Check for updates

REVIEW

REVISED Divergence and aggregation of ESG ratings: A survey

[version 2; peer review: 3 approved]

Arianna Agosto, Alessandra Tanda

Department of Economics and Management, University of Pavia, Pavia, Lombardia, 27100, Italy

v2 **First published:** 30 Jan 2025, **5**:28
https://doi.org/10.12688/openreseurope.19238.1
Latest published: 30 Jun 2025, **5**:28
https://doi.org/10.12688/openreseurope.19238.2

Abstract

Purpose

This paper reviews the existing literature on Environmental, Social, and Governance (ESG) ratings divergence and aggregation methods. It highlights the challenges posed by inconsistent ESG ratings and their implications for investment decisions.

Design/methodology/approach

The study conducts a comprehensive review of prior research focusing on ESG ratings, examining their correlation levels and the methodologies employed to assess corporate sustainability. It also investigates traditional aggregation techniques and modern machine learning approaches used to address these inconsistencies.

Findings

The review reveals that ESG ratings exhibit a low level of correlation across different providers, raising concerns about their reliability as investment indicators. Although some studies propose advanced aggregation methods to enhance accuracy, significant gaps remain in understanding how to effectively consolidate ESG information to create a dependable sustainability indicator.

Originality

This paper provides a critical analysis of the current state of ESG rating methodologies, emphasizing the need for improved aggregation strategies. It underscores the importance of future research in leveraging ESG data to develop more consistent and reliable measures of corporate sustainability.

Open Peer Review

Approval Status ✓ ✓ ✓

	1	2	3
version 2 (revision) 30 Jun 2025		✓ view	✓ view
version 1 30 Jan 2025	✓ view	? view	

1. **Vincenzo Capizzi**, University of Piemonte Orientale, Novara, Italy
2. **Luca Di Persio**, Verona University, Verona, Italy
3. **Diana Elena Vasiu**, Lucian Blaga University of Sibiu, Sibiu, Romania

Any reports and responses or comments on the article can be found at the end of the article.

Keywords
ESG ratings, divergence, aggregation, survey

This article is included in the COST Actions gateway.

Corresponding author: Alessandra Tanda (alessandra.tanda@unipv.it)

Author roles: Agosto A: Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Tanda A**: Conceptualization, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the European Union's Framework Programme for Research & Innovation as part of the COST Action CA19130 'Fintech and Artificial Intelligence in Finance' (FinAI), as supported by the COST Association (European Cooperation in Science and Technology). This project also received funding from the European Union - Next Generation EU, Component M4C2, Investment 1.1, Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN) – Notice 1409, 14/09/2022- BANDO PRIN 2022 PNRR. Project title: 'Climate risk and uncertainty: environmental sustainability and asset pricing'. Project code 'P20225MJW8', CUP: F53D23009320001. And has also received funding the European Union-NextGenerationEU, in the framework of the GRINS -Growing Resilient, INclusive and Sustainable (GRINSPE00000018). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2025 Agosto A and Tanda A. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Agosto A and Tanda A. **Divergence and aggregation of ESG ratings: A survey [version 2; peer review: 3 approved]** Open Research Europe 2025, **5**:28 https://doi.org/10.12688/openreseurope.19238.2

First published: 30 Jan 2025, **5**:28 https://doi.org/10.12688/openreseurope.19238.1

REVISED **Amendments from Version 1**

We have rewritten the introduction and several key sections to better integrate the reviewers' suggestions, providing a clearer explanation of the motivation behind our work and the main contributions it offers. In particular, we have updated the section on the literature review to incorporate more recent and relevant studies, ensuring a more comprehensive overview of the existing research. Additionally, we have added a dedicated section outlining the methodology used to select the papers included in our analysis, thereby improving the transparency and reproducibility of our literature survey.

We have also expanded the discussion on the metrics most commonly used to assess divergence in ESG scores. Specifically, in Section 5 of the revised manuscript, we clarified that most studies quantify divergence by analysing the low pairwise correlations between ESG ratings assigned by different providers. Among the most frequently used measures, the Pearson correlation coefficient is often adopted due to the continuous nature of ESG scores, although rank-based alternatives like Spearman's coefficient are also used. Furthermore, we summarised studies that incorporate unsupervised techniques such as Principal Component Analysis (PCA) to explore the structure of ESG data.

In response to the reviewer's comments, we emphasised the absence of a standardised and widely accepted framework for evaluating the performance of ESG score aggregation methods. This lack remains a significant limitation in the field. Finally, we outlined several directions for future research in the concluding section of the revised manuscript, including the application of clustering techniques to group companies based on their ESG profiles and the investigation of methods to enhance the interpretability of the weights used in aggregated scores. We also discussed the trade-offs involved in using complex, computationally intensive tools - such as Shapley values - for explainability, which may not always offer clear benefits in terms of predictive performance.

Any further responses from the reviewers can be found at the end of the article

1 Introduction and background

The environmental impact of corporate activities has become increasingly important over the last decades (Yu *et al.*, 2021). Indeed, companies are made accountable for the impact of their business activities on the planet and society through the measurement of their Corporate Social Performance (CSP).

Many companies have implemented over time sustainability reporting that started on a voluntary basis (e.g., the Global Reporting Initiative - GRI) (Monteiro *et al.*, 2023, for a review) and was later incorporated into regulatory provisions. For instance, in Europe, the Non-Financial Reporting Directive (NFRD) (Directive 2014/95/EU) introduced in 2014 the requirement to publish non-financial information related to ESG topics for large companies and public-interest entities with more than 500 employees. The directive was then revised by the Corporate Sustainability Reporting Directive (CSRD) (Directive 2022/2464/EU). The new rules expand the NFRD, to improve the quality and consistency of information provided by companies on sustainability. Additionally, it enlarges the applicability of these rules to all large companies and, gradually, to listed SMEs. The Directive also introduces European Sustainability Reporting Standards and the external assurance (verification) of sustainability information.

Besides corporate disclosure requirements, several data providers developed sustainability scores or ratings, often proxied through the ESG profile of the companies (Pollman, 2022). This information, together with companies' disclosure in annual reports or sustainability reports complements financial information available on the market and enables investors to assess the sustainability of corporates, i.e., considering how companies' behaviour affects external stakeholders and paying attention to the economic, environmental, social, and time factors (Lozano, 2012; Lozano, 2018; Muñoz-Torres *et al.*, 2019). In other words, this set of information contributes to the reduction of asymmetries of information for investors and can determine a reduction in the cost of capital for corporates (Yu *et al.*, 2021).

Despite these efforts to promote the diffusion of sustainability disclosure, the existence of multiple ESG rating providers determined to a proliferation of methods and scores that often disagree and create "aggregate confusion" on the market (Berg *et al.*, 2022). This has the perverse effect to generate additional uncertainty, rather than actually reducing asymmetries of information on the market for investors.

This paper surveys the evidence on ESG rating divergence: low correlation between ESG ratings of different providers is determined by differences in methods and definitions. Low agreement impacts the investment strategies and outcomes of portfolio allocation (Erhart, 2022).

Understanding the origin of ESG rating divergence can help regulators put in place proper measures to increase and harmonise the ESG disclosure of companies, investors to choose between alternative rating providers or to manage divergence through the aggregation of scores issued by different providers, and companies to improve disclosure of their ESG approach

The paper also reviews the existing methods to overcome divergence by aggregating different ESG scores. Traditional techniques and more advanced approaches coexist in the literature, with their advantages and limits.

The manuscript contributes to the discussion on ESG rating transparency and harmonisation and formulates research questions that warrant attention from scholars.

The paper is structured as follows: Section 2 provides a brief overview of the ESG ratings and scores; Section 3 motivates the need for developing ESG aggregating measures by presenting the relevant literature focused on ESG rating divergence and its possible sources; Section 4 presents the most relevant contributions in terms of ESG rating aggregation; Section 5 concludes.

2 ESG ratings and scores

The ESG profile of companies is often taken as a proxy of their sustainability behaviour (Eng *et al.*, 2022).

While Environment factors relate to the impact on the environment deriving from the production of goods or services, such as carbon emissions, preservation of the natural environment, biodiversity protection, waste and water management (European Commission, nd; Robeco, nd; Times, nd); Social factors measure the impact of the company on society, including issues of employee satisfaction, diversity, inequality, gender gap, protection of young and children, investment in human capital and communities, and human rights (European Commission, nd; Van Duuren *et al.*, 2016); Governance factors evaluate the "good" governance of companies, which can contribute to a more sustainable and balanced firms' growth, and thus to a more sustainable economic development (Adams & Mehran, 2012; Alkaraan, 2023; Esteban-Sanchez *et al.*, 2017).

These three factors have also become the basis for investment decisions within sustainable financing and can drive the choice of investors in terms of which companies to finance through equity or debt. Furthermore, since international authorities have included sustainability in their agenda (European Commission, 2018; European Commission, 2021; UN, 2015), policymakers and regulators have developed specific provisions. In the financial market, European regulators request financial intermediaries to include ESG aspects for lending and investment decisions (EBA, 2020; ESMA, 2020a; ESMA, 2020b) and request credit rating agencies to include ESG aspects into creditworthiness evaluations (European Action Plan for Financing Sustainable Growth) (European Commission, 2018). The aim is increase allocative efficiency and direct funds to the best-performing companies in terms of environmental and social impact.

The ESG factors are also related to economic performance. Events caused by climate change can affect profitability and produce losses able to affect companies' ability to meet their credit obligations. Companies are therefore pushed by different actors towards a more sustainable behaviour. In addition, they are incentivised to disclose more about their ESG behaviour, to reduce asymmetries of information and improve their ability to retrieve cheaper funds (Yu *et al.*, 2021).

To inform investors concerning corporate ESG performance, specialised companies (including rating agencies) publish ESG ratings or scores that express the level of sustainability and the degree of accountability of companies on ESG aspects (Avetisyan & Ferrary, 2013; Scalet & Kelly, 2010). Each rating provider collects information from different sources (company reports, news, stock exchange information, etc.) and applies proprietary methodologies to combine available inputs and produce a synthetic measure of ESG behaviour.

ESG evaluations assigned by different providers often produce divergent results (Abhayawansa & Tyagi, 2021; Berg *et al.*, 2022; Billio *et al.*, 2021; Dimson *et al.*, 2020; Dorfleitner *et al.*, 2015), yielding to potential "sustainability arbitrage" (Pollman, 2022). Indeed, divergent ESG ratings for a given company may emerge, providing opaque information about the company's ESG profile and pave the way for greenwashing practices (Agnese *et al.*, 2025). KPMG surveyed more than 160 ESG ratings and data providers (KPMG, 2020), with numerous agencies (eg. Bloomberg, Thomson Reuters, S&P, etc.) and found their ESG ratings are divergent. The outcome and impact dimensions of ESG behaviour have limited effect in determining ESG ratings (Guerrero & Viteri, 2025).

The presence of divergent scores for the same companies increases uncertainty for investors and regulators, who may need to choose among different measures and/or combine them into a single metric.

As the importance of ESG metrics will grow in the future and influence investors' decisions, and firms' ability to finance their investment, the market feels a strong need for ESG metric standardisation. This represents a major managerial and policy challenge, coupled with the growing importance of sustainable finance[1] and climate change effects on creditworthiness of companies.

3 Methods

We searched the database Scopus for manuscripts on ESG ratings divergence and aggregation in early April 2025. We employed the following search strings for divergence (TITLE-ABS-KEY ((“ESG score*” OR “ESG RATING*”) AND “*AGREE*”) OR TITLE-ABS-KEY ((“ESG score*” OR “ESG RATING*”) AND “*VERGE*”)) and for aggregation TITLE-ABS-KEY ((“ESG score*” OR “ESG RATING*”) AND “*aggr*”).

The search returned, respectively, 173 and 90 documents. We manually scanned through the titles, abstracts and keywords and selected the papers that explicitly discuss divergence of ESG ratings, the consequences on financial markets and investors, and aggregation techniques. We eliminate irrelevant studies and review the remaining to provide a bird's eye view of the literature available to researchers approaching this topic.

4 Divergence of ESG scores

Empirical evidence shows little convergence between different ESG ratings through empirical studies (Anselmi & Petrella, 2025; Bissoondoyal-Bheenick *et al.*, 2024; Vasiu, 2024).

In most works, the disagreement in ESG evaluations is expressed in terms of low pairwise correlations between the ESG ratings of different providers. As ESG scores are usually

[1] Just as a background information, the number of sustainable investment funds available in Europe grew from 200 at the end of 2016 to 3,196 at the end of 2020 (Morningstar, 2021), with a corresponding growth in the managed assets.

continuous, the most natural choice to model their correlation is the Pearson correlation coefficient.

Specifically, given two providers, i and j, each of which assigning ESG scores to a set of N companies, the Pearson correlation coefficient between the scores assigned by i and j is calculated as:

$$\rho_{ij}^{P} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$$

Where σ_{ij} is the covariance between ESG scores i and j, calculated on the N-sized sample of companies, while σ_i and σ_j are the standard deviations of scores i and j respectively.

Pearson correlation is also used by Bissoondoyal-Bheenick *et al.* (2024) to evaluate the similarity of ESG scores aggregated with different methods (see Section 5).

However, correlation can be also calculated through rank-based measures such as the Spearman coefficient, defined as:

$$\rho_{ij}^{S} = \frac{\sigma_{ij}^{R}}{\sigma_i^{R} \sigma_j^{R}}$$

where the σ_{ij}^{R} covariance and the σ_i^{R} and σ_j^{R} standard deviations are calculated using the scores converted to ranks, rather than the original ones.

Rank correlation is used by Billio *et al.* (2021), together with mean absolute error and the percentage of observed agreement among the considered rating providers.

Among the studies focused on ESG score divergence, Dorfleitner *et al.* (2015), analysing ESG rating data of publicly traded companies provided by three agencies in the 2002–2012 period, find that the ratings assigned by the considered providers exhibit low linear correlations and significantly differ in distribution, with a small percentage of companies assigned to the same score quartile. The differences in the scoring approaches - revealed in the qualitative evaluation of the three rating methodologies performed in the same paper - partially explain such inconsistencies. Similar conclusions were reached by Chatterji *et al.* (2016), who documented the lack of agreement across ESG ratings issued by six providers between 2002 and 2010, even after adjusting for explicit differences in the definition of CSP held by the different raters. In their review of the related literature, Abhayawansa and Tyagi (2021) examined the possible causes of the differences in the ESG ratings and rankings produced by different agencies, and found that divergences in ESG scores are mainly attributable to disagreements about the definition, composition and weightings for each of the E, S and G factors and to differences in the methodological approaches. In particular, according to Capizzi *et al.* (2021), the weight component is crucial in explaining ESG rating divergences. Bissoondoyal-Bheenick *et al.* (2024) find high correlations between Asset4 and Sustainalytics, but not among other providers.

The issue of divergence of ESG ratings and its possible causes were explored in depth by Berg *et al.* (2022), who provided a broad empirical study to document the ESG evaluations' divergence and identify the factors that contribute to the differences. Specifically, the authors consider data from six different ESG rating providers in 2014, including ESG ratings and the related underlying variables. The latter are grouped by imposing a taxonomy of categories, which is then used to decompose the divergence into contributions of scope, measurement, and weight. By re-estimating ESG ratings based on the identified categories, the authors found that the most important dimensions in driving ESG rating divergence are climate risk management, product safety, corporate governance, corruption, and environmental management system. Further differences among ratings arise when considering the Environmental, Social and Governance dimensions separately. Such discrepancies are connected to the nature and measurement of the three factors. Indeed, while the Environmental impact can be relatively easy to measure, the Social and Governance impacts are related to qualitative aspects, which are more difficult to assess. Muñoz-Torres *et al.* (2019) find that rating agencies, when computing ESG scores, mostly focus on the environmental dimension (e.g. waste reduction, energy consumption, water management).

The sources of divergence of ESG rating consist of a lack of commonality in the ESG score criteria used by prominent agencies, with the consequence that agencies assign even opposite ratings to a given company (Billio *et al.,* 2021). Kurbus and Rant (2025) discuss the legal drivers of ESG ratings disagreement, differentiating between civil law and common law firms: correlation disagreement is lower for civil law firms, while dispersion disagreement is low for common law corporates. Also country and company characteristics can influence the ESG ratings disagreement (Rubino *et al.*, 2024). Despite some criticisms, evidence shows that rating agencies worked to update their computational methodologies for ESG ratings, although they are not able to entirely capture into ratings (Escrig-Olmedo *et al.*, 2019; Muñoz-Torres *et al.*, 2019).

Many studies focus on the impact of ESG divergence on corporate performance, market efficiency, and portfolio decisions and performance. ESG score disclosure can encourage companies to improve their CSP behaviour (Zeng & Xu, 2019), but ESG disagreement presents possible drawbacks. Firms tend to reduce corporate green innovation (Li *et al.*, 2024; Li *et al.*, 2025; Xiao *et al.*, 2025; Yuan *et al.*, 2024; Zhu *et al.*, 2025) or total factor productivity (Li & Yang, 2025) when ESG ratings disagree. But ESG rating divergence can result in additional efforts by companies to promote the quality and quantity of green innovation (Wan *et al.*, 2025; Zhou *et al.*, 2024; Zhou *et al.*, 2025). ESG rating disagreement can also act as a moderating factor for the corporate performance, e.g., efficiency (Lin *et al.*, 2025).

Billio *et al.* (2021) investigate the implications that ESG rating disagreements may have on ESG portfolios performance.

Indeed, the identification of sustainable investments and therefore the choice of the relevant benchmarks depends on the ratings provided by agencies. ESG rating heterogeneity leads to the identification of alternative benchmarks, with the consequence that the impact of ESG profile on asset prices is weak and green investments often show no significant difference in financial performances with respect to non-ESG ones.

ESG rating disagreement improves stock pricing efficiency (Yin *et al.*, 2025), also under high analysts' attention (Tan *et al.*, 2025), but contradicting evidence highlights the benefits of reducing investor ambiguity via an improved market efficiency (Bachner, 2025).

ESG rating discordance also negatively impacts stock excess return rates; using Chinese market data, Wang *et al.* (2024) found that the smaller the distance between ESG rating issued by domestic and foreign agencies, the higher the excess returns on stocks.

5 ESG score aggregation

In the economic and financial context, scholars frequently construct indicators to synthesise multiple input variables. The need for synthesising variables in a single indicator also arises in the ESG context and is related to the empirically found divergence of ESG ratings discussed in Section 3. Indeed, when ESG scores provided by different agencies for a given company are discordant, a possible solution for the investor/regulator is to use an aggregated measure which synthesises them into a single metric.

The recent academic literature provides ways to correct the divergence in ESG ratings (Dupuy & Garibal, 2022) and solutions to build ESG-aggregated scores.

The weighted mean represents a simple and reliable solution (Bissoondoyal-Bheenick *et al.*, 2024), but the estimation of weights can be problematic and their choice can be subjective, affecting the final indicators. Indeed, the most natural mathematical formulation for an aggregated ESG indicator (ESG^i_{aggr}) is a linear combination of the scores assigned to the same company i by a set of $k = 1,...,K$ providers (ESG^i_k) and the corresponding w_k weights :

$$ESG^i_{aggr} = \sum_{k=1}^{K} w_k ESG^i_k$$

However, the estimation of weights can be problematic, and their choice can be subjective, affecting the final indicators. The issue is so relevant that the European Commission provided a set of recommendations for the construction of aggregated metrics to be used for policy purposes and indicated the steps to be followed to build composite indicators, including data management, weighting, uncertainty and sensitivity analysis (European Union and Joint Research Centre, 2008). Erhart (2022) refines the weighted mean approach using Montecarlo simulations and varying weights (1/N ±25%) to build an aggregate ESG score by the arithmetic, geometric, or harmonic mean.

Data-driven techniques can help build aggregate measures from several input variables. A method widely used in the statistical literature is Principal Component Analysis (PCA), that aims at creating one or more new factors from a larger set of observed variables, where each component is a linear combination of the original variables and orthogonally independent from the other components. Gucciardi *et al.* (2024) employ PCA (together with pairwise correlation analysis and OLS) to evaluate the "common factors" driving Environmental scores. The study finds a limited number of common factors underlying the E score of different providers, mainly linked to more quantitative variables (e.g., natural resources use). Bissoondoyal-Bheenick *et al.* (2024) apply PCA and voting average to Sustainalytics, MSCI, and Asset4 ESG ratings and conclude the methods might contribute to ESG rating divergence.

The limits of PCA emerge when building indexes with a normative value, according to Gai *et al.* (2023), who developed an alternative aggregation method based on Multidimensional Synthesis of Indicators (MSI). The authors claim this method can also help in detecting greenwashing practices in companies. Dynamic Factor Models (DFM) also represent an alternative to PCA that allows taking the temporal dimension into account. Similarly to PCA, DFM creates new factors through a system of latent variables and parameters associated with the contribution of each observed variable (see Bitetto *et al.*, 2021). Both PCA and DFM assess the role played by each variable, easing the interpretability of results. However, both approaches require a large set of variables (at least 10–15), which is not always available. Note that the above-mentioned issue of weighting is also crucial in this situation, as the weight attributed to each rating should not only be easily estimated, but also interpreted, allowing the stakeholders to be aware of the contribution of each indicator to the assessment of the company sustainability profile.

Agosto *et al.* (2023a) and Agosto *et al.* (2023b) relied on a Bayesian approach - based on Giudici *et al.* (2003) and Cerchiello and Giudici (2014) - to obtain an aggregated indicator for the ESG performance of companies by integrating ratings assigned by different providers. In both works the aggregated ESG indicator is computed starting from single ESG scores, by attributing a weight to each. The proposed weighting procedure is data-driven, as it relies on the relationship between the ESG performance and the creditworthiness measured by credit ratings issued by recognised agencies.

In Agosto *et al.* (2023b), each available ESG score receives a weight that is a function of the likelihood of the observed counts of credit rating classes, under the alternative partitions generated by the ESG scores. The likelihood weights are obtained through the application of Bayes' theorem. The relationship between ESG scores and credit rating was suggested by previous research and prompted by policymakers and regulators (EBA, 2020; Lagoarde-Segot, 2019). The proposed methodology is applied to a sample of 791 European companies for which the ESG scores issued by three different providers,

together with Moody's credit rating, are available. For each of the mentioned variables, the most recent value in the 2018–2020 period is considered. After estimating the likelihood weights for the considered ESG scores, the authors performed an out-of-sample analysis to check whether the combined score improves the accuracy of credit rating predictions, as measured by using the distance (in notches) between the observed and predicted credit rating class. The findings show that the merged score reduces the overestimation of credit risk (lower positive error) for the best rated companies (AA, A and BBB classes) with respect to the individual scores, while, for the BB and B-rated companies, the cases of credit risk underestimation (negative errors) are more frequent and higher in magnitude with the combined ESG score than with the individual ones. In other words, according to the results of Agosto *et al.* (2023b), the merged sustainability rating seems to better recognise the most credit-reliable companies.

The Bayesian approach is also used by Agosto *et al.* (2023a), who built an indicator of ESG performance of listed companies that integrates the ESG scores assigned by different providers. In this work, the credit-related target variable used to weight the single scores is a binary one which indicates whether a company's credit rating is speculative or investment grade, according to the Bloomberg Issuer Default Risk model. Specifically, the higher the capability of the ESG score to classify companies into speculative and investment grade, the higher the attributed weight. The authors apply the described methodology to a training sample of European listed companies and validate the estimated weights on a testing sample. For each company, the probability of belonging to a speculative rating class is calculated as the weighted mean of the probabilities assigned by the three considered scores, using the Bayesian likelihood-based weights. By comparing the classification performance - in terms of Receiving Operating Curve (ROC) - of the aggregated score with that of the individual ones, the authors found that the merged score performs better in the tails of the distribution, where the more extreme financial profiles (very bad or very good companies) lie. The authors also showed that the results are robust to the inclusion of financial ratios as control variables.

Besides Bayesian learning techniques, machine learning and AI algorithms are used to statistically aggregate single indicators. Tech-driven alternative methods provide a more transparent and democratic aggregation process (Hughes *et al.*, 2021). Indeed, Agosto *et al.* (2023a) applied an artificial intelligence method, XGBOOST (Chen, 2015), an AI ensemble model which works over the idea of combining several weak classifiers to create a strong one characterised by extremely good performance thanks to a regularised gradient boosting framework. The XGBOOST and the Bayesian model lead to very similar results in terms of classification performance, but, as claimed by the authors, the latter provides explainable weights which can be interpreted and used in further evaluations. Indeed, when, as in this case, a machine learning technique not explainable by design leads to a poor gain in predictive accuracy, the use of computationally expensive AI methods like Shapley values is not justified. López-García *et al.* (2025) apply Un-weighted TOPSIS - technique for order preference by similarity to ideal solution - to Top 100 EMEA companies and provide an ESG ranking of companies that solves the divergence in ratings.

It is worth mentioning that, regardless of the specific aggregation methodology, it is difficult to measure and compare the performance of the obtained ESG indicators, due to the absence of a target variable. Indeed, differently from credit scoring models, which are validated using the actual incidence of defaults, the ESG behaviour cannot be easily and objectively summarised by an observed variable. Indeed, among the presented works, some use unsupervised techniques such as PCA (Bissoondoyal-Bheenick *et al.*, 2024; Gucciardi *et al.*, 2024), while others rely on the relationship between ESG and financial performance (Billio *et al.*, 2021; Gai *et al.*, 2023) or credit worthiness measured by agency ratings (Agosto *et al.*, 2023a; Agosto *et al.*, 2023b).

6 Concluding remarks

This paper has reviewed the literature on the divergence of ESG ratings. It has become evident that the proliferation of ESG rating providers has brought a high number of ratings that investors, policymakers and markets can access to evaluate the sustainability of companies. While this can be positive, there is a general lack of consensus on the "best" rating to be used as these ratings have a high level of divergence. Hence, this can create confusion and have the perverse effect of increasing uncertainty instead of reducing asymmetries of information.

To partially overcome this issue, a few studies have tried to employ statistical aggregation techniques to leverage the information included in each rating and provide more coherent and complete information on the overall ESG profile of firms. Studies employ both traditional techniques (e.g., PCA or OLS) and more advanced algorithms (e.g., XGBOOST). There is, to date, no strictly preferable approach and further research is needed to provide additional insights into the divergence of ESG ratings and new attempts are needed to improve the utility and accuracy of aggregation methods. In addition, the unavailability of an homogeneous and easily implementable testing framework represents a limitation of the existing ESG aggregation methodologies. Applying cluster analysis to the proposed ESG scores and identifying common financial and strategic features of companies belonging to the same clusters is one of the aspects that deserve further research, together with a more in- depth analysis of the topic of explainability of the weights attributed to different scores.

For sure, the definition of a harmonised regulatory framework and an increase in the transparency of these ratings will help to solve ESG score divergence. Also, from the corporates' point of view, understanding what are the drivers of ESG ratings will help them to better plan their strategies to improve their environmental, social and governance performance in an effective way that produces a tangible effect on their performance or risk and hence that determines their ability to stay on the market.

Policy efforts to improve standardisation of ESG metrics remains vital for investors to choose among investment opportunities; companies to take appropriate measures to improve their environmental, social and governance footprint; and ESG

rating providers to improve their transparency (OICV-IOSCO, 2021).

Ethics and consent

Ethical approval and consent were not required.

Data availability

No data are associated with this article.

Acknowledgement

The paper is not published elsewhere in this current form.

References

Abhayawansa S, Tyagi S: **Sustainable investing: the black box of Environmental, Social, and Governance (ESG) ratings.** *The Journal of Wealth Management.* 2021; **24**(1): 49–54.
Publisher Full Text

Adams RB, Mehran H: **Bank board structure and performance: evidence for large bank holding companies.** *J Financ Intermed.* 2012; **21**(2): 243–267.
Publisher Full Text

Agnese P, Carè R, Cerciello M, *et al.*: **Reconsidering the impact of environmental, social and governance practices on firm profitability.** *Manag Decis.* 2025; **63**(1): 25–48.
Publisher Full Text

Agosto A, Cerchiello P, Giudici P: **Bayesian learning models to measure the relative impact of ESG factors on credit ratings.** *Int J Data Sci Anal.* 2023a; 1–12.
Publisher Full Text

Agosto A, Giudici P, Tanda A: **How to combine ESG scores? A proposal based on credit rating prediction.** *Corp Soc Responsib Environ Manag.* 2023b; **30**(6): 3222–3230.
Publisher Full Text

Alkaraan F: **Corporate governance and sustainability issues.** *Corporate Governance and Sustainability Review.* 2023; **7**(1): 4–6.
Publisher Full Text

Anselmi G, Petrella G: **ESG ratings: disagreement across providers and effects on stock returns.** *Journal of International Financial Markets, Institutions and Money.* 2025; **100**: 102133.
Publisher Full Text

Avetisyan E, Ferrary M: **Dynamics of stakeholders' implications in the institutionalization of the CSR field in France and in the United States.** *J Bus Ethics.* 2013; **115**(1): 115–133.
Publisher Full Text

Bachner F: **Decoding market reactions: analysis of divergent signals of ESG ratings.** *Int Rev Financ Anal.* 2025; **103**: 104161.
Publisher Full Text

Berg F, Kölbel JF, Rigobon R: **Aggregate confusion: the divergence of ESG ratings.** *Rev Financ.* 2022; **26**(6): 1315–1344.
Publisher Full Text

Billio M, Costola M, Hristova I, *et al.*: **Inside the ESG Ratings: (Dis)agreement and performance.** *Corp Soc Responsib Environ Manag.* 2021; **28**(5): 1426–1445.
Publisher Full Text

Bissoondoyal-Bheenick E, Bennett S, Lehnherr R, *et al.*: **ESG rating disagreement: implications and aggregation approaches.** *International Review of Economics & Finance.* 2024; **96**(Part A): 103532.
Publisher Full Text

Bitetto A, Cerchiello P, Mertzanis C: **A data-driven approach to measuring epidemiological susceptibility risk around the world.** *Sci Rep.* 2021; **11**(1): 24037.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Capizzi V, Gioia E, Giudici G, *et al.*: **The divergence of ESG ratings: an analysis of Italian listed companies.** *J Financ Manag Mark Inst.* 2021; **9**(02): 2150006.
Publisher Full Text

Cerchiello P, Giudici P: **Bayesian credit ratings.** *Commun Stat Theory Methods.* 2014; **43**(4): 867–878.

Chatterji AK, Durand R, Levine D, *et al.*: **Do ratings of firms converge? Implications for managers, investors and strategy researchers.** *Strateg Manag J.* 2016; **37**(8): 1597–1614.
Publisher Full Text

Chen T: **Xgboost: extreme gradient boosting.** *R package version 0.4-2.* 2015; **1**(4).

Dimson E, Marsh P, Staunton M: **Divergent ESG ratings.** *J Portf Manag.* 2020; **47**(1): 75–87.
Publisher Full Text

Dorfleitner G, Halbritter G, Nguyen M: **Measuring the level and risk of corporate responsibility - an empirical comparison of different ESG rating approaches.** *Journal of Asset Management.* 2015; **16**(7): 450–466.
Publisher Full Text

Dupuy P, Garibal JC: **Cross-dispersion bias-adjusted ESG rankings.** *J Asset Manag.* 2022; **23**(7): 631–643.
Publisher Full Text

EBA: **Discussion paper on management and supervision of ESG risks for credit institutions and investment firms.** EBA/DP/2020/03. 2020.
Reference Source

Eng LL, Fikru M, Vichitsarawong T: **Comparing the informativeness of sustainability disclosures versus ESG disclosure ratings.** *Sustainability Accounting, Management and Policy Journal.* 2022; **13**(2): 494–518.
Publisher Full Text

Erhart S: **Take it with a pinch of salt—ESG rating of stocks and stock indices.** *Int Rev Financ Anal.* 2022; **83**: 102308.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Escrig-Olmedo E, Fernández-Izquierdo MA, Ferrero-Ferrero I, *et al.*: **Rating the raters: evaluating how ESG rating agencies integrate sustainability principles.** *Sustainability.* 2019; **11**(3): 915.
Publisher Full Text

ESMA: **No action letter on sustainability-related disclosures for benchmarks**. 2020a.
Reference Source

ESMA: **Strategyon sustainable finance**. ESMA 2-105-1052, 2020b.
Reference Source

Esteban-Sanchez P, de la Cuesta-Gonzalez M, Paredes-Gazquez JD: **Corporate social performance and its relation with corporate financial performance: International evidence in the banking industry.** *J Clean Prod.* 2017; **162**: 1102–1110.
Publisher Full Text

European Commission: **Action plan: financing sustainable growth.** Communication from the Commission to the European Parliament, the European Council, the Council, the European Central Bank, the European Economic and Social Committee and the Committee of the regions, Brussels, 8.3.2018, COM(2018) 97 final, 2018.
Reference Source

European Commission: **Provisional agreement on the European Climate Law**. 2021.
Reference Source

European Commission: **Overview of sustainable finance**. (n.d.).
Reference Source

European Union and Joint Research Centre: **Handbook on constructing composite indicators: methodology and user guide.** OECD publishing, 2008.
Reference Source

Gai L, Bellucci M, Biggeri M, *et al.*: **Banks' ESG disclosure: a new scoring model.** *Financ Res Lett.* 2023; **57**: 104199.
Publisher Full Text

Giudici P, Mezzetti M, Muliere P: **Mixtures of products of Dirichlet processes for variable selection in survival analysis.** *J Stat Plan Inference.* 2003; **111**(1–2): 101–115.
Publisher Full Text

Gucciardi G, Ossola E, Parisio L, *et al.*: **Common factors behind companies' environmental ratings.** University of Milan Bicocca Department of Economics, Management and Statistics Working Paper Forthcoming, 2024.
Publisher Full Text

Guerrero S, Viteri JP: **What are environmental, social, and governance scores measuring? The role of outcome and impact indicators in ESG scores.** *Financ Res Lett.* 2025; **72**: 106529.
Publisher Full Text

Hughes A, Urban MA, Wójcik D: **Alternative ESG ratings: how technological**

innovation is reshaping sustainable investment. *Sustainability.* 2021; **13**(6): 3551.
Publisher Full Text

KPMG: **Sustainable investing: fast-forwarding its evolution**. Technical report, 2020.
Reference Source

Kurbus B, Rant V: **A legal origins perspective on ESG rating disagreement.** *Res Int Bus Finance.* 2025; **74**: 102702.
Publisher Full Text

Lagoarde-Segot T: **Sustainable finance. A critical realist perspective.** *Res Int Bus Finance.* 2019; **47**: 1–9.
Publisher Full Text

Li Z, Yang Z: **ESG rating disagreement and corporate Total Factor Productivity: inference and prediction.** *Financ Res Lett.* 2025; **78**: 107127.
Publisher Full Text

Li J, Ye Y, Li J: **Corporate behavior under confusion: ESG disagreement and green innovation.** *Asia-Pacific Journal of Accounting & Economics.* 2025; 1–20.
Publisher Full Text

Li L, Zhang D, Li R: **ESG rating disagreement and corporate innovation: evidence from China.** *Financ Res Lett.* 2024; **62**(Part A): 105096.
Publisher Full Text

Lin YE, Teng S, Yu B, *et al.*: **ESG rating, rating divergence and investment efficiency: international evidence.** *Q Rev Econ Finance.* 2025; **100**: 101975.
Publisher Full Text

López-García A, Liern V, Pérez-Gladish B: **Determining the underlying role of corporate sustainability criteria in a ranking problem using UW-TOPSIS.** *Ann Oper Res.* 2025; **346**: 1321–1344.
Publisher Full Text

Lozano R: **How can 'theories of the firm' help make societies more sustainable.** *SRI PAPERS.* 2012; **47**: 1–9.
Reference Source

Lozano R: **Proposing a definition and a framework of organisational sustainability: a review of efforts and a survey of approaches to change.** *Sustainability.* 2018; **10**(4): 1157.
Publisher Full Text

Monteiro AP, Pereira C, Barbosa FM: **Environmental disclosure on mandatory and voluntary reporting of Portuguese listed firms: the role of environmental certification, lucratively and corporate governance.** *Meditari accountancy research.* 2023; **31**(3): 524–553.
Publisher Full Text

Morningstar: **Sustainable funds' record-breaking year.** 2021.
Reference Source

Muñoz-Torres MJ, Fernández-Izquierdo MA, Rivera-Lirio JM, *et al.*: **Can environmental, social, and governance rating agencies favor business models that promote a more sustainable development?** *Corp Soc Responsib Environ Manag.* 2019; **26**(2): 439–452.
Publisher Full Text

OICV-IOSCO: **Environmental, Social and Governance (ESG) ratings and data products providers consultation report.** cr02/21. Technical report, 2021.
Reference Source

Pollman E: **The making and meaning of ESG.** U of Penn, Inst for Law & Econ Research Paper, (22-23). 2022.
Reference Source

Robeco: **Sustainability investing glossary: ESG definition.** Technical report. (n.d.).

Rubino M, Mastrorocco I, Garegnani GM: **The influence of market and institutional factors on ESG rating disagreement.** *Corp Soc Responsib Environ Manag.* 2024; **31**(5): 3916–3926.
Publisher Full Text

Scalet S, Kelly TF: **CSR rating agencies: what is their global impact?** *J Bus Ethics.* 2010; **94**(1): 69–88.
Publisher Full Text

Tan C, Yin K, Wu H, *et al.*: **Analysts' ESG attention and stock pricing efficiency: evidence from machine learning and text analysis.** *Journal of Accounting Literature.* ahead-of-print. 2025.
Publisher Full Text

Times F: **Definition of ESG.** (n.d.).
Reference Source

United Nations: **Global Sustainable Development Report (GSDR).** 2015.
Reference Source

Van Duuren E, Plantinga A, Scholtens B: **ESG integration and the investment management process: fundamental investing reinvented.** *J Bus Ethics.* 2016; **138**(3): 525–533.
Publisher Full Text

Vasiu DE: **ESG rates divergence on the emerging markets in the European Union.** *Studies in Business and Economics.* 2024; **19**(2): 274–289.
Publisher Full Text

Wan J, Wang Y, Wang Y: **Promoting or hindering: the impact of ESG rating differences on energy enterprises' green transformation—a causal test from double machine-learning algorithms.** *Energies.* 2025; **18**(3): 464.
Publisher Full Text

Wang H, Jiao S, Ge C, *et al.*: **Corporate ESG rating divergence and excess stock returns.** *Energ Econ.* 2024; **129**: 107276.
Publisher Full Text

Xiao Z, Shum WY, Lai F, *et al.*: **How do firms respond to divergent ESG ratings? The perspective of green innovation.** *Res Int Bus Finance.* 2025; **75**: 102741.
Publisher Full Text

Yin L, Zhu X, Su Z, *et al.*: **Is disagreement beneficial for market efficiency? Evidence from ESG ratings.** *J Int Money Finance.* 2025; **154**: 103322.
Publisher Full Text

Yu EPy, Tanda A, Luu BV, *et al.*: **Environmental transparency and investors' risk perception: cross-country evidence on multinational corporations' sustainability practices and cost of equity.** *Bus Strateg Environ.* 2021; **30**(8): 3975–4000.
Publisher Full Text

Yuan H, Luan H, Wang X: **The impact of ESG rating events on corporate green technology innovation under sustainable development: perspectives based on informal environmental regulation of social systems.** *Sustainability.* 2024; **16**(19): 8308.
Publisher Full Text

Zeng G, Xu Y: **Sustainable development and the rating effects: a strategic categorization approach.** *Corp Soc Responsib Environ Manag.* 2019; **26**(6): 1554–1564.
Publisher Full Text

Zhou C, Cai P, Wu Y: **ESG rating divergence and corporate green innovation: evidence from Chinese A-share listed companies.** *Environ Dev Sustain.* 2025; 1–29.
Publisher Full Text

Zhou J, Lei X, Yu J: **ESG rating divergence and corporate green innovation.** *Bus Strategy Environ.* 2024; **33**(4): 2911–2930.
Publisher Full Text

Zhu J, Xiong Z, Lu X, *et al.*: **Does ESG rating disagreement impede corporate green innovation?** *Glob Financ J.* 2025; **64**: 101068.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔ ✔

Version 2

Reviewer Report 18 September 2025

https://doi.org/10.21956/openreseurope.22316.r56291

© **2025 Di Persio L.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Luca Di Persio**
Verona University, Verona, Italy

To the Editorial Office

Subject: Final Assessment of Revised Manuscript "Divergence and Aggregation of ESG Ratings: A Survey"

Dear Editors,
I am writing to convey my final judgment regarding the revised (second version) manuscript entitled *"Divergence and Aggregation of ESG Ratings: A Survey"* by Arianna Agosto and Alessandra Tanda. After carefully reviewing this amended version, I am pleased to recommend the acceptance of the paper for indexing in its current form.
Indeed, the authors have undertaken a commendable effort to address my (previously written, composing the first paper's review) comments and substantially improve the manuscript. In particular, the authors have significantly strengthened the introduction by offering a clearer articulation of the motivation behind the study and the central contributions it makes to the literature on ESG score divergence and aggregation. Moreover, a new, dedicated section outlines the methodology adopted to select and screen the literature under review. The latter enhances the transparency and reproducibility of the research process. I would also like to underline that Section 5 has been expanded to offer a more thorough discussion of the metrics used to evaluate divergence among ESG ratings. The authors provide a clear exposition of correlation-based methods, such as Pearson and Spearman coefficients, and have incorporated recent empirical studies applying unsupervised learning methods like Principal Component Analysis (PCA). Analogously, They have directly acknowledged the persistent absence of a standardised framework for evaluating ESG aggregation methods, hence addressing a critical limitation in the current literature. It is also worth to mention that the concluding section has been extended to propose several promising avenues for future investigation. These include the use of clustering techniques for ESG profile classification and the exploration of interpretability tools, such as Shapley values, to evaluate aggregation weights. Importantly, the authors critically assess the trade-offs between methodological complexity and practical utility.
Summing up, the paper now offers a timely and complete contribution to a field of increasing

practical and theoretical importance. The Authors systematically review a broad and up-to-date body of literature and succeed in synthesizing a wide range of methodological approaches for ESG score aggregation, moreover, by highlighting methodological discrepancies across ESG providers and proposing pathways for methodological convergence, the paper has also increased its relevance for both academic researchers and market practitioners.
For all what above, I recommend the indexing of the submitted, amended version of the article without further revision.

Sincerely,
Luca Di Persio
College of Mathematics - Dept. Comp.Science.
University of Verona

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Stochastic analysis and applications, financial arguments included. Machine learning and application to time series forecasting, LLM-based models, calibration methodologies.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 30 August 2025

https://doi.org/10.21956/openreseurope.22316.r56575

© 2025 Vasiu D. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Diana Elena Vasiu
Lucian Blaga University of Sibiu, Sibiu, Romania

The paper titled *Divergence and Aggregation of ESG Ratings: A Survey* by Arianna Agosto and Alessandra Tanda provides an extensive literature survey on the divergence of Environmental, Social, and Governance (ESG) and the methodologies proposed for their aggregation. It emphasizes the challenges that inconsistent ratings across providers create for investors, regulators, and corporations, and critically examines both traditional statistical techniques and contemporary machine learning approaches proposed to address these discrepancies.

The authors perform an extensive and current literature review, citing over 60 academic and regulatory sources. The review integrates legal, methodological, and economic perspectives, addressing both foundational works and very recent contributions.
The paper responds to prior peer reviews by improving clarity and adding more recent and relevant studies, expanding the discussions of statistical divergence measures (Pearson, Spearman), clustering, PCA, and AI-based approaches. Furthermore, it emphasised the absence of

a standardised and widely accepted framework for evaluating the performance of ESG score aggregation methods.

The article's factual assertions are consistently substantiated by peer-reviewed research, official reports, and regulatory documentation, thereby minimizing speculative interpretations and ensuring claims are grounded in credible sources. While the citations are accurate, some referenced studies are novel or rely on proprietary data, and thus their findings may not yet have been independently validated or broadly adopted within the field. Moreover, the literature review places considerable emphasis on European regulatory frameworks, which may not fully capture global perspectives. Future research would benefit from expanding the literature base to include more diverse geographical and institutional perspectives, thereby enhancing the representativeness and global applicability of the analysis.

The review is written in a clear and accessible academic style, making it suitable for a multidisciplinary audience across finance, economics, public policy, and sustainability. Although it engages with complex methodologies such as Principal Component Analysis, Bayesian inference, and machine learning, advanced techniques are explained using accurate yet comprehensible language, maintaining analytical rigor while remaining comprehensible to readers without specialized quantitative expertise. But despite efforts to clarify, sections involving advanced statistical or machine learning methods may still present a barrier for readers lacking a strong quantitative background.

The conclusions presented in the review are closely aligned with the current state of research on ESG rating divergence and aggregation. They accurately synthesize empirical findings and theoretical insights, reflecting the fragmented methodologies and ongoing discrepancies among ESG rating providers. The authors acknowledge the persistent lack of consensus among ESG rating providers, the methodological fragmentation across studies, and the challenges of creating standardized aggregation frameworks. They underscore the need for future research focused on developing explainable, data-driven, and policy-relevant aggregation models. The conclusions are both measured and forward-looking, offering well-grounded directions for further investigation without overgeneralizing findings or making unsupported claims. By stressing the importance of transparency and harmonization, the review's conclusions are well-positioned to inform ongoing regulatory debates and initiatives.

Is the topic of the review discussed comprehensively in the context of the current literature?

Yes

Are all factual statements correct and adequately supported by citations?

Yes

Is the review written in accessible language?

Yes

Are the conclusions drawn appropriate in the context of the current research literature?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Financial Analysis, Accounting, Fundamentals of statistics, Financial markets, ESG(Environmental, Social, and Governance)

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 24 February 2025

https://doi.org/10.21956/openreseurope.20823.r50578

© **2025 Di Persio L.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Luca Di Persio
[1] Verona University, Verona, Italy
[2] Verona University, Verona, Italy

The paper titled *Divergence and Aggregation of ESG Ratings: A Survey* by Arianna Agosto and Alessandra Tanda presents an extensive review of the literature concerning the divergence of Environmental, Social, and Governance (ESG) ratings and the methodologies proposed for their aggregation.

The authors highlight the inconsistencies observed across ESG ratings provided by different agencies, emphasizing the implications of such discrepancies on investment decision-making, regulatory policies, and corporate sustainability strategies. The study adopts a methodological approach grounded in a systematic review of previous works, critically evaluating the statistical and econometric techniques employed to assess and consolidate ESG ratings into a unified indicator.

The main findings suggest that ESG ratings exhibit a low level of correlation across different providers, which undermines their reliability as a basis for investment strategies. The sources of such divergence can be traced back to variations in scope, measurement techniques, and weighting schemes employed by different agencies. In response to this issue, various aggregation techniques have been proposed, ranging from classical statistical approaches, e.g., Principal Component Analysis (PCA) and Ordinary Least Squares (OLS) regression, to more sophisticated ML-based methods, including Bayesian aggregation and gradient boosting algorithms such as XGBoost. The article underscores that no single method has emerged as universally optimal, thereby necessitating further research into the design of robust aggregation techniques that enhance the reliability of ESG ratings.

It is worth mentioning that the paper provides a well-structured discussion of ESG rating divergence and its consequences. Nevertheless, certain aspects require further refinement to strengthen its scientific rigour. In particular, I suggest the Authors to better formalise the theoretical framework to structure the discussion around ESG rating divergence. A rigorous mathematical model could be introduced to formalize the problem of ESG rating inconsistency. E.g., let R_i^j denote the ESG rating assigned to firm i by provider j. The divergence between two providers j and k for a given firm i could be measured through a divergence metric, such as the Euclidean norm:

$$
D_{ik} = \sqrt{\sum_{j=1}^{N} (R_i^j - R_i^k)^2}
$$

where N represents the total number of ESG rating providers under consideration. Analogously, as an alternative to the latter, one could consider a Mahalanobis distance approach to account for the covariance structure between ratings.

Moreover, theoretical discussions on PCA, Bayesian aggregation, and ML-methods provide valuable insights, but their effectiveness in synthesizing ESG ratings remains an open question. An empirical study could be designed in which multiple ESG rating datasets are aggregated using different methodologies, and the resulting aggregated scores are evaluated against financial market outcomes such as credit spreads or stock returns. Specifically, given a set of ESG ratings R_{i}^j, an aggregated ESG score S_i could be constructed as a linear combination:

$$
S_i = \sum_{j=1}^{N} w_j R_i^j
$$

where w_j are weights derived through various optimization approaches, such as Bayesian updating or regression models that maximize correlation with an independent measure of firm sustainability.

It seems to the underwritten that the lack of such a discussion also weakens the practical applicability of the study. Indeed, ML-techniques such as XGBoost, while powerful in their predictive capabilities, often suffer from low interpretability, making it difficult for investors to understand the underlying structure of the aggregated ESG scores. A possible solution could be to consider SHAP (SHapley Additive exPlanations) values or partial dependence plots to dissect the contribution of individual ESG components to the final score.

Let me also consider that the divergence of ESG ratings can be examined using statistical measures of dispersion, e.g. (standard deviation)

$$
\sigma_i = \sqrt{\frac{1}{N} \sum_{j=1}^{N} (R_i^j - \bar{R}_i)^2}
$$

where $\bar{R}_i$ is the mean ESG rating for firm i, provides an indication of variability. If the standard deviation is high, it implies substantial disagreement among providers. Another approach is to use clustering techniques, such as k-means or hierarchical clustering, to group firms based on rating similarity, revealing whether discrepancies arise due to fundamental

methodological differences among rating agencies.
Analogously, a regression-based approach could also be employed to quantify the determinants of ESG rating divergence. A mixed-effects model of the form:

$$
D_{ij} = \alpha + \beta_1 X_{i} + \beta_2 Y_{j} + \epsilon_{ij}
$$

where X_{i} represents firm-level characteristics, and Y_{j} represents rating-provider-specific attributes, e.g., methodology used and data sources, which could identify the primary drivers of rating inconsistencies.

Summing up, the submitted paper provides a critical review of ESG rating divergence and aggregation methods, yet it could benefit from a more rigorous theoretical and empirical foundation. It would also be interesting if the Authors could better underline/propose possible future research aiming at, e.g., developing hybrid models that integrate statistical, Bayesian, and ML-approaches while maintaining interpretability. Moreover, the relationship between ESG scores and financial performance should be further explored through dynamic asset pricing models that incorporate ESG factors as state variables.

Is the topic of the review discussed comprehensively in the context of the current literature?
Yes

Are all factual statements correct and adequately supported by citations?
Yes

Is the review written in accessible language?
Yes

Are the conclusions drawn appropriate in the context of the current research literature?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Stochastic analysis and applications, financial arguments included. Machine learning and application to time series forecasting, LLM-based models, calibration methodologies.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 06 Jun 2025

Alessandra Tanda

We thank the reviewer for his useful comments and suggestions. Below we provide a

response to his remarks.
>Response: we thank the reviewer for his comments and suggestions. In the revised version, we have defined the most used metrics in the literature regarding the divergence of ESG scores. In particular, the works analysed in our study express divergence in terms of low pairwise correlations between the ESG ratings of different providers. As ESG scores are usually continuous, the most natural choice to model their correlation is the Pearson correlation coefficient. However, correlation can be also calculated through rank-based measures such as the Spearman coefficient. The latter is used by Billio et al. (2021), together with mean absolute error and the percentage of observed agreement among the considered rating providers. Correlation is not only employed to measure convergence between the original ESG scores - to show the empirical relevance of the topic of divergence - but can also be included, in the next step of aggregating the scores, to check to what extent the results of the aggregation obtained with different methods are similar (see Bissoondoyal-Bheenick et al., 2024). Concerning the evaluation of the presented methods in producing aggregated scores able to effectively predict the ESG profile of companies, we have first defined the most natural mathematical formulation for an aggregate ESG indicator, i.e. a weighted average of different scores, that can be defined according to an expert-based approach or estimated by data (Berg et al, 2022; Agosto et al., 2023a; Agosto et al., 2023b; Bissoondoyal-Bheenick et al., 2024). In the revised version we have also discussed how difficult it is to measure and compare the performance of ESG aggregated indicators due to the absence of a target variable. Indeed, differently from credit scoring models, which are validated using the actual incidence of defaults, the ESG behaviour cannot be easily and objectively summarised by an observed variable. Indeed, some works in the literature use unsupervised techniques such as PCA (Bissoondoyal-Bheenick et al., 2024; Gucciardi et al., 2024), while others rely on the relationship between ESG and financial performance (Billio et al., 2021; Gai et al., 2023) or credit worthiness measured by agency ratings (Agosto et al., 2023a; Agosto et al., 2023b). As pointed out by the reviewer and now discussed in the revised version, the unavailability of an homogeneous and easily implementable testing framework represents a limitation of the existing ESG aggregation methodologies. Applying cluster analysis to the proposed ESG scores and identifying common financial and strategic features in companies belonging to the same clusters is now mentioned among the aspects that deserve further research, together with a more in-depth analysis of the topic of explainability of the weights attributed to different scores. Regarding this last aspect, we have emphasized how many machine learning techniques not explainable by design, such as XGBOOST, do not necessarily lead to a gain in predictive accuracy that justifies the use of a computationally expensive AI method like Shapley values, as discussed by Agosto et al. 2023a.

Competing Interests: No competing interests were disclosed.

Reviewer Report 22 February 2025

https://doi.org/10.21956/openreseurope.20823.r50969

© 2025 Capizzi V. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Vincenzo Capizzi
[1] University of Piemonte Orientale, Novara, Italy
[2] University of Piemonte Orientale, Novara, Italy

The paper reviews the existing literature on the divergence of Environmental, Social, and Governance (ESG) ratings and explores various aggregation methods. It highlights the challenges posed by inconsistent ESG ratings and their implications for investment decisions. The major finding of the analysis is the persistence of significant gaps in understanding how to effectively consolidate ESG information to create a dependable sustainability indicator.
Overall, the paper is truly interesting in that addresses an actual research topic within the reference extant literature. The research design is truly original and rigorous
However, there are some limitations in the current draft and, hence, here below I provide some guidelines to further enhance the quality of the manuscript.

- First, the Author might consider improving the introductory section, trying to better clarify the motivation behind their research and the source of originality provided .
- Second, as far as the literature review is concerned, the Authors might explain how they selected the investigated contributions.
- Furthermore, the literature selection seems not updated: try to cite all relevant and recent papers on the same topic.

Finally, the writing style might be improved
Good luck!

Is the topic of the review discussed comprehensively in the context of the current literature?
Yes

Are all factual statements correct and adequately supported by citations?
Yes

Is the review written in accessible language?
Yes

Are the conclusions drawn appropriate in the context of the current research literature?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Finance, Entrepreneurial Finance, ESG ratings, sustainable finance,

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 06 Jun 2025

Alessandra Tanda

Response to reviewers for the paper Divergence and aggregation of ESG ratings:A survey

Reviewer 1 : We thank the reviewer for his comments and suggestions. Below we have provided a response to his very useful comments.

First, the Author might consider improving the introductory section, trying to better clarify the motivation behind their research and the source of originality provided >
Response: we thank the reviewer for his comments and suggestions. We have rewritten the Introduction and different sections of the paper to integrate the suggestions and clarify the motivation and the contribution of the paper.

Second, as far as the literature review is concerned, the Authors might explain how they selected the investigated contributions >
Response: we appreciate reviewer's suggestion. In the new version we have included a short section describing the process employed to select the papers for our review.

Furthermore, the literature selection seems not updated: try to cite all relevant and recent papers on the same topic. >
Response: In the new version we have updated the literature review and included more recent papers.

Competing Interests: no competing interests