Check for updates

Corporate Social Responsibility and Environmental Management

WILEY

Corporate Social Responsibility and Environmental Management

RESEARCH ARTICLE OPEN ACCESS

Uncovering ESG Ratings: The (Im)Balance of Aspirational and Performance Features

Andrea Ferro[1] | Daniele Marazzina[1] | Davide Stocco[2]

[1]Department of Mathematics, Politecnico di Milano, Milan, Italy | [2]CREST, ENSAE, Institut Polytechnique de Paris, Palaiseau, France

Correspondence: Daniele Marazzina (daniele.marazzina@polimi.it)

Received: 20 September 2024 | **Revised:** 16 May 2025 | **Accepted:** 22 May 2025

Funding: This work was supported by European Commission, 1-CUP D53D23017690001; and Ministero dell'Università e della Ricerca, Dipartimento di Eccellenza 2023-2027.

Keywords: ESG ratings | explainability | feature selection | greenwashing | LSEG methodology | machine learning | transparency

ABSTRACT

Environmental, Social, and Governance (ESG) scores are crucial for evaluating corporate sustainability. However, the undisclosed and complex methodologies used by rating agencies have raised concerns about their reliability and consistency. This study replicates LSEG's ESG scoring methodology using machine learning to shed light on the key drivers behind ESG ratings, with a focus on the balance between forward-looking promises (aspirational) and past achievements (performance). Our analysis finds that approximately 60% of ESG scores are based on aspirational promises, while only approximately 40% reflect actual performance. This imbalance suggests a potential over-reliance on future commitments, which could inflate ESG scores and mislead investors about a company's true sustainability efforts, even accounting for LSEG's transparency stimulation mechanism, where non-disclosure of material data is penalized. The findings emphasize the need for greater transparency and a clearer distinction between aspirational and performance metrics to ensure credible ESG assessments for informed investment decisions.

1 | Introduction

Environmental, social, and governance (ESG) criteria have become central to the financial world's efforts to account for non-financial risks and sustainability concerns. The concept of ESG was formally introduced in the landmark report by the United Nations Global Compact in 2004 (UN Global Compact 2004), which emphasized the integration of ESG factors into financial analysis and investment decision-making. More precisely, the Environmental pillar evaluates corporate policies and performance on issues like climate change, pollution, and biodiversity. The Social pillar covers labor standards, human rights, and workplace safety. The Governance pillar assesses board independence, shareholder rights, executive pay, and compliance with legal and ethical standards.

Over the past two decades, interest in ESG has grown rapidly. ESG scores, that is, quantitative ratings designed to evaluate a company's exposure to and management of sustainability-related risks, have emerged as a key tool for assessing corporate non-financial performance. These ratings aim to provide investors with a comprehensive and comparable snapshot of how companies perform across a wide range of ESG dimensions, from carbon emissions to labor practices and board diversity.

The importance of ESG ratings is reflected in the sharp rise of ESG-oriented investment strategies. In the United States, assets managed under ESG mandates grew from $3 trillion in 2010 to over $17 trillion by 2020, representing roughly 12% of professionally managed assets in 2024 (US SIF Foundation 2024). Globally, sustainable investing has become mainstream, with

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Corporate Social Responsibility and Environmental Management* published by ERP Environment and John Wiley & Sons Ltd.

institutional investors increasingly required to integrate ESG considerations into their portfolio strategies, due both to investor demand and evolving regulatory expectations.

There are several ways to incorporate ESG considerations into investment strategies, among which three approaches dominate: (1) Corporate engagement and shareholder action, the most common strategy globally in 2022 (Global Sustainable Investment Alliance 2023), involves actively engaging with company management to influence behavior and improve ESG practices; (2) Negative screening, or exclusion, entails removing certain sectors, companies, or countries from investment universes based on ESG-related criteria (e.g., controversial products, labor rights violations, environmental damage); (3) ESG integration, which refers to the explicit inclusion of ESG factors in financial analysis. This is often operationalized by using ESG scores and applying constraints at the portfolio level to achieve a desired sustainability profile (Pedersen et al. 2021; Avramov et al. 2022; Azzone et al. 2024; Useche et al. 2024).

Sustainability (2023) survey reveals that 69% of investors now use ESG ratings multiple times weekly (versus 35% in 2018–2019), with a preference towards ratings coming from more active raters (e.g., CDP) rather than passive ones (e.g., LSEG—former Refinitiv). Moreover, 53% of investors report developing in-house ESG models, citing dissatisfaction with the quality and consistency of external providers (only 37% of surveyed practitioners surveyed states that ESG ratings are a credible source of information on ESG corporate performance). This trend may introduce bias into ESG assessments, potentially forming the basis of a systemic risk. Such increasing reliance and dissatisfaction on ESG scores highlight the need for trustworthy ratings. On one side, the integration of AI methodologies has been proposed in ESG assessments to improve ESG data analysis (Alshahmy and Sahiner 2024). On the regulatory side, many initiatives—such as the European Corporate Sustainability Reporting Directive (CSRD) and the Sustainable Finance Disclosure Regulation (SFDR) which aim to improve the transparency, reliability, and comparability of sustainability-related information as well as to push the convergence toward double materiality frameworks—have been introduced in the ESG framework to raise the bar for ESG data quality, objective measures, and use. The main weaknesses pointed out by such regulations concern a lack of standardization in methodologies, limited transparency of how ratings are constructed, insufficient data quality and reporting coverage. Moreover, due to the complexity of sustainability-related information, ESG rating methodologies often require advanced approaches to manage heterogeneous data from various industries, missing values, and data inaccuracies, all of which can compromise the transparency of these scores. Additionally, ESG ratings are influenced by subjective choices, such as the inclusion or exclusion of specific metrics or the assignment of weights, which can significantly impact a company's final score and lead to discrepancies across rating agencies. Therefore, understanding the construction and limitations of ESG scores is critical for three primary stakeholders. For investors, it enables a critical assessment of whether the scores align with their investment objectives. For companies, it provides guidance to prioritize meaningful ESG improvements rather than focusing solely on optimizing specific score metrics. For regulators, it helps identify instances of "greenwashing" and suggests ways to improve underlying methodologies. Our paper addresses the needs of rating agencies, investors, and regulators by offering deeper insights into the mechanisms of these scoring systems. Exploring these limitations is essential for preventing misuse or misinterpretation of ESG scores, enhancing transparency, and fostering critical evaluation.

The lack of clarity and common definitions characterizing the world of sustainable investments, which causes weak or even heterogeneous results in unveiling the financial influence of ESG ratings, reflects also in the methodology exploited in the ESG issuance exercise. These models generally rely on publicly available data, including annual reports, sustainability disclosures, and information from company websites. However, the lack of standardized criteria and methodologies, due to the proprietary nature of the algorithms employed and the aggregating rule employed that could overlook sector-specific materiality (Giese et al. 2021), results in significant discrepancies across different rating agencies, such as MSCI, Sustainalytics, Bloomberg, and LSEG, among others. The existing literature highlights these discrepancies and attempts to shed light on them, comparing different rating agencies' algorithms (Lanza et al. 2023; Billio et al. 2024; Mao et al. 2024; Rubino et al. 2024; Bernardini et al. 2024; Liu et al. 2024). Specifically, Gibson Brandon et al. (2021) prove that variations in ESG ratings are often due to differences in the underlying methodologies and data sources used by rating agencies. Billio et al. (2021) highlight the heterogeneity in rating criteria as well, pointing out that opposite opinions on the same companies create different benchmarks, complicating ESG investment practices. Berg et al. (2022) address the problem of ESG divergence among rating agencies, showing an average correlation among global ESG scores of around 60%, compared to the nearly 90% among credit ratings. Furthermore, the authors identify three main drivers behind ESG rating divergence: differences in the factors measured by each rating agency (*scope*), the relative importance assigned to each factor (*weight*), and discrepancies in how data is gathered and assessed (*measurement*). Their analysis shows that *measurement* divergence accounts for more than half of the overall divergence, while *scope* and *weight* are moderately less impactful, though still significant. Lastly, they propose various methods for aggregating individual ESG ratings and demonstrate that such aggregation enhances portfolio performance, suggesting that, despite their inherent noise, ESG ratings offer valuable insights for portfolio construction.

Delving into the methodologies of the most relevant rating agencies, many researchers have tried to unveil, understand, and explain the models exploited by data providers to issue ESG ratings. Del Vitto et al. (2023) explore the use of machine learning techniques to enhance the explainability of LSEG ESG ratings using a subset of ESG granular data points referred to three industry groups ("Financial and Insurance", "Manufacturing", and "Information & Professional, Scientific, and Technical Services") and three geographical regions (US, Europe, and China). They employ both white-box and black-box models to replicate and understand the rating process. Their findings indicate that it is possible to replicate ESG ratings with a high degree of accuracy, even though some unlearnable noise persists. They also place great emphasis on model explainability through tools

15353966, 2025, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/csr.70007 by CochraneItalia, Wiley Online Library on [05/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

like SHAP values. Berg et al. (2021) delve into the reliability and consistency of ESG ratings, exposing the ongoing and unannounced rewriting of historical ESG data by LSEG, raising concerns about the reliability of these ratings for both academic research and practical investment strategies. Bams and van der Kroft (2022) employ a non-parametric rank-ordering approach on granular LSEG ESG information to demonstrate that ESG ratings combine forward-looking promises and backward-looking realizations of sustainable performance to assess ESG scores. Furthermore, they underscore the lack of correlation between ESG ratings and realized sustainable performance in conjunction with a limited execution of such promises moving forward, indicating that ratings are an inflated measure of sustainable performance. In addition, they argue that socially responsible institutional investors tilt their portfolios toward firms with inflated ratings, consequently lowering their cost of capital and increasing their growth rates, leading to inefficient capital allocation. Other studies include also financial statement information of a company to predict its ESG rating, underlying a relationship between the economical and sustainability performance of a company. D'Amato et al. (2022) find that financial statements represent a powerful tool to explain the ESG score, employing random forest technique. Considering Bloomberg data on the STOXX Europe 600 Index to predict each company's ESG score, they find that the most important variable appear to be net profit generated by a unit of sales, negatively associated to the sustainability score of the company. Castellano et al. (2024) find significant prediction results using random forest classification models. They predict the ESG rating class of a company at each year starting from a set of financial statement variables and a dynamic measure of systemic risk observed at the previous year. They find that the most influential data are Total Equity, ROE and EBITDA, which are essential indicators for sustainable investors as they offer insights into a company's sustainability profile, along with its long-term profit generation capabilities.

Our paper contributes to the existing literature concerning ESG ratings transparency aiming to replicate the LSEG methodology of scoring issuance and quantifying the main drivers of ESG ratings in terms of forward-looking promises (aspirational) and backward-looking realizations (performance) balance. By replicating step-by-step the LSEG ESG scoring algorithm and including all ESG-related granular data available in the LSEG dataset for all companies from 2010 to 2023, as captured in a single snapshot downloaded in September 2023, we employ machine learning models to predict companies' sustainability ratings, providing suitable feature selection to highlight the set of the most relevant metrics in the ESG rating issuance procedure. We stress that our methodology is not meant to replicate fully the exact aggregation method, which is not fully disclosed, but to analyze its drivers and its functioning. We then address the problem of quantifying the importance weight of each measure in the final ESG score by replicating the proprietary (undisclosed) Materiality Matrix, which represents the relative importance of each theme to each individual industry group. Lastly, exploiting Bams and van der Kroft (2022)'s methodology to assess forward-looking promises and backward-looking realization data within LSEG ESG measure set, we quantify our prior results in terms of aspirational/performance ESG balance. We point out that the aim of our paper differs from Bams and van der Kroft (2022). While Bams and van der Kroft (2022) focus on the impact of the two classes of indicators on socially responsible investors choices, finding that the shift of investors' portfolios towards firms with high ESG ratings provides cost of capital incentives for firms to increase their ESG rating without improving sustainable performance, we examine the proprietary methodology unveiling the contributions of ESG promises and backward-looking indicators to the final ESG scores. We find that approximately 60% of ESG scores relies on average on promises, while the remaining 40% depends on backward-looking performance, implying a significant over-reliance on promised ESG policies rather than on actual past sustainability performances. This result is crucial for sustainable investors in determining whether ESG ratings align with their investment and impact goals.

Our findings emphasize the need for greater transparency and a clearer distinction between aspirational and performance metrics to ensure credible ESG assessments for informed investment decisions. Interpreting ESG scores as purely concrete progress towards sustainability could lead to resource misallocation and to dispersing the social benefits of such ratings due to information asymmetries (Hartzmark and Sussman 2019; Bams and van der Kroft 2022; Fliegel 2025). The imbalance between aspirational and performance metrics highlights a potential bias in ESG ratings towards future commitments. Within this framework, firms may continue to improve their ESG ratings even while facing ESG controversies, scaling back sustainability commitments, or operating under deteriorating labor conditions and governance structures. This imbalance can potentially overshadow actual sustainability achievements and transform ESG ratings to a "ticking-the-boxes" exercise rather than an actual long-term value creation opportunity (Edmans 2024), or even greenwashing (Kathan et al. 2025). The paper closest to ours is Guerrero and Viteri (2025) which examines the contributions of social and environmental outcome and impact indicators to the ESG scores. Similarly to us, they find that outcome and impact indicators contribute only 18%–37% of the total score, while the remaining contribution comes from processes and policies indicators. We expand their analysis considering the changes of such dependency for each pillar and year from 2010 to 2023 replicating closely the LSEG methodology, offering more granular insights on the functioning of the ESG scores. We point out that our analysis focuses exclusively on the construction and functioning of ESG ratings, and does not incorporate adjustments from LSEG's ESG controversies overlay. This exclusion is intentional, as the overlay is applied ex post and serves as a separate corrective mechanism. Including it would confound our focus on the primary ESG signal generated by the rating methodology itself, which represents the aim of this paper.

It is important to note that, following the acquisition of Refinitiv by LSEG in 2021, several updates to the ESG scoring methodology have been introduced. While the core structure of the ESG score has remained largely consistent, LSEG has progressively aligned its approach with emerging regulatory frameworks—most notably the European Union's CSRD, the SFDR, and the EU Taxonomy. These changes include the adoption of a double materiality perspective—evaluating both the financial impact of ESG risks on the company and the company's impact on society and the environment—and

15353966, 2025, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/csr.70007 by CochraneItalia, Wiley Online Library on [05/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the expansion of the indicator set, which grew from approximately 630 metrics (pre-2021) to over 870 by 2024. Given these dynamics, we clarify that our analysis is based on a single snapshot of the dataset, downloaded in September 2023. At the time of data extraction, the LSEG ESG dataset was still based on the pre-expansion methodology, comprising approximately 630 indicators, and had not yet integrated the extended indicator set introduced in 2024, nor the full transition to a double materiality framework, which LSEG reported in 2025 to be in progress. Moreover, LSEG retroactively updates historical scores when new data or methods become available. This score rewriting practice makes it difficult to isolate actual temporal shifts in firm behavior from methodological changes. Although the snapshot includes retroactive adjustments made up to that date, it does not allow us to disentangle true score evolution from methodological shifts. Nonetheless, using a single, retroactively updated dataset enables consistent comparisons across years and supports the analysis of temporal patterns in feature relevance within a unified scoring framework.

The remainder of the paper is organized as follows. Section 2 introduces LSEG ESG scores and elaborates on the disclosed aspects of their assessment methodology. Section 3 provides a detailed overview of the dataset used in this study. Sections 4 and 5 present the core results of the paper, focusing on feature selection and the balance between aspirational and performance-based ESG factors, respectively. Finally, Section 6 concludes the paper by discussing the key findings of this study.

2 | LSEG ESG Ratings

LSEG is a global provider of financial market data and infrastructure, serving financial institutions, corporations, and governments. It was formed in 2018 from the rebranding of Thomson Reuters' Financial and Risk business, but its ESG dataset originates from the work of the financial data provider ASSET4, specializing in ESG data and analytics, acquired by Thomson Reuters in 2009. As of the date of our data snapshot (September 2023), the LSEG ESG database covered over 85% of the global market capitalization, included more than 630 ESG metrics, and provided historical data dating back to 2002 (LSEG 2022). LSEG ESG ratings are available today for over 12,500 public and private companies globally, and they are expressed as percentile rank scores ranging from 0 (worst) to 100 (best). These ratings aim to provide an objective assessment of a company's relative ESG performance, commitment, and effectiveness.

LSEG asserts that its ratings are entirely data-driven, and they rely on company-disclosed data available on their websites, reports, media, and news, as well as on ESG news by global media and NGOs websites. While ESG scores are published annually, similarly to the non-financial disclosures, they undergo weekly updates and remain subject to alterations for up to five fiscal years before being finalized. This process of data rewriting is influenced by company restatements, data corrections, and post-publication changes in the aggregation rule implemented by LSEG. These changes affect tests that relate ESG ratings to returns, potentially impacting ESG-driven investments (Berg et al. 2021).

As of the date of our data snapshot (September 2023), LSEG computation started by collecting 630 data points from websites, reports, media, and news, and selecting a subset of 186 metrics, see Figure 1. To incentivize transparency and full disclosure, LSEG penalizes missing data in two distinct ways for boolean and numerical data (LSEG 2022). For boolean data, missing values are replaced with the worst boolean value for each specific variable before the ranking transformation is computed. Consequently, non-disclosed data carry the same impact as non-positive disclosed data, incentivizing companies' disclosure. For numerical values, companies that do not disclose any information are excluded from the feature ranking transformation and are subsequently assigned the worst

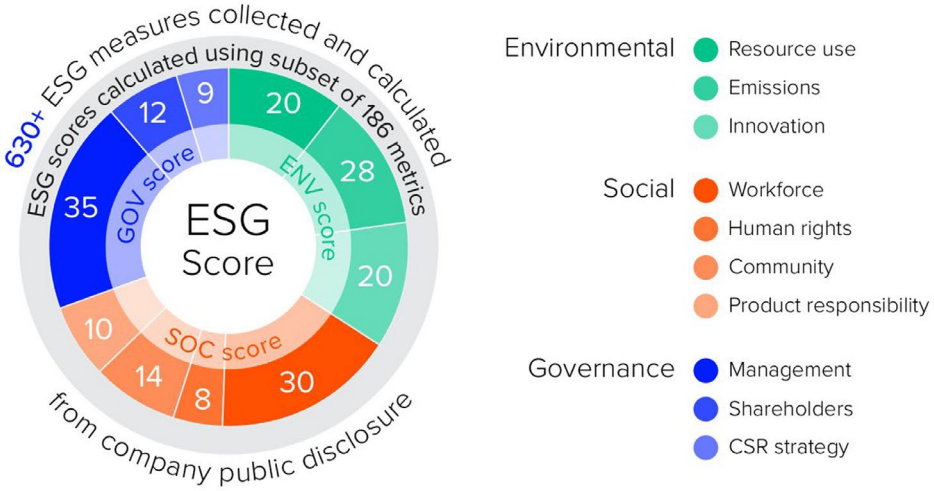


FIGURE 1 | LSEG ESG scores composition: Out of more than 630 data points, 186 are selected to contribute to the ESG score computation. Metrics are split into three pillars and 10 categories. Numbers in the pie chart represent the dimension of each category in terms of metrics used. Each category subset of metrics generates a category score for all companies evaluated. Category scores are then combined into pillar scores and finally into an overall ESG score. *Source:* LSEG (2022).

15353966, 2025, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/csr.70007 by CochraneItalia, Wiley Online Library on [05/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

possible ranking value. In this scenario, unless a company exhibits the worst performance in the entire ranking cluster, there is a penalty imposed for companies that opt not to disclose data. Notice that, under the applied weighting scheme, the absence of disclosure on immaterial data points has a limited impact on a company's overall score, whereas failure to report on highly material indicators results in a significant negative effect on the score.

Furthermore, in accordance with LSEG guidelines, each variable necessitates a minimum disclosure threshold from all companies within the same industry group or country of incorporation to be considered relevant for computing the ESG category score. Notice that industry groups are defined according to The Refinitiv Business Classifications (TRBC), for which we present a complete list in Appendix A. This threshold stands at 5% for numerical variables and at 7% for boolean variables (LSEG 2022).

Once the initial dataset is completed, all raw data, both numerical and categorical, are ranked and transformed according to the comparative formula:

$$\text{Percentile score} = \frac{\text{no. of companies with worse value} + \frac{\text{no. of companies with same value}}{2}}{\text{no. of companies with a value}}. \quad (1)$$

This ranking system allows a fairer comparison among companies of varying sizes and structures. LSEG approach does not dictate what constitutes "good" performance. Instead, it evaluates companies based on their performance relative to their industry or country peers within the ranking system. Consequently, to achieve a favorable score, companies must outperform their industry competitors. This methodology ensures that the assessment of ESG performance is contextually relevant and reflective of industry standards, rather than imposing arbitrary benchmarks (LSEG 2022).

After filling missing values and transforming data with Formula (1), LSEG identifies a set of key features within each category (Figure 1) and utilizes them to determine scores for each category. More precisely, at the end of the ranking process, each company is associated with a percentile score for each relevant metric, indicating its performance relative to other companies in the same group (TRBC sector for the Environmental and Social pillars; country of incorporation for the Governance pillar). These percentile scores are summed up at the company level, providing an overall assessment of each company's performance within each category. The ranking Formula (1) is once again utilized to convert this category performance into a percentile score, known as the ESG category score.

The final step of the LSEG algorithm involves transforming the 10 category scores for each company into a single ESG score. This transformation utilizes a Materiality Matrix, developed as a proprietary model by LSEG, which contains the importance weights of each category score in the composition of the final ESG scores. These weights vary for each industry group; therefore, the Materiality Matrix assigns the appropriate set of weights to each company based on the industry in which they operate. Specifically, TRBC sectors determine company clusters at this stage and serve as the only clustering criteria, as Countries of Incorporation are not relevant for category weights assessment.

A primary drawback of this ranking system lies in its lack of memory when assessing a company's performance. Since companies are ranked against each other based on their yearly performance, there is no consideration given to the ESG evolution of individual companies over time. Consequently, there is no direct link between a company's past ESG score and its recent one within this framework. Without such historical context, the rankings may fail to capture long-term trends or improvements in a company's ESG performance, potentially leading to inconsistencies or inaccuracies in the assessment process.

A secondary issue that affects the transparency of LSEG methodology is the disclosure of the Materiality Matrix. An example of such matrix is presented in the methodology report of LSEG (2022), however the one utilized by the agency in the ESG rating issuance algorithm remains a proprietary undisclosed information. In Section 5, we use our algorithm to reconstruct this matrix: understanding the weights associated with each category to build up the final ESG score is essential information for investors to understand if such ratings align with their sustainability goals.

Further limitations of such rating system concern (i) limited inclusion of controversies, which have been shown to bring valuable information (Aouadi and Marsat 2018; Aste et al. 2024); (ii) changes in the importance of different ESG criteria over time due to innovations in the rating issuance methodology, impacting the alignment to specific investment goals, such as climate action or social impact (Eskantar et al. 2024); (iii) LSEG's ESG ratings have been shown to be unstable over time due to rewriting changes (Berg et al. 2021); (iv) finally, the methodology appears unclear in the balance between the forward- and backward-looking metrics.

In our paper, we decide to focus on the last of these underinvestigated issues, which may limit their utility for investors aiming to make sustainable impacting investments. We decide to work on LSEG data provider to shed more light on such limitations and since, to our knowledge, LSEG comprises the most comprehensive scope of granular ESG information, ensuring a thorough understanding of the whole underlying scoring methodology. Using such a granular dataset enables us to match ESG policies, targets, activities, performance, and controversies on similar aspects of ESG, as in Bams and van der Kroft (2022). Moreover, to address the third issue, we decide to perform a yearly analysis.

3 | Data

We consider ESG yearly data from 2010 to 2023 of LSEG data provider. We download all the available ESG information published for each rated company as of September 29, 2023. In Table 1 we report the number of companies included in the analysis and the average percentage of missing ESG information by pillar. We observe an yearly increase in the market coverage of LSEG's database, passing from 2352 firms in 2010 to 11,443 in

15353966, 2025, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/csr.70007 by CochraneItalia, Wiley Online Library on [05/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 | Dataset dimensions: Number of companies covered by LSEG database by year (first row of the table) and missing data percentage average by pillar.

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Companies	2352	2775	3083	3237	3365	3526	4029	4748	5823	6685	7756	9045	10,424	11,443
Environmental	73.37%	72.89%	72.42%	72.13%	72.09%	71.46%	71.69%	68.31%	67.41%	64.33%	62.84%	62.05%	61.11%	60.59%
Social	69.71%	69.52%	69.22%	68.80%	68.52%	68.45%	68.58%	68.43%	66.97%	65.54%	64.42%	57.77%	55.84%	55.41%
Governance	39.70%	38.81%	38.13%	37.43%	37.01%	36.86%	36.44%	35.93%	35.11%	33.19%	24.63%	22.58%	21.93%	21.69%

2023. Regarding data availability, we emphasize a significant improvement from 2019 onwards, with a consistent reduction of 10% points in the prevalence of missing data across all three ESG pillars. Additionally, our analysis reveals that Governance features are less impacted by missing data compared to features in the other two pillars, with around 22% average missing data in 2023.

For each company in the dataset, we collect its TRBC Industry Group classification, which serves as a fundamental input for the computation of scores in the Environmental and Social pillars. In fact, Environmental and Social pillar scores compare companies belonging to the same industry. The TRBC classification encompasses 59 industry groups, which are clustered according to the 13 TRBC economic sectors they operate in, listed in Appendix A.

We also include the country of incorporation for each company in the dataset. This information plays a pivotal role in LSEG methodology for computing the Governance score. Given that LSEG methodology does not disclose full information about the grouping techniques by country for Governance pillar categories, we conduct an alternative clustering approach based on geographical macro-regions. To ensure that each cluster is sufficiently populated, we split our dataset into seven clusters: North America, Latin America, Europe, Africa and the Middle East, Asia, Japan, and Oceania, see Figure 2. Introducing clusters of countries, rather than considering individual countries, is necessary to prevent having too few companies in the analysis. This choice is not arbitrary, but rather a methodological necessity aimed at ensuring statistical robustness and avoiding overfitting in the reconstruction of the ESG rating logic: since the aggregation logic used by LSEG is proprietary and undisclosed, a reliable reverse-engineering requires a sufficient number of observations per group. In many country-year combinations—particularly in earlier periods—the number of rated firms is too small to support statistically stable estimation. Aggregating countries into broader macro-regions allows us to maintain statistical robustness while still capturing meaningful geographical patterns.

3.1 | Data Preprocessing

Following the data provider methodology presented in Section 2, we firstly exclude from our dataset all variables with a missing value percentage lower than the LSEG thresholds (5% for numerical variables and 7% for boolean variables). In Table 2 we report the number of metrics with data availability exceeding the minimum disclosure percentage for each year, and therefore that are considered in the methodology. It is noteworthy how categories exhibit different trends over time concerning the completeness of their feature data. Categories such as *emissions*, *CSR strategy*, *product responsibility*, and *workforce* exhibit a significant increase in data availability between 2019 and 2021, indicating a notable rise in the amount of data for these areas. In contrast, categories such as *innovation*, *resource use*, *community*, *human rights*, *management*, and *shareholders* demonstrate consistently high availability and more stable behavior, suggesting that these categories have maintained robust and well-populated data over a longer time period.

15353966, 2025, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/csr.70007 by CochraneItalia, Wiley Online Library on [05/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

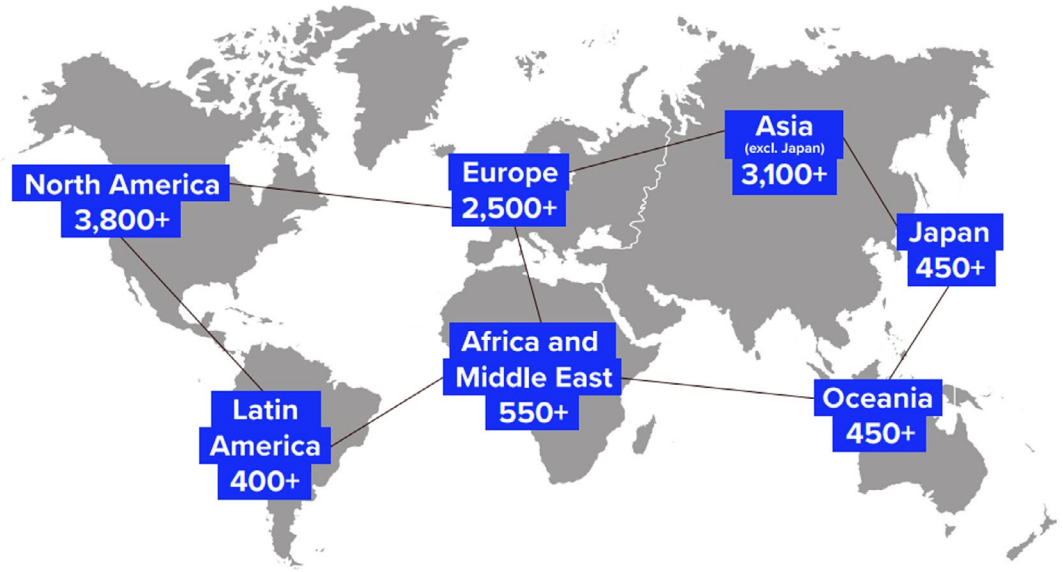


FIGURE 2 | Geographical clusters for companies' countries of incorporation. *Source:* LSEG (2022).

TABLE 2 | Number of features exceeding the minimum disclosure threshold by year.

Category	Available features	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Emissions	192	59	59	62	62	62	62	61	107	109	115	119	124	126	128
Innovation	42	31	30	30	30	30	30	29	29	29	29	29	29	29	29
Resource use	60	35	35	35	35	35	35	35	37	37	39	39	39	39	39
Community	42	27	27	28	28	28	28	23	22	22	22	22	21	21	21
Human rights	17	12	12	12	12	12	12	12	12	12	12	12	12	12	12
Prod. resp.	151	30	29	29	30	30	30	30	30	36	36	36	57	57	57
Workforce	90	58	58	60	60	60	60	60	60	61	67	68	69	75	76
Management	78	67	67	68	68	67	67	67	67	67	67	68	68	70	73
Shareholders	56	43	43	43	43	42	42	43	43	43	44	44	42	42	42
CSR strategy	32	11	12	12	12	12	12	12	12	13	13	30	30	30	30

Note: Categories are grouped according to their respective ESG pillars. The Environmental pillar includes *emissions*, *innovation*, and *resource use*. The Social pillar encompasses *community*, *human rights*, *product responsibility* (prod. resp.), and *workforce*. The Governance pillar comprises *management*, *shareholders*, and *CSR strategy*. See Figure 1 for reference.

In order to follow as closely as possible LSEG methodology, we do not perform any data transformation, except for ranking percentile scaling. This is done considering Formula (1). For assigning a meaningful polarity to our data, that is, for assessing if a high (low) value of a feature corresponds to a good (bad) sustainability performance of the company, we compute the correlation of each feature with the target category score. When a variable exhibits a positive (negative) correlation with its category score, we assume that the polarity is positive (negative), indicating that a higher value implies a better (worst) sustainability profile. Finally, for missing values of boolean features, we follow LSEG ratio and assign a negative performance of the company, that is, 0 value, while for numerical features, we assign the worst scoring during the ranking transformation, exploiting the polarity information.

4 | ESG Score Calculation and Feature Selection

Identifying the relevant variables for each ESG category is a critical step in understanding ESG ratings. This process not only clarifies which attributes companies should prioritize to enhance sustainability, but also uncovers the key drivers influencing these priorities.

Our objective is to identify a subset of relevant features for each ESG category and for each year, starting from the larger set of all variables. It is important to emphasize that LSEG does not disclose the specific features selected for each category, but only the total number of selected features, as shown in Figure 1. Therefore, our aim is to provide insights into the feature selection process utilized

15353966, 2025, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/csr.70007 by CochraneItalia, Wiley Online Library on [05/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

by LSEG. Additionally, by applying our algorithm to datasets from different years, we aim to assess the stability and evolution of LSEG ESG rating methodology over time.

To achieve this, the first step involves constructing an algorithm capable of computing the ESG category scores, closely replicating LSEG proprietary ESG algorithm. Building on the approach outlined by Del Vitto et al. (2023), we implement a random forest model for each year's dataset and each category, regressing the full set of ranking-scaled variables onto the respective category score.

A random forest is a machine learning ensemble algorithm that combines multiple decision trees to make more accurate predictions and reduce overfitting, albeit at the cost of reduced interpretability, as it is considered a black-box model. It is used for both classification and regression tasks. The latter is the case of this work, since the random forest is trained to predict the category score by averaging the predictions from all the trees. Therefore, a random forest is an ensemble of multiple decision trees, meaning that it combines the outputs of several trees to make a more robust prediction. Each tree in the forest is trained on a different subset of the original data (bagging), selected through bootstrapping (random sampling with replacement). Therefore, every time an observation is selected, it is put back into the pool before the next selection. As a result, the subset is of the same size as the original training dataset, but it contains duplicate entries of some data points, while others may not be included at all. This introduces randomness and reduces the variance of the model, preventing overfitting. In addition, random forest adds another layer of randomness by selecting a random subset of features at each split in the decision tree. This reduces the correlation between the trees, leading to a more diverse and robust ensemble.

Given that a random forest is a set of decision trees, it is important to understand how a regression decision tree works. The tree starts from a root node and grows by making decisions based on features that split the data in a way that minimizes the Mean Square Error (MSE) within the child nodes. The tree continues splitting until it reaches a stopping criterion (such as maximum depth or minimum number of samples per node). The MSE is defined as:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}\right)^2$$

where y_i is the actual value of the i-th data point, $\hat{y}$ is the predicted value (i.e., the mean of the category score in the node), and N is the number of data points in the node.

For a given split that divides the dataset D into two child nodes, D_{left} and D_{right}, the Impurity Gain is computed. It is defined as the reduction in RMSE before and after the split, that is

$$\text{Impurity gain} = \mathrm{MSE}_{\mathrm{parent}} - \mathrm{MSE}_{\mathrm{split}}$$

where MSE_{parent} is the MSE of the parent node, while

$$\mathrm{MSE}_{\mathrm{split}} = \frac{N_{\mathrm{left}}}{N}\mathrm{MSE}_{\mathrm{left}} + \frac{N_{\mathrm{right}}}{N}\mathrm{MSE}_{\mathrm{right}}$$

N_{left} (N_{right}) and $\mathrm{MSE}_{\mathrm{left}}$ ($\mathrm{MSE}_{\mathrm{right}}$) being the number of data points and the MSE in the left (right) child node respectively, with $N_{\mathrm{left}} + N_{\mathrm{right}} = N$, the number of data points in the parent node. Therefore for each node, the algorithm finds the feature and the split rule (e.g., value smaller than a threshold) maximizing the Impurity Gain.

It is important to emphasize two critical advantages unique to random forests: their built-in mechanism for evaluating feature importance—which provides insights into variable significance—and the out-of-bag (OOB) error estimation, which enables robust internal validation without the need for separate validation datasets. Therefore, once the random forest model is trained, conducting a feature importance analysis becomes straightforward. Feature importance is computed as the total decrease in node impurity brought by each feature, summed over all trees in the forest. For each split, the impurity reduction is weighted by the number of samples that reach the parent node, reflecting the impact of the split. These totals are then normalized so that the importances across all features sum to one.

Summarizing, a step-by-step description of our feature selection algorithm is described below:

1. Each dataset pertaining to a specific ESG category and a particular year of data is provided independently as input to the random forest algorithm. Initially, each dataset contains all potential features that may be influential in computing the given ESG category score.
2. The model is trained using 80% of the data points and tested on the remaining 20%. The RMSE, that is, the square root of the MSE, is computed on the test set, as the accuracy of the score prediction.
3. Once the model is trained, all input features are ranked based on their importance for predictions.
4. The least important feature for the model is dropped from the dataset, and the procedure is iteratively repeated from step 2 until only one feature remains in the dataset.

A key point in the training of a random forest is the definition of the hyperparameters of the algorithm. As a first step, the selection of the hyperparameters is performed via cross-validation, with an independent tuning for each year's dataset and each ESG category. The category-specific tuning is performed to compare parameter selection for categories with a high density of variables (e.g., *emissions*) against those with fewer initial variables (e.g., *human rights*). On the other hand, differentiation by reference year is intended to assess the effectiveness of tuning for more recent years with larger datasets compared to earlier years, which typically feature less data. Once category/year specific tuning is performed, to ensure comparability, we apply the same parameter tuning for all category-year pairs, opting for the configuration that performs best across all datasets, that is

- *n_estimators*: The number of trees in the forest is set to *200*;

15353966, 2025, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/csr.70007 by CochraneItalia, Wiley Online Library on [05/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 | Grid of hyperparameters used in the cross-validation procedure.

Hyperparameter	Values
n_estimators	100, **200**, 300, 400, 500
max_depth	10, 20, 30, **None**
min_samples_split	**2**, 5, 10
min_samples_leaf	**1**, 2, 4

Note: Selected values are indicated in bold.

- *max_depth*: The maximum depth of each decision tree in the forest is set to *none*, therefore the trees are allowed to grow without any restrictions on their depth;
- *min_samples_split*: The minimum number of samples required to split an internal node is set to 2;
- *min_samples_leaf*: The minimum number of samples required to be at a leaf node is set to 1.

Notice that the last two hyperparameters allow any possible split: a split can be done if the parent node contains at least two samples, and any split having child nodes with at least one sample each is considered. All tested hyperparameters are reported in Table 3.

The decision to use a single set of hyperparameters is also made to align as closely as possible with LSEG methodology, where all category scores are derived using the same algorithm.

4.1 | Results

In this section we present the results of the feature selection analysis. Firstly, we discuss the evolution of the RMSE by category over time and the number of features selected by our algorithm compared to the ones disclosed by LSEG. Secondly, we evaluate the stability of the selected features in each category along the time interval considered.

Figure 3 displays the main results of our feature selection analysis. More precisely, we report the evolution of RMSE by category for each year. The x-axis represents the number of input features for each random forest iteration, while the y-axis displays the RMSE computed on the test set. The dashed blue vertical line represents the number of features used in the ESG scoring process disclaimed by LSEG (see Figure 1), while the solid red line represents the minimum number of features necessary to have a small RMSE according to our algorithm, exploiting an elbow approach: the elbow of the plot is the point where adding an additional feature does not lead to a significant reduction of the RMSE.

The analysis of RMSE spots an elbow behavior (in agreement with Pérez-Martín et al. (2020)) in each category considered and for every year in the analysis. We notice that the RMSE decreases consistently up to the solid red line and then becomes flat. The features belonging to the decreasing region are those relevant for explaining the category scores, while those in the flat region are redundant data. This pattern underlines that in every category score the number of relevant features that can explain the final category score is limited and conservative, especially for categories that include a high number of initial features. We further notice that for categories in which the number of input features increases considerably over time (e.g., *emissions category*, *workforce category*, *community category*, *product responsibility category*) the elbow pattern does not change, leading to the same number of relevant data for the whole time interval considered. This implies that the new information included in the dataset by LSEG appears not to add significant variability in the final category scores.

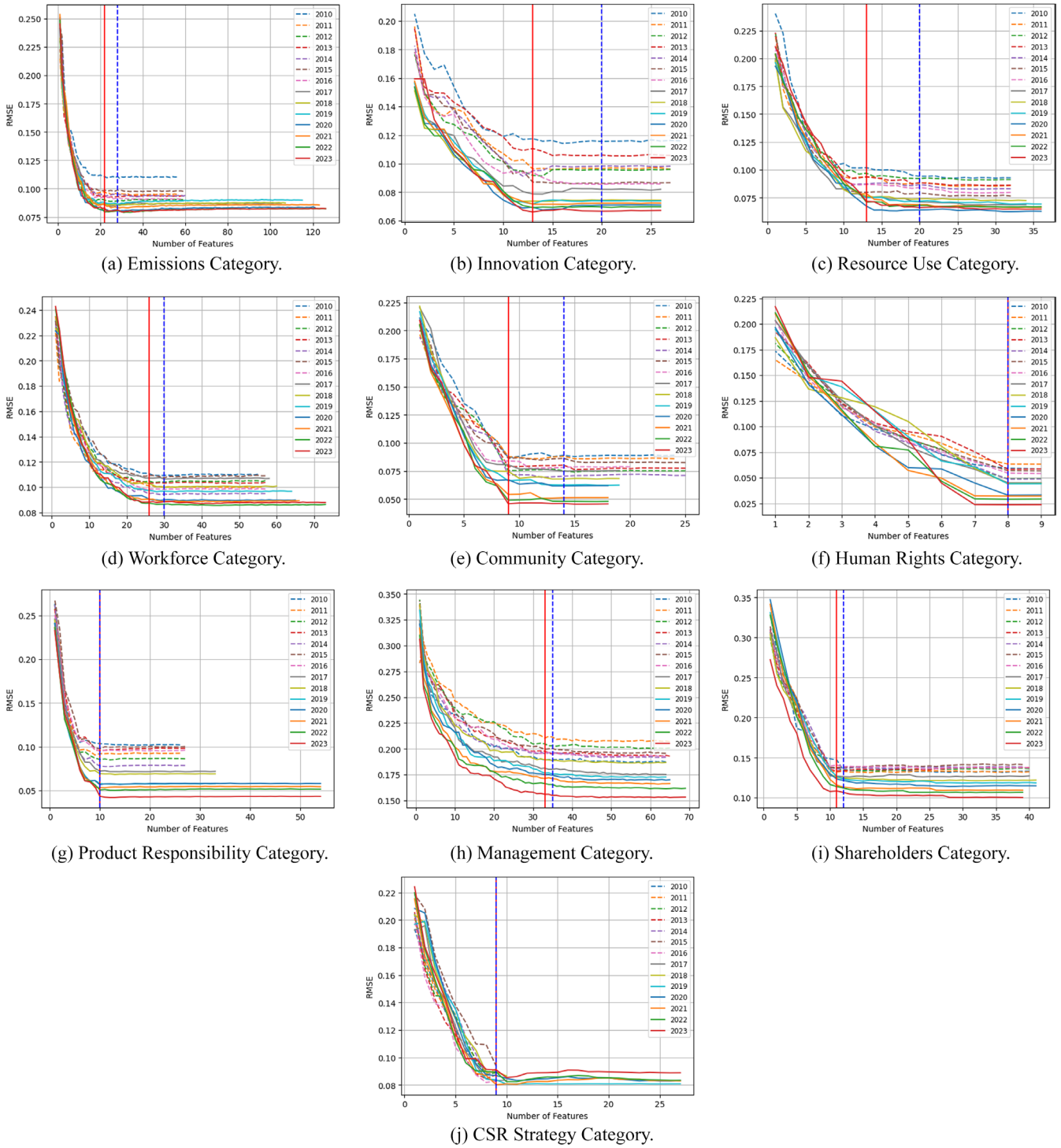
Comparing the number of features selected by the random forest algorithm (solid red line in Figure 3) to the ones declared by LSEG (dashed blue line), we notice that our model requires fewer input variables to attain comparable performance results for most of the categories, except for *product responsibility* and *CSR strategy* in which both models employ the same number of input data. This difference suggests that the set of measures strictly necessary for the LSEG algorithm is indeed smaller than claimed, meaning the usage of some redundant features in the proprietary model whose variability is already included in some others. This reduction in the number of essential variables could have important practical implications, such as lowering data collection costs, enhancing transparency, and enabling more focused and efficient ESG reporting.

Furthermore, the RMSE analysis reveals that random forest models predicting the ESG category scores within the Governance pillar, that is, *management*, *shareholders*, and *CSR strategy* categories, consistently underperform the rest of the category predictions (higher RMSE), confirming the results of Del Vitto et al. (2023).

Finally, as evident in the steeper RMSE decline for recent years in Figure 3, we notice that the random forest model returns systematically higher RMSE when processing older data (compare dotted and dashed lines of RMSE by year, associated respectively to years from 2010 to 2016 and from 2017 to 2023). Despite the similarity in RMSE general behavior across years, the most precise prediction results are consistently achieved with the latest years' data. A possible explanation for this behavior may be the fact that older datasets tend to have fewer data points (firms) and a higher prevalence of missing data or from the rewriting of past ESG scores based on newly disclosed data, restatements, or updates in the aggregation logic as stated in Berg et al. (2021).

The second step of our feature importance analysis concerns the stability of the results overtime. In particular, we assess if the set of relevant features is consistent across different years, that is, if the most important features do not change over time. Table 4 reports the total number of features marked as relevant according to our algorithm (Number of Selected Features), as well as how many of these features are selected consistently throughout all the 14 years (relevancy count). In the last column (average relevancy), we also address the stability of the selected features over time. Given the feature frequency, defined as the number of years in which the feature is selected, divided by the total number of years, the Average Relevancy

15353966, 2025, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/csr.70007 by CochraneItalia, Wiley Online Library on [05/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



(a) Emissions Category. (b) Innovation Category. (c) Resource Use Category.

(d) Workforce Category. (e) Community Category. (f) Human Rights Category.

(g) Product Responsibility Category. (h) Management Category. (i) Shareholders Category.

(j) CSR Strategy Category.

FIGURE 3 | Evolution of RMSE by Category. The *x*-axis represents the number of input features used in each random forest iteration, while the *y*-axis shows the RMSE computed on the test set. Dashed lines correspond to datasets from 2010 to 2016, while solid lines correspond to datasets from 2017 to 2023. The blue dotted vertical line indicates the number of features selected by LSEG for the given category, whereas the solid red vertical line marks the number of features selected by our algorithm.

is obtained by sorting the features from the highest frequency to the lowest, and averaging the first n feature frequencies, n being the number of selected features (second column of the table). An example for the *emissions* category is reported in Table A1 in Appendix B. An average relevancy close to 1 indicates a minimal turnover of the selected features across the years, suggesting a stable and consistent selection process over time, which is our case: we notice that the majority of the selected features remain constant over time, and that the selected features always show an average relevancy very close to 1 (and equal to 1 in one case, that is, the selected features are the same in all 14 years). This result suggests a relative stability in the logic underlying the LSEG ESG scores over time. This observed consistency could reflect both the structure of

15353966, 2025, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/csr.70007 by CochraneItalia, Wiley Online Library on [05/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 | Feature relevance over time by category.

Category	Number of selected features	Relevancy count	Average relevancy
Emissions	22	16	0.92 (0.42)
Innovation	13	10	0.93 (0.59)
Resource use	13	10	0.97 (0.52)
Community	9	8	0.98 (0.52)
Human rights	8	8	1.00 (0.77)
Product resp.	10	8	0.97 (0.44)
Workforce	26	19	0.95 (0.56)
Management	33	21	0.92 (0.63)
CSR strategy	9	7	0.98 (0.38)
Shareholders	11	9	0.97 (0.78)

Note: The table reports the number of features selected by our algorithm (number of selected features), how many of them are consistently selected across all 14 years (relevancy count), and the average frequency of the top n selected features (average relevancy). Values in parentheses indicate the expected average relevancy under random feature selection, serving as a baseline for comparison.

the scoring methodology and the stability of the underlying data, as captured in the retroactively updated snapshot.

5 | Aspirational Versus Performance

It is crucial for sustainable investors to understand if ESG ratings align with their investment goals. In the LSEG algorithm framework, this can be achieved by exploiting the information retained in the undisclosed Materiality Matrix, which represents the relative importance of each category at an industry group level. In this section, we firstly propose a methodology to estimate the Materiality Matrix and then we use it to understand how much of the final ESG ratings information is driven by forward-looking promises (aspirational) and backward-looking realizations (performance) features. This result is essential for practitioners to understand the true nature of ESG ratings, that is, to have a quantification of how much this financial tool is based on policy-oriented or promised realizations of sustainability goals of the company and how much on realized sustainability performances.

Since the Materiality Matrix provided by LSEG is defined at the level of TRBC Industry Groups, we extend our analysis from broader TRBC economic sectors to these more granular groupings in order to better reflect the structure of the underlying scoring methodology. This is also possible, since, as shown below, in this section we employ a multiple linear model, which remains statistically viable even with the smaller sample sizes associated with TRBC Industry Groups. However, fragmenting each yearly dataset into TRBC Industry Groups reveals the scarcity of company data for some sectors, particularly in the earliest years, leading to the risk of overfitting across various industry groups. To address this issue, a minimum threshold of 30 companies is set for each TRBC group: industry groups that fail to meet the minimum company threshold are excluded. This is the reason why this analysis is made using the scores from 2020 to 2023, given that we have more available data in those years: in fact, out of the 59 TRBC industry groups, only 8 fail to meet the minimum threshold of 30 companies when restricting the analysis to the 2020–2023 period. By contrast, extending the time horizon to 2010–2023 would require excluding 29 sectors, as each of these has at least 1 year in which the threshold is not met.

The first step of this analysis is to reconstruct an approximation of LSEG undisclosed Materiality Matrix. This matrix assigns to each sector group the weight of each category score in the final ESG score. Since this matrix may vary from year to year and is not disclosed by LSEG, our aim is to replicate it using the linear regression weights that link the independent variables (i.e., category scores) to the target variable, namely the ESG scores. The Materiality Matrix provides insights into the true importance of the underlying category scores, thereby shedding light on their initial features and enabling considerations regarding the entire ESG rating process, end to end.

During this phase, we opt to utilize the actual category scores disclosed by LSEG rather than our predicted ones. This decision is made to avoid introducing external noise into the model. Furthermore, it provides an opportunity to test the linearity between input and output, a framework that LSEG claims to respect according to its methodology.

We leverage many variants of linear regression, including multiple linear, Ridge, and Lasso regressions. Remarkably precise results are obtained with all of these methods. However, for the sake of model simplicity, we proceed with multiple linear regressions. Independent and identical models are developed for each year considered and each TRBC Industry Group. In each model, all 10 ESG category scores are linearly combined for every company to predict the final ESG score. Thus, the multiple linear model is defined as:

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + \ \dots \ + \beta_{10} x_{10,i} + \epsilon, \qquad (2)$$

with:

- y_i: ESG Final Score of the i-th company,
- β_0: intercept term,
- $\beta_1, \beta_2, \beta_3, \ \dots \ , \beta_{10}$: coefficients of the regression,
- $x_{1,i}, x_{2,i}, x_{3,i}, \ \dots \ , x_{10,i}$: ESG Category Scores of the i-th company,
- ϵ: error term.

To accurately represent the linear combination of category scores as a weighted sum in Equation (2), we impose the following constraints:

$$\text{Condition 1:} \quad \beta_0 = 0$$

$$\text{Condition 2:} \quad \sum_{i=1}^{10} \beta_i = 1$$

$$\text{Condition 3:} \quad \beta_i \geq 0 \quad \text{for } i = 1, 2, \ \dots \ , 10.$$

15353966, 2025, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/csr.70007 by CochraneItalia, Wiley Online Library on [05/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The second step of the analysis concerns the quantification of backward- and forward-looking drivers in the final ESG scores. We follow Bams and van der Kroft (2022)'s feature classification to categorize each feature into two distinct groups: *performance* and *aspirational*. The first denotes quantifiable results already attained by the company, as documented in its raw data. These features reflect the company's past performance or specific outcomes it has achieved. Aspirational features, on the other hand, represent the company's intentions, policies, reporting initiatives, or future targets. They indicate forward-looking statements or commitments made by the company concerning its ESG objectives.

Each feature is individually analyzed according to its description reported on the LSEG platform and labeled according to Bams and van der Kroft (2022)'s classification methodology. Note that the latter does not provide a classification for all the features selected by our algorithm, therefore we extend their methodology for the missing values[1]. We show the complete information for the *emissions* category in Table A1 of Appendix B. Information about the other categories is available in the Supporting Information.

We point out that our binary classification into aspirational and performance features may not fully capture all nuances. Nevertheless, our analysis is intended as a practical, first-order approximation of the aspirational/performance balance in ESG ratings, not a definitive statement on metric semantics. Moreover, due to the absence of granular information on how LSEG dynamically adjusts weights based on the interaction of these elements and the continuous update of the ESG regulatory framework, a more fine-grained classification was not feasible.

5.1 | Results

As a first step, we replicate the Materiality Matrix employed by LSEG in the ESG scoring assessment process. It is composed of the weights associated with each category score by industry group. Once the Materiality Matrix is defined, the final ESG score is just a linear combination of the category scores. We also provide a yearly estimation of the matrix, showing the results for 2023 in Appendix C. Results for the other years are available upon request.

In Figure 4, we present the estimated Materiality Matrix for the period 2020–2023. We observe that while the weights for Environmental and Social categories exhibit higher heterogeneity across different sectors, Governance categories display a more uniform weight distribution across various industry groups. Notably, the *management* category plays a pivotal role compared to the other two governance categories, as it is associated with significantly higher weights. Numerical results not included here for brevity indicate a very low yearly variance in the weights of the Materiality Matrix, suggesting that it remains largely stable over time with no changes or minimal changes.

We finalize our analysis by delving into the nature of each significant feature and quantifying the amount of aspirational/performance information in the final ESG ratings. Firstly, we split the initial features into aspirational and performance data. We follow Bams and van der Kroft (2022) feature classification as detailed in the previous section. Table 5 illustrates the distribution of significant features selected by the random forest algorithm, categorized as either forward-looking (aspirational) or backward-looking (performance). For each category, we report the number of features selected by the random forest algorithm, as discussed in Section 4, and the division of these features into aspirational or performance-related. Importantly, this classification is applied only to the subset of features selected by our model, that is, those identified as most relevant through the elbow-based selection criterion, see Figure 3. As such, we focus only on the features that contribute meaningfully to score prediction. While this approach does not capture the exact weight assigned to each feature in LSEG's proprietary aggregation logic—since such weights are not publicly disclosed—it ensures that our analysis reflects features with demonstrated empirical influence. Our results show a strong imbalance between the relevancy of aspirational features compared to performance, except for *emissions*, *management*, and *shareholders* categories.

The results presented above do not account for the relative importance (weight) associated with each category, meaning they do not quantify the variability in the final ESG ratings that can be attributed to aspirational or performance features. To obtain such a quantification, we compute the portion of the Materiality Matrix final weights associated with aspirational or performance features. For this purpose, we average the *materiality matrices* of similar industry groups, yielding 11 broad industrial groups, referred to as TRBC economic sectors. A complete list of TRBC economic sectors and their corresponding TRBC Industry Groups can be found in Appendix A. This approach aims to provide a more synthetic result by reducing the dimensionality of the problem, based on the assumption that similar industry groups share comparable *materiality matrices*.

We present the results in Table 6 and in Figure 5. In Table 6 we show the average weight associated with each category across every TRBC sector and its split into aspirational and performance features. Results show that, on average, each of the 10 ESG categories covers around 10% of the overall ESG score, with the exception of the Governance pillar, where the *management* category contributes to over 20%, at the expense of the *CSR strategy* and *shareholders* categories. Given that the *management* category is mainly composed of performance features, it boosts the performance contribution of the overall Governance pillar performance weight (0.190) to the final ESG score (0.396). Within the Social pillar, we highlight the underweight of performance metrics (0.094) compared to the aspirational ones (0.309). Given a high weight associated with the overall Social pillar (0.403), this imbalance results in an influential impact on the final scoring assessment. Finally, the Environmental pillar presents similar results to the Social pillar: the performance weight (0.112) contributes less than 40% to the total pillar weight (0.291). This confirms again a higher representation of the aspirational features in the final rating. These results are obtained averaging over all sectors. In Appendix D we provide results for the Industrials TRBC sector. Results for the other sectors are available upon request.

In Figure 5 we show a synthetic representation of the results obtained. On the left panel, we show a histogram representing the contribution of aspirational and performance features

15353966, 2025, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/csr.70007 by CochraneItalia, Wiley Online Library on [05/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

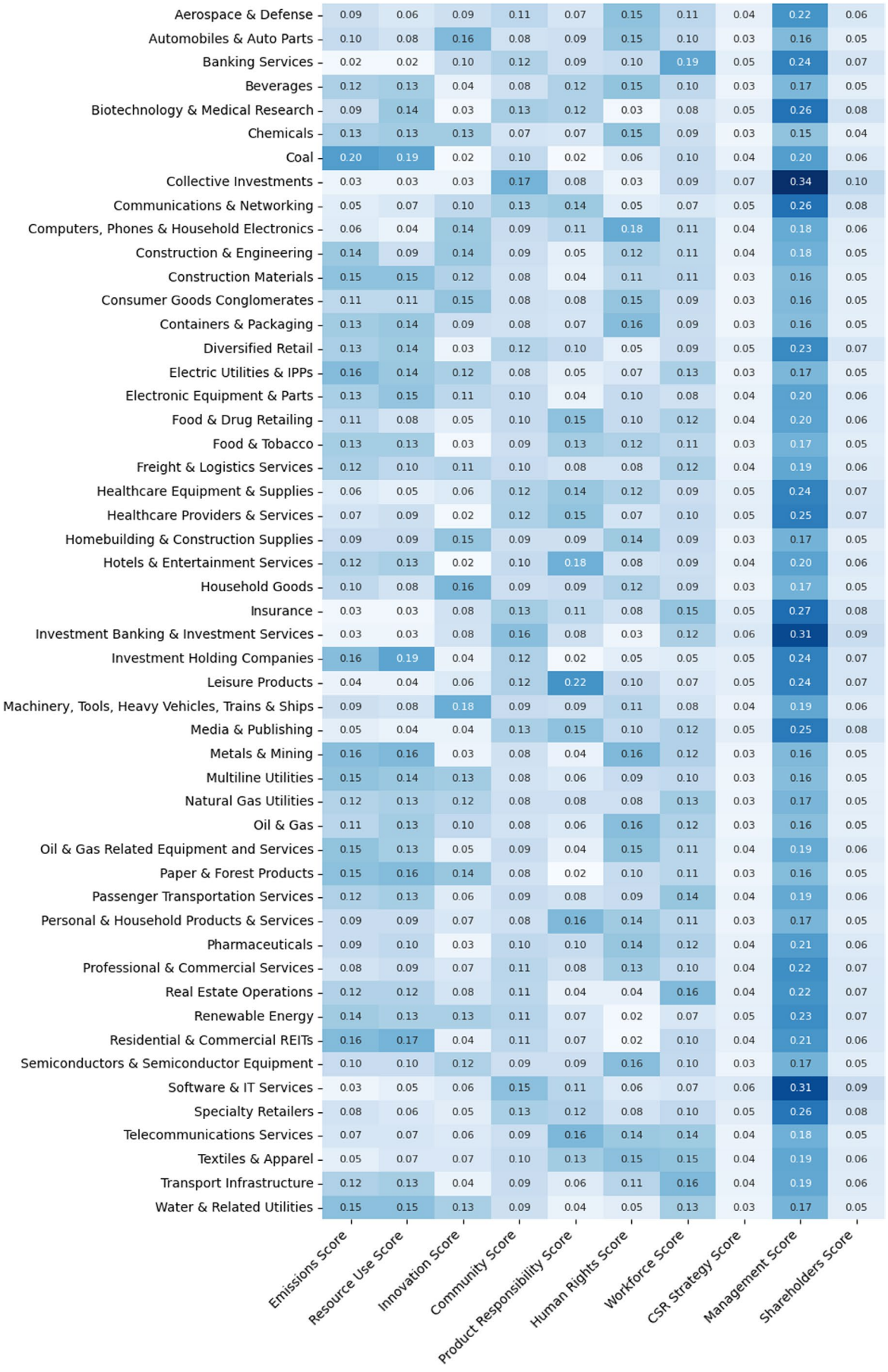


FIGURE 4 | Legend on next page.

15353966, 2025, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/csr.70007 by CochraneItalia, Wiley Online Library on [05/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

FIGURE 4 | Materiality Matrix: Average of matrices from 2020 to 2023, including TRBC industry groups with at least 30 companies for each year. The matrix represents the weight that each category score assumes for each company, depending on its industry cluster. Color intensity grows proportionally with the magnitude of each category weight.

TABLE 5 | Random forest relevant features split into aspirational and performance groups by category.

Pillar	Category	Number of selected features	Number of aspirational features	Number of performance features	Aspirational (%)	Performance (%)
E	Emissions	22	10	12	45.45%	54.55%
	Innovation	13	8	5	61.54%	38.46%
	Resource use	13	10	3	76.92%	23.08%
S	Community	9	6	3	66.67%	33.33%
	Human rights	8	8	0	100.00%	0.00%
	Product resp.	10	8	2	80.00%	20.00%
	Workforce	26	16	10	61.54%	38.46%
G	Management	33	10	23	30.30%	69.70%
	CSR strategy	9	6	3	66.67%	33.33%
	Shareholders	11	5	6	45.45%	54.55%

Note: Results are reported in absolute values and as percentages. The complete list of features, their descriptions, and their classifications into aspirational/performance are reported in Appendix B for the *emissions* category. Data for the rest of the categories is available in the Supporting Information.

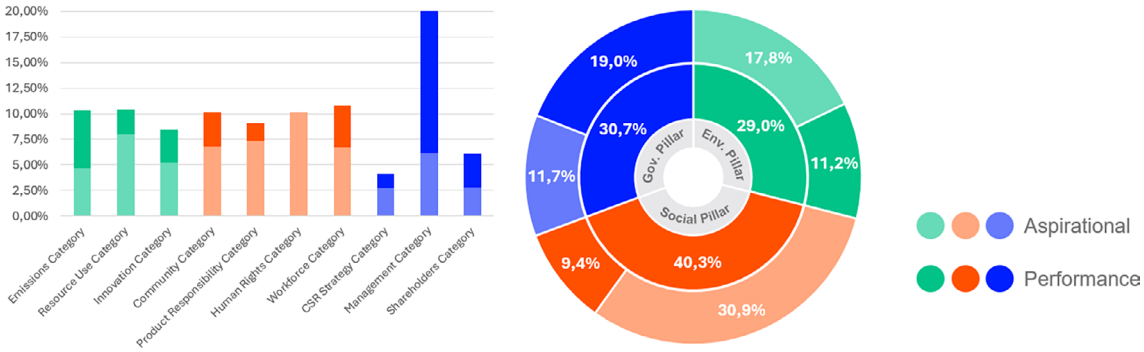
TABLE 6 | Average materiality matrix across TRBC sectors: Each column of Figure 4 is averaged into a single final weight, allowing us to compute a materiality vector that weights every category score.

Category	Materiality weight	Aspirational weight	Performance weight
Emissions score	0.103 (0.042)	0.047 (0.019)	0.056 (0.023)
Innovation score	0.083 (0.045)	0.051 (0.028)	0.032 (0.017)
Resource use score	0.104 (0.045)	0.080 (0.034)	0.024 (0.010)
Environmental pillar	0.291 (0.076)	0.178 (0.048)	0.112 (0.031)
Community score	0.102 (0.022)	0.068 (0.015)	0.034 (0.007)
Human rights score	0.102 (0.043)	0.102 (0.043)	0.000 (0.000)
Product responsibility score	0.091 (0.044)	0.073 (0.035)	0.018 (0.009)
Workforce score	0.108 (0.027)	0.066 (0.016)	0.041 (0.010)
Social pillar	0.403 (0.070)	0.309 (0.060)	0.094 (0.015)
Management score	0.204 (0.044)	0.062 (0.013)	0.143 (0.031)
CSR Strategy score	0.041 (0.009)	0.027 (0.006)	0.014 (0.003)
Shareholders score	0.061 (0.013)	0.028 (0.006)	0.033 (0.007)
Gov pillar	0.307 (0.047)	0.117 (0.016)	0.190 (0.032)
Totals	1.000 (0.114)	0.604 (0.078)	0.396 (0.047)

Note: We report, for each category, mean and standard deviation in brackets. According to the *aspirational/performance* data type split reported in Table 5, we are able to break down each weight into its *aspirational* and *performance* components. Subtotals present an overview of the materiality decomposition at the ESG pillars level.

by category, while on the right pie chart we show the overall contribution of each pillar to the final ESG score and the percentage of aspirational/performance weights for each pillar. On average, 60% of the ESG score composition is based on promises (aspirational features), while only 40% is built upon performance outcomes. Notably, the Environmental and Social pillars are mainly driven by aspirational components (respectively 61.2% and 76.7% of the Average pillar Weights), while the

15353966, 2025, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/csr.70007 by CochraneItalia, Wiley Online Library on [05/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



(a) ESG Average Category Weights: the histogram represents the contribution of each category to the final ESG score according to Table 6. Colors are chosen according to pillar differentiation, and their intensity underlines the *aspirational/performance* metrics nature.

(b) ESG Average pillar Weights: the chart represents the overall contribution of each pillar to the final ESG score according to Table 6. Color intensity underlines the *aspirational/performance* proportion for each pillar.

FIGURE 5 | ESG score composition: Starting from individual feature details, the *aspirational/performance* division has been assembled into the final ESG scores by computing the *average materiality matrix*. The charts show the impact of *aspirational* measures upon which ESG scores are built.

Governance pillar relies more on performance data. In the latter, the *management* category is the main driver of this behavior. While aspirations and intentions are undoubtedly vital aspects of corporate sustainability efforts, the disproportionate emphasis on aspirational outcomes in ESG scoring frameworks might incentivize the ESG inflation phenomenon. This comes with financial implications, leading to potentially suboptimal investment outcomes in the realm of socially responsible investing, in line with Bams and van der Kroft (2022).

We finalize our analysis following a similar procedure from an industry perspective. In order to understand if the prevalence of aspirational over performance information is associated with specific industrial sectors, we compute the average aspirational and performance weights in the final Materiality Matrix for each economics sector. We present the results in Table 7. We observe that the proportion of aspirational features consistently represents more than 50% of the ESG overall score, ranging from a minimum of 56.1% in the Financial sector to a maximum of 64.2% in the Telecommunications sector.

Our findings reveal that economic sectors exhibit similar results in terms of the aspirational/performance balance, with a higher representation of aspirational features in the final ESG scores. This imbalance between aspirational and performance data could potentially shift companies' focus from genuinely enhancing their ESG practices to merely projecting a green image on paper. As emphasized by Bams and van der Kroft (2022), high ESG ratings can lead to a perception of lower risk among investors and a reduced cost of capital. Consequently, firms with a lower cost of capital grow more rapidly and sustain their operations more easily. This dynamic poses a risk of resource misallocation, as capital may disproportionately flow toward firms with inflated ESG ratings—driven by aspirational rather than performance-based metrics—rather than toward those genuinely dedicated to sustainable practices.

6 | Conclusions

This paper replicates the LSEG ESG scoring methodology and explores the balance between forward-looking (aspirational) and backward-looking (performance) features in ESG ratings. Utilizing the full set of ESG-related data from 2010 to 2023, we replicate the proprietary algorithm using random forest to unveil the main drivers of ESG ratings.

In the first part of the paper, we identify the most relevant metrics in the rating issuance process. We observe that our feature selection algorithm considers a smaller set of features compared to the LSEG methodology and shows stable results over time. We also find that it is possible to obtain results with high accuracy using such methodology, especially in the Environmental and Social pillars.

The second part of the paper is devoted to replicating LSEG's proprietary undisclosed Materiality Matrix to determine the relative importance of each ESG theme. This approach allows us to address a critical gap in quantifying the weight of each measure in the final ESG score, enhancing transparency in the rating process. Moreover, using the classification proposed in Bams and van der Kroft (2022), we quantify the balance between aspirational and performance-based ESG data, finding that approximately 60% of ESG scores are based on forward-looking promises, while 40% rely on past performance. This significant reliance on promises highlights a potential bias in ESG ratings toward future commitments, potentially overshadowing actual sustainability achievements.

While there are similarities between LSEG and other providers such as MSCI and Sustainalytics, our analysis of LSEG ESG Scores cannot be entirely generalized due to methodological and focus differences (Berg et al. 2022). However, our choice of relying on LSEG data provider comes from several reasons. To our knowledge, LSEG comprises the most comprehensive scope of granular ESG information, ensuring a thorough understanding

15353966, 2025, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/csr.70007 by CochraneItalia, Wiley Online Library on [05/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 7 | *Aspirational/performance* decomposition comparison: Results obtained by the overall averaging procedure discussed above are compared with the TRBC economic sectors averaging results covered in Appendix D.

TRBC economic sector	TRBC industry groups averaged	Average aspirational weight	Average performance weight
Industrials	10	0.613 (0.051)	0.387 (0.024)
Consumer Cyclicals	8	0.609 (0.076)	0.391 (0.031)
Financials	5	0.561 (0.074)	0.439 (0.046)
Consumer Non-Cyclicals	5	0.631 (0.050)	0.369 (0.016)
Health Care	4	0.598 (0.051)	0.402 (0.020)
Basic Materials	3	0.624 (0.052)	0.376 (0.017)
Energy	4	0.597 (0.067)	0.403 (0.029)
Technology	5	0.603 (0.076)	0.397 (0.036)
Utilities	4	0.599 (0.025)	0.401 (0.011)
Real Estate	2	0.568 (0.034)	0.432 (0.020)
Telecommunications	1	0.642 (0.000)	0.358 (0.000)
Total	51	0.604 (0.078)	0.396 (0.047)

Note: For each economic sector, we report the number of TRBC Industry Groups incorporated into the sector, the standard deviation (in brackets) obtained by averaging all Industry Groups into their economic sector, and the *aspirational/performance* balance.

of the whole underlying scoring methodology. Using such a granular dataset enables us to match ESG policies, targets, activities, performance, and controversies on similar aspects of ESG, as in Bams and van der Kroft (2022). Considering the methodological differences among various raters, we emphasize that certain critical aspects—such as transparency and the limitations associated with sectoral materiality—emerge as common themes among different providers. These themes offer valuable insights for facilitating a broader comparison and understanding of ESG ratings.

Our findings emphasize the need for greater transparency and a clearer distinction between aspirational and performance metrics to ensure credible ESG assessments for informed investment decisions. Interpreting ESG scores as purely concrete progress towards sustainability could lead to resource misallocation and to dispersing the social benefits of such ratings due to information asymmetries (Hartzmark and Sussman 2019; Bams and van der Kroft 2022). The imbalance between aspirational and performance metrics highlights a potential bias in ESG ratings toward future commitments. Within this framework, firms may continue to improve their ESG ratings even while facing ESG controversies, scaling back sustainability commitments, or operating under deteriorating labor conditions and governance structures. This imbalance can potentially overshadow actual sustainability achievements and transform ESG ratings to a "ticking-the-boxes" exercise rather than an actual long-term value creation opportunity (Edmans 2024), or even greenwashing (Kathan et al. 2025).

This study has some limitations that should be acknowledged. First, the reverse-engineering of the ESG scoring methodology relies on publicly available LSEG data as of September 2023, which reflects historical scores that may have been retroactively adjusted. While we account for these adjustments by using a static snapshot, we cannot disentangle actual historical disclosure from methodological rewriting. Second, the exclusion of sectors and years with limited observations—while necessary for model stability—introduces a bias toward mature, disclosure-rich industries and may limit the generalizability of our findings to emerging sectors. Third, our classification of ESG features into aspirational and performance-based metrics, while grounded in the taxonomy proposed by Bams and van der Kroft (2022), necessarily involves interpretative decisions for a small share of variables.

Future research could extend this work in several directions. One promising direction involves studying the evolution of aspirational and performance metrics over longer time horizons to assess the alignment of ESG scores with real-world impact and regulatory trends, such as the EU's CSRD and SFDR frameworks. In terms of policy implications, our findings highlight the persistent dominance of forward-looking (aspirational) metrics in ESG ratings. Regulators and institutional investors may wish to scrutinize the weighting schemes used by rating agencies and push for greater transparency in score construction. The increasing reliance on ESG ratings in portfolio allocation and regulatory compliance underscores the importance of explainability, consistency, and comparability across providers.

Acknowledgments

This research has been funded by the European Union—Next Generation EU, Mission 4-Component 1-CUP D53D23017690001—Project PRIN PNRR 2022 P20228CHNL "Measuring, managing and hedging indirect climate-transition risk", and by Italian MUR, grant Dipartimento di Eccellenza 2023-2027. Daniele Marazzina and Davide Stocco are members of the Gruppo Nazionale Calcolo Scientifico-Istituto Nazionale di Alta Matematica (GNCS-INdAM). Open access publishing facilitated by Politecnico di Milano, as part of the Wiley - CRUI-CARE agreement.

15353966, 2025, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/csr.70007 by CochraneItalia, Wiley Online Library on [05/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Endnotes

[1] For transparency and reproducibility, the full mapping of ESG indicators into aspirational and performance-based categories—based on the taxonomy proposed by Bams and van der Kroft (2022) and extended where necessary—is provided in the Supporting Information.

References

Alshahmy, S., and M. Sahiner. 2024. "Enhancing the Issuance and Monitoring of Sustainable Finance Instruments Through Ai." In *Artificial Intelligence, Finance, and Sustainability: Economic, Ecological, and Ethical Implications*, 217–241. Springer.

Aouadi, A., and S. Marsat. 2018. "Do ESG Controversies Matter for Firm Value? Evidence From International Data." *Journal of Business Ethics* 151: 1027–1047.

Aste, T., E. Barucci, M. L. D. Nicolas, and D. Stocco. 2024. "Resilience of Market Returns Around ESG Controversies: Insights From the S&P100." SSRN 5031012.

Avramov, D., S. Cheng, A. Lioui, and A. Tarelli. 2022. "Sustainable Investing With ESG Rating Uncertainty." *Journal of Financial Economics* 145, no. 2: 642–664.

Azzone, M., E. Barucci, and D. Stocco. 2024. "Asset Management With an ESG Mandate." *arXiv* preprint, arxiv:2403.11622. https://doi.org/10.48550/arXiv.2403.11622.

Bams, D., and B. van der Kroft. 2022. "Tilting the Wrong Firms? How Inflated ESG Ratings Negate Socially Responsible Investing Under Information Asymmetries." MIT Center for Real Estate Research Paper, 22/12.

Berg, F., K. Fabisik, and Z. Sautner. 2021. "Is History Repeating Itself? The (Un)predictable Past of ESG Ratings." *SSRN Electronic Journal*: 3722087.

Berg, F., J. F. Koelbel, and R. Rigobon. 2022. "Aggregate Confusion: The Divergence of ESG Ratings." *Review of Finance* 26, no. 6: 1315–1344.

Bernardini, E., M. Fanari, E. Foscolo, and F. Ruggiero. 2024. "Environmental Data and Scores: Lost in Translation." *Corporate Social Responsibility and Environmental Management* 31, no. 5: 4796–4818.

Billio, M., M. Costola, I. Hristova, C. Latino, and L. Pelizzon. 2021. "Inside the ESG Ratings: (Dis)agreement and Performance." *Corporate Social Responsibility and Environmental Management* 28, no. 5: 1426–1445.

Billio, M., A. C. Fitzpatrick, C. Latino, and L. Pelizzon. 2024. "Unpacking the ESG Ratings: Does One Size Fit All?" SAFE Working Paper.

Castellano, R., F. Cini, and A. Ferrari. 2024. "Machine Learning for ESG Rating Classification: An Integrated Replicable Model With Financial and Systemic Risk Parameters." In *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, 87–92. Springer.

D'Amato, V., R. Ecclesia, and S. Levantesi. 2022. "ESG Score Prediction Through Random Forest Algorithm." *Computational Management Science* 19, no. 2: 347–373.

Del Vitto, A., D. Marazzina, and D. Stocco. 2023. "ESG Ratings Explainability Through Machine Learning Techniques." *Annals of Operations Research*: 1–30.

Edmans, A. 2024. "Rational Sustainability." *SSRN Electronic Journal*: 4701143.

Eskantar, M., C. Zopounidis, M. Doumpos, E. Galariotis, and K. Guesmi. 2024. "Navigating ESG Complexity: An In-Depth Analysis of Sustainability Criteria, Frameworks, and Impact Assessment." *International Review of Financial Analysis* 95: 103380.

Fliegel, P. 2025. "How (Not) to Measure Companies' Climate Transition Risk: A Framework and Categorized Literature Review." *Corporate Social Responsibility and Environmental Management* 32, no. 3: 3049–3077. https://onlinelibrary.wiley.com/doi/full/10.1002/csr.3094

Gibson Brandon, R., P. Krueger, and P. S. Schmidt. 2021. "ESG Rating Disagreement and Stock Returns." *Financial Analysts Journal* 77, no. 4: 104–127.

Giese, G., Z. Nagy, and L.-E. Lee. 2021. "Deconstructing ESG Ratings Performance: Risk and Return for E, S, and G by Time Horizon, Sector, and Weighting." *Journal of Portfolio Management* 47, no. 3: 94–111.

Global Sustainable Investment Alliance. 2023. "Global Sustainable Investment Review." https://www.gsi-alliance.org/wp-content/uploads/2023/12/GSIA-Report-2022.pdf.

Guerrero, S., and J. P. Viteri. 2025. "What Are Environmental, Social, and Governance Scores Measuring? The Role of Outcome and Impact Indicators in ESG Scores." *Finance Research Letters* 72: 106529.

Hartzmark, S. M., and A. B. Sussman. 2019. "Do Investors Value Sustainability? A Natural Experiment Examining Ranking and Fund Flows." *Journal of Finance* 74, no. 6: 2789–2837.

Kathan, M. C., S. Utz, G. Dorfleitner, J. Eckberg, and L. Chmel. 2025. "What You See Is not What You Get: ESG Scores and Greenwashing Risk." *Finance Research Letters* 74: 106710.

Lanza, A. A. G., E. Bernardini, and I. Faiella. 2023. "Machine Learning, ESG Indicators, and Sustainable Investment." In *Financial Risk Management and Climate Change Risk: The Experience in a Central Bank*, 223–250. Springer.

Liu, X., J. Dai, X. Dong, and J. Liu. 2024. "ESG Rating Disagreement and Analyst Forecast Quality." *International Review of Financial Analysis* 95: 103446.

LSEG. 2022. "Environmental, Social and Governance Scores From LSEG." https://www.lseg.com/content/dam/marketing/en_us/documents/methodology/refinitiv-ESG-scores-methodology.pdf.

Mao, Z., S. Wang, and Y.-E. Lin. 2024. "ESG, ESG Rating Divergence and Earnings Management: Evidence From China." *Corporate Social Responsibility and Environmental Management* 31, no. 4: 3328–3347.

Pedersen, L. H., S. Fitzgibbons, and L. Pomorski. 2021. "Responsible Investing: The ESG-Efficient Frontier." *Journal of Financial Economics* 142, no. 2: 572–597.

Pérez-Martín, A., A. Pérez-Torregrosa, A. Rabasa, and M. Vaca. 2020. "Feature Selection to Optimize Credit Banking Risk Evaluation Decisions for the Example of Home Equity Loans." *Mathematics* 8, no. 11: 1971.

Rubino, M., I. Mastrorocco, and G. M. Garegnani. 2024. "The Influence of Market and Institutional Factors on ESG Rating Disagreement." *Corporate Social Responsibility and Environmental Management* 31, no. 5: 3916–3926.

Sustainability. 2023. "Rate the Raters 2023: ESG Ratings at a Crossroads." https://www.sustainability.com/globalassets/sustainability.com/thinking/pdfs/2023/rate-the-raters-report-april-2023.pdf.

UN Global Compact. 2004. "Who Cares Wins: Connecting Financial Markets to a Changing World." Technical Report.

US SIF Foundation. 2024. "US Sustainable Investing Trends 2024/2025." US SIF Foundation. https://www.ussif.org/research/trends-reports/us-sustainable-investing-trends-2024-2025-executive-summary.

Useche, A. J., J. Martínez-Ferrero, and J. L. Alayón-Gonzales. 2024. "Socially Responsible Portfolios, Environmental, Social, Corporate Governance (ESG) Efficient Frontiers, and Psychic Dividends." *Corporate Social Responsibility and Environmental Management* 31, no. 2: 1323–1339.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

15353966, 2025, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/csr.70007 by CochraneItalia, Wiley Online Library on [05/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Appendix A

TRBC Classification

In this section we provide the complete list of TRBC Industry Groups, which are clusterized by Refinitiv macro-sectors called TRBC economic sectors.

1. Basic Materials:

   Chemicals, METALS & Mining, Paper & Forest Products, Coal

2. Consumer Cyclicals:

   Automobiles & Auto Parts, Diversified Retail, Hotels & Entertainment Services, Household Goods, Leisure Products, Specialty Retailers, Textiles & Apparel, Media & Publishing

3. Consumer Non-Cyclicals:

   Beverages, Food & Drug Retailing, Food & Tobacco, Personal & Household Products & Services, Consumer Goods Conglomerates

4. Energy:

   Oil & Gas, Oil & Gas Related Equipment and Services, Renewable Energy, Coal

5. Financials:

   Banking Services, Collective Investments, Insurance, Investment Banking & Investment Services, Investment Holding Companies, Financial Technology (Fintech) & Infrastructure

6. Health Care:

   Biotechnology & Medical Research, Healthcare Equipment & Supplies, Healthcare Providers & Services, Pharmaceuticals

7. Industrials:

   Aerospace & Defense, Construction & Engineering, Construction Materials, Freight & Logistics Services, Homebuilding & Construction Supplies, Machinery, Tools, Heavy Vehicles, Trains & Ships, Professional & Commercial Services, Transport Infrastructure, Passenger Transportation Services, Containers & Packaging, Diversified Industrial Goods Wholesale

8. Institutions, Organizations & Associations:

   Schools, Colleges & Universities, Miscellaneous Educational Service Providers

9. Technology:

   Communications & Networking, Computers, Phones & Household Electronics, Electronic Equipment & Parts, Semiconductors & Semiconductor Equipment, Software & IT Services, Integrated Hardware & Software

10. Utilities:

    Electric Utilities & IPPs, Multiline Utilities, Natural Gas Utilities, Water & Related Utilities

11. Real Estate:

    Real Estate Operations, Residential & Commercial REITs

12. Academic & Educational Services:

    Professional & Business Education

13. Telecommunications:

    Telecommunications Services

15353966, 2025, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/csr.70007 by CochraneItalia, Wiley Online Library on [05/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Appendix B

Feature Selection and Data Type by Category

In this section, we present a set of filtered features that we consider valid candidates for final selection, since selected by our random forest algorithm with the highest frequencies. We recall that the frequency is given by the number of years in which the feature is selected, divided by the total number of years. Table A1 reports results for the *emissions* category, which measures a company's commitment and effectiveness towards reducing environmental emissions in its production and

TABLE A1 | Emissions category features.

Feature	Description	Data type	Frequency
Environmental partnerships	Does the company report on partnerships or initiatives with specialized NGOs, industry organizations, governmental or supra-governmental organizations, which are focused on improving environmental issues?	Aspirational	1.00
Biodiversity impact reduction	Does the company report on its impact on biodiversity or on activities to reduce its impact on native ecosystems and species?	Aspirational	1.00
Environmental restoration initiatives	Does the company report or provide information on company-generated initiatives to restore the environment?	Aspirational	1.00
CO_2 equivalent emissions indirect, scope 3	Total CO_2 and CO_2 Scope Three equivalent emission in tonnes.	Performance	1.00
Target emissions	Has the company set targets or objectives to be achieved on emission reduction?	Aspirational	1.00
Environmental expenditures investment	Does the company report on its environmental expenditures or does the company report to make proactive environmental investments?	Aspirational	1.00
Total CO_2 equivalent emissions to revenues	Total CO_2 and CO_2 equivalents emission in tonnes divided by net sales or revenue in US dollars in million.	Performance	1.00
e-waste reduction	Does the company report on initiatives to recycle, reduce, reuse, substitute, treat, or phase out e-waste?	Aspirational	1.00
Climate change commercial risk opportunities	Is the company aware that climate change can represent commercial risks and/or opportunities?	Performance	1.00
Policy emissions	Does the company have a policy to improve emission reduction?	Aspirational	1.00
Estimated CO_2 equivalent emissions total	The estimated total CO_2 and CO_2 equivalents emission in tonnes.	Performance (estimation)	1.00
Waste reduction initiatives	Does the company report on initiatives to recycle, reduce, reuse, substitute, treat, or phase out total waste?	Aspirational	1.00
NOx emissions to revenues	Total amount of NOx emissions emitted in tonnes divided by net sales or revenue US dollars in million.	Performance	1.00
Staff transportation impact reduction	Does the company report on initiatives to reduce the environmental impact of transportation used for its staff?	Aspirational	1.00
Total waste to revenues	Total amount of waste produced in tonnes divided by net sales or revenue in US dollars in million.	Performance	1.00
ISO 14000 or EMS	Does the company claim to have an ISO 14000 or EMS certification? Certification is required.	Aspirational	1.00
Waste total	Total amount of solid waste produced in tonnes.	Performance	0.79
CO_2 equivalent emissions total	Total Carbon dioxide (CO_2) and CO_2 equivalents emission in tonnes.	Performance	0.79
SOx emissions to revenues	Total amount of SOx emissions emitted in tonnes divided by net sales or revenue US dollars in million.	Performance	0.71
Total hazardous waste to revenues	Total amount of hazardous waste produced in tonnes divided by net sales or revenue in US dollars in million.	Performance	0.64
GHG emissions indirect to revenues	Greenhouse gas emissions indirect, in tonnes, divided by net sales or revenue US dollars in million.	Performance	0.64
Waste recycled to total waste	Total recycled and reused waste produced in tonnes divided by total waste produced in tonnes.	Performance	0.57

15353966, 2025, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/csr.70007 by CochraneItalia, Wiley Online Library on [05/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

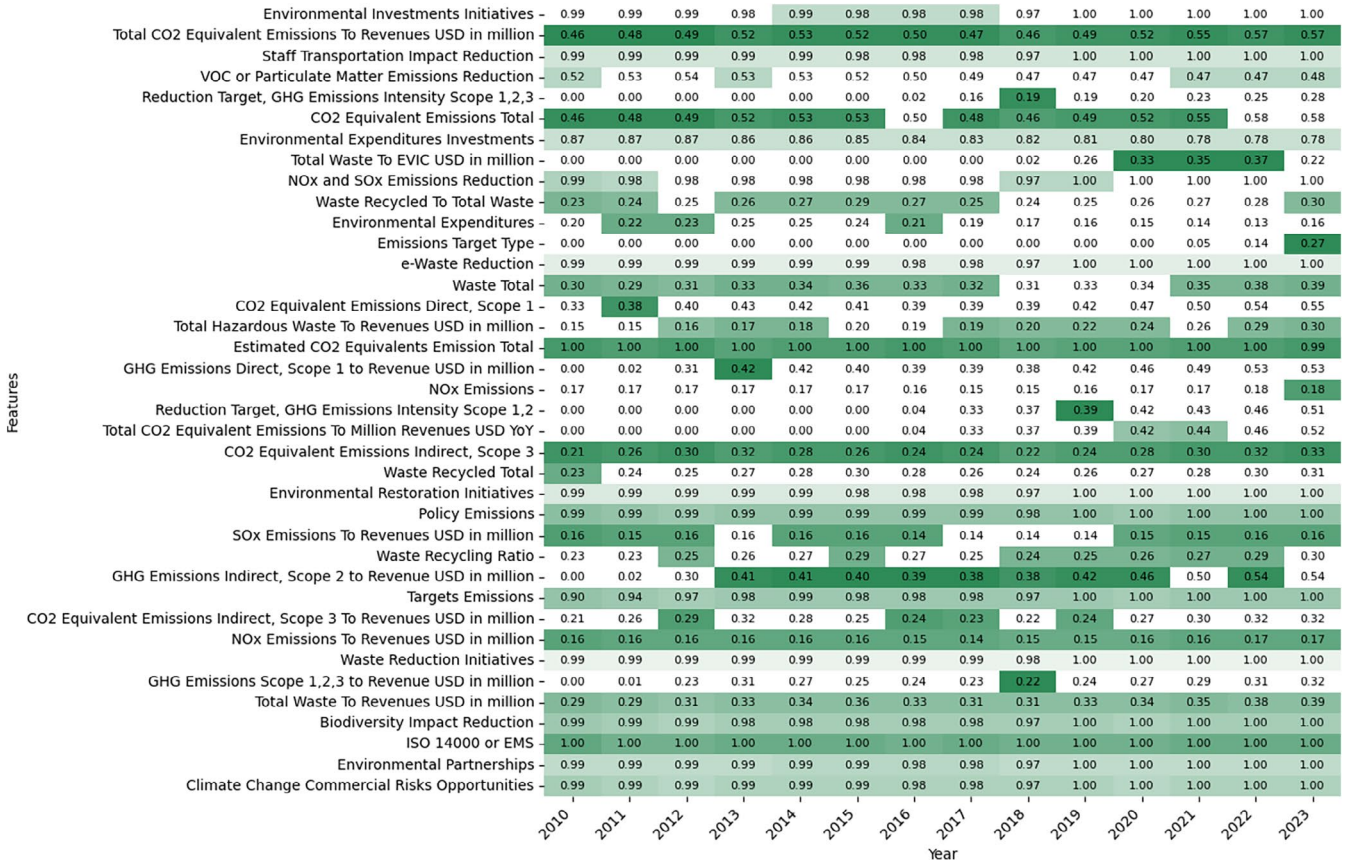


FIGURE A1 | Emissions Feature Selection: Each row represents a feature, and each column represents a year. The intensity of the color indicates the importance that the feature has in the corresponding year according to the random forest algorithm. Numbers represent the completeness percentage of each variable in terms of missing data.

operational processes. The results for the remaining categories are available in the Supporting Information.

Among the 22 features reported in Table A1, one deals with LSEG's estimation of companies' emissions, namely "*Estimated CO_2 Equivalent Emissions Total*," while 11 represent performance outcomes and 10 represent aspirations or policies reported by the companies themselves. Additionally, all variables related to performance exhibit a high percentage of missing data, with an average exceeding 50%, as shown in Figure A1, where all the features selected at least once are considered.

Finally, when comparing the most recent features (from 2020 to 2023) with earlier ones, it is worth noting a slight shift in variable importance. For example, "*VOC Emissions Reduction*" and "*Total Waste to EVIC*" have become relevant. Notably, the latter was not measured until 2019, as its values were initially missing.

Appendix C

Materiality Matrix by Year

Figure A2 shows the 2023 estimated Materiality Matrix.

We also report the TRBC groups that are left out from our analysis, due to the threshold set on the minimum number of companies per sector of our algorithm: "*Diversified Industrial Goods Wholesale*," "*Integrated Hardware & Software*," "*Miscellaneous Educational Service Providers*," "*Office Equipment*," "*Professional & Business Education*," "*Schools, Colleges & Universities*," "*Uranium*."

Finally, for each regression model, dataset dimensions and accuracy metrics are summarized in Table A2. Results from 2010 to 2022 are made available upon request.

15353966, 2025, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/csr.70007 by CochraneItalia, Wiley Online Library on [05/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

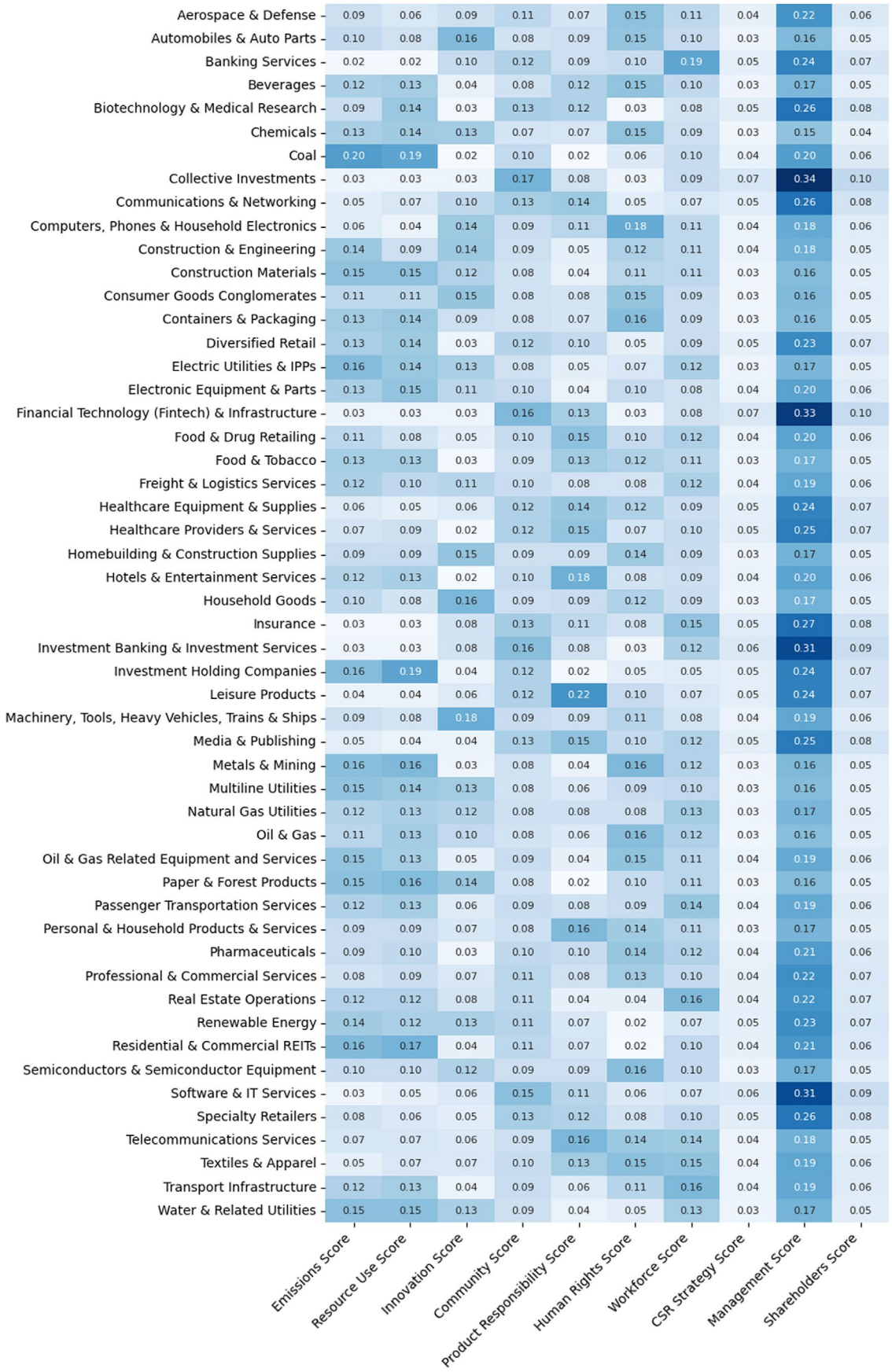


FIGURE A2 | Materiality matrix: 2023.

15353966, 2025, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/csr.70007 by CochraneItalia, Wiley Online Library on [05/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE A2 | Linear regression results by TRBC industry groups (2023).

TRBC sector name	Number of companies	RMSE	R^2
Aerospace & Defense	110	4.39e-16	1.000
Automobiles & Auto Parts	265	7.73e-04	1.000
Banking Services	936	5.00e-16	1.000
Beverages	108	3.09e-16	1.000
Biotechnology & Medical Research	477	4.20e-16	1.000
Chemicals	401	4.05e-04	1.000
Coal	50	2.80e-16	1.000
Collective Investments	92	8.52e-16	1.000
Communications & Networking	95	3.70e-16	1.000
Computers, Phones & Household Electronics	111	4.52e-16	1.000
Construction & Engineering	269	3.07e-16	1.000
Construction Materials	109	4.25e-16	1.000
Consumer Goods Conglomerates	63	9.26e-16	1.000
Containers & Packaging	81	7.06e-16	1.000
Diversified Retail	96	4.59e-16	1.000
Electric Utilities & IPPs	251	2.74e-16	1.000
Electronic Equipment & Parts	148	2.81e-16	1.000
Financial Technology (Fintech) & Infrastructure	39	5.81e-16	1.000
Food & Drug Retailing	127	3.47e-16	1.000
Food & Tobacco	396	4.68e-16	1.000
Freight & Logistics Services	181	2.92e-16	1.000
Healthcare Equipment & Supplies	293	3.93e-16	1.000
Healthcare Providers & Services	166	1.31e-15	1.000
Homebuilding & Construction Supplies	166	3.28e-16	1.000
Hotels & Entertainment Services	285	3.69e-16	1.000
Household Goods	87	3.86e-16	1.000
Insurance	277	3.41e-16	1.000
Investment Banking & Investment Services	355	6.84e-16	1.000
Investment Holding Companies	48	3.42e-16	1.000
Leisure Products	54	4.57e-16	1.000
Machinery, Tools, Heavy Vehicles, Trains & Ships	601	7.70e-16	1.000
Media & Publishing	210	5.19e-16	1.000
Metals & Mining	521	3.19e-16	1.000
Multiline Utilities	43	5.33e-16	1.000
Natural Gas Utilities	55	3.80e-16	1.000
Oil & Gas	261	3.61e-16	1.000
Oil & Gas Related Equipment and Services	176	3.76e-16	1.000
Paper & Forest Products	71	4.58e-16	1.000
Passenger Transportation Services	101	4.09e-16	1.000
Personal & Household Products & Services	89	5.35e-16	1.000
Pharmaceuticals	384	5.28e-16	1.000
Professional & Commercial Services	333	4.80e-16	1.000
Real Estate Operations	423	4.89e-16	1.000
Renewable Energy	80	4.60e-16	1.000
Residential & Commercial REITs	375	6.17e-16	1.000
Semiconductors & Semiconductor Equipment	216	3.63e-16	1.000
Software & IT Services	744	4.58e-16	1.000
Specialty Retailers	245	9.52e-16	1.000
Telecommunications Services	186	4.30e-16	1.000
Textiles & Apparel	144	3.50e-16	1.000
Transport Infrastructure	93	3.21e-16	1.000
Water & Related Utilities	37	4.53e-16	1.000

15353966, 2025, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/csr.70007 by CochraneItalia, Wiley Online Library on [05/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Appendix D

Aspirational/Performance Decomposition by Industry

In Table A3 we report the average materiality weights over time for the Industrials sector. We also estimate the *aspirational/performance* decomposition of the category scores. Results for the other sectors are available upon request.

TABLE A3 | Average materiality matrix for industrials sector. All TRBC industry groups belonging to *industrials* TRBC economic sector of Figure 4 are averaged into a single final weight.

Category	Materiality weight	Aspirational weight	Performance weight
Emissions score	0.114 (0.024)	0.052 (0.011)	0.062 (0.013)
Innovation score	0.104 (0.028)	0.064 (0.017)	0.040 (0.011)
Resource use score	0.107 (0.042)	0.083 (0.033)	0.025 (0.010)
Environmental pillar	0.325 (0.056)	0.198 (0.039)	0.127 (0.020)
Community score	0.093 (0.010)	0.062 (0.007)	0.031 (0.003)
Human rights score	0.121 (0.015)	0.121 (0.015)	0.000 (0.000)
Product responsibility score	0.071 (0.024)	0.056 (0.019)	0.014 (0.005)
Workforce score	0.112 (0.025)	0.069 (0.015)	0.043 (0.010)
Social pillar	0.396 (0.039)	0.308 (0.030)	0.088 (0.011)
Management score	0.186 (0.004)	0.056 (0.001)	0.130 (0.003)
CSR strategy score	0.037 (0.020)	0.025 (0.014)	0.012 (0.007)
Shareholders score	0.056 (0.006)	0.025 (0.003)	0.030 (0.003)
Gov pillar	0.279 (0.022)	0.107 (0.014)	0.172 (0.008)
Totals	1.000 (0.072)	0.613 (0.051)	0.387 (0.024)

Note: Specifically, a total of 10 industry groups are included in this economic sector. We report, for each category, mean and standard deviation, in brackets. According to the *aspirational/performance* data type split reported in Table 5 we are able to break down each weight into its "*aspirational*" and "*performance*" components. Subtotals present an overview of the materiality decomposition at ESG pillars level.

15353966, 2025, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/csr.70007 by CochraneItalia, Wiley Online Library on [05/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License