Ecological Indicators 127 (2021) 107740

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Does uncertainty in single indicators affect the reliability of composite indexes? An application to the measurement of environmental performances of Italian regions

Vincenzo Mauro [a], Caterina Giusti [b], Stefano Marchetti [b,*], Monica Pratesi [b]

[a] Department of Political Science, Communication and International Relations, University of Macerata, Macerata, Italy
[b] Department of Economics and Management, University of Pisa, Pisa, Italy

ARTICLE INFO

Keywords:
Composite indexes
Sampling variability
Bootstrap
Montecarlo simulations
2020 MSC:
62-08
62P12
62P25

ABSTRACT

In recent decades, the measurement and evaluation of important social and natural phenomena has significantly evolved, with many traditional measurements based on single variables increasingly being replaced by multi-dimensional approaches. One key aspect of these approaches is the development of composite indexes, usually real-value functions of multiple achievements of a group of units. The achievements in each of the selected dimensions are generally synthesised through one or more variables, often referred to as indicators. When indicators are obtained through an estimation process, it is crucial to understand if and how their estimation error – for example, sampling error – affects the resulting composite index.

This paper presents a methodology based on a parametric bootstrap technique that evaluates to what extent uncertainty in indicators affects the reliability of the aggregate composite index. The method is applied to four composite indexes measuring the environmental performances of Italian regions based on real population and survey data.

To our knowledge, this is the first attempt to measure the impact of indicators' sampling error on composite indexes. If adequately generalised, our methodology could be used in the presence of measurement errors, non-response issues, or other kinds of non-sampling errors.

1. Introduction

In recent decades, the measurement and evaluation of important social and natural phenomena has witnessed a significant revolution, due largely to the general agreement in the scientific community regarding the multifaceted nature of complex subjects like well-being, poverty, or the environment in the way they are defined and measured. Traditional measurements based on a single variable are now often replaced by multidimensional methods (Sen, 1999; Chakravarty and DAmbrosio, 2006; Alkire and Foster, 2011; Bossert et al., 2012; Nicholas et al., 2017). These new methods approach the phenomena under study from a broader perspective through the use of a large number of dimensions. Depending on the theoretical model chosen, these dimensions (i.e. "achievements" or "indicators") are interpreted either as the result of latent concepts that need to be analysed, or as their cause (Grace and Bollen, 2008; Bollen and Bauldry, 2011).

A large number of studies grounded in a multidimensional approach rely on the use of a composite index (CI), an operational tool to reduce the dimensionality of data (Booysen, 2002; OECD, 2008), whose use has significantly increased in recent years. Bandura (2011) and Yang (2014) cite the use of hundreds of composite indices covering an extensive range of topics, from well-being to the environment and child poverty. The surge in the amount of structured and unstructured information – often referred to as 'big data' – is a key driver that is accelerating the development of multidimensional frameworks. The availability of a large amount of data has boosted the attention on multidimensional approaches, leading to new theoretical and methodological challenges.

Sometimes, when using sample survey data, all the indicators used to define the CI are measured on subgroups of the target population. In other situations, they are obtained from different datasets, resulting in a mixed framework in which some indicators are measured from the overall population and others from samples of the population. As a result, the aggregated CIs obtained in these scenarios are likely to be affected by sampling variability in a way that can be very difficult to

* Corresponding author at: Department of Economics and Management, University of Pisa, Via C. Ridolfi, 10, 56124 Pisa, PI, Italy.
E-mail address: stefano.marchetti@unipi.it (S. Marchetti).

https://doi.org/10.1016/j.ecolind.2021.107740
Received 20 August 2020; Received in revised form 16 April 2021; Accepted 16 April 2021
Available online 5 May 2021

measure.

This is especially true for complex aspects like environmental and ecological measures, where the dimensions included in the analysis can be extremely diverse and belong to very different areas of study. For example, a single ecological study can simultaneously take into account variables such as biological diversity, resource depletion, food production, pollution, global temperature, and human population growth. For example, the sustainable development index provides a one-dimensional metric to evaluate country-specific information on different dimensions: economic, environmental, and social (Schmidt-Traub et al., 2017). Moreover, measures based on subjective processes (e.g. opinions or perceptions of a group of people) are usually derived from sample surveys and represent an additional source of heterogeneity between indicators.

Sampling variability is just one of the possible sources of error that should be accounted for. For example, non-sampling errors like non-response and measurement errors may also significantly affect the data used to calculate single indicators that comprise the CI. In this paper, we focus only on sampling error as a source of uncertainty affecting the indicators; nevertheless, our methodology can be extended to consider other sources of error, as we explain later.

The literature on the development of CIs has focused on the effects of normalisation procedures (Drewnowski, 1972; McGranahan et al., 1972; Nardo et al., 1972), combination of different indicators (linear or non-linear) (Moldan et al., 1997; Saaty, 1972; Lebart et al., 1984), and sensitivity of CIs to methods of aggregation and weighting (Gan et al., 2017; Bohringer and Jochem, 2007). Some authors consider the variability among the indicators included in CIs (Mazziotta and Pareto, 2013; Biggeri and Mauro, 2018), while others discuss the uncertainty in the construction process of a CI (Burgass et al., 2017). Nonetheless, the sensitivity of CIs due to the uncertainty in indicators has often been neglected in the literature, and 'providing measures of accuracy for composite index is a problem even more far from the solution' [p. 2] Ceccarelli et al. (2020).

In this paper, we try to fill this gap, analysing how uncertainty about the true values of the indicators – due to sampling error – affects the final aggregate score. We propose a method to obtain estimates of CIs as well as their variability. The method is applied to real data indicators that are combined to create a CI describing the environmental performances of different Italian regions. The proposed method empirically captures how the magnitude of the sampling error in single indicators influences the corresponding error of the CI. In discussing the results, we also suggest how errors in the regional CIs should be interpreted – for example, in terms of ranking the environmental performance of the regions.

Specifically, we propose accounting for the sampling error in the aggregation phase of CIs using a parametric bootstrap technique. We replicate the sampling distribution of each indicator considered to obtain a distribution, which is used to derive standard errors and confidence intervals for the CI. We apply the proposed technique to four aggregation methods: arithmetic means, geometric means, Mazziotta-Pareto index, and multidimensional synthesis of indicators.

To our knowledge, there is no literature on estimating the impact of indicators' sampling error on CIs. The methodology proposed in this paper must therefore be considered as a first attempt in this direction, with the caveat that the preliminary results provided rely on hypotheses that could be revised, extended, or relaxed in order to embrace more general situations.

The paper is organised as follows. In Section 2, the proposed methodology is illustrated, focusing first on the aggregating functions used to define the CIs and then on the proposed estimation method. Section 3 presents the single indicators used to define the example environmental CI used in the study, the results of applying our method, and a discussion. Finally, in Section 4, we present the main conclusions and future research directions.

2. Methodology

2.1. The aggregating functions for Composite Indexes

Before describing our method for handling sampling error in a CI, we introduce the four functions used to aggregate indicator dimensions, namely the arithmetic mean (A), geometric mean (G), Mazziotta-Pareto [1]index (De Muro et al., 2011)[MPI], and multidimensional synthesis of indicators (Mauro et al., 2018)[MSI].

We choose these aggregating functions because, except the arithmetic mean, the rest take into account – explicitly or implicitly – the different degrees of substitutability[2] between the indicators of a single unit (e.g. a country, an area, or an individual). As a result, they should be able to capture changes in the heterogeneity of the generic row of the data matrix due to the sampling error of the single indicators in the CI.

Since 2010, the United Nations Development Programme (UNDP) has adopted the geometric mean to calculate the Human Development Index (HDI), because this function has a desirable limited substitutability property for aggregation across heterogeneous variables (UNDP, 2010). However, if at least one outcome is zero, the combination of all indicators through the calculation of the geometric mean will collapse to zero (Klugman et al., 2011). To overcome this significant problem, De Muro et al. (2011); Mazziotta and Pareto (2013) and Mauro et al. (2018) proposed alternative functions to manage the substitutability between indicators.

We describe how to compute these four functions when the sampling error is zero.

The procedure of building a composite index is made of many steps, that include, among the others, crucial decisions on functions, weighting, or degree of substitutability between dimensions. As the aim of this paper is to provide a first contribution to the study of the source of uncertainty mentioned above, we decided to set these steps at a basic level (e.g. assuming equal weighting) and leave for future research the analysis of all the possible generalisations. In particular, the standardisation phase of the variables – with its many different varieties – is still debated in the scientific community, Pollesch and Dale (2016), so we limit our analysis to the widely used method of normalisation described below. Nonetheless, the results can easily be replicated by adopting a wider range of decisions on the various step, as well as on standardisation methods.

First, data are transformed for the variables to be comparable. Let $\boldsymbol{X}$ be the standard data matrix with generic entry x_{ij} the j-th indicator for unit i, with $j = 1, \ldots, p$ and $i = 1, \ldots, n$, and let $\boldsymbol{x}_j$ be the j-th column of matrix $\boldsymbol{X}$. For A, G, and MSI, the j-th generic column of the standardised matrix $\boldsymbol{Z}$ when the j-th indicator has a positive polarity (the higher the value, the higher the environmental performance) is defined as

$$z_j = \frac{\boldsymbol{x}_j - \min(\boldsymbol{x}_j)}{\max(\boldsymbol{x}_j) - \min(\boldsymbol{x}_j)},$$

when the j-th indicator has a negative polarity (red the smaller the value, the lower the environmental performance), let

$$z_j = \frac{\max(\boldsymbol{x}_j) - \boldsymbol{x}_j}{\max(\boldsymbol{x}_j) - \min(\boldsymbol{x}_j)}.$$

[1] This index is meant for static analyses, and it is therefore adopted here as the application refers to only one data matrix. A test analysis using the index AMPI (Mazziotta and Pareto, 2016), generally used for comparisons over time, reported similar results, suggesting a good degree of robustness of the methods.

[2] Let $\mathbf{x}$ be a fixed vector of p indicators with generic element x_j, let $\Delta_j > 0$ be the quantity that needs to be added to x_j to compensate a decrease $\Delta_i > 0$ in x_i to leave the index unchanged (i.e. $f(x_1, \ldots, x_i, \ldots, x_j, \ldots, x_p) = f(x_1, \ldots, x_i - \Delta_i, \ldots, x_j + \Delta_j, \ldots, x_p)$). The degree of substitutability between the two generic dimensions i and j is defined as $\lim_{\Delta_i \to 0} \Delta_j / \Delta_i$

The functions A, G and MSI for unit i are:

$$A_i = \frac{1}{p}\sum_{j=1}^{p} z_{ij},$$
$$G_i = \left(\prod_{j=1}^{p} z_{ij}\right)^{\frac{1}{p}}, \quad (1)$$
$$MSI_i = 1 - \left[\frac{1}{p}\sum_{j=1}^{p}\left(1 - z_{ij}\right)^{g(\cdot_i)}\right]^{\frac{1}{g(\cdot_i)}},$$

where $g(\cdot)$ in the MSI formula is a generic real-value function of the i-th row of matrix $\boldsymbol{Z}$, with $g(\cdot)\geqslant 1$. In this work, we assume $g(\cdot) = A_i^{-1}$.

The normalisation method for the MPI is based on a standardised variable with mean 100 and standard deviation 10, as suggested in De Muro et al. (2011). This standardisation methods, on the basis of the Bienaymè-Cebycev theorem, ensure the terms of the distribution within the range (70; 130) are at least 89% of total terms. The entries of the normalised matrix $\boldsymbol{Z}$ are:

$$z_{ij} = 100 \pm \frac{x_{ij} - \overline{x}_j}{s_{x_j}} 10,$$

where $\overline{x}_j = n^{-1}\sum_{i=1^n} x_{ij}, s_{x_j} = \{n^{-1}\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2\}^{1/2}$ and $\pm$ depends on the relation of the j-th indicator with the phenomenon to be measured: '$+$' for dimensions with positive polarity, and '$-$' for dimensions with negative polarity.

The final value of the MPI is obtained through a penalisation proportional to its *horizontal variability*:

$$MPI_i^{\pm} = \overline{z}_i \pm s_{z_i} cv_i, \quad (2)$$

where $\overline{z}_i = p^{-1}\sum_{j=1}^{p} z_{ij}, s_{z_i} = \{p^{-1}\sum_{j=1}^{p}(z_{ij} - \overline{z}_i)^2\}^{1/2}$ and $cv_i = s_{z_i}/\overline{z}_i$. The $\pm$ is set according to the positive or negative interpretation of the MPI: the '$-$' is used when an increase of the MPI positively affects the unit (e.g. when measuring well-being), and '$+$' when an increase of the MPI negatively affects the unit (e.g. when measuring poverty).

2.2. Estimation of CI variance

Let X be a matrix with generic entry x_{ij}, an *estimate* of the j-th indicator for unit i (country, region, area, etc.), $j = 1,\dots,p$, and let $\boldsymbol{\Sigma}$ be a matrix with generic entry σ_{ij}^2, the variance of the estimate x_{ij}. The values for x_{ij} and σ_{ij}^2 are typically obtained from aggregation of elementary units in the sample. If a generic entry is measured without error, we simply assume $\sigma_{ij}^2 = 0$.

Let us also assume that: 1. the sampling errors of the indicators are independent of each other within units, i.e. the sampling error of indicator j for unit i does not depend on that of the other $k-1$ indicators of unit i, 2. the estimates x_{ij} are unbiased with respect to the design, and 3. σ_{ij}^2 is known $\forall i, j$. Usually, σ_{ij}^2 is unknown and it is estimated from the sample data. However, the estimator of σ_{ij}^2 is often smoothed, and smoothed estimators are often treated as true sampling variances [examples can be found in] (Rao and Molina, 2015).

Assumption 1 is required to estimate the sampling error of the CIs without taking into account the potential correlation among sampling errors of the indicators (e.g. indicators obtained from the same surveys may have correlated sampling errors). Assumption 2 is needed for the unbiasedness of the CI estimators. We shortly discuss the sample properties of point estimators of CIs at the end of this section. Assumption 3 is required to generate the sampling variability of each indicators through the bootstrap technique. This is a baseline work with the goal to start exploring how uncertainty in single indicators might affect the reliability of composite indexes. Future research will aim to address more complex scenarios where one or more of these assumptions are violated.

The next step is to estimate the four CIs and their variances as a function of indicators that are possibly affected by sampling error.

Let $\widehat{A}, \widehat{G}, \widehat{MSI}$, and $\widehat{MPI}$ be the estimators for the unknown values A, G, MSI, and MPI, respectively. Let $\boldsymbol{\Xi}$ be a matrix with generic entry ξ_{ij}, the standardised value of x_{ij}. The four estimators are obtained according to (1) and (2) using ξ_{ij} instead of z_{ij}. We need the standardised values ξ_{ij}s instead of the usual z_{ij}, because in the standardisation phase, there is also a need to take into account the sampling error of the indicators.

Analytically estimating the variances of the CIs is not trivial. Therefore, we propose a parametric bootstrap technique:

1. generate B bootstrap matrix $\boldsymbol{X}^{*,b}$, where the entry $x_{ij}^{*,b}$ is a realisation of the sampling distribution of x_{ij} (the unbiased estimator of the indicator j for unit i), where $b = 1,\dots,B$; usually, $x_{ij}^{*,b} = x_{ij} + e_{ij}$ where, if the indicator j is a mean, a total, or a proportion, then by the central limit theorem, e_{ij} tends to be a Normal random variable with mean zero and variance σ_{ij}^2 (the variance of indicator – i.e. unbiased estimator – x_{ij});
2. starting from $\boldsymbol{X}^{*,b}$, compute B standardised matrix $\boldsymbol{\Xi}^{*,b}$; the minimum x_j^{min} and maximum x_j^{max} are held fixed; for the generic indicator j the x_j^{min} is the minimum among the lower bounds of $1-\alpha$ confidence intervals for indicators x_{ij}s, i.e. $x_j^{min} = \min\{x_{ij} - c_{\alpha/2}\sigma_{ij}\}$, where $c_{\alpha/2}$ is the $(1-\alpha/2)$th quantile of the standard Normal distribution; $x_j^{max} = \max\{x_{ij} + c_{\alpha/2}\sigma_{ij}\}$; if an indicator has natural bounds, like a proportion that must be between zero and one, then the minimum and maximum values must lie within the bounds;
3. using $\boldsymbol{\Xi}^{*,b}$, we obtain B estimates of target CIs, which we collectively denote by $\tau_i^{*,b}, b = 1,\dots,B$;
4. obtain the estimated variance from the B-vector of CIs:

$$\widehat{V}(\tau)_i = \frac{1}{B}\sum_{b=1}^{B}\left(\tau_i^{*,b} - \overline{\tau}_i^*\right)^2,$$

where $\overline{\tau}_i^* = B^{-1}\sum_{b=1}^{B}\tau_i^{b,*}$.

If some indicators are not means, totals or proportions, then the normal distribution used in step 1 of the bootstrap to generate the sampling variability may not be appropriate. For example, this is the case of indicators based on quantiles, inequality indexes or other statistics for which the sampling distribution is not known or is not normal. In these cases, the step 1 of the proposed bootstrap technique must be changed. Of course, many bootstrap techniques can be applied, depending on the data availability. When survey micro data are available, non-parametric bootstrap can be used to generate the sampling variability. Nevertheless, the idea to generate the sampling variability of single indicators in a bootstrap technique remains valid.

In step 2 of the bootstrap procedure, the x_{ij}s are standardised using the lower and upper bounds of a $1-\alpha$ confidence interval as minimum and maximum values for two reasons: 1. for each indicator, the units corresponding to the minimum and maximum values are fixed, and 2. bootstrap values $x_{ij}^{(*,b)}$ (step 1) are almost surely in the standardized $[0, 1]$ bound. These considerations apply to A, G and MSI, while MPI requires further investigation that is let to future work.

The sample properties of the estimators $\widehat{A}$, $\widehat{G}$, $\widehat{MSI}$, and $\widehat{MPI}$ were assessed using a Monte Carlo simulation study. Moreover, by simulation, we assessed the properties of the bootstrap estimator of the variance of the CIs. However, to keep the paper concise and readable, the simulation settings and the results are not reported here (they are available upon request). The simulations results show that point estimators are unbiased with very small variability around the target. The bootstrap estimator of the variance shows a small negative bias that is judged to be negligible.

3. Application

3.1. The data

As mentioned in Section 1, many sources of information can be used to develop environmental CIs. To propose a CI to measure the environmental performance of 20 Italian regions, we based the choice of data sources and indicators on those currently used in the Italian National Statistical Institute's (Istat) Equitable and Sustainable Well-being (BES) domains 'Landscape and cultural heritage' and 'Environment' (Ciommi et al., 2017). These are the two BES domains -' out of its total 12 – dealing with the environmental aspects of sustainable development. These indicators are measured using data from administrative archives and from the Istat sample survey on Aspects of daily life. In addition, we include three additional indicators in the CI, obtained from two national sample surveys: the EU-SILC (European Union - Statistics on Income and Living Conditions) and the HBS (Household Budget Survey). In total, we use 14 different indicators. The complete list of elementary indicators is presented in Table 1.

The indicator 'Urban green areas per inhabitant' is measured in squared meters, with no fixed upper limit. All other indicators are expressed in percentage terms and specify the share of the population affected by the measured phenomenon. Indicators computed using data from administrative archives do not exhibit sampling error because archives usually cover the entire population. This study uses seven such indicators, which provide information about environmental variables such as air quality, water treatment, green areas, and waste sorting.

Information obtained from sample surveys is an important source to account for the perception of environmental problems among Italians, as is the case for the indicators defined using data from the Aspects of daily life and EU-SILC surveys. For each region, we also measure the share of car transportation expenses in overall transportation expenses (train, metro, etc.) using HBS data. Thus, the number of indicators obtained using sample survey data is also seven; since these dimensions are affected by sampling error, we estimated both the value and standard error for each indicator in each region. Specifically, for the indicators coming from the EU-SILC and HBS surveys the standard errors were computed by the authors using the design weights as unit level data with the regional reference. For the indicators from the Istat survey on Aspects of daily life, the standard errors were computed instead using the information available in Istat methodological reports (Istat, 2017).

Table 1 reports a brief summary of the indicators and their polarity.

3.2. Application results

In this section, we estimate the four CIs presented in Section 2, aggregating the 14 indicators described in Table 1 for the 20 Italian regions.

CIs can be computed as a synthesis of indicators measured with or without sampling errors. National-level estimates are often based on surveys with large sample sizes that usually exhibit negligible sampling errors, while analyses based on smaller areas usually rely on smaller samples with significant sampling errors. Although regional estimates can be considered reliable, their sampling error may have a significant impact on the aggregate CI. In our application, seven indicators are measured without sampling error and seven are obtained from survey samples. The estimates of the different indicators have a coefficient of variation (CV) between 1.2% (very small) to approximately 50% (very large). This aspect must be taken into account to make credible inference through the four CIs (ESS, 2013).

Point estimates, estimated standard errors, and confidence intervals for the four CIs are obtained using the method introduced in Section 2. The assumption of independence between the CIs' sampling errors, as specified in Section 2, appears reasonable since most of the estimates are obtained from different surveys. An example could help to clarify this concept. Consider the indicator j of region i with estimate x_{ij} that has a positive sampling error, say ε_{ij}. There is no reason to think that another indicator for region i has an estimate with an error related to ε_{ij}, if the two indicator estimates come from different surveys.

Data are normalised according to the method described in Section 2. The minimum and maximum values to compute the normalised indicators are obtained setting $\alpha = 7E^{-5}$, which turns out in $c_{3.5E^{-5}} = 3.976$ (in the next paragraph, we argue about the impact of the choice of $c_{\alpha/2}$ on the CI variance estimator). In each bootstrap replication, we perform a check of the values simulated and discard out-of-bounds replications: with the value of $c_{\alpha/2}$ we set, the probability to generate values out of bounds (x_j^{min}, x_j^{max}) for each indicator is $2 - 2\Phi(c_{\alpha/2})$. In our application, with seven indicators affected by sampling error, setting $c_{3.5E^{-5}} = 3.976$, we expect to discard $1 - \{2\Phi(c_{3.5E^{-7}}) - 1\}^7 = 0.05\%$ replications. We run 1000 replications (Efron, 1987) and discard only one replication.

The effect of $c_{\alpha/2}$ on the estimated standard errors was tested using Monte Carlo simulations (available upon request). The estimated variances $\widehat{V}(\widehat{A}_i)$ and $\widehat{V}(\widehat{MSI}_i)$ of CIs estimators $\widehat{A}_i$ and $\widehat{MSI}_i$ show a similar

Table 1
Indicators used to define the CIs: name, source, polarity, minimum observed value, maximum observed value, and information on sampling error in terms of percentage coefficient of variation (CV).

Indicator	Source	Polarity	Min	Max	Min CV (%)	Max CV (%)
Dissatisfaction for the landscape deterioration (%)	Istat survey on Aspects of daily life.	-	7.70	32.80	2.38	7.88
Concern for landscape deterioration (%)	Istat survey on Aspects of daily life.	-	8.00	21.80	2.71	7.87
Satisfaction with environmental quality (%)	Istat survey on Aspects of daily life.	+	54.30	90.50	1.25	2.36
Concern for biodiversity loss (%)	Istat survey on Aspects of daily life.	-	14.90	23.70	2.69	5.33
Noise from neighbours or from the street (%)	Istat survey EU-SILC	-	5.04	25.54	5.51	31.42
Pollution, grime or other environment problems (%)	Istat survey EU-SILC	-	2.78	25.18	5.25	49.74
Share of car transportation expenses on overall transportation expenses (%)	Istat survey HBS	-	84.98	96.70	0.69	4.37
Losses in drinking water supply network (%)	Istat census on water availability for civil use	-	18.70	56.30	-	-
Exceeding of the NO2 annual limit for the protection of human health (%)	Istat environmental data on regional capital cities	-	0.00	50.00	-	-
Urban green areas per inhabitant	Istat environmental data on regional capital cities	+	7.10	571.80	-	-
Wastewater treatment failures (%)	Istat census on water availability for civil use	-	43.90	78.90	-	-
Protected natural areas (%)	Istat data based on Ministry for the Environment information	+	12.20	36.60	-	-
Energy from renewable resources (electricity, %)	Istat data based on Ministry for the Environment information	+	8.60	323.10	-	-
Waste sorting (%)	Istat data based on Ministry for the Environment information	+	12.80	68.80	-	-

behaviour with respect to the constant $c_{\alpha/2}$, while $\widehat{G}$ responds in a different way. The estimated variances of $\widehat{MSI}$ and $\widehat{A}$ tend to underestimate the true (empirical) variances as $c_{\alpha/2}$ increases, flattening to a relative bias of about −20%. The estimated variance of $\widehat{G}$ is highly positively biased for small $c_{\alpha/2}$s (about 60% for $c_{\alpha/2}$ about 1.3), and it decreases as $c_{\alpha/2}$ increases, obtaining a negative biased estimator (about −15%) for $c_{\alpha/2}$ about 4. The $\widehat{MPI}$ is less sensitive by construction to the choice of $c_{\alpha/2}$, and its estimated variance is not sensitive to changes in $c_{\alpha/2}$, with a very small negative relative bias of about −6%. The choice of α also affects the correctness of point estimates; however, its impact is limited because the relative bias is between −1% and 1.5% for all the four estimators. To balance the performance of the four CIs and compare them using the same parameters, we set $c_{\alpha/2}$ to about 4. We suggest setting α smaller than 0.001 for the geometric mean and 0.15 or 0.2 for the average and the *MSI* (for the *MPI* it is irrelevant).

Table 2 shows point estimates and estimated standard errors of the four CIs. To avoid the well-known limit of the geometric mean *G* collapsing to zero, we replaced the zero values in the standardised matrix with 10^{-4}.

As expected, the $\widehat{MSI}$ – which is supposed to induce milder penalisation than the geometric mean – lies between $\widehat{A}$ and $\widehat{G}$. The CVs vary between 0.3% (for the $\widehat{MPI}$ in Veneto region) and 5.1% (for the geometric mean in Trentino-Alto Adige). The average CV for $\widehat{A}$ is 1.89%; for $\widehat{G}$, 2.86%; for $\widehat{MPI}$, 0.47%; and for $\widehat{MSI}$, 2.60%. From this, $\widehat{G}$ shows the highest average variability, while $\widehat{MPI}$ shows the lowest. Even if the variability of the CIs is limited, given that point estimates are sometimes very close to each other, their variability cannot be ignored.

To visualise the uncertainty of the estimates, we plot in Fig. 1 the regional estimates of the four CIs with 95% confidence bounds obtained using the normal distribution as reference – i.e. $\widehat{\tau}_i \pm 1.96\sqrt{\widehat{V}(\widehat{\tau}_i)}$, where $\widehat{\tau}_i$ is the estimate of one of the CIs among $\widehat{A}, \widehat{G}, \widehat{MSI}$, and $\widehat{MPI}$ in region *i* and $\widehat{V}(\widehat{\tau}_i)$ is its estimated variance.

As we can see from Fig. 1, the error in the estimates does not allow for a straightforward interpretation of the rank of the regions according to CIs. When two confidence intervals overlap, in fact, the null hypothesis that two regions share the same rank cannot be rejected. For example, focusing on the $\widehat{MSI}$, it is not trivial to identify the worst-performing regions, as the last four regions exhibit widely overlapping confidence intervals. Although this is a very important issue, data users such as policymakers and stakeholders often base their analyses only on point estimates even when standard errors of estimates are provided. A simple graphical representation can help to visualise the uncertainty of the estimated CIs.

Policymakers and socio-economic practitioners prefer to rank units to highlight the best and the worst performances. Our method enables easy estimation of the probability that a region occupies a given rank by using the bootstrap replications – while capturing the impact of sampling error – and also providing data users with an alert on the dangers of incorrect inferences if they focus only on point estimates. As an example, in Table 3, we show the rank probability distribution for the $\widehat{MSI}$.

From Table 3, we can see, for example, that Valle d'Aosta region is the best region with probability 0.99. Meanwhile, the Sicilia region is the worst with probability 0.56, and Trentino-Alto Adige is 4th with probability 0.18, 5th with probability 0.18, and so on.

The four CIs used in our application have drawbacks and advantages. More complex CIs, like the *MPI* and the *MSI*, can achieve a better synthesis than that of simple indicators like the arithmetic mean *A* and the geometric mean *G*. However, given their simplicity, arithmetic and geometric means are often used (Klugman et al., 2011; Sachs et al., 2018; Schmidt-Traub et al., 2017).

One of the main issues in using the geometric mean is that it requires strictly positive values, while some widely used methods to standardise an indicator assign zero to the smallest value of the indicator. Some authors suggest removing the units with zero values and then computing the geometric mean (Chakrabartty, 2017); others suggest alternative solutions (Mauro et al., 2018). In this paper, we added a small constant (e.g. 0.001) to every zero after normalisation. This choice is due to the small number of units in our analysis. As there are only 20 regions, dropping six or seven units could significantly bias the results. In the case of indicators measured with error, the use of the proposed normalisation method avoids the zeros in the normalised values (if *c* is large enough).

Table 2

Point estimates and estimated standard error $\widehat{se}$ of CIs: $\widehat{A}$ is the arithmetic mean, $\widehat{G}$ is the geometric mean, $\widehat{MSI}$ is the multidimensional synthesis of indicators and $\widehat{MPI}$ is the Mazziotta-Pareto index.

Region	$\widehat{A}$	$\widehat{se}(\widehat{A})$	$\widehat{G}$	$\widehat{se}(\widehat{G})$	$\widehat{MSI}$	$\widehat{se}(\widehat{MSI})$	$\widehat{MPI}$	$\widehat{se}(\widehat{MPI})$
Piemonte	0.370	0.0061	0.300	0.0053	0.327	0.0065	96.930	0.3590
Valle d'Aosta	0.601	0.0105	0.457	0.0151	0.523	0.0144	107.884	0.4756
Liguria	0.391	0.0099	0.129	0.0032	0.336	0.0097	99.608	0.6315
Lombardia	0.391	0.0059	0.291	0.0050	0.328	0.0066	98.368	0.3709
Trentino-AltoAdige	0.509	0.0140	0.156	0.0079	0.422	0.0174	105.677	0.8802
Veneto	0.479	0.0057	0.365	0.0059	0.412	0.0068	101.203	0.3011
Friuli-VeneziaGiulia	0.487	0.0072	0.386	0.0068	0.419	0.0085	102.125	0.3968
Emilia-Romagna	0.422	0.0062	0.193	0.0029	0.353	0.0067	99.456	0.3678
Toscana	0.463	0.0082	0.351	0.0065	0.404	0.0090	101.433	0.5658
Umbria	0.454	0.0103	0.373	0.0162	0.383	0.0125	100.635	0.5833
Marche	0.526	0.0082	0.389	0.0186	0.454	0.0108	103.113	0.4347
Lazio	0.318	0.0075	0.207	0.0045	0.258	0.0071	95.804	0.4757
Abruzzo	0.490	0.0091	0.385	0.0153	0.428	0.0113	102.640	0.5252
Molise	0.555	0.0083	0.399	0.0069	0.482	0.0082	106.949	0.4788
Campania	0.374	0.0081	0.224	0.0090	0.272	0.0095	99.272	0.5460
Puglia	0.350	0.0071	0.232	0.0071	0.277	0.0086	96.072	0.3657
Basilicata	0.508	0.0089	0.260	0.0068	0.430	0.0092	105.093	0.4933
Calabria	0.443	0.0084	0.346	0.0117	0.358	0.0102	100.079	0.4157
Sicilia	0.357	0.0061	0.137	0.0038	0.256	0.0062	97.843	0.3742
Sardegna	0.409	0.0104	0.282	0.0101	0.335	0.0121	97.976	0.8396

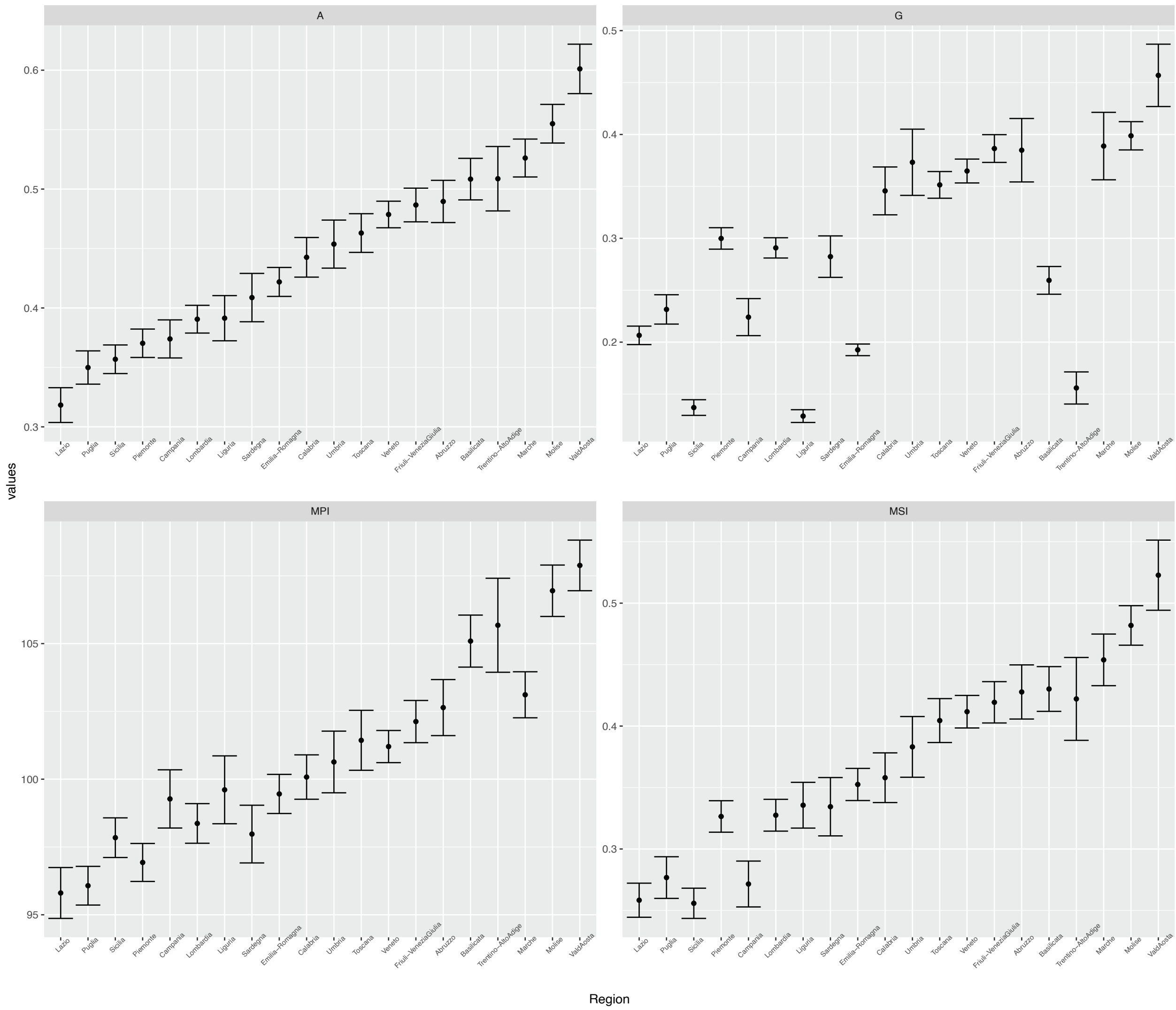

Fig. 1. Point estimates and 95% confidence intervals for CIs: $\widehat{A}$ (arithmetic mean), $\widehat{G}$ (geometric mean), $\widehat{MPI}$ (Mazziotta-Pareto index), $\widehat{MSI}$ (multidimensional synthesis of indicators). In all the graphs, the regions are sorted by increasing values of the arithmetic mean $\widehat{A}$.

Another issue that may arise in applying our bootstrap method in the estimation of G is that negative or positive deviates of an indicator do not have the same impact on the geometric mean. A positive deviation has a smaller impact than a negative deviation of equal magnitude. However, through a Monte Carlo simulation, we checked that the different impacts of adding or subtracting deviates offset one another, and the $\widehat{G}$ estimator is nearly unbiased. Deeper investigations into the properties of the estimator of G are beyond the aims of this paper and are left to future inquiry.

4. Conclusions

In this paper, we proposed a methodology to estimate the standard error of a CI when one or more of its indicators are affected by sampling error. The methodology, based on a parametric bootstrap technique applicable to many functions, was tested on the arithmetic mean, geometric mean, MPI, and MSI using real data on the environmental performances of the 20 Italian regions.

We obtained point estimates and standard error estimates for each CI, showing that the presence of (non-negligible) sampling error is a relevant issue affecting the reliability of the CIs. Since analyses that ignore the presence of sampling error can be especially misleading when CIs are used to rank units under study (i.e. the Italian regions in this paper), we showed how the proposed methodology enables estimation of the probability that a unit belongs to a given rank.

The results show that all the CIs identified Val d'Aosta and Molise as the regions with the best environmental performances in Italy. This is also confirmed by the estimated probability of rank 1 for Val d'Aosta (99%) and rank 2 for Molise (99%) according to the MSI. Meanwhile, the regions with the poorest performances are not clearly identified by any of the four CIs, as the estimates' confidence intervals overlap.

To our knowledge, the methodology proposed in this paper is the first attempt to measure the impact of indicators' sampling error on CIs. Our preliminary results rely on hypotheses that could be revised, extended, or modified to embrace more general situations. For example, sampling error is not the only kind of error affecting the accuracy of indicators. If adequately generalised, our methodology could be used in the presence of measurement errors, non-response issues, or other kinds of non-sampling errors. Other important extensions of the methodology concern the management of situations where sampling errors cannot be considered independent, for example when they are obtained from the same survey, as well as for more general cases that include, but are not limited to, frameworks where the dimensions are first aggregated within single pillars and then between pillars. Finally, possible extensions can be also presented in case of alternative normalisation methods, with the aim of evaluating the sensitivity of the procedure to different

Table 3
Rank probability distribution for the $\widehat{MSI}$

Region	r1	r2	r3	r4	r5	r6	r7	r8	r9	r10	r11	r12	r13	r14	r15	r16	r17	r18	r19	r20
Piemonte	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.08**	**0.22**	**0.32**	**0.37**	0.00	0.00	0.00	0.00
Valle d'Aosta	**0.99**	**0.01**	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	0.00	0.00	0.00	0.00
Liguria	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.01**	**0.07**	**0.34**	**0.31**	**0.15**	**0.12**	0.00	0.00	0.00	0.00
Lombardia	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.01**	**0.09**	**0.22**	**0.37**	**0.31**	0.00	0.00	0.00	0.00
Trentino-Alto Adige	0.00	0.00	**0.04**	**0.18**	**0.18**	**0.16**	**0.14**	**0.13**	**0.14**	**0.03**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Veneto	0.00	0.00	0.00	0.00	**0.03**	**0.13**	**0.28**	**0.41**	**0.14**	**0.01**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Friuli-Venezia Giulia	0.00	0.00	0.00	**0.07**	**0.17**	**0.26**	**0.30**	**0.14**	**0.05**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Emilia-Romagna	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.01**	**0.29**	**0.58**	**0.11**	**0.01**	0.00	0.00	0.00	0.00	0.00	0.00
Toscana	0.00	0.00	0.00	0.00	**0.01**	**0.04**	**0.11**	**0.22**	**0.55**	**0.06**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Umbria	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.02**	**0.09**	**0.83**	**0.05**	**0.01**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Marche	0.00	**0.02**	**0.89**	**0.07**	**0.02**	**0.01**	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
Lazio	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	0.00	0.00	**0.02**	**0.12**	**0.48**	**0.38**
Abruzzo	0.00	0.00	**0.03**	**0.31**	**0.25**	**0.22**	**0.11**	**0.05**	**0.02**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Molise	**0.01**	**0.97**	**0.02**	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	0.00	0.00	0.00
Campania	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	0.00	0.00	**0.34**	**0.50**	**0.11**	**0.05**
Puglia	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	0.00	0.00	**0.64**	**0.30**	**0.05**	**0.01**
Basilicata	0.00	0.00	**0.02**	**0.36**	**0.34**	**0.18**	**0.06**	**0.03**	**0.01**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Calabria	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.05**	**0.62**	**0.25**	**0.06**	**0.02**	0.00	0.00	0.00	0.00	0.00	0.00
Sicilia	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	0.00	0.00	0.00	**0.08**	**0.35**	**0.56**
Sardegna	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.03**	**0.08**	**0.33**	**0.22**	**0.15**	**0.19**	0.00	0.00	0.00	0.00

normalisation techniques.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The work of Giusti, Marchetti and Pratesi was supported by Project InGRID-2, H2020 European Project, G.A. No. 730998.http://www.inclusivegrowth.eu.

References

Alkire, S., Foster, J.E., 2011. Counting and multidimensional poverty measurement. J. Public Econ. 95, 476–487.
Bandura, R. (2011). Composite indicators and rankings: Inventory 2011. Technical report. New York: Office of Development Studies, United Nations Development Programme (UNDP) (2011th ed.).
Biggeri, M., Mauro, V., 2018. Towards a more sustainable human development index: Integrating the environment and freedom. Ecol. Ind. 91, 220–231.
Bohringer, C., Jochem, P.E., 2007. Measuring the immeasurable – a survey of sustainability indices. Ecol. Econ. 63, 1–8.
Bollen, K.A., Bauldry, S., 2011. Three cs in measurement models: causal indicators, composite indicators, and covariates. Psychological Methods 16, 265–284. https://doi.org/10.1037/a0024448.
Booysen, F., 2002. An overview and evaluation of composite indices of development. Soc. Indic. Res. 59, 115–151.
Bossert, W., Chakravarty, S., D'Ambrosio, C., 2012. Poverty and time. J. Economic Inequality 10, 145–162.
Burgass, M.J., Halpern, B.S., Nicholson, E., Milner-Gulland, E., 2017. Navigating uncertainty in environmental composite indicators. Ecol. Ind. 75, 268–278.
Ceccarelli, C., Guandalini, A., Martini, A., Pontecorvo, M.E., 2020. Accuracy evaluation of lfs-bes indicators: A regional assessment. Soc. Indic. Res. https://doi.org/10.1007/s11205-020-02532-3.
Chakrabartty, S., 2017. Composite index: Methods and properties. Journal of Applied Quantitative Methods 12, 25–33.
Chakravarty, S.R., DAmbrosio, C., 2006. The measurement of social exclusion. Review of Income and Wealth, 52, 377–398.
Ciommi, M., Gigliarano, C., Emili, A., Taralli, S., Chelli, F., 2017. A new class of composite indicators for measuring well-being at the local level: An application to the equitable and sustainable well-being (bes) of the italian provinces. Ecol. Ind. 76, 281–296. https://doi.org/10.1016/j.ecolind.2016.12.050.
De Muro, P., Mazziotta, M., Pareto, A., 2011. Composite indices of development and poverty: An application to mdgs. Soc. Indic. Res. 104, 1–18. https://doi.org/10.1007/s11205-010-9727-z. URL:https://doi.org/10.1007/s11205-010-9727-z.
Drewnowski, J., 1972. Social indicators and welfare measurement: Remarks on methodology. J. Dev. Stud. 8, 77–90.
Efron, B., 1987. Better bootstrap confidence intervals. Journal of the American Statistical Association 82, 171–185. URL:http://www.jstor.org/stable/2289144.
ESS, 2013. Handbook on precision requirements and variance estimation for ESS households surveys. eurostat (2013th ed.).
Gan, X., Fernandez, I.C., Guo, J., Wilson, M., Zhao, Y., Zhou, B., Wu, J., 2017. When to use what: Methods for weighting and aggregating sustainability indicators. Ecol. Ind. 81, 491–502.
Grace, J.B., Bollen, K.A., 2008. Representing general theoretical concepts in structural equation models: The role of composite variables. Environ. Ecol. Stat. 15, 191–213.
Istat, 2017. Aspetti della vita quotidiana: Aspetti metodologici dell'indagine. Technical Report Istat.
Klugman, J., Rodriguez, F., Choi, H.J., 2011. The hdi 2010: new controversies, old critiques. J. Econ. Inequality 9, 249–288.
Lebart, L., Morineau, A., Warwick, K., 1984. Multivariate descriptive statistical analysis. Wiley Series in Survey Methodology. Wiley, New York.
Mauro, V., Biggeri, L., Maggino, F., 2018. Measuring and monitoring poverty and well-being: A new approach for the synthesis of multidimensionality. Social Indicator Res. 135, 75–89.
Mazziotta, M., Pareto, A., 2013. A non-compensatory composite index for measuring well-being over time. Cogito, Multidisciplanary Res. J. 5, 93–104.
Mazziotta, M., Pareto, A., 2016. On a generalized non-compensatory composite index for measuring socio-economic phenomena. Soc. Indic. Res. 127, 983–1003.
McGranahan, D., Richard-Proust, C., Sovani, N., Subramanian, M., 1972. Contents and measurement of socioeconomic development. Technical Report A Staff Study of the United Nations Research Institute for Social Development (UNRISD).
Moldan, B., Billharz, S., Matravers, R., 1997. Sustainability indicators: Report of the project on indicators of sustainable development. Wiley Series in Survey Methodology. Wiley, Chichester and New York.
Nardo, M., Saisana, M., Tarantola, S., Hoffman, A., Giovannini, E., 1972. Handbook on Constructing Composite Indicators: Methodology and User Guide. Technical Report

EC Joint Research Centre and OECD Statistics Directorate and the Directorate for Science. Technology and Industry.
Nicholas, A., Ray, R., Sinha, K., 2017. Differentiating between dimensionality and duration in multidimensional measures of poverty: Methodology with an application to china. Rev. Income Wealth 60, 48–74.
OECD, 2008. Handbook on Constructing Composite Indicators. Methodology and user guide. OECD Publications, Paris (2008th ed.).
Pollesch, N., Dale, V., 2016. Normalization in sustainability assessment: Methods and implications. Ecol. Econ. 130, 195–208. https://doi.org/10.1016/j.ecolecon.2016.06.018. URL: http://www.sciencedirect.com/science/article/pii/S0921800915305899.
Rao, J., Molina, I., 2015. Small Area Estimation. Wiley Series in Survey Methodology. Wiley. URL:https://books.google.it/books?id=i1B_BwAAQBAJ.
Saaty, R., 1972. The analytic hierarchy process – what it is and how it is used. Math. Modelling 9, 161–176.
Sachs, J., Schmidt-Traub, G., Kroll, C., Lafortune, G., Fuller, G., 2018. SDG Index and Dashboards Report 2018. Technical Report New York: Bertelsmann Stiftung and Sustainable Development Solutions Network.
Schmidt-Traub, G., Kroll, C., Teskoz, K., Durand-Delacre, D., Sachs, J., 2017. National baselines for the sustainable development goals assessed in the sdg index and dashboards. Nat. Geosci. 10, 547–555.
Sen, A.K., 1999. Development as freedom. Oxford University Press, Oxford.
UNDP, 2010. Human Development Report 2010. The Real Wealth of Nations Pathways to Human Development. UNDP. New York: Palgrave Macmillan (2010th ed.).
Yang, L., 2014. An inventory of composite measures of human progress. Occasional Paper on Methodology. United Nations Development Programme Human Development Report Office (2014th ed.).