Article:

Balan, P. orcid.org/0009-0005-1899-0043, Antunes, J., Wanke, P. orcid.org/0000-0003-1395-8907 et al. (2 more authors) (2025) The Black-Box of ESG scores from rating agencies: do they genuinely reflect sustainability practices, or are they disproportionately shaped by financial performance? International Journal of Finance & Economics. ISSN: 1076-9307

https://doi.org/10.1002/ijfe.70043

International Journal of Finance & Economics

WILEY

RESEARCH ARTICLE OPEN ACCESS

The Black-Box of ESG Scores From Rating Agencies: Do They Genuinely Reflect Sustainability Practices, or Are They Disproportionately Shaped by Financial Performance?

Philipe Balan[1] | Jorge Antunes[2] | Peter Wanke[2] | Yong Tan[3] | Ali Meftah Gerged[4,5]

[1]FGV EBAPE—Edifício Roberto Campos—R. Jorn. Orlando Dantas, Rio de Janeiro, Brazil | [2]COPPEAD Graduate Business School, Federal University of Rio de Janeiro, Rua Paschoal Lemme, Rio de Janeiro, Brazil | [3]School of Management, University of Bradford, Bradford, West Yorkshire, UK | [4]Sheffield University Management School, The University of Sheffield, Sheffield, UK | [5]Faculty of Economics, Misurata University, Misurata, Libya

Correspondence: Ali Meftah Gerged (a.m.gerged@sheffield.ac.uk)

Received: 23 January 2025 | **Revised:** 4 August 2025 | **Accepted:** 30 August 2025

Keywords: data envelopment analysis | ESG | rating agencies

ABSTRACT

This study examines the environmental, social and governance (ESG) scoring methodologies used by Bloomberg and S&P Global through the lens of Data Envelopment Analysis (DEA). It addresses a notable gap in the literature by identifying the underlying factors that shape ESG scores and providing practical insights for companies seeking to understand or improve their sustainability ratings. Our comparative analysis reveals clear differences between the two rating agencies. While Bloomberg's raw ESG scores are generally higher than those of S&P Global, the DEA-normalised results tell a different story. Bloomberg applies stricter internal benchmarks, resulting in lower efficiency scores. In contrast, S&P's lower raw scores convert into higher DEA efficiencies, suggesting a more lenient, peer-based benchmarking approach that tends to cluster firms near the top regardless of their absolute ESG performance. A particularly striking finding is that 99% of ESG scores from both agencies correlate with net income, highlighting a strong connection between financial performance and ESG ratings. Our regression analysis supports this, showing that firms with better financial outcomes tend to receive higher ESG scores. However, we also find that companies with growing cash reserves—often indicative of reinvestment and expansion—may be penalised, receiving lower ESG scores. This suggests a potential bias against firms prioritising long-term growth over immediate returns. This study lays the groundwork for future research aimed at refining ESG datasets and expanding the scope of analysis.

1 | Introduction

Imagine two investors striving to build sustainability-focused portfolios. They evaluate the same company but consult different environmental, social and governance (ESG) ratings—one from Bloomberg, the other from S&P Global—and arrive at radically different conclusions. This is not an anomaly but a symptom of a deeper, unresolved tension in today's ESG ecosystem. Recent studies have shown that ESG scores often diverge significantly across providers due to differences in underlying indicators, weighting schemes and assessment philosophies (Berg et al. 2022; Gibson Brandon et al. 2022). As ESG factors become central to corporate responsibility and investment decision-making, the lack of transparency in how these scores are constructed remains both perplexing and consequential (Liu 2022; Clementino and Perkins 2021). Despite their growing influence in directing capital and shaping firm behaviour, ESG ratings still function largely as black boxes, with proprietary methodologies rarely disclosed in full (Mayer and Ducsai 2023). This raises pressing questions:

What exactly do these scores measure? Why do they vary so drastically across agencies? And to what extent do they reflect genuine sustainability performance rather than merely mirroring traditional financial metrics (Mahanta et al. 2024; Duque-Grisales and Aguilera-Caracuel 2021)?

This study arises from the intersection of such practical puzzles and empirical uncertainties. While ESG frameworks aim to capture a firm's non-financial performance—ranging from carbon emissions and board diversity to ethical supply chains and transparency—they are often reduced to single numerical scores by rating agencies. These scores are intended to guide investors, inform regulators and drive corporate change (Barko et al. 2022; Duque-Grisales and Aguilera-Caracuel 2021). Yet research has consistently flagged significant inconsistencies in the methodologies behind them. Liu (2022) observes that the metrics used to assign ESG scores are rarely disclosed, while Mayer and Ducsai (2023) highlight the unreliability that emerges from such methodological opacity. Berg et al. (2022) further document the divergent practices and weighting schemes that drive variation in ESG ratings, raising serious questions about their comparability and utility.

At a practical level, these disparities create confusion for stakeholders and raise the risk of misallocated capital. Corporations, unsure of how they are being assessed, may adopt inefficient or performative strategies. Investors, attempting to align their portfolios with ethical goals, confront inconsistent signals. The lack of transparency in ESG scoring methodologies thus undermines the accountability and trust that these metrics are supposed to engender. In this context, the present study seeks to reveal how two leading rating agencies—Bloomberg and S&P Global—evaluate and rank firms using ESG criteria. Our aim is not only to expose the mechanics beneath their assessments but also to critically examine whether ESG ratings genuinely reflect sustainability practices or are disproportionately shaped by financial performance.

To investigate this, we employ a hybrid methodological framework that combines Data Envelopment Analysis (DEA) and the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS[1]). This approach allows us to infer the implicit priorities embedded in ESG scores. Drawing on a panel of 67 firms across the United States, Hong Kong and South Korea, spanning 24 industries over 4 years (2018–2021), our analysis reveals substantial divergence in how ESG ratings are constructed. Bloomberg's ESG scores have higher raw averages (e.g., 54.10 overall) compared to S&P's lower averages (29.10), suggesting a more generous baseline. However, when DEA-normalised, the picture reverses—Bloomberg's scores yield lower efficiency ratings, reflecting a stricter, innovation-driven evaluation. This implies that although Bloomberg awards higher raw ESG scores, it applies tougher internal benchmarks when transformed into efficiency metrics. S&P, despite its lower raw averages, shows higher DEA efficiency clustering (80%–100%), revealing a more lenient, benchmark-oriented method that places firms nearer the top, regardless of absolute sustainability performance. Thus, DEA-normalised results expose Bloomberg's critical stance and S&P's relatively tolerant scoring logic. Yet perhaps most strikingly, we find that nearly 99% of ESG scores from both agencies correlate with net income. This raises a provocative implication, indicating that ESG ratings may be more reflective of profitability than previously acknowledged.

Moreover, our findings expose a paradox. Companies accumulating larger cash reserves—typically indicative of growth and reinvestment—often receive lower ESG ratings. This pattern suggests that specific scoring frameworks may inadvertently penalise expansion-oriented firms, perhaps based on an assumption of higher environmental footprints or delayed ESG compliance. Such insights challenge the normative expectations surrounding ESG metrics and invite deeper scrutiny into how they are applied.

This study contributes to the literature by providing a comparative, empirical dissection of ESG methodologies—something that has been largely missing to date. It introduces a novel, data-driven approach to understanding the priorities behind ESG assessments and reveals how financial variables may be disproportionately influencing sustainability scores. In doing so, it offers practical insights for investors seeking transparency, for firms aiming to align strategies with rating criteria, and for policymakers considering how best to regulate and standardise ESG disclosures.

The structure of this paper is as follows: Section 1 introduces the study; Section 2 provides an overview of ESG and DEA and reviews the literature on ESG and DEA applications; Section 3 details the methodology; Section 4 discusses the findings and Section 5 concludes the study.

2 | ESG, Efficiency and the Evolving Landscape of Corporate Evaluation

Over the last decade, the concept of sustainability has transformed from a peripheral concern into a central force reshaping the ethos of global business. At the heart of this shift lies the ESG framework—ESG—which now acts as a guiding lens for understanding corporate behaviour beyond traditional financial performance. While this framework promises a more holistic view of a firm's impact, the growing institutionalisation of ESG has also raised new questions about how these metrics are assessed, rated and ultimately trusted.

ESG, in its most elemental form, captures a company's alignment with ecological responsibility, social fairness and ethical governance. The environmental pillar focuses on how firms manage natural resources, reduce emissions, develop eco-innovation capabilities and comply with regulations like the Paris Agreement (Fuente et al. 2022; Luo and Tang 2023). The social dimension is concerned with labour practices, stakeholder relationships, philanthropic activities, diversity and broader societal well-being (Pelosi and Adamson 2016; O'Riordan and Fairbrass 2014; Zanten and Tulder 2021). Governance, the often underexamined pillar, addresses how companies are structured and led, highlighting transparency, executive compensation and anti-corruption efforts (Arjoon 2006; Veldman et al. 2023; Sancak 2023). Together, these components form a triad intended to steer businesses towards long-term value creation that benefits shareholders, communities and the environment alike (Sandberg et al. 2023; Chen et al. 2023).

As the ESG movement matures, however, the tools used to measure and compare ESG performance remain deeply fragmented. Although numerous reporting frameworks like GRI, SASB and TCFD have emerged to encourage transparency (Arvidsson and Dumay 2022), there is still no universally accepted benchmark. As a result, firms often approach ESG disclosure as a compliance exercise rather than a strategic imperative, leading to variability in reporting quality and materiality assessments (Atkins et al. 2023; Bouten et al. 2011). This lack of consistency undermines the credibility of ESG assessments and leaves stakeholders with a challenging question: How can one trust the scores that claim to reflect a firm's sustainability profile?

Amid this uncertainty, the role of efficiency analysis becomes increasingly relevant. Efficiency, in this context, is not simply a matter of cost-cutting or output maximisation but a broader evaluation of how well firms convert financial, human and environmental resources into meaningful ESG outcomes. DEA has emerged as a valuable method for benchmarking such performance. As a non-parametric technique rooted in operations research, DEA evaluates the relative efficiency of decision-making units (DMUs) by comparing multiple inputs and outputs across peers (Paradi and Zhu 2013; Antunes et al. 2024; Yang et al. 2018). This approach is particularly suited to ESG evaluation, where performance is multi-dimensional and cannot be reduced to a single metric.

DEA's flexibility enables researchers to model variable returns to scale (VRS) or constant returns to scale (CRS), capturing the unique growth trajectories of firms (Teixeira et al. 2023). Extensions such as the Malmquist-DEA Index allow for tracking technological progress and productivity over time, while Free Disposal Hull (FDH) models accommodate undesirable outputs like pollution, crucial in sectors where sustainability involves mitigating negative externalities (Chen and Ali 2004; Cherchye et al. 2002; Ferreira et al. 2020; Halkos and Tzeremes 2011). Scale efficiency (Lozano and Villa 2010; Liou and W 2011) and catch-up dynamics (Shao and Lin 2016; Ndicu et al. 2023) further deepen the interpretive power of DEA, especially in rapidly evolving industries where firms must continuously adapt to new standards.

To complement DEA, multi-criteria decision-making tools such as TOPSIS are increasingly employed. TOPSIS ranks firms based on their geometric distance from an ideal ESG performance vector, offering an intuitive way to assess trade-offs across competing criteria (Rouyendegh et al. 2020). Unlike DEA, which determines weights endogenously, TOPSIS allows for external optimisation of weights, revealing latent priorities in rating practices. This makes it particularly effective for unpacking the black box of ESG ratings, especially when different agencies use opaque and inconsistent methodologies (Behzadian et al. 2012; Kim et al. 2013).

Despite the analytical promise of these tools, the literature remains uneven in its application. ESG has been praised for its positive associations with financial resilience, especially during economic shocks like the COVID-19 pandemic (Broadstock et al. 2021; Zhou and Zhou 2021). It has been linked to improved regulatory compliance, customer loyalty and even access to capital (Paraschi 2022; Linnenluecke 2022; Lee and Kim 2022; Keeley et al. 2022). Yet, deeper questions persist around whether high ESG scores truly reflect superior sustainability performance or simply echo familiar patterns of financial robustness.

Recent studies employing DEA provide some insight into this complexity. Xie et al. (2018) and Iazzolino et al. (2023) identify a non-linear, bell-shaped relationship between ESG disclosure and efficiency, suggesting that more is not always better. Other region-specific studies reveal substantial heterogeneity. For example, Ali et al. (2022a, 2022b) highlight the positive social efficiency of ESG firms in East and Southeast Asia, while Pham et al. (2022) find stronger business performance among ESG-aligned transport firms. In contrast, Su and Xue (2023) observe that ESG boosts labour efficiency in China, whereas Moskovics et al. (2023) document similar benefits in Brazil's environmental performance. Yet Karginova-Gubinova (2022) challenges this optimism by showing that ESG initiatives had little effect on market efficiency in Russia. This finding underscores how geopolitical and institutional context can mediate ESG outcomes.

Collectively, these studies point to an unresolved challenge: ESG metrics are only as reliable as the frameworks and methodologies that underpin them. Rating agencies like Bloomberg and S&P Global—arguably two of the most influential ESG score providers—apply divergent and largely opaque methods. What remains poorly understood is how these agencies weigh various performance indicators and whether their scores meaningfully capture differences in sustainability performance or merely reflect traditional financial metrics such as profitability and leverage.

This is the critical gap the current study seeks to address. By applying DEA and TOPSIS to firm-level data scored by both Bloomberg and S&P Global, we investigate the underlying architecture of ESG scores—how they are constructed, what factors drive them, and whether they reward genuine sustainability or mask financial biases under the ESG label. In doing so, we hope to offer a more transparent, comparative and data-driven perspective on the effectiveness and credibility of ESG rating systems in a global context.

3 | Methodology

In this study, similar to Wanke et al. (2021), we use a hybrid method for the Multi-Criteria Decision Making (MCDM) approach that involves implementing three distinct steps. The first step entails testing various nonparametric frontier models, such as DEA, FDH and Malmquist Productivity Index (MPI), as well as their underlying specifications, such as CRS and VRS, to establish the boundaries of epistemic uncertainty regarding ESG scores' efficiency. The second step uses Kullback–Leibler divergence measures to evaluate the distributional similarities of efficiency score vectors in the first step and ascertain if distinct assumptions yield relevant variations in outcomes. This step is helpful in analysing and interpreting

ESG efficiency levels resulting from decisions made by firms. The third step involves utilising a quadratic programming model to maximise the correlation matrix among TOPSIS scores and efficiency measures computed in the first step. By optimising TOPSIS weights, the latent importance levels assigned by the agencies to distinct performance criteria are revealed.

3.1 | Efficiency

DEA is a non-parametric linear programming technique that was introduced by Charnes et al. (1978). This method aims to establish a connection between the estimation of technology and the calculation of performance related to this technology (Bogetoft and Otto 2010). In simpler terms, DEA is a methodology that is employed to determine the best-practice production frontiers and to estimate the relative efficiency of different decision-making units (DMUs) based on the observations of inputs and outputs. It is important to note that DEA differs from the parametric methods in that it allows for the consideration of multiple inputs and outputs, as well as the inclusion of specific functional forms that do not dictate the efficient frontier. Consequently, DEA is less constrained as it assumes that variations in the data contain information about efficiency and technology conditions. The estimation of the efficiency score for a given DMU is conducted by utilising the efficient frontier constructed by the DMUs with the highest performance (Paradi et al. 2011). Additionally, a set of DMUs is utilised to evaluate one another, and the DEA approach combines the methods of minimal extrapolation and the Farrell efficiency of a firm, resulting in a proportional improvement.

$$\begin{gathered} \min \ \theta \\ s.t. \\ \sum_{j=1}^{n} \lambda_j x_{ij} \le \theta x_{io}, \forall i \\ \sum_{j=1}^{n} \lambda_j y_{rj} \ge y_{ro}, \forall r \\ \lambda_j \ge 0, \forall j \\ \left(\text{Add } \sum_{j=1}^{n} \lambda_j = 1 \quad \text{when} \quad \text{using} \quad \text{VRS}\right) \end{gathered} \tag{1}$$

$$\begin{gathered} \max \ \phi \\ s.t. \\ \sum_{j=1}^{n} \lambda_j x_{ij} \le x_{io}, \forall i \\ \sum_{j=1}^{n} \lambda_j y_{rj} \ge \phi y_{ro}, \forall r \\ j \ge 0, \forall j \\ \left(\text{Add } \sum_{j=1}^{n} \lambda_j = 1 \quad \text{when} \quad \text{using} \quad \text{VRS}\right) \end{gathered} \tag{2}$$

This methodology incorporates various approaches to scaling, including CRS, increasing or decreasing returns to scale and VRS. The VRS specification, also known as BCC, assumes free disposability, convexity, and a $\gamma = 1$. This model ensures that each observation is benchmarked only against similar-sized observations. In contrast, a CRS considers $\gamma \ge 0$. The DEA envelopment model computation is presented above for both input-oriented and output-oriented approaches. Furthermore, Equations (1) and (2) depict the envelopment model for both CRS and VRS frontier types.

3.2 | Multi-Criteria Decision Making (MCDM)

MCDM is a sophisticated procedure encompassing numerous divergent criteria. It is common to encounter predicaments characterised by non-square matrices, rendering classical inversion techniques inapplicable. Consequently, the Moore-Penrose pseudoinverse emerges as an invaluable tool in such circumstances. Suppose A is a matrix with m alternatives and n criteria, representing how each alternative performs against each criterion. It is an $m \times n$ matrix. ω is a column vector of size n, representing the optimal weights for each criterion. These weights reflect the importance or priority of each criterion. B is a column vector of performance scores for each alternative, possibly obtained through methods like DEA, SFA (Stochastic Frontier Analysis), TOPSIS, or COPRAS (Complex Proportional Assessment). We want to find the vector ω that best satisfies the equation $A\omega = B$, given A and B.

However, if $m \neg n$, A is not square, and we cannot use the classical inversion method. Instead, we can use the Moore-Penrose pseudoinverse to find a solution that minimises the least squares error. First, we compute the Moore-Penrose Pseudoinverse, calculating the pseudoinverse of the matrix A. This can be done using Singular Value Decomposition (SVD), where $A^{+} = V \sum^{+} U^{T}$. Then we solve for weights by multiplying the pseudoinverse with vector B to obtain the optimal weights: $w = A^{+}B$. This gives a solution that minimises the least squares error $\|Aw - B\|$. The resultant vector w encompasses the optimal coefficients for each criterion, reflecting their relative significance in the decision-making procedure. The solution obtained through the utilisation of the Moore-Penrose pseudoinverse might not be unique if the system is underdetermined (i.e., if there are more criteria than alternatives). Nonetheless, it does furnish a solution that minimises the sum of squared differences, thereby presenting a sensible approach for numerous MCDM problems encountered in the real world. Moreover, this approach presupposes that the matrix A possesses a complete column rank. If it does not, the regularisation techniques may be necessary to ensure the identification of a meaningful solution. Lastly, the implementation of this method is relatively straightforward in most numerical computing environments such as MATLAB, Python with SciPy, or R, as they typically offer built-in functions for the computation of the Moore-Penrose pseudoinverse. To ensure the weight of vector w sums to 1, we normalise the weights through the Moore-Penrose pseudoinverse, dividing each weight by the sum of all weights:

$$w_i = \frac{w_i}{\sum_{j=1}^{n} w_j} \tag{3}$$

3.3 | Free Disposal Hull (FDH)

FDH is a nonparametric mathematical programming technique used for efficiency assessment and benchmarking (Fukuyama et al. 2016). It is a frontier method that allows for non-convexity in production possibilities, unlike other methods such as DEA (Fukuyama et al. 2016). The FDH model assumes free input–output disposability, meaning that any excess inputs or outputs can be disposed of without cost (Fukuyama et al. 2016). The FDH model is used to evaluate the efficiency of DMUs by comparing their performance to that of the observed units (Dibachi and Izadikhah 2023). It envelops a given sample of data with a piecewise linear hull, which represents the production possibility set (Tauchmann 2012). The efficiency of a DMU is determined by its distance to the hull, with closer distances indicating higher efficiency (Abbasi et al. 2014).

One advantage of the FDH model is its ability to handle super-efficient observations, which are located beyond the estimated production possibility frontier (Tauchmann 2012). This allows for a more comprehensive assessment of efficiency, as it considers the potential for DMUs to achieve higher levels of performance (Tauchmann 2012). Additionally, the FDH model is less sensitive to outliers compared to other nonparametric approaches (Tauchmann 2012). The FDH model has been applied in various sectors, including container ports, transit services, zakat management organisations and maintenance groups (Cullinane et al. 2005; Ryandono et al. 2021; Abbasi et al. 2014; Dibachi and Izadikhah 2023). It has been used to estimate container port production efficiency, analyse the efficiency of zakat management organisations in Indonesia, and evaluate the performance of maintenance groups under uncertainty conditions (Cullinane et al. 2005; Ryandono et al. 2021; Dibachi and Izadikhah 2023). The FDH model can be implemented using different software and commands, such as R (Tauchmann 2012). Statistical inference based on subsampling bootstrapping can be used to assess the uncertainty of the efficiency estimates (Tauchmann 2012). Furthermore, improvements have been made to the finite sample approximation of the FDH model, which enhances the accuracy of confidence intervals for individual efficiency scores (Simar and Zelenyuk 2020). The mathematical expression for calculating the efficiency score (ρ) of a DMU within FDH is presented as a linear programming problem, aiming to maximise the efficiency while adhering to the input–output constraints. For a single DMU scenario involving m desirable outputs and s undesirable outputs, the formulation proceeds as follows:

$$\text{Maximize: } \rho \tag{4}$$

Subject to:

$$\text{Input constraints: } \sum_{i=1}^{m} \lambda_i y_i \leq \rho y \tag{5}$$

where λ_i represents the weights assigned to the desirable outputs and y_i denotes the amounts of desirable outputs for the efficient DMUs.

$$\text{Undesirable output constraints: } \sum_{j=1}^{s} \mu_j z_j \leq \rho z \tag{6}$$

Where μ_j signifies the weights allocated to the undesirable outputs and z_j stands for the amounts of undesirable outputs for the efficient DMUs.

$$\text{Non}-\text{negativity constraints: } \rho \geq 0, \lambda_i \geq 0, \quad \mu_j \geq 0 \quad \text{for all } i, j \tag{7}$$

$$\text{Normalisation constraint: } \sum_{i=1}^{m} \lambda_i y_i + \sum_{j=1}^{s} \mu_j z_j = 1 \tag{8}$$

This constraint normalises the weights, ensuring their collective sum equals 1. The objective revolves around determining the maximum value of ρ while upholding the outlined constraints. The weights λ_i and μ_j play a pivotal role, signifying the relative significance or contribution of each desirable and undesirable output, respectively, to the efficiency score. This formulation empowers the FDH model to optimise concerning both output categories, facilitating a more comprehensive evaluation of efficiency.

3.4 | Technique for Order Preference by Similarity to Ideal Solution (TOPSIS)

We constructed a TOPSIS ranking employing the TOPSIS package in R (Yazdi 2013). TOPSIS, a ranking technique for MCDM, calculates positive ideal solutions that maximise the benefit criteria and minimise the cost criteria, as well as negative ideal solutions that maximise the cost criteria and minimise the benefit criteria (Roy and Shaw 2023). Naturally, there exist several alternative MCDM models, such as VIseKriterijumsko KOmpromisno Rangiranje (VIKOR) and Multi-Attributive Border Approximation Area Comparison (MABAC). Nevertheless, TOPSIS presents four primary advantages over other methods, as highlighted by Kim et al. (2013). Firstly, it is founded on sound logic that embodies the rationale behind human choice. Secondly, it relies on a scalar value that concurrently accounts for both the best and worst alternatives (Kim et al. 2013). Thirdly, it entails a straightforward computational process that can be readily implemented in a spreadsheet. Fourthly, it enables the visualisation of performance measures for all alternatives on attributes in a polyhedron, at least for any two dimensions (Kim et al. 2013). Furthermore, TOPSIS possesses the benefit of employing the computational procedure irrespective of problem size (Ic 2012). Lastly, TOPSIS also generates cardinal or scale metrics for positive and negative ideal solutions, which are obtained through linear combinations of the original criteria (Aye et al. 2018; de Andrade et al. 2020). Additionally, it represents one of several multicriteria models resembling established nonparametric and parametric efficiency measurement methods, such as DEA and stochastic frontier analysis, respectively (Bogetoft and Otto 2010; Galan and Tan 2024). Considering the ideal solutions presented by TOPSIS, each alternative is assessed by taking into account its distance from the positive ideal solution and the negative ideal solution (Behzadian et al. 2012). As a result, an evaluation matrix is formed, consisting of m alternatives and n criteria. The elements of this matrix, represented by x_{ij}, correspond to the intersections of $m \times n$. The first step in the procedure proposed by Tzeng and Huang (2011) entails normalising the initial decision matrix.

$$r_{ij} = \frac{x_{ij}}{\sqrt{\left(\sum x^2_{(ij)} \right)}} \quad \textit{for } i = 1, \ldots, m \text{ and } j = 1, \ldots, n \tag{9}$$

The initial values of the decision matrix, x_{ij}, have been transformed into r_{ij} (i.e., the standardised values of the decision matrix). Subsequently, a weighted standardised decision matrix is constructed, which serves as a valuable tool for performance evaluation, wherein w_j represents the assigned weight for the criterion j.

$$v_{ij} = \left(w_j r_{ij} \right)_{m \times n'} \tag{10}$$

In (9), we have a very well-known result.

Following the discussion of this preliminary TOPSIS procedure, we revisit this step to explain the approach employed for calculating the variable weights. The third step involves the determination of both the positive (A*) and negative (A0) ideal solutions, utilising the following equations:

$$A^{*} = \left\{v_1^{*},\ \dots,v_n^{*}\right\} \tag{11}$$

Where $v_1^{*} = \left\{\max\left(v_{ij}\right)\ if j \in J; \min\left(v_{ij}\right)\ if j \in J'\right\}$, and

$$A' = \left\{v_1',\ \dots,v_n'\right\} \tag{12}$$

Where $v_1' = \left\{\min\left(v_{ij}\right)\ if j \in J; \max\left(v_{ij}\right)\ if j \in J'\right\}$.

Defining the ideal solutions, we calculate the distances of the targets for each alternative i. The first is the separation from the positive ideal:

$$S_i^{*} = \left[\sum\left(v_i^{*} - v_{ij}\right)^2\right]^{1/2} \quad i = 1,\ \dots,m \tag{13}$$

The same is done with the negative ideal:

$$\grave{S}_i = \left[\sum\left(\grave{v}_i - v_{ij}\right)^2\right]^{1/2} \quad i = 1,\ \dots,m \tag{14}$$

Finally, we need to compute the relative closeness coefficient. All alternatives can be cardinally ranked according to the closeness coefficient:

$$C_i^{*} = \frac{\grave{S}_i}{\left(S_i^{*} + \grave{S}_i\right)} \quad 0 < C_i^{*} < 1 \tag{15}$$

There exists a notable distinction between DEA models and TOPSIS methods, as previously mentioned. In DEA models, the initial step involves the calculation of the weights assigned to each criterion within the model. Conversely, TOPSIS methods employ externally defined values for these criteria weights, thus allowing for the utilisation of various techniques to achieve this objective. DEA models themselves establish a well-suited set of weights for both inputs and outputs. On the contrary, TOPSIS calculates the Euclidean distances on the normalised vectors of the positive outputs and negative inputs criteria, employing the weights that were previously determined by the decision-maker (Aye et al. 2018).

3.5 | Weight Optimisation

The establishment of the weighting is an essential undertaking in the subsequent portion of this investigation. It incorporates the construction of a covariance matrix connecting the measures of efficiency with the indicators employed by S&P and Bloomberg. The optimisation of the weights assigned to each variable is carried out to capture the orientation of the association more effectively between the two collections of scores. The description of the method employed to optimise the weights is presented in (16). The initial group of variables consists of four frontier analyses: FDH, VRS, CRS and SE, along with three dynamic efficiency measures: MPI, technological change, and efficiency change, and a TOPSIS score. These variables utilise total assets, total liabilities, total debt, working capital and operating cash flow as inputs and revenue, net income and Earnings Before Interest and Taxes (EBIT) as outputs on an annual basis from 2018 to 2021. Furthermore, each method is computed in terms of a Farrell output-oriented efficiency measure.

$$\max\ |\mathrm{cor(TOPSIS\_\{scores\}}, X\_\{\mathrm{agency}\})|\_F\ s.t.\ 0 \le w\ \mathrm{TOPSIS}\_i \le 2,$$
$$0\,|\,\forall i \quad \sum\_\{i\}\ (w\ \mathrm{TOPSIS}\_i) = 1 \tag{16}$$

4 | Empirical Analysis

4.1 | Data

Our empirical investigation draws on a panel dataset comprising 67 publicly listed firms operating across Hong Kong, South Korea and the United States, covering a broad spectrum of 24 industries. Spanning the years 2018–2021, this dataset yields 268 firm-year observations—each one representing a distinct snapshot of how financial structures and ESG scores evolve over time. The financial data were retrieved from Yahoo Finance, while the ESG scores—our primary measure of sustainability performance—were obtained from Bloomberg and S&P Global, two of the most widely used and influential ESG rating providers in both academic and practitioner contexts (Berg et al. 2022; Gibson Brandon et al. 2022).

The variables used in the analysis were carefully chosen based on well-established research connecting financial structure to sustainability outcomes (Chatterji et al. 2016; Tamimi and Sebastianelli 2017; Duque-Grisales and Aguilera-Caracuel 2021). We treated total assets, total liabilities, total debt, working capital and operating cash flow as input variables, reflecting the resource base and financial obligations that shape a firm's operational context (Zhou and Zhou 2021; Lozano and Villa 2010). On the output side, we focused on revenue, net income and EBIT, which serve as outcome proxies for how effectively firms transform these financial inputs into performance, including their capacity to generate sustainable value (Halkos and Tzeremes 2011). This input–output configuration allowed us to operationalise a DEA framework that captures both financial efficiency and its latent influence on ESG ratings, building on methods established in prior DEA-ESG research.

The 2018–2021 time frame was chosen deliberately. It offered not only comprehensive and consistent data coverage across all three countries but also captured a unique composition of economic conditions. The years 2018 and 2019 represent a period of relative macroeconomic stability—a critical baseline—while 2020 and 2021 reflect the profound disruption brought on by the COVID-19 pandemic. This temporal span allowed us to observe whether ESG scoring remained stable or adaptive in the face of systemic shocks, drawing on Broadstock et al. (2021), who similarly explored ESG dynamics during financial crises, as well as Albuquerque et al. (2020), who examined ESG resilience during market downturns.

Constructing this panel involved several steps. From an initial pool of 200 firms identified through Bloomberg and S&P Global's ESG databases, we undertook a multi-stage filtering protocol to refine the sample. Firms with missing ESG scores, incomplete financial data, or extreme outliers that distorted comparative analysis were carefully excluded (Berg et al. 2022). This rigorous screening process, summarised in Table 1, ensured analytical integrity while preserving diversity across sectors and geographies. The final sample of 67 firms reflects a balance between data quality and representativeness, providing a robust foundation for both the DEA model and subsequent regression analyses.

Table 2 provides an overview of the descriptive statistics of ESG scores and the financial conditions of the firms in this study. What immediately stands out is the marked divergence between the two major ESG rating agencies. Bloomberg's scores suggest a broader and more generous evaluation framework, with firms averaging a score of 54.10 on the overall ESG index. The ESG components follow a similarly distributed pattern—54.60, 54.90 and 52.70, respectively—each with a standard deviation of about 28. These figures suggest not only internal consistency across the sub-pillars but also a degree of variation that reflects meaningful differentiation between firms.

S&P Global, by contrast, offers a more conservative view. The average overall ESG score across firms stands at just 29.10, nearly half that of Bloomberg and its subcomponents—environmental (29.80), social (28.70) and governance (29.00)—track closely, each with slightly narrower dispersion. This difference is not just a random variation; it highlights fundamental discrepancies in how rating agencies interpret and score ESG-related disclosures. Where Bloomberg appears to reward technological leadership and forward-looking initiatives, S&P seems linked more to relative benchmarking, leading to a more conservative assessment landscape.

Beyond the ESG scores themselves, the financial backdrop of the sample firms reveals a complex and uneven landscape. On average, firms posted a Return on Equity (ROE) of −0.72, with a standard deviation of 13.88, reflecting a highly dispersed set of outcomes—from deeply negative returns to strong equity gains. The mean return on assets (ROA) was more stable at 0.04, though modest, suggesting constrained efficiency in capital usage across the sample. These financial metrics, while narrow in average terms, point to considerable heterogeneity in performance—a pattern seen across other indicators as well.

Firm size varied dramatically. The average total assets came in at just over \$10 billion, but this figure masks an extraordinary range, as shown by a standard deviation exceeding \$27 billion. Similarly, revenue averaged \$6.71 billion, but with a spread wide enough to include firms of vastly different scales and maturities. Core profitability indicators tell a similar story. Operating cash flow, for example, averaged \$588 million, yet varied widely (SD = \$1.53 billion) and EBIT averaged \$349 million with similarly large deviations. The data show a clear contrast: some firms are financially strong, while others are struggling with instability or slow growth. Net income, a key variable in our analysis, averaged \$152 million but varied widely—with a standard deviation of \$693 million—highlighting the uneven financial landscape behind ESG evaluations. On the debt side, firms held an average of \$2.42 billion in total debt, but again, the dispersion was wide (SD = \$5.27 billion), reflecting a spectrum of capital structures and leverage strategies.

What this descriptive profile reveals is far more than just numerical variability. It tells the story of an ESG rating environment situated within a field of financial contrasts. These statistics are not simply background—they are active components in shaping how rating agencies perceive firms. As our findings later show, ESG scores may not be as neutral as they seem—they often reflect a firm's financial situation, highlighting the need for greater transparency and clearer understanding of what these scores really mean.

To contextualise our sample characteristics, we compare our descriptive statistics with those reported in prior studies of a similar nature, as detailed in Table 3. For instance, the mean ESG score in our sample (Bloomberg: 54.1; S&P: 29.1) falls within the mid-range of those observed by Mahanta et al. (2024) and Basdekidou and Papapanagos (2024), who report average ESG scores of 58.0 and 73.3, respectively, using Refinitiv data. Similarly, our financial variables, such as ROA (mean = 4%) and ROE (mean = −0.72), are broadly aligned with the ranges documented in Agarwala et al. (2024) and Cheng et al. (2024), though our sample shows slightly more variation, particularly in profitability and asset

TABLE 1 | Sample selection process.

Step	Criteria/description	Numbers of firms/observations	Rationale
1	Initial firm pool from Bloomberg and S&P ESG datasets (across HK, Korea, and the US)	200 firms (800 firm-year observations)	Firms listed in at least one ESG rating system between 2018 and 2021
2	Removed firms with incomplete ESG data for any year (Bloomberg or S&P)	−60 firms	Ensures consistent ESG score availability over the full period
3	Removed firms with missing financial data (e.g., ROA, total assets, liabilities, interest expense)	−35 firms	Required for DEA, TOPSIS, and regression models
4	Removed firms with inconsistent or outlier values (extreme financial anomalies or duplicated records)	−38 firms	Data cleaning to enhance robustness and reliability
5	Final sample retained for analysis	67 firms (268 firm-year observations)	Dataset with complete ESG and financial data over 2018–2021

TABLE 2 | Summary statistics for variables.

Variables	Min	Max	Median	Mean	SD
ROE	−227	7.41	0.08	−0.72	13.88
ROA	−0.37	0.26	0.03	0.04	0.05
Equity	-6.45×10^{5}	7.10×10^{10}	1.20×10^{8}	4.57×10^{9}	1.09×10^{10}
Total assets	9.52×10^{5}	2.11×10^{11}	3.37×10^{8}	1.02×10^{10}	2.75×10^{10}
Total liabilities	3.37×10^{5}	1.45×10^{11}	1.75×10^{8}	5.71×10^{9}	1.72×10^{10}
Revenue	4.97×10^{4}	7.63×10^{10}	1.01×10^{8}	6.71×10^{9}	1.37×10^{10}
Operating cash flow	-2.07×10^{9}	1.32×10^{10}	1.33×10^{7}	5.88×10^{8}	1.53×10^{9}
Cash resources	1.64×10^{4}	4.77×10^{9}	3.57×10^{7}	4.17×10^{8}	7.96×10^{8}
Non-current assets	3.10×10^{5}	1.89×10^{13}	2.18×10^{8}	7.38×10^{9}	2.30×10^{10}
Current assets	4.75×10^{4}	4.66×10^{10}	1.58×10^{8}	2.89×10^{9}	6.20×10^{9}
EBIT	-5.17×10^{9}	9.85×10^{9}	1.71×10^{7}	3.49×10^{8}	1.02×10^{9}
Interest expense	1.41×10^{3}	2.04×10^{9}	2.37×10^{6}	9.39×10^{7}	2.75×10^{8}
Cost of revenue	1.60×10^{3}	6.44×10^{10}	7.82×10^{7}	5.66×10^{9}	1.23×10^{10}
Current liabilities	8.92×10^{3}	3.17×10^{10}	8.85×10^{7}	2.30×10^{9}	4.79×10^{9}
Net income	-5.31×10^{9}	6.61×10^{9}	7.83×10^{6}	1.52×10^{8}	6.93×10^{8}
Total debt	1.00×10^{3}	2.82×10^{10}	9.72×10^{7}	2.42×10^{9}	5.27×10^{9}
Working capital	-9.68×10^{9}	2.55×10^{10}	4.66×10^{6}	5.98×10^{8}	2.88×10^{9}
ESG bloomberg	4.06	100	56	54.10	24.30
E bloomberg	1.60	100	53.40	54.60	27.50
S bloomberg	1.60	100	57.60[1]	54.90	28.10
G bloomberg	0	100	53.60	52.70	27.90
ESG SP	0	89	19.50	29.10	25
E SP	0	96	19	29.80	28.10
S SP	0	91	19.60	28.70	24.60
G SP	0	103	20	29	24.10

scales. For example, Cheng et al. (2024) report mean ROA and ROE of 5.98% and 12.41%, respectively, in a Chinese context. In contrast, our global sample includes firms with both positive and negative performance across multiple regions and sectors. Our firms also exhibit greater heterogeneity in total assets and revenues, indicating broader cross-industry representation. These comparisons confirm that our sample is not only representative of ESG research norms but also adds value through its methodological novelty and the diverse institutional contexts covered.

4.2 | Density

Figure 1 presents a clear but concerning picture. Most firms in our sample are performing around the 50% efficiency mark, regardless of the method used. At first glance, this might seem like a positive sign of consistency. But when we look closer, it suggests something else—a kind of stagnation. Rather than improving or falling behind, many companies appear to be stuck in the middle, not pushing forward towards stronger ESG performance. This idea aligns with previous findings that firms often aim for the 'safe middle', choosing not to lead or lag significantly in sustainability performance (Kotsantonis and Serafeim 2019).

The results become more revealing when we focus on the TOPSIS method, which shows a peak around 25% efficiency. This suggests that many firms are not close to the top performers, but rather trying to catch up. These companies may be aiming high but face barriers—such as a lack of resources, complex ESG standards, or unclear rating systems—that keep them from reaching the leaders. This pattern supports arguments by Berg et al. (2022), who found that inconsistent ESG rating methods often confuse rather than encourage firms working to improve.

Our productivity and frontier-shift results tell a similar story. The MPI shows little movement over time, meaning that both firm performance and the standards used to judge that performance

TABLE 3 | Comparative descriptive statistics from similar ESG-related studies.

Author(s), year	Variable	Mean	SD
Mahanta et al. (2024)	ESG (refinitiv)	58.00	19.82
	E score	53.57	25.70
	S score	58.23	24.14
	G score	60.79	21.01
	Total assets (million USD)	21,791	53,963
	Debt (million USD)	5358	14,178
	Revenue (million USD)	14,118	40,120
Basdekidou and Papapanagos (2024)	ESG	73.31	7.84
Agarwala et al. (2024)	ESG	33.97	11.56
	ROA	8.55	7.04
	ROE	16.68	13.12
Cheng et al. (2024)	ROA	5.98	8.35
	ROE	12.41	16.43
	Total assets (Bil RMB)	801.52	3443
	Total debt (Bil RMB)	170.00	560.51
	Net income (Bil RMB)	2.26	8.23
Xie et al. (2018)	ESG	22	13
	E score	22	16
	S score	27	15
	G score	48	7
	Revenue (million USD)	4378	15,248

Note: Mean and standard deviation of variables of similar studies.

remained largely unchanged during the study period. This suggests that while the global ESG conversation moves forward, the tools we use to measure progress may not be keeping up. This has been a concern raised by Eccles and Stroehle (2018), who warned that ESG rating systems can become outdated or disconnected from real sustainability goals.

In Figure 2, we see an essential difference between the two rating agencies. S&P's ESG scores tend to fall between 80% and 100%, while Bloomberg's scores cluster lower, around 65%. This suggests that while S&P may appear more generous in efficiency terms, this stems from a comparative framework that compresses scores towards the top, rather than truly rewarding high sustainability performance. Bloomberg, by contrast, offers a more discriminating lens, possibly making it tougher for firms to appear efficient, even if their raw scores are numerically higher. These differences are not small. They confirm concerns raised in the literature (Chatterji et al. 2016; Gibson Brandon et al. 2022) that ESG scores can vary widely depending on who is doing the rating, even when the firm's behaviour remains the same.

Altogether, these findings raise important questions. If most firms score similarly and the standards for judging them barely shift, what are ESG scores really telling us? And if different agencies give the same company different results, can we trust these scores as indicators of real sustainability performance? What we see in these figures is more than data—it's a reflection of how unclear and uneven the ESG landscape remains for firms, investors and policymakers alike.

4.3 | Weight

Figure 3 and Table 4 reveal the internal mechanics of ESG score construction by two of the field's most influential rating agencies, namely Bloomberg and S&P Global. Though both operate under the shared goal of quantifying corporate sustainability, the pathways they take diverge in ways that are both methodologically telling and ideologically significant.

In Bloomberg's case, the overwhelming reliance on the CRS model—accounting for 85% of its scoring logic—suggests a deeply entrenched assumption, indicating that firms, regardless of size, operate under similar efficiencies when it comes to transforming financial resources into ESG performance. This dominant logic is complemented by a 10% weight on frontier shift methods, showing that Bloomberg values technological change and innovation, reflecting the evolving nature of ESG progress. The small but notable inclusion of VRS at 4% indicates some room for recognising disparities in operational efficiency, particularly among firms of differing sizes or maturity levels. Interestingly, the MPI is almost entirely sidelined at 1%, suggesting limited emphasis on productivity changes over time in Bloomberg's evaluative approach.

S&P Global, by contrast, crafts its assessments through a more comparative lens. CRS still dominates at 73%, yet the agency assigns a significantly larger share (25%) to Scale Efficiency (SE). This structure implies a strong emphasis on benchmarking firms against industry leaders—those that define what ‘best practice’ looks like in ESG performance. Such an approach aligns with prior research suggesting that ESG scores often act as signalling mechanisms, reflecting not just internal practices but relative positioning within peer groups (Chatterji et al. 2016; Berg et al. 2022). The density plots bear this out: they exhibit sharp peaks at the upper bounds, consistent with rating methodologies that prioritise outperformance and comparison against a shifting frontier.

Bloomberg's incorporation of frontier shift models further underlines its forward-leaning stance. By capturing changes in technological capabilities and process improvements over time, the agency appears to reward firms that innovate, adapt and stay ahead of regulatory or market-driven ESG expectations (Eccles and Stroehle 2018). The limited but present use of VRS complements this, suggesting that Bloomberg is willing to acknowledge

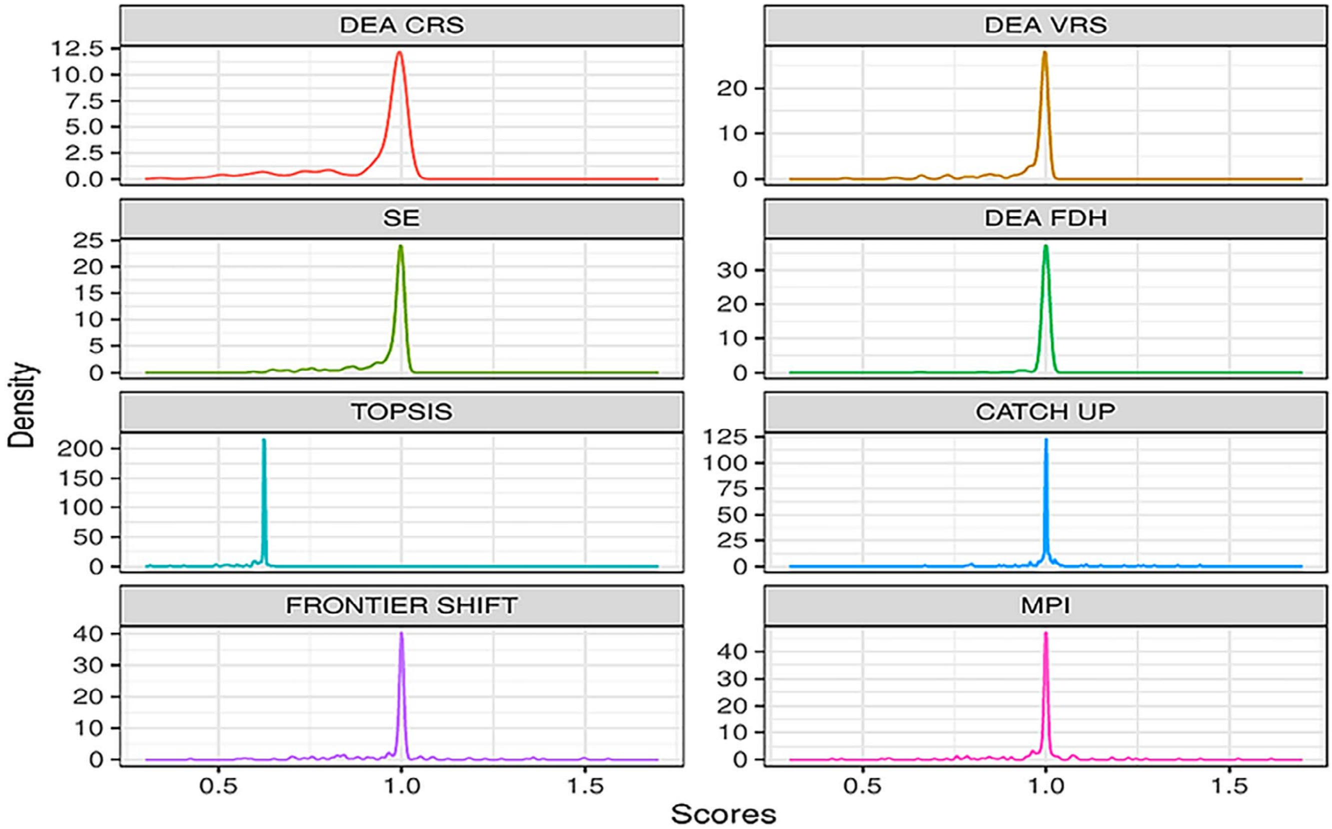

FIGURE 1 | Density plot scores. [Colour figure can be viewed at wileyonlinelibrary.com]

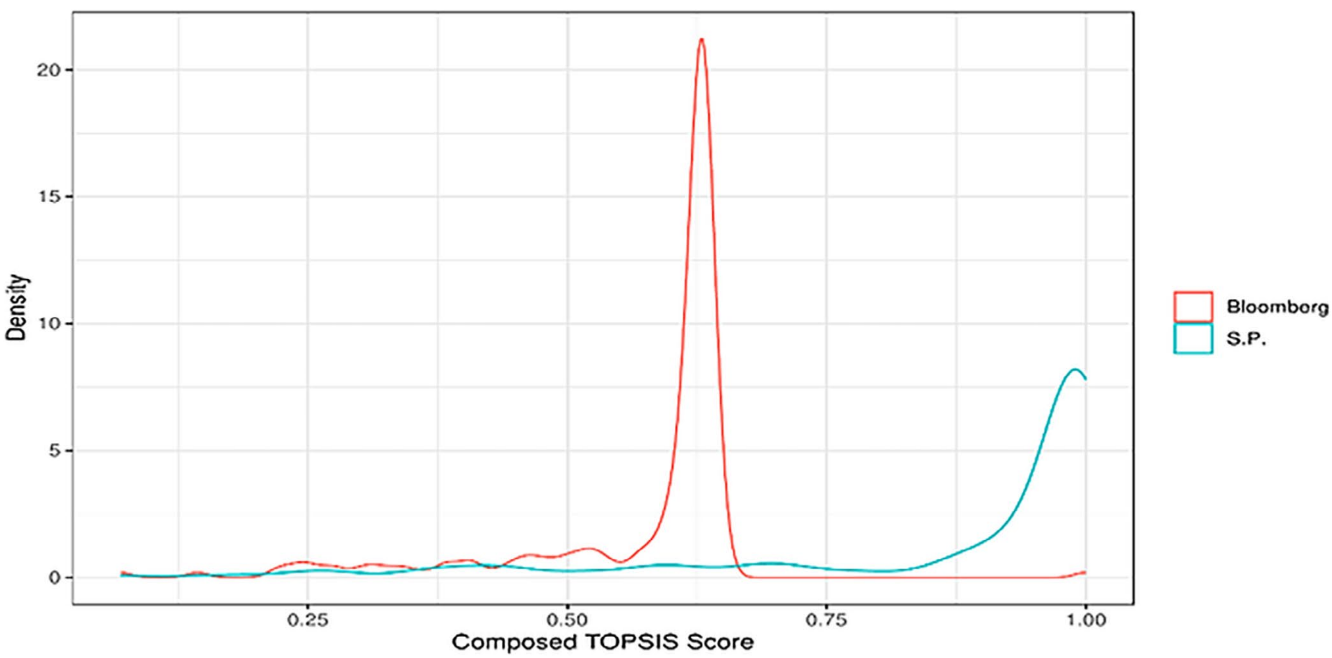

FIGURE 2 | Density plot TOPSIS of scores. [Colour figure can be viewed at wileyonlinelibrary.com]

differentiated performance levels that may not fit a one-size-fits-all efficiency assumption.

What is most striking, however, is not the divergence in scoring composition but a common thread running through both agencies' frameworks: the centrality of net income. With a weighting of 99% in both models, financial profitability emerges as the silent architect behind ESG scoring. This finding resonates with recent critiques in the literature pointing to the financialisation of ESG metrics, where performance is often inferred from fiscal health rather than concrete sustainability outcomes (Gibson Brandon et al. 2022). This implies that firms with greater financial slack are more capable of allocating resources towards sustainability reporting, technology adoption, or compliance, thus

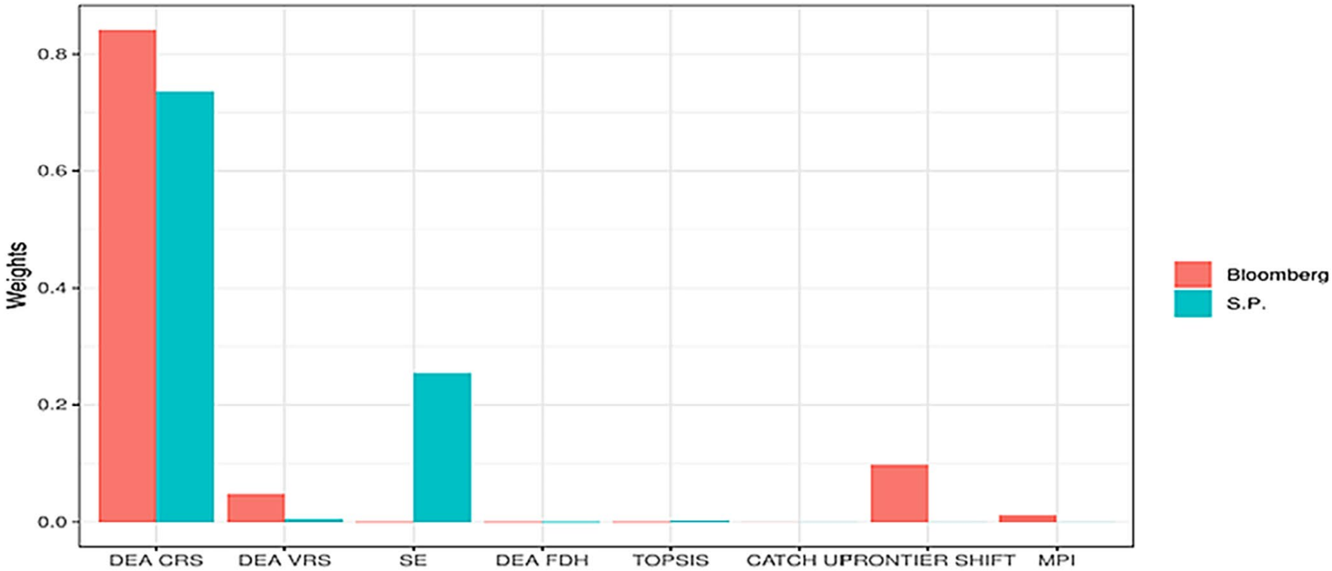

FIGURE 3 | TOPSIS weights of scores. [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 4 | Optimised weights.

Ratio	Minimum	Maximum
Total assets	1.94×10^{-9}	1.14×10^{-5}
Total liabilities	2.06×10^{-6}	8.57×10^{-3}
Total debt	1.93×10^{-7}	1.58×10^{-7}
Working capital	1.36×10^{-6}	1.27×10^{-6}
Operating cash flow	3.33×10^{-6}	2.41×10^{-4}
Revenue	4.09×10^{-10}	3.84×10^{-8}
Net income	0.99	0.99
EBIT	2.15×10^{-8}	6.29×10^{-6}

scoring higher, not necessarily because they are more sustainable, but because they are more solvent.

In short, while Bloomberg and S&P diverge in method—one leaning into innovation and flexibility, the other into benchmarking and comparison—they converge on a fundamental assumption, which argues that financial performance, particularly profitability, is a reliable proxy for ESG commitment. Whether that assumption holds under greater scrutiny is a question this study raises but does not presume to resolve. What it does suggest, however, is the urgent need for a more transparent, pluralistic and context-sensitive approach to ESG evaluation—one that resists reducing sustainability to a financial footnote.

4.4 | Regression

Table 5 reveals that the regression coefficients show no significant differences across countries, suggesting a uniform adoption of ESG practices globally, despite varying cultural contexts. This is supported by the insignificant trend variable, implying stable ESG scores over the years. Surprisingly, the Oil & Gas Refining & Marketing industry—usually seen as highly polluting—has a positive and significant coefficient of 0.14, likely reflecting investments in carbon credits, setting it apart from other sectors when compared to the real estate benchmark. Financial metrics, such as ROA and Current Liabilities Ratio, positively correlate with ESG scores, while Log Interest Expense and Log Cash Resources have a negative impact on scores. This pattern suggests that larger, more established firms with substantial assets receive higher ESG ratings, potentially due to their capacity to invest in ESG practices, unlike growing firms, which may prioritise expansion over strict ESG compliance. The impact of ROA and current liabilities on ESG scores is partly supported by Tang et al. (2024). In contrast, the negative effects of cash resources on ESG scores are in line with Jabbouri and Almustafa (2021), who focused primarily on financial performance rather than ESG. Finally, the negative impact of interest expense on ESG scores is supported by Andersson et al. (2022), who argue that interest expense may lead enterprises to prioritise financial returns over investments in environmental protection and social responsibility, resulting in lower ESG performance. This implies that agencies might impose penalties on expanding companies, assuming they have a larger environmental footprint, which results in lower ESG scores for these firms.

To ensure robustness, we conducted two additional regression analyses, namely a Panel Regression with random effects and a generalised linear model (GLM). As Table 6 indicates, the GLM was employed to accommodate potential non-normality in the distribution of the dependent variable (TOPSIS scores) and to allow for a flexible link function between predictors and outcomes. While key variables such as ESG scores and Interest Expense remained significant, others, including ROA, Current Liabilities Ratio, Construction Materials, Food Products, Household Products, Industrial Conglomerates, Cost of Revenue (log) and Cash Resources (log), lost significance. At the same time, Trend and Construction and Engineering emerged as significant. These results suggest some model-specific sensitivities, though the consistency of the main relationships reinforces the robustness of our findings.

TABLE 5 | Regression coefficients.

Variable	Coef.	SE	*z*	*p*
(Intercept)	1425	3180	0.44	0.654
ROE	0.00	0.00	1.57	0.115
ROA	0.39	0.14	2.66	0.007**
Current assets ratio	−1424	3180	−0.44	0.654
Non-current assets ratio	−1424	3180	−0.44	0.654
Current liabilities ratio	0.12	0.06	2.00	0.044*
ESG	0.00	0.00	−3.22	0.001***
S ratio	−0.07	0.05	−1.36	0.173
G ratio	0.00	0.06	−0.06	0.950
Trend	0.00	0.00	0.56	0.569
Country Hong Kong	−0.02	0.03	−0.65	0.516
Country Korea	−0.08	0.05	−1.64	0.100
Industry aluminium	0.02	0.03	0.58	0.562
Industry chemicals	−0.14	0.03	−4.46	0.000***
Industry coal & consumable fuels	−0.01	0.03	−0.30	0.764
Industry construction materials	−0.06	0.03	−2.29	0.021*
Industry construction & engineering	0.06	0.05	1.36	0.174
Industry personal products	−0.05	0.03	−1.48	0.137
Industry electric utilities	0.01	0.03	0.35	0.724
Industry food & staples retailing	0.05	0.05	0.87	0.385
Industry food products	−0.06	0.03	−2.06	0.039*
Industry paper & forest products	−0.06	0.04	−1.33	0.184
Industry gas utilities	−0.09	0.04	−2.64	0.008**
Industry household products	−0.10	0.05	−2.18	0.029*
Industry industrial conglomerates	−0.11	0.04	−2.94	0.003**
Industry machinery and electrical equipment	−0.03	0.04	−0.65	0.513
Industry electronic equipment, instruments & components	−0.18	0.05	−3.99	0.000***
Industry metals & mining	−0.04	0.04	−0.97	0.329
Industry oil & gas refining & marketing	0.14	0.04	4.00	0.000***
Industry oil & gas upstream & integrated	0.06	0.07	0.85	0.394
Industry energy equipment & services	−0.06	0.03	−1.74	0.081
Industry retailing	−0.24	0.05	−4.56	0.000***
Industry steel	−0.38	0.05	−8.05	0.000***
Industry tobacco	−0.18	0.05	−3.80	0.000***
ESG score B	−0.27	0.01	−24.30	0.000***
Interest expense (log)	−0.04	0.01	−4.44	0.000***
Cost of revenue (log)	0.02	0.01	2.06	0.039*
Cash resources (log)	−0.02	0.01	−2.16	0.030*
Log (scale)	−2.39	0.04	−64.43	0.000***

Note: The dependent variable is the TOPSIS-derived ESG scores. Standard errors (SE) are reported alongside *z* values and *p* values. Significance levels: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

The Panel Regression, as shown in Table 7, accounts for heterogeneity across firms and time, thereby improving the reliability of inference. A Hausman test indicated that random effects were the optimal choice. The results from this model closely mirrored those of the original regression, with most core variables maintaining their statistical significance.

5 | Discussion and Conclusion

This study is driven by a key question that challenges the sustainability field. What do ESG scores really measure? ESG scores are widely interpreted as signals of ethical or responsible business conduct, yet their construction remains elusive, often concealing more than they disclose. This concern is not new; studies such as Berg et al. (2022) and Mayer and Ducsai (2023) have shown that ESG ratings suffer from notable divergence, stemming from inconsistent methodologies, variable indicator weightings and a lack of transparency across providers. Liu (2022) echoes this criticism, emphasising that the proprietary nature of ESG score construction leaves stakeholders with limited ability to compare firms meaningfully.

In response to this ambiguity, our investigation set out to understand the mechanics behind ESG scores assigned by two of the most influential rating agencies—Bloomberg and S&P Global. Employing a hybrid approach that combines DEA and the TOPSIS, we aimed not merely to add methodological sophistication but to address a critical empirical gap in the literature. While ESG scores increasingly shape investment flows (Friede et al. 2015; Barko et al. 2022) and corporate positioning

TABLE 6 | Robustness analysis—GLM regression coefficients.

Variable	Coef.	SE	z	p
(Intercept)	14,018	27,333	0.51	0.608
ROE	0.00	0.00	1.15	0.251
ROA	2.43	1.26	1.93	0.054
Current assets ratio	−14,012	27,333	−0.51	0.608
Non-current assets ratio	−14,012	27,333	−0.51	0.608
Current liabilities ratio	0.74	0.54	1.37	0.170
ESG	−0.01	0.00	−3.35	0.001***
S ratio	−0.46	0.46	−0.98	0.326
G ratio	0.36	0.50	0.71	0.475
Trend	0.08	0.05	1.66	0.097*
Country Hong Kong	−0.29	0.26	−1.11	0.268
Country Korea	−0.51	0.43	−1.21	0.228
Industry aluminium	0.09	0.27	0.33	0.739
Industry chemicals	−0.58	0.26	−2.28	0.023*
Industry coal & consumable fuels	−0.11	0.28	−0.41	0.681
Industry construction materials	−0.21	0.22	−0.96	0.335
Industry construction & engineering	0.83	0.40	2.08	0.038*
Industry personal products	−0.27	0.28	−0.96	0.339
Industry electric utilities	0.15	0.25	0.60	0.548
Industry food & staples retailing	0.54	0.45	1.18	0.237
Industry food products	−0.23	0.24	−0.96	0.335
Industry paper & forest products	−0.26	0.37	−0.71	0.480
Industry gas utilities	−0.30	0.30	−0.99	0.320
Industry household products	−0.43	0.41	−1.03	0.304
Industry industrial conglomerates	−0.52	0.31	−1.67	0.094

(Continues)

TABLE 6 | (Continued)

Variable	Coef.	SE	z	p
Industry machinery and electrical equipment	−0.12	0.35	−0.36	0.719
Industry electronic equipment, instruments & components	−0.80	0.40	−2.01	0.045*
Industry metals & mining	0.13	0.32	0.40	0.691
Industry oil & gas refining & marketing	1.65	0.30	5.45	0.000***
Industry oil & gas upstream & integrated	0.31	0.57	0.54	0.586
Industry energy equipment & services	−0.11	0.30	−0.37	0.708
Industry retailing	−1.25	0.46	−2.74	0.006**
Industry steel	−1.86	0.41	−4.50	0.000***
Industry tobacco	−1.00	0.38	−2.60	0.009**
ESG score B	−1.84	0.10	−18.28	0.000***
Interest expense (log)	−0.23	0.07	−3,03	0.002**
Cost of revenue (log)	0.07	0.07	1,03	0.301
Cash resources (log)	−0.07	0.06	−1,21	0.226

Note: The dependent variable is the TOPSIS-derived ESG scores. Standard errors (SE) are reported alongside z values and p values. Significance levels: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

(Duque-Grisales and Aguilera-Caracuel 2021), relatively few studies have sought to reverse-engineer the criteria and preferences that silently govern how these scores are assigned (Gibson Brandon et al. 2022; Clementino and Perkins 2021). Our study seeks to shed light on that very complicated process.

What emerged from our analysis is a more complex—and arguably more concerning—landscape than initially expected. Bloomberg gives higher average ESG scores than S&P (54.10 vs. 29.10), which suggests it is more generous at first glance. But after applying the DEA, Bloomberg's scores lead to lower efficiency ratings, showing it actually uses stricter standards. Meanwhile, S&P's lower scores result in higher DEA efficiencies, pointing to a more lenient, peer-based rating approach that pushes firms closer to the top, even if their actual performance is modest (Behzadian et al. 2012; Antunes et al. 2024).

TABLE 7 | Robustness analysis—panel regression coefficients.

Variable	Coef.	SE	z	p
(Intercept)	1425	3180	0.45	0.654
ROE	0.00	0.00	1.58	0.115
ROA	0.40	0.15	2.66	0.008**
Current assets ratio	−1424	3180	−0.45	0.654
Non-current assets ratio	−1424	3180	−0.45	0.654
Current liabilities ratio	0.13	0.06	2.01	0.045*
ESG	0.00	0.00	−3.22	0.001***
S ratio	−0.08	0.06	−1.36	0.173
G ratio	0.00	0.06	−0.06	0.950
Trend	0.00	0.01	0.57	0.569
Country Hong Kong	−0.02	0.03	−0.65	0.516
Country Korea	−0.08	0.05	−1.64	0.100
Industry aluminium	0.02	0.03	0.58	0.562
Industry chemicals	−0.14	0.03	−4.46	0.000***
Industry coal & consumable fuels	−0.01	0.03	−0.30	0.764
Industry construction materials	−0.06	0.03	−2.29	0.022*
Industry construction & engineering	0.06	0.05	1.36	0.175
Industry personal products	−0.05	0.03	−1.48	0.138
Industry electric utilities	0.01	0.03	0.35	0.724
Industry food & staples retailing	0.05	0.05	0.87	0.386
Industry food products	−0.06	0.03	−2.06	0.040*
Industry paper & forest products	−0.06	0.04	−1.33	0.185
Industry gas utilities	−0.09	0.04	−2.64	0.008**
Industry household products	−0.10	0.05	−2.18	0.029*
Industry industrial conglomerates	−0.11	0.04	−2.94	0.003**
Industry machinery and electrical equipment	−0.03	0.04	−0.65	0.514
Industry electronic equipment. Instruments & components	−0.18	0.05	−4.00	0.000***
Industry metals & mining	−0.04	0.04	−0.97	0.330
Industry oil & gas refining & marketing	0.14	0.04	4.00	0.000***
Industry oil & gas upstream & integrated	0.06	0.07	0.85	0.394
Industry energy equipment & services	−0.06	0.03	−1.74	0.081
Industry retailing	−0.24	0.05	−4.56	0.000***
Industry steel	−0.38	0.05	−8.05	0.000***
Industry tobacco	−0.18	0.05	−3.80	0.000***
ESG score B	−0.27	0.01	−24.30	0.000***
Interest expense (log)	−0.04	0.01	−4.44	0.000***
Cost of revenue (log)	0.02	0.01	2.06	0.039*
Cash resources (log)	−0.02	0.01	−2.16	0.031*

(Continues)

TABLE 7 | (Continued)

Note: The dependent variable is the TOPSIS-derived ESG scores. Standard errors (SE) are reported alongside z values and p values. Significance levels: $^{*}p < 0.05$. $^{**}p < 0.01$. $^{***}p < 0.001$.

These methodological differences carry significant implications. They influence how firms are ranked, how investors interpret ESG credentials and how policy frameworks might standardise scoring mechanisms (Berg et al. 2022; Arvidsson and Dumay 2022).

Yet beyond the divergent logic of these agencies, another finding emerges—one that raises profound questions about what ESG scores truly measure. We observed that approximately 99% of ESG score variation could be explained by firms' net income. This relationship, supported by our regression analysis and aligned with insights from Cheng et al. (2024) and Mahanta et al. (2024), suggests that financial strength may disproportionately drive ESG evaluations. On one level, this is logical as firms with healthier financials likely have more slack resources to allocate towards ESG initiatives (Richardson and Cragg 2010; Broadstock et al. 2021). However, a more detailed reading complicates this narrative. Our data show that firms with higher cash reserves and growing interest expenses—often indicative of reinvestment, R&D, or expansion—receive systematically lower ESG scores. This mirrors the concerns raised by Jabbouri and Almustafa (2021) and Andersson et al. (2022), who argue that financial leverage and liquidity dynamics can lead agencies to penalise firms despite forward-looking sustainability efforts.

These findings are not just statistical anomalies—they have real consequences. For investors, they underscore the need for a more critical reading of ESG scores. Although these ratings are frequently used to align investment portfolios with ethical values, the influence of financial metrics—particularly profitability—means that the moral signal may be distorted (Gibson

Brandon et al. 2022). Two firms with comparable ESG practices may receive markedly different ratings, not because of differences in ethical conduct, but because of differences in capital structure or operational maturity.

For companies, our analysis provides a more precise roadmap for navigating ESG assessments. Instead of treating ESG scores as opaque or externally imposed, firms can begin to understand how their internal financial architecture—net income, working capital, operating cash flow—shapes their sustainability profile. This finding supports recent calls to integrate financial performance with sustainability strategy in a more deliberate, evidence-based way (Ali et al. 2022a; Iazzolino et al. 2023). It also reinforces the notion that ESG disclosure should not be treated merely as a box-ticking exercise, but as a strategic narrative underpinned by the financial realities of the firm (Atkins et al. 2023).

From a regulatory standpoint, the divergence in scoring logic between Bloomberg and S&P Global signals a troubling absence of standardisation. If high-growth, innovation-heavy firms are being penalised based on assumptions about risk or opacity, then ESG scores may be systematically underrepresenting future-oriented sustainability commitments (Moskovics et al. 2023; Karginova-Gubinova 2022). Regulators might consider this a call to action, encouraging greater transparency in ESG methodology and possibly mandating disclosure of rating criteria to safeguard comparability and investor confidence (Arvidsson and Dumay 2022; Veldman et al. 2023).

This study also contributes methodologically by offering a replicable model for uncovering latent ESG scoring logic through the integration of DEA and TOPSIS. These tools allow for the exploration of efficiency and closeness-to-ideal profiles while simultaneously accounting for the weights that ESG agencies implicitly assign to financial and operational variables (Aye et al. 2018; Antunes et al. 2023). We believe this approach can serve as a foundation for future inquiries into ESG scoring construction across other providers, industries and regulatory regimes.

Importantly, the broader social relevance of this work should not be overlooked. ESG scores have moved beyond niche financial metrics to become influential drivers of corporate legitimacy. If they are internally inconsistent or financially biased, the consequences extend well beyond firm rankings—they ripple into capital markets, employment outcomes, innovation incentives and public trust (Zhou and Zhou 2021; Sandberg et al. 2023). By exposing the financial undercurrents within ESG scores, this research contributes to a more grounded and transparent dialogue on sustainability assessment in practice.

That said, our study is not without limitations. Our data span only three countries—Hong Kong, South Korea and the United States—across the years 2018 to 2021. While this cross-national design introduces comparative diversity, it may limit generalisability to emerging markets or more regulated ESG environments (Bouten et al. 2011). Future research should expand this scope to include underrepresented economies, where ESG adoption varies more widely (Su and Xue 2023).

Moreover, we did not account explicitly for exogenous shocks—such as changes in environmental regulations, political risk, or climate events—which may significantly influence scoring outcomes. Integrating these contextual variables into future DEA frameworks could enrich the explanatory power of ESG scoring studies (Zhao et al. 2022). Similarly, our reliance on annual data may mask short-term fluctuations. Researchers might consider using higher-frequency data to assess how financial volatility, ESG scandals, or regulatory shifts are absorbed into ESG ratings in real time.

Finally, while our DEA-TOPSIS hybrid offers a solid analytical foundation, there is ample scope for methodological triangulation. Alternative MCDM techniques—such as Preference Ranking Organisation Method for Enrichment Evaluation (PROMETHEE) or Analytic Hierarchy Process (AHP)—may yield complementary or competing insights into how agencies rank firms (Roy and Shaw 2023; de Andrade et al. 2020). With ESG methodologies evolving rapidly—particularly with the incorporation of AI and machine learning—future research could explore how advanced models capture dynamic, non-linear relationships that conventional frameworks may overlook (Tan and Tsionas 2022).

As ESG continues to move from the periphery to the core of financial decision-making, the need for robust, transparent and accountable scoring methodologies becomes ever more urgent. The present study offers one step towards that end, bridging the normative promise of ESG with the operational mechanics that too often remain in the shadows.

Ethics Statement

The authors have nothing to report.

Consent

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Endnotes

[1] The Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) is a widely used multi-criteria decision-making (MCDM) method that ranks alternatives based on their geometric proximity to an ideal solution. In the context of this study, TOPSIS enables the reverse-engineering of ESG scoring patterns by identifying how closely each firm's financial and operational profile aligns with the optimal ESG performance as implicitly defined by Bloomberg and S&P Global. Its methodological strength lies in simultaneously considering both the best and worst possible scenarios for each criterion, thus offering a more comprehensive evaluation framework. Unlike DEA, which derives weights endogenously, TOPSIS relies on pre-determined or optimised weights, making it particularly effective for uncovering latent prioritizations within ESG assessments (Behzadian et al. 2012; Kim et al. 2013; Aye et al. 2018; de Andrade et al. 2020).

References

Abbasi, M., G. Jahanshahloo, M. Rostamy-Malkhlifeh, and F. H. Lotfi. 2014. "Estimation of Congestion in Free Disposal Hull Models Using Data Envelopment Analysis." *Scientific World Journal* 2014: 427673.

Agarwala, N., S. Jana, and T. N. Sahu. 2024. "ESG Disclosures and Corporate Performance: A Non-Linear and Disaggregated Approach." *Journal of Cleaner Production* 437, no. 1: 140517. https://doi.org/10.1016/j.jclepro.2023.140517.

Albuquerque, R., Y. Koskinen, S. Yang, and C. Zhang. 2020. "Resiliency of Environmental and Social Stocks: An Analysis of the Exogenous COVID-19 Market Crash." *Review of Corporate Finance Studies* 9, no. 3: 593–621.

Ali, M. H., N. A. Rahim, M. H. Yahya, and F. Kamarudin. 2022a. "Efficiency in Giving Back to the Masses: Insights From ESG and Non-ESG Firms in Selected East Asian Countries." *Asia-Pacific Management Accounting Journal* 17: 247–277.

Ali, M. H., N. A. Rahim, M. H. Yahya, and F. Kamarudin. 2022b. "Determinants of Social Performance Efficiency of ESG and Non-ESG Firms: Evidence From Southeast Asian Countries." *Management and Accounting Review* 21: 129–165.

Andersson, E., M. Hoque, M. L. Rahman, G. S. Uddin, and R. Jayasekera. 2022. "ESG Investment: What Do We Learn From Its Interaction With Stock, Currency and Commodity Markets?" *International Journal of Finance and Economics* 27: 3623–3639.

Antunes, J., A. Hadi-Vencheh, A. Jamshidi, Y. Tan, and P. Wanke. 2023. "TEA-IS: A Hybrid DEA-TOPSIS Approach for Assessing Performance and Synergy in Chinese Health Care." *Decision Support Systems* 171: 113916.

Antunes, J., A. Hadi-Vencheh, A. Jamshidi, Y. Tan, and P. Wanke. 2024. "Cost Efficiency of Chinese Banks: Evidence From DEA and MLP-SSRP Analysis." *Expert Systems With Applications* 237: 121432.

Arjoon, S. 2006. "Striking a Balance Between Rules and Principles-Based Approaches for Effective Governance: A Risks-Based Approach." *Journal of Business Ethics* 68: 53–82.

Arvidsson, S., and J. Dumay. 2022. "Corporate ESG Reporting Quantity, Quality and Performance: Where to Now for Environmental Policy and Practice?" *Business Strategy and the Environment* 31: 1091–1110.

Atkins, J., F. Doni, A. Gasperini, S. Artuso, I. L. Torre, and L. Sorrentino. 2023. "Exploring the Effectiveness of Sustainability Measurement: Which ESG Metrics Will Survive COVID-19?" *Journal of Business Ethics* 185: 629–646.

Aye, G. C., R. Gupta, and P. Wanke. 2018. "Energy Efficiency Drivers in South Africa: 1965–2014." *Energy Efficiency* 11: 1465–1482.

Barko, T., M. Cremers, and L. Renneboog. 2022. "Shareholder Engagement on Environmental, Social, and Governance Performance." *Journal of Business Ethics* 180: 777–812.

Basdekidou, V., and H. Papapanagos. 2024. "The Use of DEA for ESG Activities and DEI Initiatives Considered as "Pillar of Sustainability" for Economic Growth Assessment in Western Balkans." *Digital* 4, no. 3: 572–598. https://doi.org/10.3390/digital4030029.

Behzadian, M., S. K. Otaghsara, M. Yazdani, and J. Ignatius. 2012. "A State-of the-Art Survey of Topsis Applications." *Expert Systems With Applications* 39: 13051–13069.

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

Bogetoft, P., and L. Otto. 2010. *Benchmarking With Dea, Sfa, and r.* Springer Science & Business Media.

Bouten, L., P. Everaert, L. V. Liedekerke, L. D. Moor, and J. Christiaens. 2011. "Corporate Social Responsibility Reporting: A Comprehensive Picture?" *Accounting Forum* 35: 187–204.

Broadstock, D. C., K. Chan, L. T. Cheng, and X. Wang. 2021. "The Role of ESG Performance During Times of Financial Crisis: Evidence From COVID-19 in China." *Finance Research Letters* 38: 101716.

Charnes, A., W. W. Cooper, and E. Rhodes. 1978. "Measuring the Efficiency of Decision Making Units." *European Journal of Operational Research* 2: 429–444.

Chatterji, A. K., R. Durand, D. I. Levine, and S. Touboul. 2016. "Do Ratings of Firms Converge? Implications for Managers, Investors and Strategy Researchers." *Strategic Management Journal* 37, no. 8: 1597–1614.

Chen, S., Y. Song, and P. Gao. 2023. "Environmental, Social, and Governance (ESG) Performance and Financial Outcomes: Analysing the Impact of ESG on Financial Performance." *Journal of Environmental Management* 345: 118829.

Chen, Y., and A. I. Ali. 2004. "DEA Malmquist Productivity Measure: New Insights With an Application to Computer Industry." *European Journal of Operational Research* 159: 239–249.

Cheng, L. T. W., C. K. Tsang, and S. K. Lee. 2024. "Comparing the Financial Performance Effect of International and Local ESG Ratings: A Two-Stage DEA Approach." *Annals of Financial Economics* 19, no. 4: 1–21. https://doi.org/10.1142/S2010495225500010.

Cherchye, L., T. Kuosmanen, and T. Post. 2002. "FDH Directional Distance Functions With an Application to European Commercial Banks." *Journal of Productivity Analysis* 15: 201–215.

Clementino, E., and R. Perkins. 2021. "How Do Companies Respond to Environmental, Social and Governance (ESG) Ratings? Evidence From Italy." *Journal of Business Ethics* 171: 379–397.

Cullinane, K., D.-W. Song, and T. Wang. 2005. "The Application of Mathematical Programming Approaches to Estimating Container Port Production Efficiency." *Journal of Productivity Analysis* 24: 73–92.

de Andrade, L. H., J. J. M. Antunes, and P. Wanke. 2020. "Performance of TV Programs: A Robust MCDM Approach." *Benchmarking: An International Journal* 27: 1188–1209.

Dibachi, H., and M. Izadikhah. 2023. "Maintenance Groups Evaluation Under Uncertainties: A Novel Stochastic Free Disposal Hull in the Presence of Log-Normally Distributed Data." *RAIRO—Operations Research* 57: 1843–1876.

Duque-Grisales, E., and J. Aguilera-Caracuel. 2021. "Environmental, Social and Governance (ESG) Scores and Financial Performance of Multilatinas: Moderating Effects of Geographic International Diversification and Financial Slack." *Journal of Business Ethics* 168: 315–334.

Eccles, R. G., and J. C. Stroehle. 2018. "Exploring Social Origins in the Construction of ESG Measures." SSRN Electronic Journal.

Ferreira, D. C., R. C. Marques, M. I. Pedro, and C. Amaral. 2020. "Economic Inefficiency Levels of Urban Solid Waste Management Services in Portugal." *Sustainability* 12: 4170.

Friede, G., T. Busch, and A. Bassen. 2015. "ESG and Financial Performance: Aggregated Evidence From More Than 2000 Empirical Studies." *Journal of Sustainable Finance & Investment* 5, no. 4: 210–233.

Fuente, G., M. Ortiz, and P. Velasco. 2022. "The Value of a Firm's Engagement in ESG Practices: Are We Looking at the Right Side?" *Long Range Planning* 55: 102143.

Fukuyama, H., J. Leth Hougaard, K. Sekitani, and J. Shi. 2016. "Efficiency Measurement With a Non-Convex Free Disposal Hull Technology." *Journal of the Operational Research Society* 67: 9–19.

Galan, J. E., and Y. Tan. 2024. "Green Light for Green Credit? Evidence From Its Impact on Bank Efficiency." *International Journal of Finance and Economics* 29, no. 1: 531–550.

Gibson Brandon, R., S. Glossner, P. Krueger, P. Matos, and T. Steffen. 2022. "Do Responsible Investors Invest Responsibly?" *Review of Finance* 26, no. 6: 1389–1432.

Halkos, G. E., and N. G. Tzeremes. 2011. "A Conditional Nonparametric Analysis for Measuring the Efficiency of Regional Public Healthcare Delivery: An Application to Greek Prefectures." *Health Policy* 103: 73–82.

Iazzolino, G., M. E. Bruni, S. Veltri, D. Morea, and G. Baldissarro. 2023. "The Impact of Esg Factors on Financial Efficiency: An Empirical Analysis for the Selection of Sustainable Firm Portfolios." *Corporate Social Responsibility and Environmental Management* 30: 1917–1927.

Ic, Y. T. 2012. "An Experimental Design Approach Using Topsis Method for the Selection of Computer-Integrated Manufacturing Technologies." *Robotics and Computer-Integrated Manufacturing* 28: 245–256.

Jabbouri, I., and H. Almustafa. 2021. "Corporate Cash Holdings, Firm Performance and National Governance: Evidence From Emerging Markets." *International Journal of Managerial Finance* 17, no. 5: 783–801.

Karginova-Gubinova, V. 2022. "Impact of the ESG Focus on Any Efficiency and Fairness of the Stock Market." *World of Economics and Management* 22: 21–34.

Keeley, A. R., C. Li, S. Takeda, T. Gloria, and S. Managi. 2022. "The Ultimate Owner of Environmental, Social, and Governance Investment." *Frontiers in Sustainability* 3: 909239.

Kim, Y., E.-S. Chung, S.-M. Jun, and S. U. Kim. 2013. "Prioritizing the Best Sites for Treated Wastewater Instream Use in an Urban Watershed Using Fuzzy Topsis." *Resources, Conservation and Recycling* 73: 23–32.

Kotsantonis, S., and G. Serafeim. 2019. "Four Things no One Will Tell You About ESG Data." *Journal of Applied Corporate Finance* 31, no. 2: 50–58.

Lee, E., and G. Kim. 2022. "Analysis of Domestic and International Green Infrastructure Research Trends From the Esg Perspective in South Korea." *International Journal of Environmental Research and Public Health* 19: 7099.

Linnenluecke, M. K. 2022. "Environmental, Social and Governance (Esg) Performance in the Context of Multinational Business Research." *Multinational Business Review* 30: 1–16.

Liou, J.-L., and P.-I. W. 2011. "Will Economic Development Enhance the Energy Use Efficiency and CO_2 Emission Control Efficiency?" *Expert Systems With Applications* 38: 12379–12387.

Liu, M. 2022. "Quantitative Esg Disclosure and Divergence of Esg Ratings." *Frontiers in Psychology* 13: 936798.

Lozano, S., and G. Villa. 2010. "Gradual Technical and Scale Efficiency Improvement in DEA." *Annals of Operations Research* 173: 123–136.

Luo, L., and Q. Tang. 2023. "The Real Effects of ESG Reporting and GRI Standards on Carbon Mitigation: International Evidence." *Business Strategy and the Environment* 32: 2985–3000.

Mahanta, A., N. C. Sahu, P. K. Behera, and P. Kumar. 2024. "Variations in Financial Performance of Firms With ESG Integration in Business: The Mediating Role of Corporate Efficiency Using DEA." *Green Finance* 6, no. 3: 518–562. https://doi.org/10.3934/GF.2024020.

Mayer, R., and A. R. Ducsai. 2023. "ESG: Credibility Behind the Scores–the Reliability and Transparency of Esg Ratings." *PRO* 10: 1–14.

Moskovics, P., P. Wanke, Y. Tan, and A. M. Gerged. 2023. "Market Structure, ESG Performance, and Corporate Efficiency: Insights From Brazilian Publicly Traded Companies." *Business Strategy and the Environment* 33: 241–262. https://doi.org/10.1002/bse.3492.

Ndicu, S., D. Ngui, and L. Barasa. 2023. "Technological Catch-Up, Innovation, and Productivity Analysis of National Innovation Systems in Developing Countries in Africa 2010–2018." *Journal of the Knowledge Economy* 15: 7941–7967. https://doi.org/10.1007/s13132-023-01327-4.

O'Riordan, L., and J. Fairbrass. 2014. "Managing CSR Stakeholder Engagement: A New Conceptual Framework." *Journal of Business Ethics* 125: 121–145.

Paradi, J. C., S. Rouatt, and H. Zhu. 2011. "Two-Stage Evaluation of Bank Branch Efficiency Using Data Envelopment Analysis." *Omega* 39: 99–109.

Paradi, J. C., and H. Zhu. 2013. "A Survey on Bank Branch Efficiency and Performance Research With Data Envelopment Analysis." *Omega* 41: 61–79.

Paraschi, E. P. 2022. "Why Esg Reporting Is Particularly Important for the Airlines During the Covid-19 Pandemic." *Journal of Business and Management Studies* 4: 63–67.

Pelosi, N., and R. Adamson. 2016. "Managing the "S" in ESG: The Case of Indigenous Peoples and Extractive Industries." *Journal of Applied Corporate Finance* 28: 87–95.

Pham, T. N., P. Tran, M. Le, H. N. P. Vo, C. S. Pham, and H. X. Nguyen. 2022. "The Effects of Esg Combined Score on Business Performance of Enterprises in the Transportation Industry." *Sustainability* 14: 8354.

Richardson, B. J., and W. Cragg. 2010. "Being Virtuous and Prosperous: SRI'S Conflicting Goals." *Journal of Business Ethics* 92: 21–39.

Rouyendegh, B. D., A. Yildizbasi, and P. Üstünyer. 2020. "Intuitionistic Fuzzy TOPSIS Method for Green Supplier Selection Problem." *Soft Computing* 24: 221–2228.

Roy, P. K., and K. Shaw. 2023. "A Credit Scoring Model for SMEs Using AHP and TOPSIS." *International Journal of Finance and Economics* 28, no. 1: 372–391.

Ryandono, M. N. H., A. S. Qulub, E. F. Cahyono, et al. 2021. "Analysis of Efficiency for Zakat Management Organization in Indonesia: A Com-Parison Study of Super Efficiency and Free Disposal Hull." *Hayula: Indonesian Journal of Multidisciplinary Islamic Studies* 5: 147–166.

Sancak, I. E. 2023. "Change Management in Sustainability Transformation: A Model for Business Organizations." *Journal of Environmental Management* 330: 117165.

Sandberg, H., A. Alnoor, and V. Tiberius. 2023. "Environmental, Social, and Governance Ratings and Financial Performance: Evidence From the European Food Industry." *Business Strategy and the Environment* 32: 2471–2489.

Shao, B. B. M., and W. T. Lin. 2016. "Assessing Output Performance of Information Technology Service Industries: Productivity, Innovation and Catch-Up." *International Journal of Production Economics* 172: 43–53.

Simar, L., and V. Zelenyuk. 2020. "Improving Finite Sample Approximation by Central Limit Theorems for Estimates From Data Envelopment Analysis." *European Journal of Operational Research* 284: 1002–1015.

Su, J., and L. Xue. 2023. "ESG Performance, Demographic Trend, and Labour Investment Efficiency in China." *Applied Economics Letters* 31: 2207–2213. https://doi.org/10.1080/13504851.2023.2212956.

Tamimi, N., and R. Sebastianelli. 2017. "Transparency Among S&P 500 Companies: An Analysis of ESG Disclosure Scores." *Management Decision* 55, no. 8: 1660–1680.

Tan, Y., and M. Tsionas. 2022. "Modelling Sustainability Efficiency in Banking." *International Journal of Finance and Economics* 27: 3754–3772.

Tang, S., L. He, F. Su, and X. Zhou. 2024. "Does Directors' and Officers' Liability Insurance Improve Corporate ESG Performance? Evidence From China." *International Journal of Finance and Economics* 29, no. 3: 3713–3737.

Tauchmann, H. 2012. "Partial Frontier Efficiency Analysis." *Stata Journal* 12: 461–478.

Teixeira, R., J. J. M. Antunes, P. Wanke, H. L. Correa, and P. Wanke. 2023. "Customer Satisfaction and Airport Efficiency in Brazil: A Hybrid NDEA-AHP Approach." *Benchmarking: An International Journal* 31: 2239–2266. https://doi.org/10.1108/BIJ-11-2022-0702.

Tzeng, G.-H., and J.-J. Huang. 2011. *Multiple Attribute Decision Making: Methods and Applications*. CRC press.

Veldman, J., T. Jain, and C. Hauser. 2023. "Virtual Special Issue on Corporate Governance and Ethics: What's Next?" *Journal of Business Ethics* 183: 329–331.

Wanke, P. F., J. J. Antunes, V. Y. Miano, C. L. do Couto, and F. G. Mixon. 2021. "Measuring Higher Education Performance in Brazil: Government Indicators of Performance vs Ideal Solution Efficiency Measures." *International Journal of Productivity and Performance Management* 71: 2479–2495.

Xie, J., W. Nozawa, H. Fujii, and M. Yagi. 2018. "Do Environmental, Social and Governance Activities Improve Corporate Financial Performance?" *Business Strategy and the Environment* 28, no. 2: 286–300. https://doi.org/10.1002/bse.2224.

Yang, G.-l., H. Fukuyama, and Y.-y. Song. 2018. "Measuring the Inefficiency of Chinese Research Universities Based on a Two-Stage Network DEA Model." *Journal of Informetrics* 12: 10–30.

Yazdi, M. 2013. "Topsis: Topsis Method for Multiple-Criteria Decision Making (mcdm)." R package version 1. https://rdrr.io/cran/topsis/.

Zanten, J. A., and R. Tulder. 2021. "Improving Companies' Impacts on Sustainable Development: A Nexus Approach to the SDGS." *Business Strategy and the Environment* 30: 3703–3720.

Zhao, Y., J. Antunes, Y. Tan, and P. Wanke. 2022. "Demographic Efficiency Drivers in the Chinese Energy Production Chain: A Hybrid Neural Multi-Activity Network Data Envelopment Analysis." *International Journal of Finance and Economics* 29: 1762–1780. https://doi.org/10.1002/ijfe.2765.

Zhou, D., and R. Zhou. 2021. "Esg Performance and Stock Price Volatility in Public Health Crisis: Evidence From the COVID-19 Pandemic." *International Journal of Environmental Research and Public Health* 19: 202.