the review of income and wealth

Review of Income and Wealth
Series 67, Number 2, June 2021
DOI: 10.1111/roiw.12475

SUSTAINABLE DEVELOPMENT IN EUROPE: A MULTICRITERIA DECISION ANALYSIS

BY GIULIANO RESCE*

Italian National Research Council (CNR)
SOSE S.p.A. (Italian Ministry of Economy and Finance)

AND

FRITZ SCHILTZ

Leuven Economics of Education Research (LEER), Faculty of Economics and Business, KU Leuven
McKinsey & Company

This paper proposes a novel approach to evaluate European countries on Sustainable Development Goals (SDG) by means of Hierarchical Stochastic Multicriteria Acceptability Analysis (HSMAA). HSMAA produces rankings with Monte Carlo generation of weights, overcoming the need to choose one specific set of weights. The main contribution of this paper lies in the possibility of quantifying the probability by which each country receives a given ranking. Furthermore, HSMAA allows to take into account the hierarchical nature of SDG measurement given that each of the 17 Goals also consists of several indicators. Our results show that Denmark outperforms other European countries, while lower levels of performance are observed in Romania and Bulgaria. In between bottom and top performers, we also find that many countries' rankings vary widely by the chosen set of weights, exemplifying the need to rank countries based on multiple weightings and to quantify the probabilities of each ranking.

JEL Codes: C43, O13, Q01

Keywords: composite indicators, hierarchical stochastic multicriteria acceptability analysis, sustainable development goals

1. INTRODUCTION

In 2015, more than 190 world leaders committed to 17 Sustainable Development Goals (SDGs).[1] In contrast to their predecessors, the Millennium Development Goals, the SDGs are designed to tackle the cause of the problems, taking into account their interconnectedness (Hák *et al.*, 2016). Another key feature is their focus on the mobilization of financial resources, as well infrastructure and technology (Zilberman *et al.*, 2018). The goals and corresponding sub-goals were set for 2030. As shown in Table A8, the majority of the goals agreed upon does not include time-bound targets. This lack of targets for each (sub-)goal does not allow a straightforward comparison of countries' position to the level of ambition put

*Correspondence to: Giuliano Resce, Italian National Research Council (CNR); SOSE S.p.A. (Italian Ministry of Economy and Finance), Via Mentore Maggini n. 48C00143 Roma, Italy (giuliano.resc@gmail.com).

[1]See Tables A7 and A8 for a full description of the goals, indicators and targets.

forward in the SDG agenda (Miola and Schiltz, 2019).[2] An alternative approach to hold governments accountable is by benchmarking performance between countries. To do so, one needs to define a set of indicators, and develop a weighting method to combine the indicators into one composite indicator, either at the goal level or overall.

To define a set of indicators when measuring SDG performance, a major difficulty is picking one that is relevant for all countries being benchmarked—irrespective of the aggregation method chosen to construct a composite indicator. For example, in the SDG index,[3] several targets for hunger and poverty were already achieved in most EU countries before the goals were adopted. This follows directly from the fact that the prevalence of stunting, undernourishment or people living on a budget below \$1.9 is very low in EU countries, and not because all forms of poverty are already defeated. As is clear from this example, different indicators will be needed depending on the context where the benchmarking is done. Therefore, several local initiatives followed the ongoing implementation phase, a process often referred to as "Localizing the SDGs." At the European level, Eurostat provided a policy framework to monitor SDG performance by constructing the EU SDG indicator set, containing 145[4] indicators for all European countries. This indicator set facilitates the monitoring of SDG targets by the European Commission while publicly available data improves transparency and allows researchers to construct measures of performance and rankings.

To measure SDG performance, the chosen set of indicators needs to be aggregated into one metric. Measuring SDG performance has been a topic of fierce debate among researchers, policymakers and other stakeholders (Radermacher, 2015). The intrinsic multidimensionality of SDGs poses new challenges for economic analysis, decision-making, and policy-making. The issue on how to synthesize multidimensional information into one metric has recently paved the way for the development of Composite Indicators (CIs) (Nardo *et al.*, 2008; Costanza *et al.*, 2016; Burgass *et al.*, 2017; Greco *et al.*, 2019). Despite their advantages, CIs may send simplistic and misleading messages, if the construction process is not transparent and/or lacks conceptual and statistical principles (Nardo *et al.*, 2008). The main concern is directly related to the choice of weights as only slight changes may give rise to significant differences in the final evaluation (Sharpe, 2004; Saisana *et al.*, 2005; Cherchye *et al.*, 2008; Permanyer, 2011; Decancq and Lugo, 2013; Foster *et al.*, 2013; Costanza *et al.*, 2016; Becker *et al.*, 2017; Greco *et al.*, 2019).

A common approach to compare countries' SDG performance is to impose a set of weights for each goal and indicator. For example, this approach is adopted when constructing the Bertelsmann Index where all components are equally weighted, both within and among goals (Lafortune *et al.*, 2018, p. 20). An important limitation of this conventional approach is that countries are not allowed to prioritize goals. Although equal weighting is often supported by the argument that

[2]OECD (2017): Measuring Distance to the SDG Targets 2019. An Assessment of Where OECD Countries Stand.

[3]The SDG index was constructed by the Bertelsmann Stiftung, in cooperation with the UN to measure how all 193 member states are performing relative to the 2030 Agenda.

[4]Eurostat (2019). Note that some indicators are considered "multi-purpose," and hence the number of unique indicators is equal to 100.

all indicators and all goals are equally important (formally in the 2030 Agenda), some countries might still value some dimensions more than others, especially when it comes to prioritizing policies over time. Assuming fixed weights for all countries does not grasp the contextual heterogeneity present in different countries and the importance of autonomy in setting priorities (Resce and Maynard, 2018; Greco *et al.*, 2019). Hence, different weightings of individual SDGs can have important implications on countries' performance and relative rankings (Booysen, 2002). In sum, the performance of different countries should be measured considering that policy-makers and citizens may attach varying importance to different dimensions (Helliwell, 2003; Decancq and Schokkaert, 2016). In the absence of information on policy priorities, earlier work has extensively employed nonparametric aggregation techniques to construct CIs. In particular, Data Envelopment Analysis (DEA–Charnes *et al.*, 1978) with equal inputs (Benefit of Doubt–BoD) (Decancq and Lugo, 2013; Shen *et al.*, 2013; Greco *et al.*, 2019; Patrizii *et al.*, 2017; Hatefi and Torabi, 2018). The BoD approach estimates a CI based on the combination of weights that is most convenient to the evaluated country—that is, the country is granted "the benefit of the doubt." Different countries can have differentiated weights in BoD, but each country is assigned a unique vector of weights in order to obtain a ranking (Vidoli *et al.*, 2015; Zanella *et al.*, 2015; Rogge *et al.*, 2017). This uniqueness in the weighting system cannot take into account the fact that different decision-makers and different citizens may have differentiated priorities even within countries, resulting in heterogeneity in the weighting scheme (Greco *et al.*, 2018).[5]

In this paper we present an innovative approach to measure SDG performance and rank countries. We do so by considering all possible weighting combinations, to obtain a distribution over potential rankings for each country. This way, the dependency of rankings on weight choice can be visualized, while calculating the "expected rank"[6] allows a comparison of countries. We apply Hierarchical Stochastic Multi-Objective Acceptability Analysis (HSMAA), first developed in Angilella *et al.* (2016) which allows to estimate a CI considering the whole set of feasible weights and the hierarchical nature of SDG measurements. The proposed method improves existing approaches to obtain an overall SDG measure by accounting for the nested structure of indicators within goals, and offers several advantages. In particular, HSMAA (1) constructs the best practice frontier in a nonparametric manner, avoiding restrictive assumptions on the functional form of the production process, (2) allows to consider the hierarchical structure of SDGs, and (3) allows to take into account the uncertainty in weights.

Our results show that Denmark clearly comes out on top, regardless of the set of weights given to single indicators and goals. Lower levels of performance are instead observed in Romania and Bulgaria. For several countries in between these top and bottom performers, we observe a strong discrepancy in rankings depending on the chosen set of weights. This corroborates the relevance of considering the whole space of feasible weights in the aggregation process as choosing a fixed set

[5]An alternative theoretical approach was proposed by Sen (1992). It allows for a certain degree of underspecification for some weights, since, as argued by Foster and Sen (1997, p. 206), "uniqueness is not really necessary to make agreed judgements in many situations."

[6]Note that the ranking obtained for each country from this "expected rank" does not equal the ranking that results from averaging all indicators within goals (see Figure A1).

of weights will inevitably benefit certain countries over others, hindering political consensus.

2. Literature Review on Composite Indicators

Synthetizing information is essential for economic analysis, decision-making, and policy-making (Costanza *et al.*, 2016). Following this need, the construction of composite indicators has received considerable attention (Nardo *et al.*, 2008). The weighing process represents one of the most important steps in the building of a composite indicator, as weights determine the trade-offs among the dimensions, and to this respect they reflect value judgments, and hence policy priorities. Also the choice of the list of candidate dimensions can be traced back to weighting schemes, since weights associated to dimensions to left out can be considered to have zero value. Decancq and Lugo (2013) distinguish three classes of approaches to weight vectors: data-driven, normative, and hybrid.

First, the weights in data-driven approaches depend solely on the distribution of the elementary indices. Into the family of data-driven methods, there are three main subcategories: frequency, statistical, and most favorable weights approaches. Frequency techniques have been extensively used in multidimensional poverty indices, they rely on the idea of an inverse correlation between the frequency of the deprivation in one dimension and the weight (importance) to be given to that dimension (Deutsch and Silber, 2005). Statistical approaches rely on multivariate methods like the Principal Component Analysis (PCA), and the factor analysis (among others, Noorbakhsh, 1998; Klasen, 2000; Krishnakumar and Nagar, 2008). The most favourable weights approaches are based on DEA models (Shen *et al.*, 2013; Patrizii *et al.*, 2017; Greco *et al.*, 2018).

Second, normative approaches set the weights on the basis of value judgments. In the family of normative methods, Decancq and Lugo (2013) distinguish the equal, the expert opinion, and the price-based weights. The use of equal weights is the most common approach in multidimensional well-being, also the Human Development Index launched by the UNDP in 1990 is estimated by means of equal weights (Ravallion, 1997). Other normative approaches use expert opinions as weights. In these cases, the opinions are translated into weights either by means of budget allocation (Macherini and Hoskins, 2008) or by Analytic Hierarchy Process developed by Saaty (1987). Many studies rely on price-based weights to the most common dimension of well-being like health and education (Card, 1999; Murphy and Topel, 2006).

Third, the hybrid approaches are subdivided in hedonic and stated preferences weights. The hedonic approach suggests deriving weights by the correlation between different dimensions to be included into the composite index and a proxy of self-reported outcome (Nardo *et al.*, 2008). Stated preferences weights are usually derived by a representative group of individuals in the society (see Resce and Maynard, 2018; Greco *et al.*, 2019).

Different techniques have been proposed to take into account the whole set of feasible weight vectors in the evaluation process, away from one vector of weights in the three approaches described above (Greco *et al.*, 2018; Lagravinese *et al.*, 2019). This move reflects the understanding that weights are likely to change according to individual preferences and needs, and in the absence of *a priori* information and without making recourse to a set of weights reflecting a merit-good approach on part of the policy-maker. Three main approaches have been explored in the literature to make comparisons in a multidimensional framework, while remaining agnostic about the weighting schemes: (1) Geometric approaches considering the entire simplex formed by the weighting schemes (see Decancq and Ooghe, 2010; Foster *et al.*, 2013); (2) Stochastic dominance approaches imposing only conditions on the derivatives of the composite indicator (Atkinson and Bourguignon, 1982; Trannoy, 2006); (3) Monte Carlo approaches drawing weightings schemes from the simplex (Zhou *et al.*, 2010; Greco *et al.*, 2019, 2018; Lagravinese *et al.*, 2019).

Decancq and Ooghe (2010) apply a geometric approach by considering the entire simplex formed by the weighting schemes to international data from World Development Indicators (World Bank, 2009). They focus on three dimensions of well-being very similar to the main components of the Human Development Index (HDI): income, measured by GDP per capita; longevity, measured by life expectancy at birth; and education, measured by the gross secondary enrolment rate and the literacy rate. The evaluation is proposed by a graphical technique that allows to visualize results in triangles in which every point represents a different weighting scheme. The main drawback of this visualizing method is that the approach cannot be employed when the dimensions to consider are more than three.

Foster *et al.* (2013) propose a method and a number of metrics to evaluate to what extent ranking by composite index obtained by a specific vector of weights is robust to changes in weighting schemes. They apply the methods to three indices: the HDI published by the United Nations Development Program (UNDP, 2009, 2010), the Index of Economic Freedom (IEF) obtained from the Heritage Foundation (2008), and the Environmental Performance Index (EPI) (Esty *et al.*, 2008). Results show that the HDI has the highest rank robustness, while the EPI was the least robust. This method requires a specific vector that serves as a starting point for the estimate, and robustness estimates are mainly based on the distance between the evaluation obtained by selected weights and the evaluation obtained with the extreme weight.

Zhou *et al.* (2010) proposed a similar estimate, which extends the multiplicative DEA method by a combination of optimistic and pessimistic optimization. In the optimistic case the weights are the most convenient to the evaluated country as the standard DEA, while the pessimistic case the weights are the most inconvenient to the evaluated country. From an operational perspective, pessimistic estimates are obtained inverting the objective function in the linear program originally proposed by Charnes *et al.* (1978). Uncertainty about weights is considered according to the degree of which the final composite index takes more from the optimistic or from the pessimistic optimization.

With regard to the previous abovementioned literature, the use of the Stochastic Multi-Objective Acceptability Analysis allows to move forward along three main lines:

1. SMAA allows to explore the whole set of feasible weights since it does not need a specific vector from which the estimates can start

(as Foster *et al.*, 2013), and does not need assumptions needed by DEA-based models (as Zhou *et al.*, 2010).

2. SMAA allows to quantify the volume of vectors of weight by which ranking changes as it is not only based on the distance between evaluation obtained by actual weight and the evaluation obtained with the extreme weight (as Foster *et al.*, 2013; and Zhou *et al.*, 2010). In other words, the Foster *et al.* (2013)'s and DEA based methods are not able to quantify how frequent the extreme rankings attained by countries would be according to a predetermined distribution of vector of weights.
3. SMAA does not have curse of dimensionality problems. This is a clear advantage when aggregating a high number of indicators when measuring SDG performance (our estimates below are based on 106 indicators). All the above mentioned models have problems when the dimensionality increases. The main problem is that as the number of indicators increases, the probability that all observations dominate the others in at least one of the considered dimension (Pareto optimality) increases, and there will be no possibility to rank at all. In other words, the curse of dimensionality does not affect SMAA results and outcomes for evaluation.

To consider the hierarchical (H) structure of SDG indicators, SMAA needs to be extended further, making the analysis independent from how many indicators are in each goal. Indeed, as shown in Table A8 in Appendices, the number of indicators is not the same in each goal. The use of standard SMAA on indicators ignoring the hierarchical structure of SDG produces different rank acceptability indices (see Table A4 in the appendices in which non-hierarchical results are reported). By means of HSMAA the uncertainness in weights is considered, but the probability for each goal to be important into the composite index is independent from the number of indicators under the goal.[7]

3. Measuring SDG Performance

We propose a flexible composite indicator (CI) to compare performance of countries using Eurostat indicators.[8] Formally, the set of 28 countries $A\left(a_1, \dots, a_m\right)$ is evaluated on a set of 17 SDGs $\left(g_1, \dots, g_n\right)$. The SDG composite indicator can be seen as the average of the seventeen goals weighted by the weights (w) associated to each of the seventeen goals:

$$CI\left(a_k, w\right) = \sum_{i=1}^{n} w_i g_i\left(a_k\right) \tag{1}$$

[7]Nevertheless, the Intraclass Correlation Coefficients (Shrout and Fleiss, 1979; McGraw and Wong, 1996; Wolak *et al.*, 2012) between rank acceptability indices obtained with standard SMAA and rank acceptability indices obtained by HSMAA shows that the two results are significantly correlated (see Table A5 in the Appendices).

[8]Data was collected using the Eurostat R package for EU28 countries. After removing missing values, the final dataset includes the latest observation for each of 106 (out of 145) SDG indicators, divided over 17 Goals (see Table A3).

where w_i reflects the importance given to the goal g_i, and $g_i\left(a_k\right)$ is the achieved result in country a_k for goal g_i, normalized by min max method so that all $g_i\left(a_k\right), i=1,\ldots,n$, end up in the [0:1] range (Nardo *et al.*, 2008, p. 30). Nardo *et al.* (2008) and Decancq and Lugo (2013) list several procedures in the construction of a composite index. The most frequent solution is equal weighting or data-driven weighting (Paruolo *et al.*, 2013; Greco *et al.*, 2019). The majority of approaches produce a single weight vector overall or a single weight vector for each Decision Making Unit (in the case of DEA). The main problem is that the order of importance given to different dimension is a subjective choice, which implies that one single vector of w does not exist. This is particularly important in the EU context, considering the political relevance of member states' autonomy. Also, choosing weights is relevant in an accountability context, as changing the set of weights directly affects the ranking of countries.

3.1. *Stochastic Multi-Objective Acceptability Analysis*

In order to embody unknown preferences on the weights assigned to each criteria, SMAA considers the probability distributions $f_W(w)$ in the set of the feasible weights W (Lahdelma and Salminen, 2001):

$$W=\left\{\left(w_1,\ldots,w_n\right)\in R_+^n, w_1+\ldots+w_n=1\right\} \tag{2}$$

The set of feasible weights is a $n-1$ dimensional simplex. A uniform weight distribution is assumed in the set of feasible weights W.

Defining ξ_{ik} as the value of goal g_i in country a_k, from the probability distributions $f_\chi(\xi)$ on χ, where χ is the evaluation space (in our case the space of the values assumed by the goals g_i in G), Lahdelma and Salminen (2001) introduce a ranking function relative to the country a_k:

$$\text{rank}(k,\xi,w)=1+\sum_{h\neq k}\rho\left[CI\left(\xi_h,w\right)>CI\left(\xi_k,w\right)\right] \tag{3}$$

where $\rho(true)=1$, and $\rho(false)=0$. Hence, the rank of country a_k, given a vector of weights w, is one plus how many times the weighted average of SDG performances of a_k ($CI\left(\xi_k,w\right)$) is dominated by the weighted average of SDG performances of the other countries ($CI\left(\xi_h,w\right)$). Thus, the value assumed by the variable *rank* (k,ξ,w) in equation (3) is one plus the number of countries that performs better than country a_k in terms of SDGs. It follows that the lower the value of *rank* (k,ξ,w) the better the performance of the country a_k.

Accordingly, for each country a_k and for each value that can be taken by SDGs performances $\xi\in\chi$, SMAA computes the set of weights for which country a_k assumes rank r:

$$W_k^r(\xi)=\{w\in W:\text{rank}(k,\xi,w)=r\} \tag{4}$$

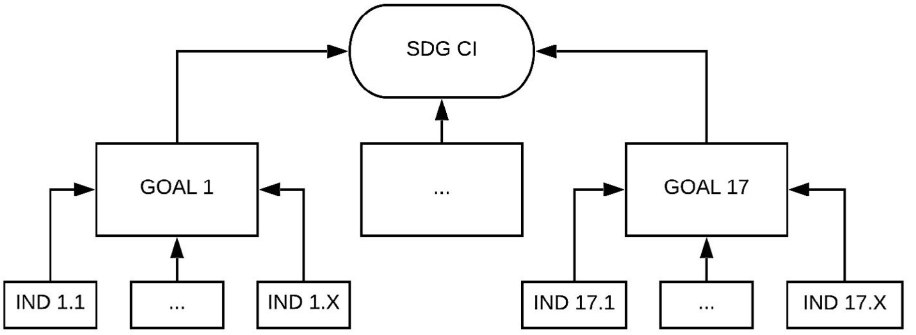

Figure 1. Hierarchical Structure of a Composite Indicator of SDG Performance
Note: This figure depicts the hierarchical structure of (Eurostat) indicators, used to monitor SDG performance. Overall SDG performance is obtained by aggregating performance on individual goals, while performance for each of the 17 goals is obtained by aggregating all indicators corresponding to that goal

From equation (4), one can then compute the rank acceptability index:

(5)
$$b_k^r = \int_{\xi \in \chi} f_\chi(\xi) \int_{w \in W_k^r(\xi)} f_W(w)\, dw d\xi$$

Equation (5) indicates the probability that the country a_k has the r-th position in the ranking. In other words, b_k^r is the ratio of the number of the vector of weights by which country a_k gets rank r to the total amount weights considered.

3.2. *Hierarchical Stochastic Multi-Objective Acceptability Analysis*

The structure of the SDG measurement problem is hierarchical: goals are in the first level and the different indicators are in the second level (Figure 1).

In the SMAA context, Angilella *et al.* (2016) proposed the inclusion of a hierarchical structure. In our problem, each criterion (goal) $g_i \in G$ is given by the weighted sum of sub-criteria (indicators) $q_{ij} \in Q_i$:

(6)
$$g_i = \sum_{j=1}^{s_i} v_{ij} q_{ij}$$

In this case, the CI of SDGs becomes the weighted average of goals, which are the weighted average of Eurostat indicators. The new value function to aggregate the evaluations of a country, from A with respect to the g_i goals from G, with respect to the indicators from Q_i, is a double weighted average. For each country $a_k \in A$, we can estimate the following CI:

(7)
$$CI\left(a_k, w, v_k\right) = \sum_{i=1}^{n} w_i \sum_{j=1}^{s_i} v_{ij} q_{ij}\left(a_k\right)$$

where w_i is the weight given to the goal i, and v_{ij} is the weight given to the Eurostat indicator j.

In order to consider the hierarchy of this problem, we use HSMAA. This approach allows us to take into account: (1) the uncertainty with respect to the weights assigned to the goals (as in the standard SMAA) and within goals (2) the uncertainty with respect to the weights assigned to the Eurostat indicators (sub-criteria).

The HSMAA considers three probability distributions: $f_W(w)$, $f_V(v)$; $f_\chi(\xi)$ on W, V; and χ (De Matteis *et al.*, 2019), respectively, where:

$$W=\left\{\left(w_1,\dots,w_n\right)\in R_+^n, w_1+\dots+w_n=1\right\} \tag{8}$$
$$V=\left\{\left(v_{i1},\dots,v_{is_i}\right)\in R_+^{s_i}, v_{i1}+\dots+v_{is_i}=1, i=1,\dots n\right\}$$

and χ is the space of the value that can be taken by the Eurostat indicators $q_{ij}\in Q_i\ (i=1,\dots,n)$.

We introduce a ranking function relative to the country a_k:

$$rank\,(k,\xi,w,v)=1+\sum_{h\neq k}\rho\left(CI\left(\xi_h,w,v_h\right)>CI\left(\xi_k,w,v_k\right)\right) \tag{9}$$

where $\rho\,(true)=1$, and $\rho\,(false)=0$. Then, for each country a_k, for each evaluation of countries $\xi\in\chi$, and for each rank $r=1,\dots,m$, HSMAA computes the set of weights of SDGs for which country a_k assumes rank r:

$$W_k^r(\xi,v)=\{w\in W: rank\,(k,\xi,w,v)=r\} \tag{10}$$

HSMAA evaluation is based on the computation of the rank acceptability index, which is the relative measure of the set of weight vectors for which the country a_k gets rank r:

$$b_k^r=\int_{w\in W_k^r(\xi)} f_W(w)\int_{\xi\in\chi} f_\chi(\xi)\int_{v\in V} f_V(v)\,dv\,d\xi\,dw \tag{11}$$

b_k^r is the probability that country a_k gets the r-th position in the preference ranking.

Within SMAA, the pairwise winning index (Leskinen *et al.*, 2006) gives the frequency that an alternative a_k is preferred to an alternative a_h in the space of possible weight vectors and possible evaluations on single criteria. In a hierarchical structure the pairwise winning index can be represented as follows:

$$p_{kh}=\int_{w\in W} f_W(w)\int_{\xi\in\chi: CI\left(\xi_k,w,v_k\right)\geq CI\left(\xi_h,w,v_h\right)} f_\chi(\xi)\int_{v\in V} f_V(v)\,dv\,d\xi\,dw \tag{12}$$

Since it is related to comparisons of couples of alternatives, the pairwise winning index is in the line of the Condorcet rule. One of the most successful Condorcet methods is the Copeland's method, in which candidates are ordered by the number of pairwise victories, minus the number of pairwise defeats. In the case of pairwise winning indices presented in (12), victories and defeats are not binary variables, but they are defined in the continuous: the frequency that an alternative is preferred to another and the frequency that another is preferred to the alternative. The differences between the sum of all pairwise winning indices for alternative a_k and the sum of all pairwise winning indices of all other alternative compared with a_k can be seen as the generalized Copeland's method when the outcomes are multiple and all the space of feasible weights is considered:

$$GCM_k = \sum_{i=1}^{m} p_{ki} - \sum_{i=1}^{m} p_{ik} \tag{13}$$

From a computational perspective, the multidimensional integrals defining the considered indices are estimated using Monte Carlo simulations. In our application, we consider uniform probability distributions $f_W(w)$ on W and $f_V(v)$ on V. To rank European countries (spatial alternatives), we apply the HSMAA technique to 10,000 extractions of w and v vectors from a uniform distribution. To this regard, Tervonen and Ladhelma (2007) show that 10,000 extractions are enough to get an error limit of 0.01 with a confidence interval of 95 percent. As a further robustness check, to the aim of considering all weights on the simplex, we repeated the analysis taking 10,000 weights from the Halton sequence. As shown in Tables A2 and A3 in the Appendices, results remain unchanged. The Intraclass Correlation Coefficients (Shrout and Fleiss, 1979; McGraw and Wong, 1996; Wolak *et al.*, 2012) between rank acceptability indices obtained with uniform distribution and rank acceptability indices obtained by Halton sequence are higher than .99 in all ranks (see Table A5 in the Appendices).

4. Results

In this paper the HSMAA is used to build a country level composite index measuring performance in terms of SDGs. In what follows we focus on overall performances (Section 4.1), in goal-specific performances (Section 4.2), and in pairwise comparisons (Section 4.3).

4.1. *Overall SDG performance*

Figure 2 reports the country level Rank acceptability indices (b_k^r) for each of the 28 ranks (Rank acceptability indices are in Table A1 in the Appendix).[9] Countries are ranked according to the expected rank (the average of ranks obtained out of 10,000 Monte Carlo simulated weights). Overall, countries with a high probability to be in the top ranks are Denmark, Sweden, and the Netherlands. This means that in these countries, there is a generalized positive performance in almost all the dimensions considered in the 106 sub-indicators of SDGs. Their narrow

[9]For the sake of comparison, Table A1 in the Appendix displays the rank changes relative to the "expected rank" presented here. Hence, we rescale all indicators and aggregate them first at goal level, and next over all SDGs. Using the resulting value, we rank all EU countries and contrast this ranking to the one proposed here.

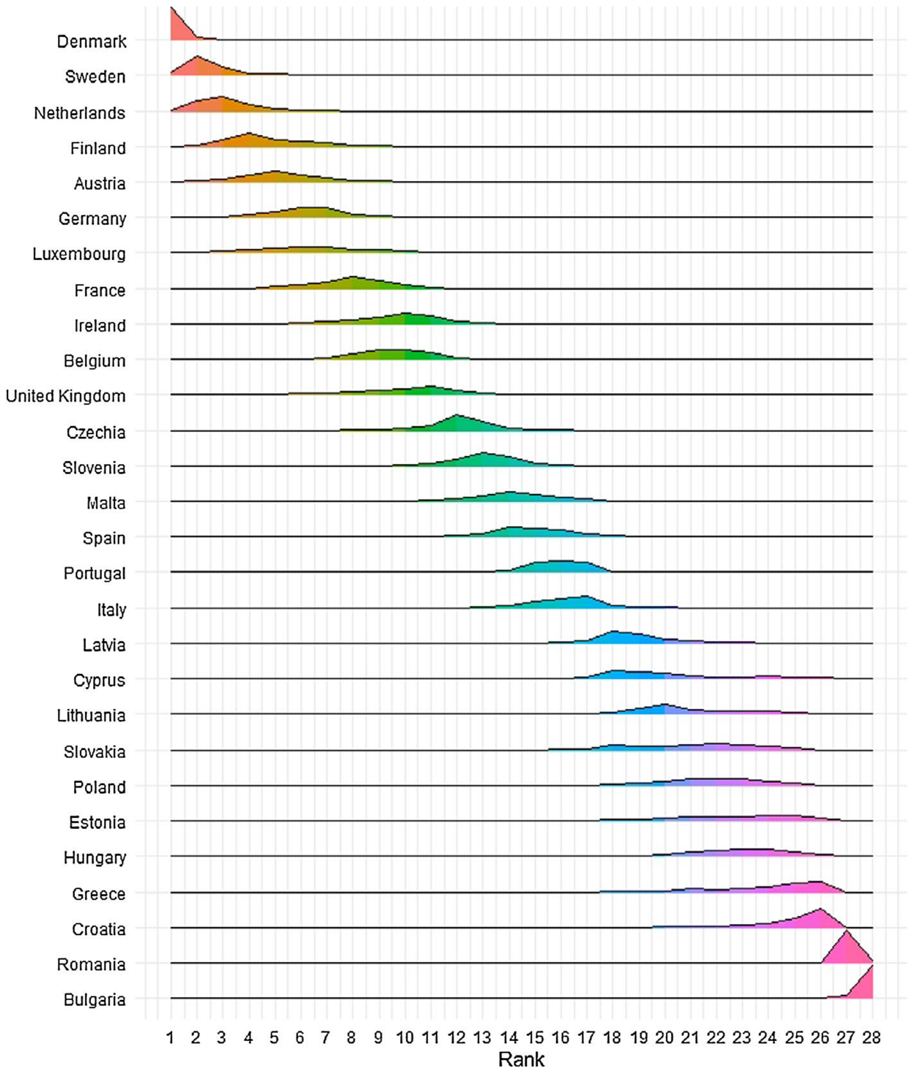

Figure 2. Rank Acceptability Index

Note: This figure depicts the Rank Acceptability Index for each of the 28 ranks, obtained by 10,000 Monte Carlo generation of weights from a uniform distribution for each goal. Countries are ranked according to the expected rank (i.e. the average of ranks obtained out of 10,000 Monte Carlo generation of weights from a uniform distribution). [Colour figure can be viewed at wileyonlinelibrary.com]

distribution of ranks reflects this overall strong performance, as changing the set of weights does not seem to affect their relative position with respect to other EU member states. The opposite can be seen in Bulgaria, Croatia, Greece, Hungary, and Romania, displaying both a high probability to be in the lowest ranks, and a narrow distribution around this low rank. Hence, these countries cannot improve their relative position as either kind of weight set will classify them at the bottom.

In between top- and bottom-ranked countries, the distribution of mean ranks reveals interesting information about countries' "weight-dependency." For example, Luxemburg, the UK, Malta, Slovakia, and Poland are characterized by a wide rank distribution, suggesting that on some goals (or indicators), their performance is superior to adjacent countries. When imposing one common set of weights, it would be unfair for those countries as their ranking could be affected drastically. Reporting the full range of possible outcomes allows a more just comparison, as countries that outperform others when all weights were considered, can be reasonably be labelled better performers on achieving the SDGs.

Other countries, where distributions overlap to some extent cannot be ranked vis-à-vis each other without making some value judgment on the priorities assigned to indicators and goals. In other countries like the Czech Republic and Romania, the ranking appears to be rather independent of weights and display a consistent performance across goals and indicators.

Overall, the variance of the distributions in Figure 2 connects with the work of Foster *et al.* (2013), where the robustness of rankings obtained from composite indices is analyzed in depth. The main contribution to this literature is here in quantifying the volume of vectors of weight by which ranking would change. As shown in Table A1 in the Appendix, both Sweden and the Netherlands have some probabilities of getting the 7th rank with extreme pessimistic weights, but at the same time Sweden gets the second rank with 53 percent of weights while the Netherlands gets the second rank with 30 percent. Both Latvia and Slovakia have some probability of getting the 25th rank for some extreme pessimistic weights, but for Latvia the volume of these weights is 1 percent of the whole space, compared to 5 percent for Slovakia (see Table A1 in the Appendix).

The metrics proposed in this analysis allow to add information on what has been shown in previous approaches making comparisons in a multidimensional framework, while remaining agnostic about the weighting schemes (see Decancq and Ooghe, 2010; Foster *et al.*, 2013). In particular, previous literature has been focused on checking all weights, which has been done either with optimistic and pessimistic DEA, or with the method proposed in Decancq and Ooghe (2010), or with the Foster *et al*. (2013) technique. The analysis proposed here focuses on measuring the volume of weights assigned to specific rankings. The measurement of volumes allows to move from a deterministic and dual ranking, where a country is uniquely associated to a specific rank, to a more flexible probabilistic ranking, where each country has some probability to obtain each rank position. From a policy perspective, this metric can be used to make comparisons considering the heterogeneity on individual preferences and needs (Greco *et al.*, 2018; Lagravinese *et al.*, 2019). In some cases, as in Denmark, the heterogeneity in individual preferences does not really affect the rank position: Denmark is first according to 93 percent of weights and it is second according to the remaining 7 percent (see Table A1). In contrast, there are countries like Estonia, Luxemburg, and Slovakia where vector of weights associated to a specific ranking is lower than the 20 percent. This means that in these countries each deterministic ranking would consider no more than one fifth of the all feasible preferences and needs of citizens. Furthermore, it is worth mentioning that DEA-based approaches like BoD compare countries based on the top of the distribution shown in Figure 2. Like any other approach focused on a single vector of weights, DEA evaluations shed little light on what would happen when more weight to the bottom tail of the distributions is given. To this regard, the use of HSMAA, by Monte Carlo generation, allows to make benchmarking more transparent in terms of weighing choice issues.

4.2. *Goal-Specific SDG Performance*

HSMAA allows to disentangle the overall performance shown in Figure 2 into specific goals thanks to its hierarchical structure. In Table 1 the expected rank (the average of ranks obtained out of 10,000 Monte Carlo generation of weights from a uniform distribution) for each goal is reported. As can be noted, there is substantial variability in the goal-specific performance within each country. Focusing on the top performers, Denmark is positioned at the bottom when looking solely at Goal 15 (Life on land). Sweden, the second best performer overall, is also ranked low in two goals: 6 and 14 (Clean water and sanitation and Life below water). Similar results can be found at the bottom of the ranking: Romania, the last but one in terms of overall performance, received the third expected rank in terms of goal 15 (Life in land), essentially ending up as being ranked a top performer. Likewise, Bulgaria, the worst overall performer, has an expected rank between 4 and 5 in terms of goal 7 (Affordable and lean energy). Overall, results in Table 1 show that the performance with respect to Sustainable Development Goals in Europe is not uniform at all, even within countries among different goals.

4.3. *The Pairwise Winning Indices*

In this section we make inference on pairwise comparisons by equation (12). Table 2 contains all pairwise combinations, it shows the probability that a country on the row gets a higher rank than a country on the column in terms of SDGs. As an example there is 42 percent probability that Belgium gets a higher rank than the UK, and there is 58 percent probability that the UK gets a higher rank than Belgium. Confirming the good performance observed in Figure 2, Denmark has 100 percent probability to get higher rank than all other countries with the exception of Sweden for which this probability is 93 percent. In contrast, Bulgaria has no probability to get higher rank than any other country with the exception of Romania for which the probability is 7 percent. With the exception of these two stratified countries, other countries show a differentiated path in pairwise comparison, with higher pairwise winning index for countries in the upper side on rank in Figure 2 and lower level of pairwise winning index for countries in the lower side of rank.

In Table 3 for any country we show the sum of all pairwise winning indices, the sum of all pairwise winning indices of all other countries compared with it, and the generalized Copeland's ranking presented in equation (13). The ranking is consistent in all the three metrics, and shows Denmark, Sweden, and the Netherlands in the first three positions, with a generalized Copeland's ranking of 26.85, 23.67, and 22.72, respectively. On the bottom side of the rankings are Bulgaria, Romania, Croatia which have a generalized Copeland's ranking of −26.85, −25.15, and −21.07 respectively. It is worth noting that the Copeland's ranking is also consistent with the expected rank (i.e. the average of ranks obtained out of 10,000

TABLE 1
Expected Rank for Each Goal

Country	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Austria	8.23	9.83	11.23	8.49	9.12	4	15.08	7.43	5.60	11.08	11.03	12.70	18.32	4	22.71	3.62	22.70
Belgium	13.21	17.95	7.54	13.29	14.48	8	24.80	6.32	6.43	6.65	12.20	16.10	18.61	8	16.48	12.80	19.41
Bulgaria	27.35	13.78	23.96	25.79	19.98	27	24.56	26.16	26.91	27.82	25.78	26.05	21.76	27	4.61	27.54	9.65
Croatia	19.87	22.65	20.28	24.45	15.12	26	10.25	18.15	21.71	23.30	15.77	7.56	7.95	26	15.43	18.09	18.51
Cyprus	19.55	25.53	7.01	20.78	19.19	3	26.80	24.20	21.70	4.76	23.28	24.07	27.23	3	23.88	19.39	22.29
Czechia	1.13	12.28	16.40	15.30	22.99	12	13.76	11.05	14.52	5.41	3.95	20.29	15.93	12	11.31	20.43	19.27
Denmark	6.34	5.39	7.08	3.05	5.81	7	2.56	2.67	1.20	11.28	14.09	2.04	2.18	7	24.01	4.75	4.61
Estonia	18.47	5.49	23.96	8.78	10.09	23	19.93	22.16	22.70	20.71	13.38	27.70	23.66	23	13.37	13.29	8.97
Finland	2.43	3.30	9.21	8.86	7.45	14	17.53	6.04	5.04	3.76	20.66	19.45	10.59	14	11.85	3.75	14.79
France	5.90	7.97	13.17	13.11	9.07	16	12.25	9.61	7.23	5.19	7.55	10.69	10.13	16	9.99	20.41	15.30
Germany	9.56	12.92	13.21	12.87	13.37	9	21.57	4.80	10.94	4.28	20.87	21.09	23.84	9	6.17	5.78	5.12
Greece	27.36	18.53	21.13	25.35	23.42	13	11.80	25.56	15.59	19.88	21.47	10.35	2.63	13	15.44	24.91	23.10
Hungary	19.82	20.87	22.79	20.58	26.96	19	17.44	16.03	15.75	16.32	6.36	23.44	16.35	19	17.47	21.16	23.17
Ireland	15.92	7.81	3.40	5.06	18.06	15	12.02	12.74	6.31	13.89	8.04	9.29	20.15	15	23.71	5.85	15.85
Italy	23.15	17.92	3.90	25.40	24.35	6	11.72	20.18	12.95	19.38	13.26	2.69	5.76	6	22.43	22.30	12.06
Latvia	24.53	11.39	27.96	15.85	6.43	5	11.12	24.25	24.63	25.35	22.69	14.22	11.34	5	9.20	23.78	7.53
Lithuania	23.69	11.97	25.92	10.73	2.83	10	23.86	17.91	26.48	25.69	22.43	14.15	16.39	10	18.88	19.21	21.00
Luxembourg	13.01	24.56	14.50	5.48	10.74	1	22.88	10.82	5.66	14.12	11.90	15.59	22.40	1	18.92	1.60	9.14
Malta	4.31	28.00	17.46	18.94	25.72	2	7.87	16.26	13.46	6.91	14.90	14.44	13.54	2	24.96	12.67	14.24
Netherlands	3.86	19.03	4.77	2.37	8.06	18	19.11	1.72	5.03	2.10	9.95	10.05	17.68	18	19.58	9.11	1.70
Poland	15.20	20.07	20.39	13.23	12.38	21	23.85	19.35	23.25	19.53	13.34	26.59	27.35	21	3.28	10.92	13.36
Portugal	19.77	15.61	19.96	18.68	15.12	11	8.48	16.18	17.10	20.72	24.50	2.40	5.69	11	13.80	12.67	25.47
Romania	26.09	13.44	25.47	27.41	24.87	28	3.03	27.59	26.45	27.08	24.22	12.81	5.73	28	9.77	22.34	10.61
Slovakia	8.72	10.01	17.10	23.66	24.57	25	9.87	19.51	19.75	15.77	1.85	21.01	12.71	25	7.74	24.98	24.81
Slovenia	10.78	25.39	15.02	7.58	2.11	17	16.91	9.80	17.12	8.42	20.26	13.95	20.36	17	17.67	11.95	12.09
Spain	17.88	7.17	5.08	20.95	17.89	20	6.22	19.65	16.37	21.61	2.05	7.67	12.42	20	4.19	13.96	18.58
Sweden	6.30	1.63	1.11	1.54	2.10	24	5.13	1.98	3.56	12.60	12.94	10.69	5.76	24	6.22	6.98	7.94
United Kingdom	13.61	15.51	6.99	8.43	13.69	22	5.60	7.86	12.58	12.39	7.28	8.90	9.55	22	12.93	11.77	4.73

Note: 1. No poverty; 2. Zero hunger; 3. Good health and well-being; 4. Quality education; 5. Gender equality; 6. Clean water and sanitation; 7. Affordable and clean energy; 8. Decent work and economic growth; 9. Industry, innovation and infrastructure; 10. Reduced inequalities; 11. Sustainable cities and communities; 12. Responsible consumption and production; 13. Climate action; 14. Life below water; 15. Life on land; 16. Peace, justice and strong institutions; 17. Partnerships for the goals.

TABLE 2
HSMAA Pair-Wise Winning Indices From Sampled Rankings

	Austria	Belgium	Bulgaria	Croatia	Cyprus	Czechia	Denmark	Estonia	Finland	France	Germany	Greece	Hungary	Ireland
Austria	0.00	0.99	1.00	1.00	1.00	1.00	0.00	1.00	0.38	0.87	0.70	1.00	1.00	0.96
Belgium	0.01	0.00	1.00	1.00	1.00	0.93	0.00	1.00	0.02	0.19	0.03	1.00	1.00	0.52
Bulgaria	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Croatia	0.00	0.00	1.00	0.00	0.12	0.00	0.00	0.13	0.00	0.00	0.00	0.35	0.14	0.00
Cyprus	0.00	0.00	1.00	0.88	0.00	0.00	0.00	0.74	0.00	0.00	0.00	0.94	0.79	0.00
Czechia	0.00	0.07	1.00	1.00	1.00	0.00	0.00	1.00	0.00	0.01	0.01	1.00	1.00	0.10
Denmark	1.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Estonia	0.00	0.00	1.00	0.87	0.27	0.00	0.00	0.00	0.00	0.00	0.00	0.67	0.52	0.00
Finland	0.62	0.98	1.00	1.00	1.00	1.00	0.00	1.00	0.00	0.95	0.75	1.00	1.00	0.99
France	0.13	0.81	1.00	1.00	1.00	0.99	0.00	1.00	0.05	0.00	0.19	1.00	1.00	0.77
Germany	0.30	0.97	1.00	1.00	1.00	0.99	0.00	1.00	0.25	0.81	0.00	1.00	1.00	0.89
Greece	0.00	0.00	1.00	0.65	0.07	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.31	0.00
Hungary	0.00	0.00	1.00	0.86	0.21	0.00	0.00	0.48	0.00	0.00	0.00	0.69	0.00	0.00
Ireland	0.04	0.48	1.00	1.00	1.00	0.90	0.00	1.00	0.01	0.23	0.11	1.00	1.00	0.00
Italy	0.00	0.00	1.00	1.00	0.94	0.02	0.00	0.96	0.00	0.00	0.00	1.00	0.99	0.00
Latvia	0.00	0.00	1.00	0.95	0.63	0.00	0.00	0.88	0.00	0.00	0.00	0.96	0.91	0.00
Lithuania	0.00	0.00	1.00	0.90	0.39	0.00	0.00	0.72	0.00	0.00	0.00	0.88	0.77	0.00
Luxembourg	0.32	0.91	1.00	1.00	1.00	0.98	0.00	1.00	0.29	0.71	0.50	1.00	1.00	0.87
Malta	0.00	0.02	1.00	1.00	1.00	0.16	0.00	0.99	0.00	0.02	0.00	1.00	1.00	0.03
Netherlands	0.84	1.00	1.00	1.00	1.00	1.00	0.00	1.00	0.78	1.00	0.98	1.00	1.00	1.00
Poland	0.00	0.00	1.00	0.95	0.29	0.00	0.00	0.62	0.00	0.00	0.00	0.76	0.67	0.00
Portugal	0.00	0.00	1.00	1.00	0.98	0.01	0.00	0.99	0.00	0.00	0.00	1.00	1.00	0.00
Romania	0.00	0.00	0.93	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Slovakia	0.00	0.00	1.00	0.98	0.36	0.00	0.00	0.66	0.00	0.00	0.00	0.78	0.74	0.00
Slovenia	0.00	0.06	1.00	1.00	1.00	0.28	0.00	1.00	0.00	0.01	0.01	1.00	1.00	0.05
Spain	0.00	0.01	1.00	1.00	0.96	0.08	0.00	1.00	0.00	0.00	0.00	1.00	1.00	0.01
Sweden	0.87	1.00	1.00	1.00	1.00	1.00	0.07	1.00	0.93	1.00	0.95	1.00	1.00	1.00
United Kingdom	0.08	0.42	1.00	1.00	1.00	0.80	0.00	1.00	0.03	0.15	0.07	1.00	1.00	0.43

TABLE 2 (CONTINUED)

	Italy	Latvia	Lithuania	Luxembourg	Malta	Netherlands	Poland	Portugal	Romania	Slovakia	Slovenia	Spain	Sweden	United K
Austria	1.00	1.00	1.00	0.68	1.00	0.16	1.00	1.00	1.00	1.00	1.00	1.00	0.13	0.92
Belgium	1.00	1.00	1.00	0.09	0.98	0.00	1.00	1.00	1.00	1.00	0.94	0.99	0.00	0.58
Bulgaria	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00
Croatia	0.00	0.05	0.10	0.00	0.00	0.00	0.05	0.00	1.00	0.02	0.00	0.00	0.00	0.00
Cyprus	0.06	0.37	0.61	0.00	0.00	0.00	0.71	0.02	1.00	0.65	0.00	0.04	0.00	0.00
Czechia	0.98	1.00	1.00	0.02	0.84	0.00	1.00	0.99	1.00	1.00	0.72	0.92	0.00	0.20
Denmark	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.93	1.00
Estonia	0.04	0.12	0.28	0.00	0.01	0.00	0.38	0.01	1.00	0.34	0.00	0.00	0.00	0.00
Finland	1.00	1.00	1.00	0.71	1.00	0.22	1.00	1.00	1.00	1.00	1.00	1.00	0.07	0.97
France	1.00	1.00	1.00	0.29	0.98	0.00	1.00	1.00	1.00	1.00	0.99	1.00	0.00	0.85
Germany	1.00	1.00	1.00	0.50	1.00	0.02	1.00	1.00	1.00	1.00	0.99	1.00	0.05	0.93
Greece	0.00	0.04	0.12	0.00	0.00	0.00	0.24	0.00	1.00	0.22	0.00	0.00	0.00	0.00
Hungary	0.01	0.09	0.23	0.00	0.00	0.00	0.33	0.00	1.00	0.26	0.00	0.00	0.00	0.00
Ireland	1.00	1.00	1.00	0.13	0.97	0.00	1.00	1.00	1.00	1.00	0.95	0.99	0.00	0.57
Italy	0.00	0.88	0.96	0.00	0.17	0.00	0.97	0.41	1.00	0.95	0.04	0.24	0.00	0.01
Latvia	0.12	0.00	0.83	0.00	0.02	0.00	0.83	0.04	1.00	0.75	0.00	0.05	0.00	0.00
Lithuania	0.04	0.17	0.00	0.00	0.01	0.00	0.67	0.00	1.00	0.60	0.00	0.01	0.00	0.00
Luxembourg	1.00	1.00	1.00	0.00	1.00	0.10	1.00	1.00	1.00	1.00	0.98	0.99	0.11	0.83
Malta	0.83	0.98	0.99	0.00	0.00	0.00	0.99	0.79	1.00	0.98	0.30	0.59	0.00	0.10
Netherlands	1.00	1.00	1.00	0.90	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	0.36	1.00
Poland	0.03	0.17	0.33	0.00	0.01	0.00	0.00	0.01	1.00	0.41	0.00	0.00	0.00	0.00
Portugal	0.59	0.96	1.00	0.00	0.21	0.00	1.00	0.00	1.00	0.99	0.02	0.28	0.00	0.00
Romania	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Slovakia	0.05	0.25	0.40	0.00	0.02	0.00	0.59	0.01	1.00	0.00	0.00	0.00	0.00	0.00
Slovenia	0.96	1.00	1.00	0.02	0.70	0.00	1.00	0.98	1.00	1.00	0.00	0.84	0.00	0.10
Spain	0.76	0.95	0.99	0.01	0.41	0.00	1.00	0.72	1.00	1.00	0.16	0.00	0.00	0.00
Sweden	1.00	1.00	1.00	0.89	1.00	0.64	1.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00
United Kingdom	0.99	1.00	1.00	0.17	0.90	0.00	1.00	1.00	1.00	1.00	0.90	1.00	0.00	0.00

TABLE 3
Generalised Copeland's Ranking

	$\sum_{i=1}^{m} p_{ki}$	$\sum_{i=1}^{m} p_{ik}$	GCM_k
Austria	22.78	4.22	18.56
Belgium	18.30	8.70	9.59
Bulgaria	0.07	26.93	−26.85
Croatia	2.97	24.03	−21.07
Cyprus	7.79	19.21	−11.42
Czechia	15.87	11.13	4.74
Denmark	26.93	0.07	26.85
Estonia	5.51	21.49	−15.99
Finland	23.25	3.75	19.50
France	20.06	6.94	13.12
Germany	21.69	5.31	16.38
Greece	3.97	23.03	−19.05
Hungary	5.16	21.84	−16.68
Ireland	18.38	8.62	9.77
Italy	11.53	15.47	−3.93
Latvia	8.98	18.02	−9.03
Lithuania	7.16	19.84	−12.68
Luxembourg	21.60	5.40	16.19
Malta	13.77	13.23	0.54
Netherlands	24.86	2.14	22.72
Poland	6.24	20.76	−14.51
Portugal	12.02	14.98	−2.96
Romania	0.93	26.07	−25.15
Slovakia	6.85	20.15	−13.30
Slovenia	15.01	11.99	3.01
Spain	13.04	13.96	−0.93
Sweden	25.34	1.66	23.67
United Kingdom	17.95	9.05	8.90

Monte Carlo generation of weights from a uniform distribution) used in Figure 2. The rank correlation between the generalized Copeland's method and the expected rank is 1.

5. Conclusion

This paper proposed Hierarchical Stochastic Multi-Objective Acceptability Analysis as an innovative composite indicator capturing the relevant context of EU28 member states. This was done using Eurostat data. In order to overcome differing policy priorities between countries, and corresponding heterogeneity in weights attached to individual indicators, the proposed method allows countries to be ranked according to the whole feasible set of weights. This approach quantifies the volume of vectors of weights by which each country obtains a specific ranking, and does not suffer from the curse of dimensionality when measuring SDG performance using a high number of indicators. Furthermore, HSMAA allows to consider the hierarchical structure of SDG measurement: that is, each goal is measured by means of different indicators.

Our results indicate weight-independent top (Denmark) and bottom (Romania, Bulgaria) performers in the EU, while many countries' ranks are weight-dependent.

The latter motivates the use of several weights, as resulting rankings are highly dependent on the chosen set. Disentangling the overall performance into specific goals, a substantial variability within each country was found. Countries ranked at the top of overall SDG performance can be ranked at the bottom for specific goals, and vice versa for overall low performing countries. This suggests that reporting at goal level remains relevant when developing composite indicators.

Embedding the proposed approach when monitoring SDG performance could strengthen accountability of member states. By considering all possible weights, accountability towards low-performing countries will be less sensitive to political maneuvering. Countries with a dense distribution at the bottom are robustly low-ranked, making it impossible for them to contest their ranking. Also, countries with wide distributions could still be compared by quantifying the probability by which each country receives a given ranking.

References

Angilella, S., S. Corrente, S. Greco, and R. Słowiński, "Robust Ordinal Regression and Stochastic Multiobjective Acceptability Analysis in Multiple Criteria Hierarchy Process for the Choquet Integral Preference Model," *Omega*, 63, 154–69, 2016.

Atkinson, A. B., and F. Bourguignon, "The Comparison of Multi-Dimensioned Distributions of Economic Status," *The Review of Economic Studies*, 49(2), 183–201, 1982.

Becker, W., M. Saisana, P. Paruolo, and I. Vandecasteele, "Weights and Importance in Composite Indicators: Closing the Gap," *Ecological Indicators*, 80, 12–22, 2017.

Booysen, F., "An Overview and Evaluation of Composite Indices of Development," *Social Indicators Research*, 59(2), 115–51, 2002.

Burgass, M. J., B. S. Halpern, E. Nicholson, and E. J. Milner-Gulland, "Navigating Uncertainty in Environmental Composite Indicators," *Ecological Indicators*, 75, 268–78, 2017.

Card, D., "The Causal Effect of Education on Earnings," *Handbook of Labor Economics*, 3, 1801–63, 1999.

Charnes, A., W. W. Cooper, and E. Rhodes, "Measuring the Efficiency of Decision Making Units," *European Journal of Operational Research*, 2(6), 429–44, 1978.

Cherchye, L., W. Moesen, N. Rogge, T. Van Puyenbroeck, M. Saisana, A. Saltelli, R. Liska, and S. Tarantola, "Creating Composite Indicators with DEA and Robustness Analysis: The Case of the Technology Achievement Index," *Journal of the Operational Research Society*, 59(2), 239–51, 2008.

Costanza, R., L. Daly, L. Fioramonti, E. Giovannini, I. Kubiszewski, L. F. Mortensen, and R. Wilkinson, "Modelling and Measuring Sustainable Wellbeing in Connection with the UN Sustainable Development Goals," *Ecological Economics*, 130, 350–5, 2016.

De Matteis, D., A. Ishizaka, and G. Resce, "The 'Postcode Lottery' of the Italian Public Health Bill Analysed with the Hierarchy Stochastic Multiobjective Acceptability Analysis," *Socio-Economic Planning Sciences*, 68, 100603, 2019.

Decancq, K., and M. A. Lugo, "Weights in Multidimensional Indices of Wellbeing: An Overview," *Econometric Reviews*, 32(1), 7–34, 2013.

Decancq, K., and E. Ooghe, "Has the World Moved Forward? A Robust Multidimensional Evaluation," *Economics Letters*, 107(2), 266–269, 2010.

Decancq, K., and E. Schokkaert, "Beyond GDP: Using Equivalent Incomes to Measure Well-Being in Europe," *Social Indicators Research*, 126(1), 21–55, 2016.

Deutsch, J., and J. Silber, "Measuring Multidimensional Poverty: An Empirical Comparison of Various Approaches," *Review of Income and Wealth*, 51(1), 145–74, 2005.

Esty, D. C., C. Kim, T. Srebotnjak, M. A. Levy, A. de Sherbinin, and V. Mara, *Environmental Performance Index. Yale Center for Environmental Law and Policy*, Yale University, New Haven, CT, 2008.

Eurostat, *Sustainable Development in the European Union: Monitoring Report on Progress towards the SDGS in an EU Context*, Publications Office of the European Union, 2019.

Foster, J., and A. Sen, "On Economic Inequality: After a Quarter Century," in A. Sen (ed.), *On Economic Inequality*, Extended Edition, Oxford University Press, New York, 107–220, 1997.

Foster, J. E., M. McGillivray, and S. Seth, "Composite Indices: Rank Robustness, Statistical Association, and Redundancy," *Econometric Reviews*, 32(1), 35–56, 2013.

Greco, S., A. Ishizaka, B. Matarazzo, and G. Torrisi, "Stochastic Multi-Attribute Acceptability Analysis (SMAA): An Application to the Ranking of Italian Regions," *Regional Studies*, 52(4), 585–600, 2018.

Greco, S., A. Ishizaka, G. Resce, and G. Torrisi, "Measuring Well-Being by a Multidimensional Spatial Model in OECD Better Life Index Framework," *Socio-Economic Planning Sciences*, 70, 100684, 2019.

Greco, S., A. Ishizaka, M. Tasiou, and G. Torrisi, "On the Methodological Framework of Composite Indices: A Review of the Issues of Weighting, Aggregation, and Robustness," *Social Indicators Research*, 141(1), 61–94, 2019.

Hák, T., S. Janoušková, and B. Moldan, "Sustainable Development Goals: A Need for Relevant Indicators," *Ecological Indicators*, 60, 565–73, 2016.

Hatefi, S. M., and S. A. Torabi, "A Slack Analysis Framework for Improving Composite Indicators with Applications to Human Development and Sustainable Energy Indices," *Econometric Reviews*, 37(3), 247–59, 2018.

Helliwell, J. F., "How's Life? Combining Individual and National Variables to Explain Subjective Well-Being," *Economic Modelling*, 20(2), 331–60, 2003.

Heritage Foundation, *Index of Economic Freedom*, The Heritage Foundation, Washington, DC, 2008.

Klasen, S., "Measuring Poverty and Deprivation in South Africa," *Review of Income and Wealth*, 46(1), 33–58, 2000.

Krishnakumar, J., and A. L. Nagar, "On Exact Statistical Properties of Multidimensional Indices Based on Principal Components, Factor Analysis, MIMIC and Structural Equation Models," *Social Indicators Research*, 86(3), 481–96, 2008.

Lafortune, G., G. Fuller, J. Moreno, G. Schmidt-Traub, and C. Kroll, *SDG Index and Dashboards: Detailed Methodological Paper*, SDSN, New York, 2018.

Lagravinese, R., P. Liberati, and G. Resce, "Exploring Health Outcomes by Stochastic Multicriteria Acceptability Analysis: An Application to Italian Regions," *European Journal of Operational Research*, 274(3), 1168–79, 2019.

Lahdelma, R., and P. Salminen, "SMAA-2: Stochastic Multicriteria Acceptability Analysis for Group Decision Making," *Operations Research*, 49(3), 444–54, 2001.

Leskinen, P., J. Viitanen, A. Kangas, and J. Kangas, "Alternatives to Incorporate Uncertainty and Risk Attitude in Multicriteria Evaluation of Forest Plans," *Forest Science*, 52, 304–12, 2006.

Mascherini, M., and B. Hoskins, *Retrieving Expert Opinion on Weights for the Active Citizenship Composite Indicator*, Scientific and Technical Research Series, Office for Official Publications of the European Communities, Luxembourg, 29 pp., 2008.

McGraw, K. O., and S. P. Wong, "Forming Inferences About Some Intraclass Correlation Coefficients," *Psychological Methods*, 1, 30–46, 1996.

Miola, A., and F. Schiltz, "Measuring Sustainable Development Goals Performance: How to Monitor Policy Action in the 2030 Agenda Implementation?," *Ecological Economics*, 164, 106373, 2019.

Murphy, K. M., and R. H. Topel, "The Value of Health and Longevity," *Journal of Political Economy*, 114(5), 871–904, 2006.

Nardo, M., M. Saisana, A. Saltelli, S. Tarantola, A. Hoffman, and E. Giovannini, *Handbook on Constructing Composite Indicators*, OECD-JRC, Paris, 2008.

Noorbakhsh, F., "The Human Development Index: Some Technical Issues and Alternative Indices," *Journal of International Development*, 10(5), 589–605, 1998.

OECD, *Glossary of Statistical Terms*, OECD, Paris, 2017.

Paruolo, P., M. Saisana, and A. Saltelli, "Ratings and Rankings: Voodoo or Science?," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(3), 609–34, 2013.

Patrizii, V., A. Pettini, and G. Resce, "The Cost of Well-Being," *Social Indicators Research*, 133(3), 985–1010, 2017.

Permanyer, I., "Assessing the Robustness of Composite Indices Rankings," *Review of Income and Wealth*, 57(2), 306–26, 2011.

Radermacher, W. J., "Recent and Future Developments Related to 'GDP and Beyond'," *Review of Income and Wealth*, 61(1), 18–24, 2015.

Ravallion, M., "Good and Bad Growth: The Human Development Reports," *World Development*, 25(5), 631–8, 1997.

Resce, G., and D. Maynard, "What Matters Most to People Around the World? Retrieving Better Life Index Priorities on Twitter," *Technological Forecasting and Social Change*, 137, 61–75, 2018.

Rogge, N., S. De Jaeger, and C. Lavigne, "Waste Performance of NUTS 2-Regions in the EU: A Conditional Directional Distance Benefit-of-the-Doubt Model," *Ecological Economics*, 139, 19–32, 2017.

Saaty, R. W., "The Analytic Hierarchy Process—What it is and How it is Used," *Mathematical Modelling*, 9(3–5), 161–76, 1987.
Saisana, M., S. Tarantola, and A. Saltelli, "Uncertainty and Sensitivity Techniques as Tools for the Analysis and Validation of Composite Indicators," *Journal of the Royal Statistical Society A*, 168(2), 307–23, 2005.
Sen, A., *Inequality Reexamined*, Clarendon Press, Oxford, 1992.
Sharpe, A., *Literature Review of Frameworks for Macro-indicators*, Centre for the Study of Living Standards, Ottawa, 2004.
Shen, Y., E. Hermans, T. Brijs, and G. Wets, "Data Envelopment Analysis for Composite Indicators: A Multiple Layer Model," *Social Indicators Research*, 114(2), 739–56, 2013.
Shrout, P. E., and J. L. Fleiss, "Intraclass Correlations: Uses in Assessing Rater Reliability," *Psychological Bulletin*, 86, 420–8, 1979.
Tervonen, T., and R. Lahdelma, "Implementing Stochastic Multicriteria Acceptability Analysis," *European Journal of Operational Research*, 178, 500–13, 2007.
Trannoy, A., "Multidimensional Egalitarianism and the Dominance Approach: A Lost Paradise." in Farina, F., and E. Savaglio, (eds), *Inequality and Economic Integration*. Routledge, London, 284–302, 2006.
United Nations Development Program, *Overcoming barriers: Human mobility and development. Human Development Report*, Palgrave-Macmillan, New York, 2009.
______, *The Real Wealth of Nations: Pathways to Human Development. Human Development Report*, Palgrave-Macmillan, New York, 2010.
Vidoli, F., E. Fusco, and C. Mazziotta, "Non-Compensability in Composite Indicators: A Robust Directional Frontier Method," *Social Indicators Research*, 122(3), 635–52, 2015.
Wolak, M. E., D. J. Fairbairn, and Y. R. Paulsen, "Guidelines for Estimating Repeatability," *Methods in Ecology and Evolution*, 3(1), 129–37, 2012.
World Bank, *World Development Indicators*. 2009. http://www.worldbank.org/. Accessed December 2009.
Zanella, A., A. S. Camanho, and T. G. Dias, "Undesirable Outputs and Weighting Schemes in Composite Indicators Based on Data Envelopment Analysis," *European Journal of Operational Research*, 245(2), 517–30, 2015.
Zhou, P., B. W. Ang, and D. Q. Zhou, "Weighting and Aggregation in Composite Indicator Construction: A Multiplicative Optimization Approach," *Social Indicators Research*, 96(1), 169–181, 2010.
Zilberman, D., B. Gordon, G. Hochman, and J. Wesseler, "Economics of Sustainable Development and the Bioeconomy," *Applied Economic Perspectives and Policy*, 40(1), 22–37, 2018.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web site:

Figure A.1: Comparison of rankings, by country

Table A.1: The rank acceptability indices (obtained out of 10,000 Monte Carlo generation of weights from a uniform distribution) for all Sustainable Development Goals

Table A.2: The rank acceptability indices (obtained out of 10,000 Monte Carlo generation of weights from the Halton sequence) for all Sustainable Development Goals

Table A.3: The rank acceptability indices (the average of ranks obtained out of 10,000 Monte Carlo generation of weights from a uniform distribution) with standard SMAA on 106 indicators under Sustainable Development Goal

Table A.4: The expected rank (the average of ranks obtained out of 10,000 Monte Carlo generation of weights from a uniform distribution) for each Sustainable Development Goal

Table A.5: Intraclass correlations between Uniform distribution and Halton sequence

Table A.6: Intraclass correlations between HSMAA and SMAA on the 106 indicators under Sustainable Development Goals

Table A.7: Sustainable development goals and indicators, and summary statistics

Table A.8: Description of Sustainable Development Goals and corresponding target