis of the coefficients (see Figure 7) underlines that some of the most meaningful variables are common across providers, such as the presence of reduction targets for emissions, waste, and resource use, as well as partnerships for reducing the environmental impacts, and policies for renewable energy and green buildings. However, the coefficients of several raw variables vary remarkably across providers, especially in connection with the policies for reducing the supply chain's environmental issues and the presence of emissions targets. 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.[9] Provider-A particularly relies on the largest number of significant raw variables. The number of identified variables is lower for Provider-C, Provider-F and Provider-B. Moreover, the quantile regression analysis reveals a higher explanatory power in the tails of E-scores distribution,

[9]For extensive results of the quantile regression, see Section C.1 in Appendix.

particularly for companies performing *worst* on environmental issues. In other words, raw data better explains E-scores on the worst companies than those that exhibit the best environmental practices. This outcome is relevant for investors seeking E-scores for *best-in-class* and *exclusion* strategies.

Moving additionally in this direction, in the next Section, we investigate whether lasso regression residuals hide an unobserved component for each provider's E-score that is not explained by raw data. This component could reflect variables' non-linear relationships or a provider's qualitative judgement.

4.1 Latent variable analysis

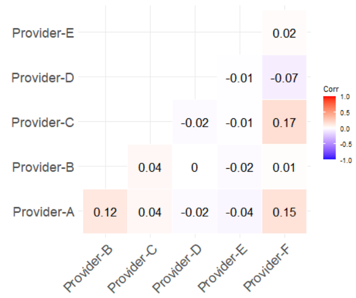
To further investigate the unexplained variance in our regression analysis, we specify a state-space model, where estimated lasso regression residuals are assumed to be noisy measurements of an unobserved state or latent variable evolving over time (Durbin and Koopman, 2012). We aim to estimate the smoothing distribution of all latent states based on the data and determine an average trajectory for these states. The Kalman filter provides an exact solution for linear Gaussian state-space models. However, for more general state-space models, analytical solutions are usually unavailable. Therefore, we resort to MCMC methods – specifically, on Gibbs sampling – to conduct inference (see Section C.4 in Appendix). The Gibbs sampler allows us to waive the assumption of normality required for the standard Kalman filter and to accommodate more flexible functional forms beyond linearity. This approach avoids more complex techniques, such as the extended Kalman filter (EKF) or unscented Kalman filter (UKF). The findings from this section indicate a low correlation between the latent variables estimated by different providers. Specifically, the average correlation is 0.02 when using the Kalman filter and 0.08 with the Gibbs sampler (see Figure 8, panels (a) and (b), respectively), both lower than the correlation of the actual E-scores (average 0.28; see Figure 8, panel (c)). The highest correlations among states are observed between Provider-A, Provider-C, Provider-E and Provider-F. Conversely, the correlation between lasso-estimated E-scores is slightly higher, reaching an average value of 0.29 (see Figure 8, panel (d)).

This evidence suggests that environmental scoring based only on raw data can lead to more similar E-scores. In contrast, the residual component – potentially judgemental or non-linear – introduces greater heterogeneity across providers' scores. This heterogeneity likely arises from varying definitions and focus on different sustainability issues across providers, as noted by Billio et al. (2021).[10] Providers may differ in their perspective of environmental assessment: some analyse the impact of environmental issues on corporate financial conditions, others examine corporate impact on environmental conditions, some evaluate environmental performance, and others adopt a dual perspective ("double materiality"). Additionally, some providers prioritise environmental risk implications for firms, while others consider associated opportunities (Larcker et al., 2022).

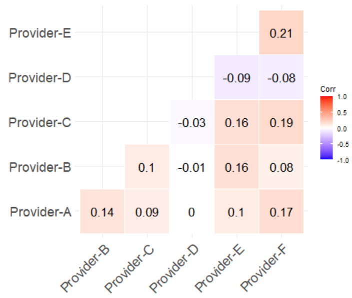
In summary, we conclude that the unexplained component remains significant for each provider, although it is quite heterogeneous across providers. This evidence suggests that raters use diverse assessment approaches, incorporating non-linear relationships or judgemental factors when interpreting raw data.

[10] Billio et al. (2021) identify the lack of a globally accepted standard methodology and a minimum level of technical requirements as the two major issues.

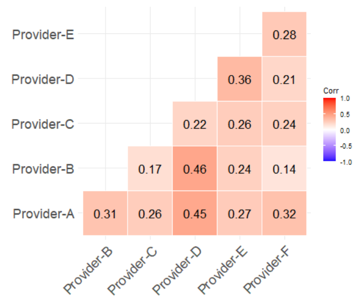
Figure 8. Correlation among Latent variables (estimated by standard Kalman filter and MCMC methods, respectively), actual, and lasso-ruled E-scores.



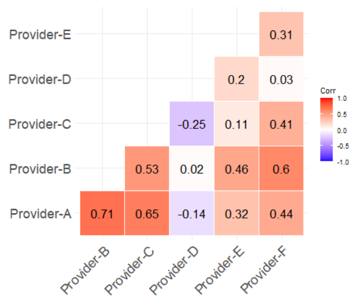
(a) Latent variables (Kalman filter)



(b) Latent variables (MCMC)



(c) Actual E-scores



(d) Lasso-estimated E-scores

5 Classification systems for sustainable investment strategies

In this section, we design an environmental assessment relying on raw data only. Specifically, we explore how the environmental assessment can be translated from a continuous variable (i.e. the environmental framework provided by external suppliers) into a discrete grading trained on the provider's E-scores. In other words, we use classification techniques applied to raw variables to rate each company in a grade consistent with that resulting from the actual E-score.[11] The application of the classification technique to raw data only allows us to circumvent the potentially distorting effects of the unexplained component, highlighted in the previous regression analysis. As a byproduct, the classification system enables the extension of the environmental assessment to companies that disclose environmental information, although providers do not rate them.

Finally, we test the classification rules through portfolio simulations. We exploit the grading system to implement sustainable investment strategies, i.e. *best-in-class* or *exclusion*, which are among the most adopted strategies at the global level (GSIA, 2022). Since classification systems inevitably entail statistical errors, we conduct portfolio simulations to examine the ex-post risk-return profiles and their alignment with providers' E-scores-based results. The bias could hinder the ex-ante financial outcomes of portfolio choices based on classes built upon ESG scores or classification rules.

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

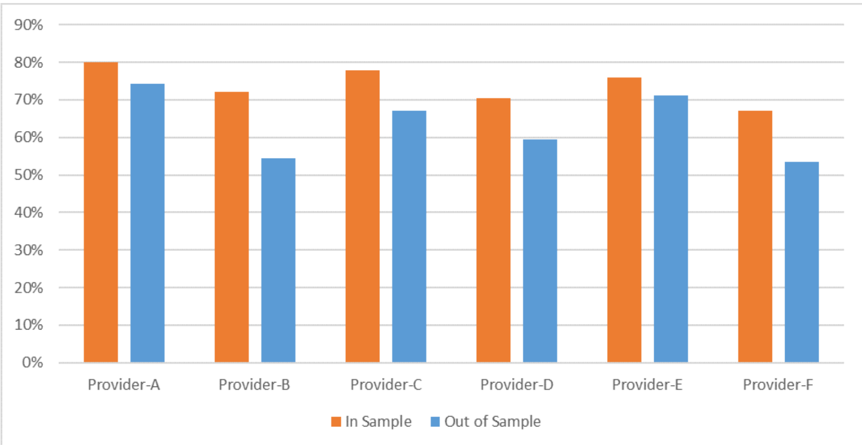
5.1 Classification rules

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, i.e., worst, intermediate, and best.[12] Companies are assigned to three environmental rating grades according to thresholds corresponding to the first and the fourth quintile of the distribution of each E-score (i.e., 20 and 80%). We split the data set into training and test samples to prevent overfitting issues and produce more accurate estimates. For both techniques, the training set consists of 2-year rolling samples, while each test set is based on the observations in the year following the training sample. Therefore, we will train the LDA and the KNN from 2017 to 2020[13] and test our system from 2019 to 2021. Differently from the regression analysis, we omit raw variables exhibiting no variability over time. In particular, we exclude 15 variables,[14] 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 from 67% to 80%. The higher accuracy ratio is found for the rule mimicking E-scores of Provider-A, Provider-C and Provider-E (around 80% on average), holding through the considered time windows (see Figure 9).

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



Note: The accuracy ratios are computed as average values of the three-year windows and classes for each provider.

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

[12] We also tested multi-logit regressions that exhibit weaker results than the LDA and the KNN, both in-sample and out-of-sample.

[13] 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 running the LDA and the KNN estimation.

[14] The disregarded variables are 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.

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 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.

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 the grades according to E-scores and those obtained with the LDA allocation rule.

Finally, the estimated coefficients of raw data for the LDA rule across data providers[15] seem consistent with those of the lasso regression analysis. The most significant variables are related to carbon, energy and water intensities.[16]

5.1.2 K-Nearest Neighbours

The second classification rule is developed by applying the KNN algorithm. First, exploiting information for 2015-2020, we determine the appropriate number of neighbour points, denoted as k, to classify observations into different classes based on information derived from each provider's Area Under the Receiver Operating Characteristic (AUROC) (see Figure 10).[17] Common choices of k 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 balance comparison needs, concise description, and satisfactory accuracy.[18]

[15]These coefficients can not be interpreted as in the standard regression framework; instead, they project all the information of the 46 variables into the one-dimensional space of the E-score. Nevertheless, thanks to the standardisation 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.

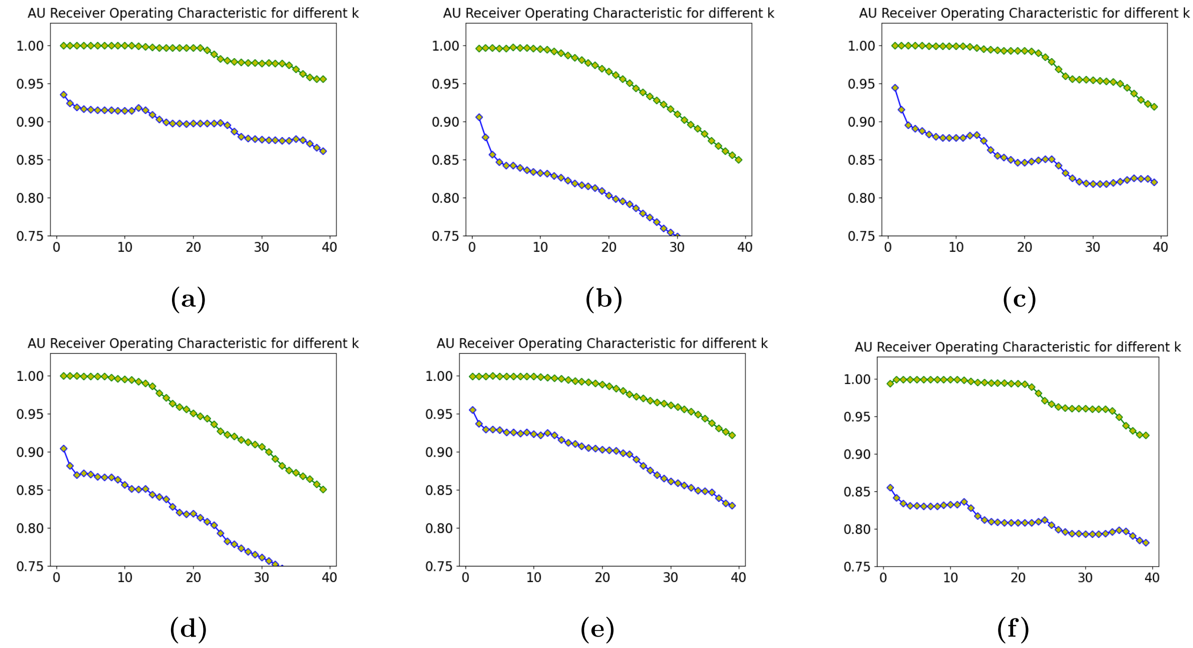
[16]Such as total CO2 equivalent emissions to revenues in USD million, CO2 equivalent direct emissions (Scope 1 and 2) total energy use, total amount of water withdrawn from a surface water or groundwater source, total energy use to revenues, and water use to revenues, respectively.

[17]The AUROC is the area under the Receiver Operating Characterization Curve (ROC). A ROC curve shows the trade-off between true positive rate (TPR) and false positive rate (FPR) across different decision thresholds. The AUROC is thus a performance metric that can be used to evaluate classification models. Generally speaking, an AUROC of 0.5 corresponds to a coin flip, i.e. a useless model; an AUROC less than 0.7 represents sub-optimal performance; an AUROC of 0.70 – 0.80 signals good performance; an AUROC greater than 0.8 means excellent performance; finally, an AUROC of 1.0 corresponds to a perfect classifier.

[18]Since KNN is a majority voting algorithm, it is standard practice to select odd values for k, enabling the resolution of potential ties.

Second, as for the LDA, the KNN assigns each issuer to one of three grades (i.e. worst, best, and intermediate) based on the similarities detected among the features of the k nearest peers.

Figure 10. AUROC for different k per provider.



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

The KNN achieves higher accuracy ratios than the LDA rule, both in-sample and out-of-sample. Notably, the highest accuracy is observed for classification rule calibrated on 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 considering the needs of comparison, parsimonious description, and satisfactory accuracy.

Moreover, the correlation between the allocation in classes based on E-scores and the relevant allocation estimated 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, while LDA 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. This result is expected and essentially due to the change in the environmental assessment (from a continuous to a discrete framework). Moreover, LDA and KNN benefit from the sample split into training and test sets.

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. As a test, the next Section finally shows two common strategies at work through portfolio simulations based on the estimated discretised E-scores.

5.2 Portfolio simulations

Once trained on actual E-scores and raw variables, the two classification techniques allow us to assign each company to one of the three environmental performance classes (i.e., worst, intermediate, and best). The previously discussed promising results in accuracy thus provide a solid ground for classifying those companies without the providers' E-scores. Since both rules are based on raw data, the proposed classification system offers a more applicable assessment methodology supporting sustainable investment decisions. In this Section, we test the classification techniques from a financial viewpoint by simulating portfolios constructed to pursue 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 built from actual providers' E-scores. Simulations are repeated for each provider. Noteworthy, this exercise aims not to compare the financial performance of sustainable strategies with the whole market portfolio but to test the rating systems' alignment based on the classification techniques as an alternative to the E-scores of professional providers.

The first investment strategy, known as *exclusion*, aims to tilt the standard market portfolio by filtering out the lower 20th percentile of the E-score distribution, i.e. the group of companies rated as the *worst*.[19] The E-grade of each company 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.

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

The second strategy, known as *best-in-class*, aims to invest 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. Likewise, for the previous investment strategy, the financial and environmental features of the portfolios built on the LDA and the KNN classification rules are compared with those built using the six providers' E-scores.

Portfolio simulations follow the same time-splitting policy adopted to train and test the classification rules (LDA and the KNN). 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 a training sample. The financial results are assessed through 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 following tables show the outcome of the environmental profile as the average of annual data and the financial results based on 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 suggests a systematic prediction bias of the classification rule compared with the provider's scores.

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

Environmental scores	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

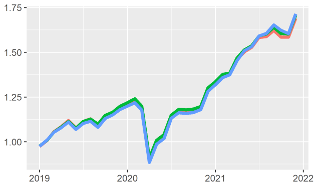
Concerning the financial performance, the portfolios built on the KNN technique achieve higher risk-adjusted returns (measured by the Sharpe ratio) than those built on the providers' scores, with limited tracking error volatility (see Table 10).

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

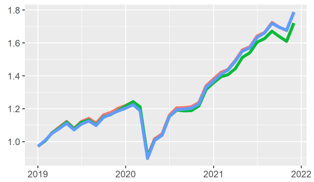
	Provider-A			Provider-B			Provider-C		
	LDA	KNN	prov.	LDA	KNN	prov.	LDA	KNN	prov.
TEV vs. prov.	1.1	1.5		1.5	1.0		0.9	1.9	
Sharpe R.	101.6	99.4	96.0	101.2	110.1	108.8	101.5	103.3	99.7
	Provider-D			**Provider-E**			**Provider-F**		
	LDA	KNN	prov.	LDA	KNN	prov.	LDA	KNN	prov.
TEV vs. prov.	0.9	1.2		1.4	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 corroborated 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.

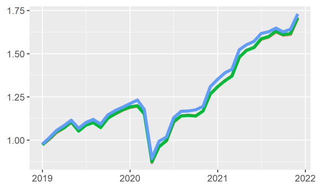
Figure 11. Portfolio cumulative returns with *exclusion* strategy.



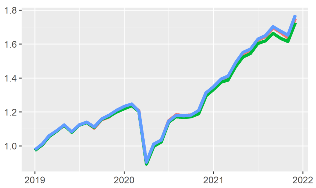
(a) Provider-A



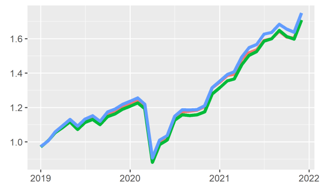
(b) Provider-B



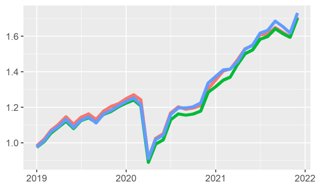
(c) Provider-C



(d) Provider-D



(e) Provider-E



(f) Provider-F

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

Regarding the *best-in-class* strategy, overall, we find that portfolios based on the classification rules have weighted average E-scores slightly lower than those of the portfolios built on providers' scores, arguably due to misclassification of actual best companies (see Table 11). The outcome of the *exclusion* strategy is comparatively better as it considers two out of three classes (i.e. the best and the intermediate).

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

Regarding the financial indicators, portfolios built on the KNN and the LDA achieve higher risk-adjusted returns than those based on E-scores, with the KNN exhibiting better performance than the LDA. However, they show higher tracking error levels than the *exclusion* strategy, given that the less diversified portfolios (concentrated on high-grade firms) amplify the differences between the provider' and the classification rules (see Table 12).

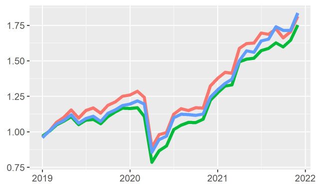
Finally, the pattern of cumulative returns among the three simulated portfolios is more differentiated for the *best-in-class strategy* (see Figure 12).

It should also be noted that financial results vary broadly across providers, even considering the same strategy – either *exclusion* or *best-in-class*. This evidence confirms that the financial materiality of the environmental assessment by the providers can vary significantly. Eventually, it will lead to different market performances of sustainable investment strategies based on their scores. In conclusion, the portfolio simulations clearly show that the discrete rating system works financially and can support sustainable investment decisions.

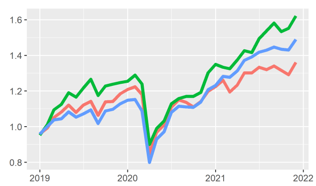
Table 12. Financials of the portfolios with *best-in-class* strategy.

	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

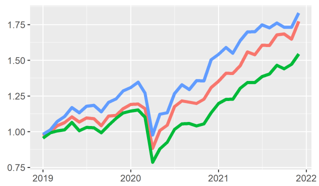
Figure 12. Portfolio cumulative returns with *best-in-class* strategy.



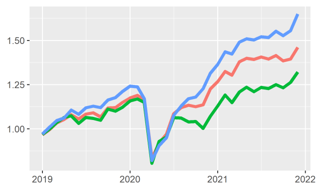
(a) Provider-A



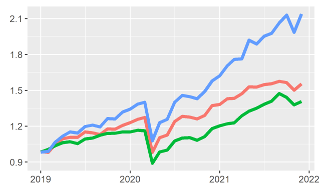
(b) Provider-B



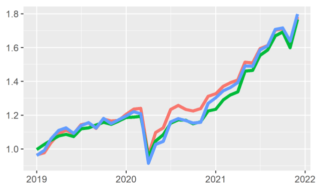
(c) Provider-C



(d) Provider-D



(e) Provider-E



(f) Provider-F

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

6 Conclusions

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

Environmental data have meaningful, although limited, explanatory power to predict the E-scores. The quantile and lasso regression analysis offer consistent evidence that only the scores of some providers are highly related to raw data. We identify some variables as significant and common across several providers, such as forward-looking measures like the presence of reduction targets for emissions, waste, and resource use, as well as partnerships for reducing the environmental impacts and policies for renewable energy and green buildings. According to these findings, the regulatory request to enhance corporate disclosure should focus on such variables. 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 attributed by the providers to environmental issues. Indeed, some providers focus their analysis on how corporate financial conditions are affected by environmental issues. In contrast, others consider how corporate conduct can affect environmental conditions, and others consider both perspectives ("double materiality"). Notwithstanding the different origins, the missing – likely judgemental – component plays a significant role for all the providers. In this regard, greater transparency by providers is a compelling need to increase the credibility of E-scores, thus fostering their wider use.

The paper also proposes two classification rules to allocate each company among three environmental performance classes, consistent with the grading based on E-scores of providers. Such classification rules rely only on raw data. Therefore, they allow the environmental assessment extension to companies that disclose environmental information, but providers do not rate them. For instance, a potential application could be offered by the broader scope of sustainability disclosure envisaged by the EU Corporate Sustainability Reporting Directive, which is expected to affect up to 50,000 companies, including unlisted SMEs, with a simplified disclosure framework starting from the 2026 financial year. According to our simulations to test their usability for investment purposes, the proposed classification rules deliver portfolios with environmental profiles similar to those obtained using the providers' E-scores. On the other hand, the return-risk characteristics of the classification-based portfolios can exhibit superior outcomes in some circumstances. This result holds under two well-established investment strategies. These results emphasise that the challenges associated with environmental scores should not hinder 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 the Taskforce on Nature-related Financial Disclosures (TNFD, 2023). The issues raised by the study regarding E-scores call for further research and regulation to enhance 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 improving corporate environmental disclosure and for financial authorities, including central banks and supervisors, to detect the most relevant variables for their environmental assessment. Expanding this line of research would benefit financial authorities in assessing the reliability and consistency of the scores and data used by supervised intermediaries or counterparts in their strategies. It has eventually implications for the stability and resilience of the financial system.

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/Angelini-11-luglio-2022.pdf*

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'.

Bams, D. and van der Kroft, B. (2022), 'Divestment, information asymmetries, and inflated ESG ratings', *Information Asymmetries, and Inflated ESG Ratings (June 3, 2022)* .

Berg, F., Koelbel, J. F. and Rigobon, R. (2022), 'Aggregate confusion: The divergence of ESG ratings', *Review of Finance* **26**(6), 1315–1344.

Berg, F., Kölbel, J. F., Pavlova, A. and Rigobon, R. (2021), 'ESG confusion and stock returns: Tackling the problem of noise', *Available at SSRN 3941514* .

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.

Del Vitto, A., Marazzina, D. and Stocco, D. (2023), 'ESG ratings explainability through machine learning techniques', *Annals of Operations Research* pp. 1–30.

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).

Durbin, J. and Koopman, S. J. (2012), *Time series analysis by state space methods*, Vol. 38, OUP Oxford.

D'Amato, V., D'Ecclesia, R. and Levantesi, S. (2022), 'ESG score prediction through random forest algorithm', *Computational Management Science* **19**(2), 347–373.

ECB (2020), 'Guide on climate-related and environmental risks. Supervisory expectations relating to risk management and disclosure'.
URL: *https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.202011finalguideonclimate-relatedandenvironmentalrisks 58213f6564.en.pdf*

Ehlers, T., Elsenhuber, U., Jegarasasingam, K. and Jondeau, E. (2023), *Deconstructing ESG Scores: How to Invest with your own Criteria?*, number 1008, International Monetary Fund.

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 (2022), 'Global sustainable investment review'.
URL: *https://www.gsi-alliance.org/wp-content/uploads/2023/12/GSIA-Report-2022.pdf*

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.
URL: *https://www.ngfs.net/sites/default/files/medias/documents/ngfs_first_comprehensive_report_-_17042019_0.pdf*

NGFS (2020), 'Guide for supervisors. integrating climate-related and environmental risks into prudential supervision'.
URL: *https://www.ngfs.net/sites/default/files/medias/documents/ngfs_guide_for_supervisors.pdf*

NGFS (2022), Final report on bridging data gaps, Technical report.
URL: *https://www.ngfs.net/sites/default/files/medias/documents/final_report_on_bridging_data_gaps.pdf*

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 (2016), 2016 Global Sustainable Investment Review, 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.
URL: *https://racetozero.unfccc.int/system/race-to-zero/*

Visco, I. (2021), 'The G20 presidency programme on sustainable finance'.
URL: *https://www.bancaditalia.it/pubblicazioni/interventi-governatore/integov2021/Visco_30092021.pdf*

Zanin, L. (2022), 'Estimating the effects of ESG scores on corporate credit ratings using multivariate ordinal logit regression', *Empirical Economics* **62**(6), 3087–3118.

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*

Disclaimer on MSCI ESG Research data

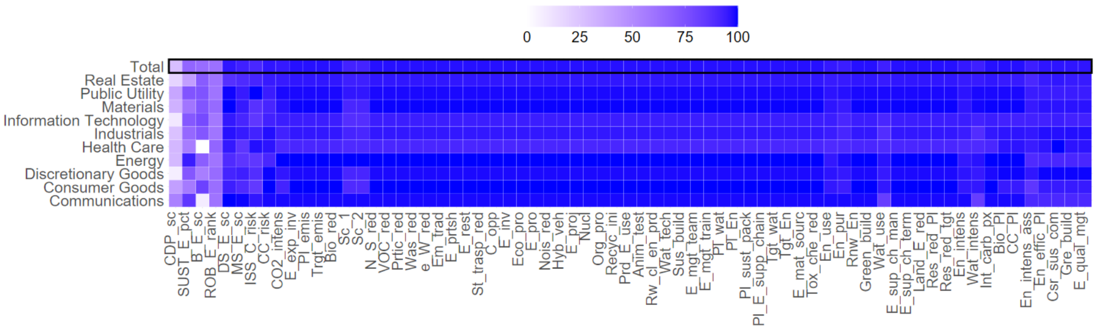
A Environmental raw variables

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 Data completeness

Figure 2. Coverage percentages per each environmental score and variable.



Note: The first row reports the whole dataset coverage percentages for E-scores in the first seven columns and raw E-variables in the remaining columns, respectively, from the left side. The subsequent rows show the sectoral coverage percentages.

C Methodologies applied for dissecting E-scores

C.1 Results of quantile regression

Quantile regression has been performed over 2015-2021 for seven quantiles simultaneously: 5, 20, 25, 50, 75, 80 and 95 per cent (see Table 1). Except for Provider-G, the number of significant variables across quantiles is, on average, between 10 and 19 across the providers. The larger the number of variables, the wider the dispersion across quantiles. Notably, more significant variables are identified 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 a larger number of data points are available and needed to discriminate the environmentally worst-performing firms. When we consider the 80-20 quantile - which investors often use 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. Number 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

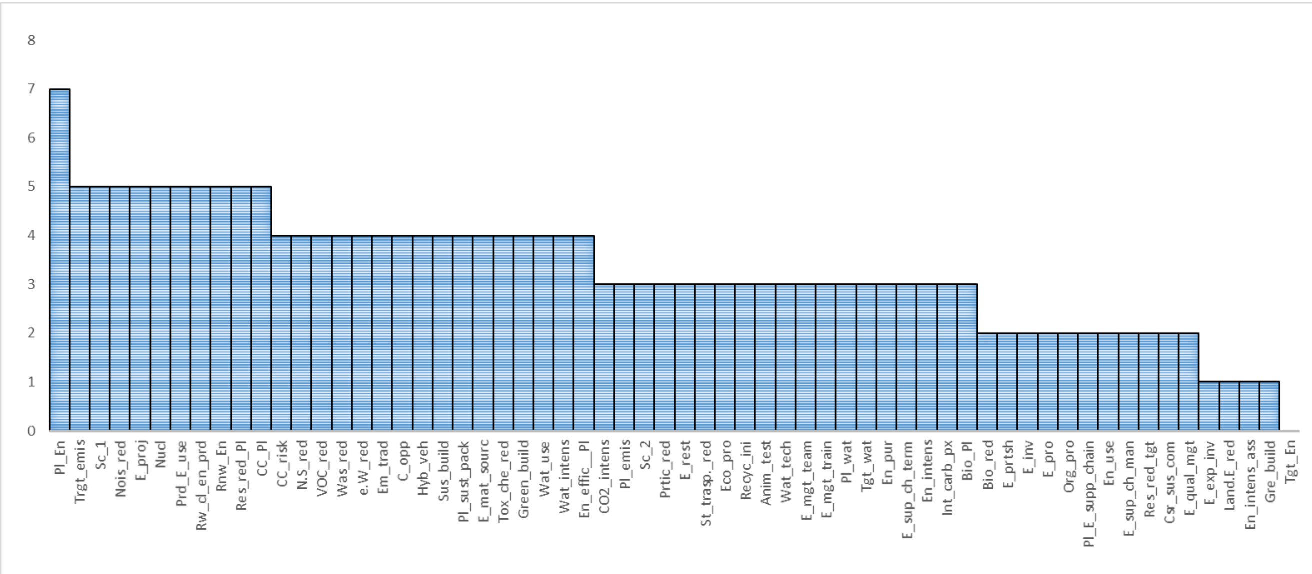
The pseudo R^2 in Table 2, computed for the pooled quantile regressions, points to higher explanatory power for the E-scores of those providers where more variables were found significant (Provider-A, Provider-C and Provider-E).

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 meaningful across different E-scores, as shown in Figure 3. Noticeably, 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 significant variables for several E-scores. The sectoral dummies are not very meaningful, arguably because the carbon emission variables can explain part of the sector-specific variance. Moreover, we exclude from the investigation some financial indicators commonly used for fundamental analysis (e.g., EBITDA margin, financial leverage, ROE, and ROI) because they do not seem directly related to environmental performance, although an inverse relation between such indicators and environmental scores 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, nor do we investigate how the judgemental component could predict future changes of raw data. Such investigation might warrant follow-up research.

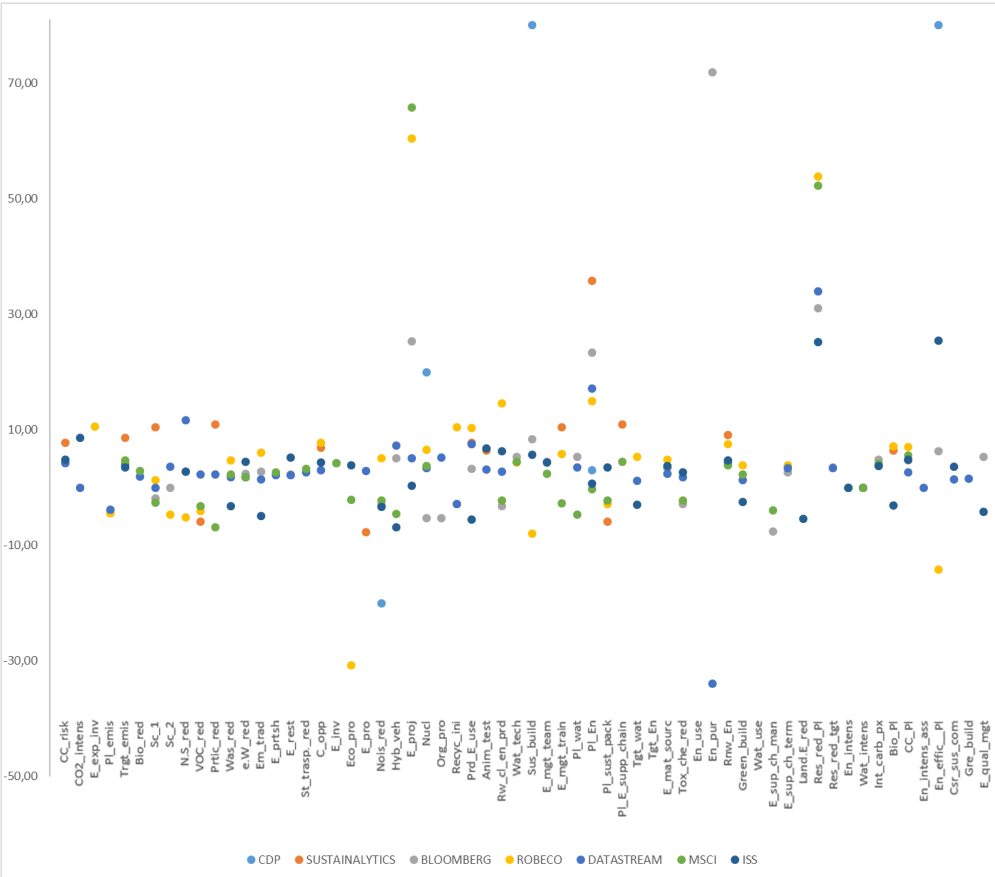
Figure 3. Quantile regression – variables frequently found relevant across E-scores.



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

Finally, we note that 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. At the same time, there are some remarkable differences as regards variables such as sustainable buildings, environmental projects, energy efficiency and resource reduction policies, and staff transportation impact reduction (Figure 4).

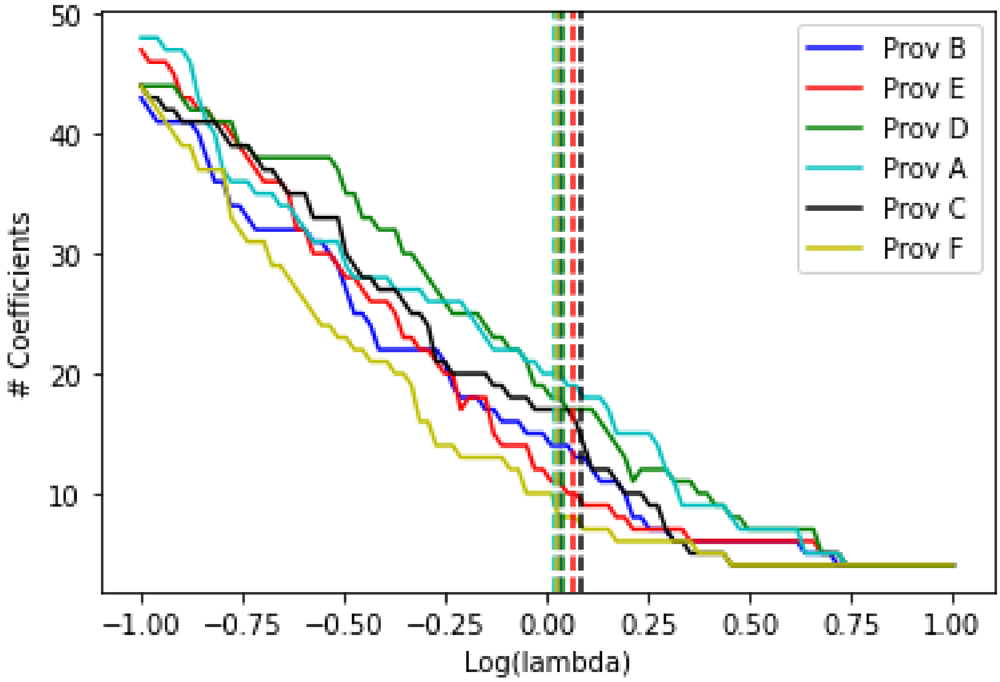
Figure 4. Quantile regression – average variables' coefficients per E-score.



C.2 Sensitivity analysis of Lasso regression

Within the Lasso regression, we select the tuning parameter λ by cross-validation. Below, we show the number of significant coefficients across possible penalisation parameters.

Figure 5. Lasso sensitivity analysis (λ).



Note: On the horizontal axis, we represent the logarithm of the penalisation parameter λ, while the vertical axis denotes the number of non-zero coefficients corresponding to each choice of the penalisation parameter. Vertical lines indicate the optimal λ selected for each series.

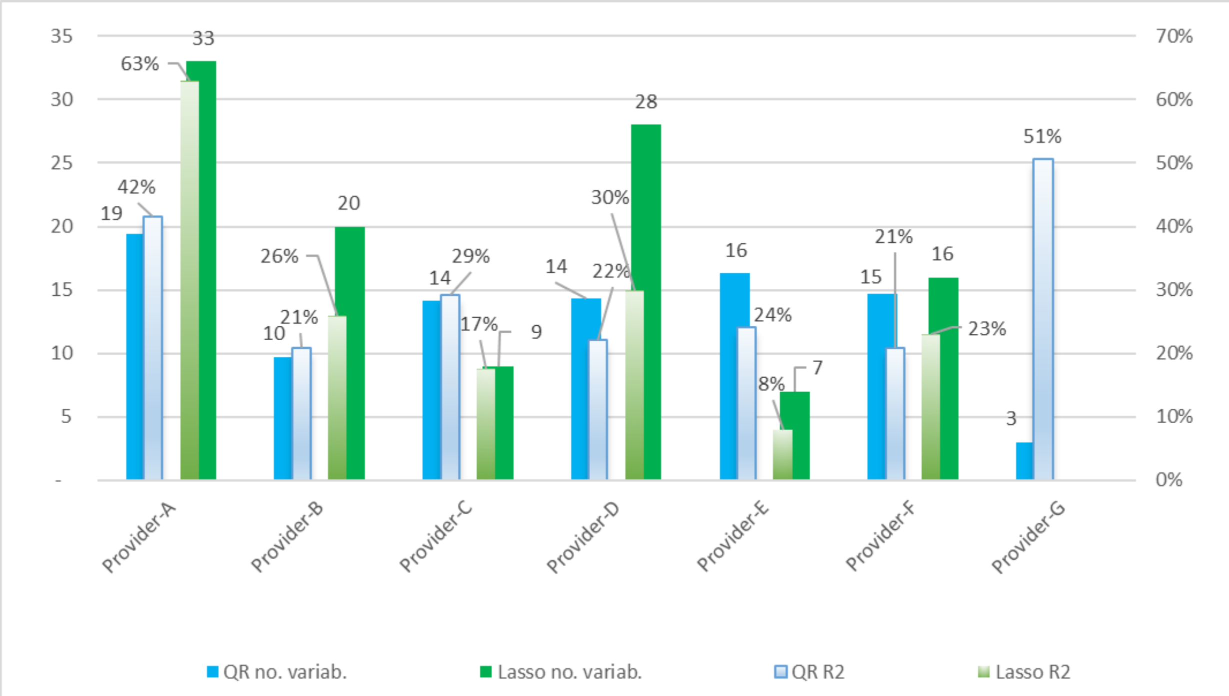
The figure illustrates the relationship between the number of significant coefficients (and relevant variables) and the logarithm of the penalisation parameter λ. The figure shows an inverse relationship. As the log of λ increases along the horizontal axis, the penalisation applied is more severe and the number of significant coefficients on the vertical axis tends to decrease. This inverse relationship between the penalisation parameter and the number of non-zero coefficients aligns with expectations in penalised regression methods, where smaller values of λ lead to less regularisation and, consequently, more non-zero coefficients. Notably, vertical lines on the plot mark the optimal λ values selected for each parameter. The close proximity of these optimal values across all providers confirms the robustness of the estimated model. This consistency suggests that the chosen penalisation parameter effectively balances model complexity and predictive accuracy across different providers, enhancing the model's reliability.

C.3 Joint analysis of results from quantile and lasso regression

Looking jointly at the results of both regression techniques, the number of regressors found per provider varies significantly between the quantile regression and lasso regression (see Figure 6). For three out of seven providers, the two techniques identify a similar number of significant variables, while for Provider-A and Provider-D there is a remarkable difference (from 19 to 33 and from 14 to 28, respectively) and R^2 (from 42% to 63% and from 22% to 30%, respectively). Both techniques consistently give 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 averages around

29%. Except for Provider-G, when the number of indicators decreases (less or equal to 10 regressors), the explained part of E-scores diminishes accordingly. The R^2 varies between 21% with 10 variables for Provider-B in case of quantile regression and 8% with 7 variables for Provider-E in case of lasso regression – the worst case we observe. Since the identified indicators can not 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.

Figure 6. Number of regressors and R^2 per E-score.



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 other providers could be due to the provider's specific focus rather than on a broader environmental assessment. Moreover, for this provider, we do not find significant variables in the lasso framework, mainly due to the scarcity of data. All these considerations suggest excluding Provider-G from the analysis.

C.4 Latent variable estimation

The relatively low R^2 observed in the quantile and lasso regression estimations calls for further investigation into the unexplained components of the regression models. A potentially significant variable, omitted from the regression equation due to its latent nature, must be considered to fully understand the relationship between E-scores and their underlying elements. This latent variable can be interpreted as the providers' judgmental component, which is not captured by quantitative raw data but significantly influences the E-scores. State-space models are employed to model time-series data where the observed data vector at each time t, denoted as Y_t, is assumed to be related to the observed variables Z_t and an unobserved latent state X_t through an observation distribution $Y_t = g_t\left(y_t \,|x_t, z_t, \theta\right)$. The latent state depends on the previous state X_{t-1} through a transition distribution $X_t = f_t\left(x_t \,|x_{t-1}, \theta\right)$. Here, θ is a vector of hyper-parameters assumed to remain constant over time.

The objective is to estimate the smoothing distribution $p\left(x_{1:t}\,|y_{1:t},\theta\,\right)$ for X_t, which provides information about the latent states from all time points given the most recent data. In our framework, Y_t represents the E-score of each provider at time t, while Z_t is the vector of 62 explanatory variables employed in the regression models, common to all providers. Therefore, $Y_t - Z_t D$ are the lasso-estimated residuals we seek to explain using the latent component, where D contains the lasso-estimated coefficients for each provider. Specifically, we conduct exact inference via the Kalman filter. We also apply the Gibbs sampler MCMC algorithm to draw samples from the smoothing distribution and estimate θ. Upon estimating θ for each provider, we finally simulate n_p paths for the state per provider and take the related average as the estimation of the latent variable.

Recently published papers in the 'Markets, Infrastructures, Payment Systems' series

n. 16 Cross-Currency Settlement of Instant Payments in a Multi-Currency Clearing and Settlement Mechanism, *by Massimiliano Renzetti*, *Fabrizio Dinacci and Ann Börestam* (Research Papers)

n. 17 What's ahead for euro money market benchmarks?, *by Daniela Della Gatta* (Institutional Issues) (in Italian)

n. 18 Cyber resilience per la continuità di servizio del sistema finanziario, *by Boris Giannetto and Antonino Fazio* (Institutional Issues) (in Italian)

n. 19 Cross-Currency Settlement of Instant Payments in a Cross-Platform Context: a Proof of Concept, *by Massimiliano Renzetti, Andrea Dimartina, Riccardo Mancini, Giovanni Sabelli, Francesco Di Stasio, Carlo Palmers, Faisal Alhijawi, Erol Kaya, Christophe Piccarelle, Stuart Butler, Jwallant Vasani, Giancarlo Esposito, Alberto Tiberino and Manfredi Caracausi* (Research Papers)

n. 20 Flash crashes on sovereign bond markets – EU evidence, *by Antoine Bouveret, Martin Haferkorn, Gaetano Marseglia and Onofrio Panzarino* (Research Papers)

n. 21 Report on the payment attitudes of consumers in Italy: results from ECB surveys, *by Gabriele Coletti, Alberto Di Iorio, Emanuele Pimpini and Giorgia Rocco* (Institutional Issues)

n. 22 When financial innovation and sustainable finance meet: Sustainability-Linked Bonds, *by Paola Antilici, Gianluca Mosconi and Luigi Russo* (Institutional Issues) (in Italian)

n. 23 Business models and pricing strategies in the market for ATM withdrawals, *by Guerino Ardizzi and Massimiliano Cologgi* (Research Papers)

n. 24 Press news and social media in credit risk assessment: the experience of Banca d'Italia's In-house Credit Assessment System, *by Giulio Gariano and Gianluca Viggiano* (Research Papers)

n. 25 The bonfire of banknotes, *by Michele Manna* (Research Papers)

n. 26 Integrating DLTs with market infrastructures: analysis and proof-of-concept for secure DvP between TIPS and DLT platforms, *by Rosario La Rocca, Riccardo Mancini, Marco Benedetti, Matteo Caruso, Stefano Cossu, Giuseppe Galano, Simone Mancini, Gabriele Marcelli, Piero Martella, Matteo Nardelli and Ciro Oliviero* (Research Papers)

n. 27 Statistical and forecasting use of electronic payment transactions: collaboration between Bank of Italy and Istat, *by* Guerino Ardizzi and Alessandra Righi (Institutional Issues) (in Italian)

n. 28 TIPS: a zero-downtime platform powered *by* automation, by *Gianluca Caricato, Marco Capotosto, Silvio Orsini and Pietro Tiberi* (Research Papers)

n. 29 TARGET2 analytical tools for regulatory compliance, *by Marc Glowka, Alexander Müller, Livia Polo Friz, Sara Testi, Massimo Valentini and Stefano Vespucci* (Institutional Issues)

n. 30 The security of retail payment instruments: evidence from supervisory data, *by Massimiliano Cologgi* (Research Papers)

n. 31 Open Banking in the payment system: infrastructural evolution, innovation and security, supervisory and oversight practices, *by Roberto Pellitteri, Ravenio Parrini, Carlo Cafarotti and Benedetto Andrea De Vendictis* (Institutional Issues) (in Italian)

n. 32 Banks' liquidity transformation rate: determinants and impact on lending, *by Raffaele Lenzi, Stefano Nobili, Filippo Perazzoli and Rosario Romeo* (Research Papers)

n. 33 Investor behavior under market stress: evidence from the Italian sovereign bond market, *by Onofrio Panzarino* (Research Papers)

n. 34 Siamese neural networks for detecting banknote printing defects, *by Katia Boria, Andrea Luciani, Sabina Marchetti and Marco Viticoli* (Research Papers) (in Italian)

n. 35 Quantum safe payment systems, *by Elena Bucciol and Pietro Tiberi*

n. 36 Investigating the determinants of corporate bond credit spreads in the euro area, *by Simone Letta and Pasquale Mirante*

n. 37 Smart Derivative Contracts in DatalogMTL, *by Andrea Colombo, Luigi Bellomarini, Stefano Ceri and Eleonora Laurenza*

n. 38 Making it through the (crypto) winter: facts, figures and policy issues, *by Guerino Ardizzi, Marco Bevilacqua, Emanuela Cerrato and Alberto Di Iorio*

n. 39 The Emissions Trading System of the European Union (EU ETS), *by Mauro Bufano, Fabio Capasso, Johnny Di Giampaolo and Nicola Pellegrini* (in Italian)

n. 40 Banknote migration and the estimation of circulation in euro area countries: the italian case, *by Claudio Doria, Gianluca Maddaloni, Giuseppina Marocchi, Ferdinando Sasso, Luca Serrai and Simonetta Zappa* (in Italian)

n. 41 Assessing credit risk sensitivity to climate and energy shocks, *by Stefano Di Virgilio, Ivan Faiella, Alessandro Mistretta and Simone Narizzano*

n. 42 Report on the payment attitudes of consumers in italy: results from the ecb space 2022 survey, *by Gabriele Coletti, Alberto Di Iorio, Emanuele Pimpini and Giorgia Rocco*

n. 43 A service architecture for an enhanced Cyber Threat Intelligence capability and its value for the cyber resilience of Financial Market Infrastructures, *by Giuseppe Amato, Simone Ciccarone, Pasquale Digregorio and Giuseppe Natalucci*

n. 44 Fine-tuning large language models for financial markets via ontological reasoning, *by Teodoro Baldazzi, Luigi Bellomarini, Stefano Ceri, Andrea Colombo, Andrea Gentili and Emanuel Sallinger*

n. 45 Sustainability at shareholder meetings in France, Germany and Italy, *by Tiziana De Stefano, Giuseppe Buscemi and Marco Fanari* (in Italian)

n. 46 Money market rate stabilization systems over the last 20 years: the role of the minimum reserve requirement, *by Patrizia Ceccacci, Barbara Mazzetta, Stefano Nobili, Filippo Perazzoli and Mattia Persico*

n. 47 Technology providers in the payment sector: market and regulatory developments, *by Emanuela Cerrato, Enrica Detto, Daniele Natalizi, Federico Semorile, Fabio Zuffranieri*

n. 48 The fundamental role of the repo market and central clearing, *by Cristina Di Luigi, Antonio Perrella and Alessio Ruggieri*

n. 49 From Public to Internal Capital Markets: The Effects of Affiliated IPOs on Group Firms, *by Luana Zaccaria, Simone Narizzano, Francesco Savino and Antonio Scalia Perrella and Alessio Ruggieri*

n. 50 Byzantine Fault Tolerant consensus with confidential quorum certificate for a Central Bank DLT, *by Marco Benedetti, Francesco De Sclavis, Marco Favorito, Giuseppe Galano, Sara Giammusso, Antonio Muci and Matteo Nardelli*