

Highlighting Methodological Limitations in the Steps of Composite Indicators Construction

Issaka Dialga¹ · Le Thi Hang Giang¹

Accepted: 8 February 2016 / Published online: 13 February 2016
© Springer Science+Business Media Dordrecht 2016

Abstract The paper opens the debate on the need to find a stable methodological framework in the construction of composite indicators (CIs) in order to address the methodological challenges including those of sensitivity and uncertainties related to methods used. As CIs are well-known to be essential in public debate, their methodological construction must be known by a large public. Illustrating CIs' construction steps by a simple indicator, the paper aims to “democratize” this disciplinary field which is still a black box for some researchers but also to show how composite scores are sensitive to methods used and then, its impacts on policies. For example, in the Sustainable Development Indicator case, the geometric aggregation system is favorable to emerging countries which lead the ranking table whereas high income countries (which are leaders in the linear and equal weight system) except Australia, are misclassified. Uncertainty and sensitivity analysis confirm these results showing that the indexes' scores seem to be influenced by the orientation (implied theoretical framework) given by its sponsors including policy makers. Regarding the validity of the index, correlation tests with some lights and well known indicators, reveal very consistent results.

Keywords Composite indexes · Methodology · Sustainable development · Public policies

✉ Issaka Dialga
issaka.dialga@univ-nantes.fr

Le Thi Hang Giang
thi-hang-giang.le@univ-nantes.fr

¹ Laboratory of Economics and Management of Nantes-Atlantique (LEMNA), Section Economics, University of Nantes, Chemin de la Censive du Tertre, 44322 Nantes Cedex 3, France

1 Introduction

The necessity to find better alternatives to GDP (see Ruta et al. 2005; Stiglitz et al. (2009); Thiry 2010; Dialga 2015) has given way to a flourishing range of composite indicators (CIs) in the nineties.¹ Some studies including those of Stiglitz et al. (2009) and UNU-IHDP² (2015) however insist that the initiatives in terms of CIs construction must be accompanied by dashboards allowing to take into account some qualitative aspects such as “inclusive wealth”. Looking at initiatives opened in the CIs construction field (see Fig. 1 and Bornand et al. 2011), it is likely that this trend, already perceptible, will become ever greater, the social demand being rather in favor of these multidimensional measures. Today, it is easy to notice that actions, in terms of public policies, are largely dependent on these synthetic tools, at least at two levels. Upstream, they can serve as guiding lights for a policy maker on what he needs to know about social aspirations. Downstream, the same tools are relevant to evaluate performances of these policies.

Nevertheless, these synthetic indicators are not without their critics (see Saisana and Saltelli 2010; Klugman et al. 2011; Chiappini 2012) as most of them focus on methodological aspects in the CIs construction. For Council of Europe (2005) and Chiappini (2012), the choice of certain system weights could be very subjective, with no empirical evidence nor defendable theoretical foundation. These persistent criticisms tend to question the local legitimacy³ of indicators as guidance tools for monitoring public actions. This lack of confidence in figures is even more likely since the profusion of these indicators makes confusion in users’ choices (see Bandura 2008). Which indicators to adopt? For which actions? And to which finalities?

As composite indicators are well-known to be essential in the public debate, their methodological construction must be known by the largest audience. If this is not the case, these synthetic tools would lack visibility and public actions would not be able to convince national or local elected representatives. Yet the composite indicators are one of the best communication materials and supports for pragmatic actions. The role of scientists in this context is to support their development by explicitly expressing reservations about these tools and by disseminating new research results and the evolution of debates in the field (Gadrey 2002). This research paper takes part of this goal.

The OECD⁴ and JRC⁵ Handbook (2008) provides a comprehensive but technical introduction to the construction of CIs.⁶ A reader may find difficult to use this handbook as a simple guide for action. As a complement, the aim of this paper is to provide a brief and accessible synthesis of the different required steps of construction, based on a simple example that will be the leading thread.

More precisely, the paper brings two major contributions in the CIs literature beyond its contribution in methodological issues by presenting exhaustively the construction steps of

¹ See Bandura (2008).

² United Nations University International Human Dimensions Programme.

³ While HPI ranks Nord-pas de Calais as one of poorest regions of France, regional HDI puts it in a higher position than some developed regions of France. See «Programme “Indicateurs 21” région Nord-pas de Calais, sept.2010».

⁴ Organization for Economic Cooperation and Development.

⁵ Joint Research Centre.

⁶ The reader can also refer to Nardo et al. (2005). For technical discussions on the steps to conduct Principal Component Analysis (PCA) and sensitivity and uncertainty analysis, the reader can refer to Saisana and Saltelli (2010).

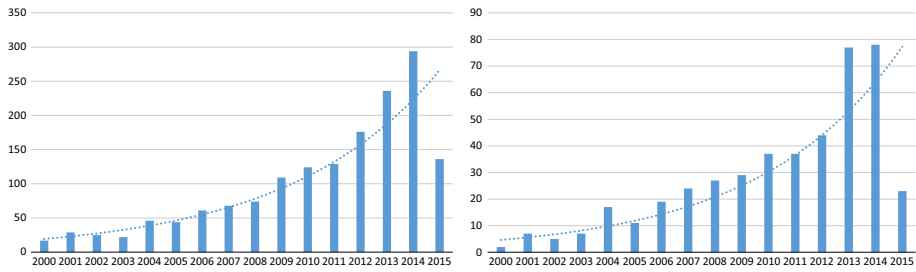


Fig. 1 Evolution of the number of research articles on “Composite Indicators” published in science direct between 2000 and 2015 (18/03) in all journals (*left side*) and in economics, econometrics and finance journals. *Source* Reproduced from Kutin et al. (2015)

a composite indicator. First, on the academic level, the paper launches the debate on the need to have a standard methodological framework in the construction of CIs in order to address the methodological challenges including both sensitivity and uncertainty on composite scores related to the methods used. Moreover, by illustrating CIs construction’s steps by a simple Sustainable Development Indicator (SDI), the paper aims to “democratize” this disciplinary field which is still a black box for some researchers. The aim is to involve more young researchers in this field given the stakes of both methodological and practical issues related to CIs. Secondly, from the perspective of CIs use, the paper highlights the need to make the construction of CIs non-technical. Given their growing use in public debates (housing policy, transport policy, sustainable development policy, social cohesion policy), the appropriation of CIs methods construction by a wider audience, becomes a major democratic challenge.

The paper is organized as follows: Sect. 2 discusses steps of a CI construction. We pay particular attention to the most problematic aspects such as choice of the theoretical framework, selection of variables, normalization, weighting and aggregation, in showing implications they can have in terms of uncertainty and of the credit we can give to the CIs. We illustrate each step (only the most used methods have been described) by constructing a very simple SDI following the scheme initially presented. In Sect. 3, we discuss results from the approaches used and then, analyze their implications in terms of sustainable development grounds. We conduct the indicator’s sensitivity and uncertainty analysis and test the SDI validity by comparing its correlation with some well-known CIs. Section 4 concludes by summarizing the most important points in this illustrated review of CIs construction.

2 Steps of a CI’s Construction

A composite indicator is a mathematical combination of many indicators representing different dimensions of the same concept (OECD and JRC 2008). From this definition, CIs don’t have measurement units. A CI can result from the combination of at most three types of variables. According to Council of Europe (2005) definition, the three types of variables are defined as follows:

- *Quantitative-objective indicators* are quantitative variables that are directly measurable values. Example: per capita income, unemployment rate, emissions of CO₂.

- *Qualitative-objective indicators* are not directly measurable but call for objectively verifiable variables such as presence or absence of a quality norm.
- *Qualitative-subjective indicators* are matters of opinion and appreciation such as satisfaction, trust.

2.1 Definition of the CI and Choice of Variables (Steps 1 and 2)

The definition step is a crucial one, since an indicator can give space to some ambiguities and create dubious or erroneous interpretations (OECD and JRC 2008). The definition of CI should be coherent with the objective and the phenomenon that it aims to represent. Dimensions of the phenomenon should be defined by most relevant variables; the latter are chosen according to criteria that can be objective or subjective (however following a coherent logic). These criteria should meet four requirements to ensure their quality (Council of Europe 2005):

- (1) Representative of the issue they deal with,
- (2) Informative and univocal,
- (3) Allowing a clear and accepted normative interpretation,
- (4) Not excessively onerous.

Thus, a sustainable development index should for example include at least three dimensions: economic, social and environmental ones and these dimensions are themselves broken down into easily identifiable and interpretable variables. In our illustration, only one variable is used for each dimension: Gross National Income (GNI) per capita based on purchasing power parity (PPP constant 2005 \$US), Gini coefficient and per capita carbon dioxide emissions (metric tons per capita). Note that our SDI differs from the existing ones on both theoretically and empirically levels. However, the SDI is not intended to replace the existing indices. For example, the new SDI differs from the Sustainable Human Development Index (HSDI) of Togtokh (2011) and Bravo (2015) in its social dimension. Even if education and health are essential to human well-being, their achievement may be compromised as long as the inequalities are important in the country. Wealth inequality may therefore induce a phenomenon of poverty trap (Dialga 2015) in which only individuals earning a minimum income level can have access to basic services such as education and health. As suggested by Talberth et al. (2006) who proposed to weight negatively income inequality using the Gini Index, and although reducing the social dimension, we use the Gini index to take into account this social dimension. As for the choice of the economic variable, we follow Stiglitz et al. (2009) who suggested that “to measure well-being, the national income is more suitable than GDP”. Finally, the unavailability of environmental data leads us to retain the measurement of carbon emissions. Both economic and environmental variables are reported on population in order to take into account the country size effect.

As shown by the analysis of the correlations summarized in Table 9, our SDI index, although based on three simple variables, provides enough information to be considered as a non-redundant index compared to the usual ones, like HDI.

The refinement into sub-indicators depends on the degree of detail of the information that we would like to provide via the CI. However we are acutely aware of a risk of “information overload”.

Nonetheless, as highlighted above, while seeking a high degree of information detail, one can come to combine theoretically incompatible concepts in one CI and thus not give a

convincing interpretation. Indeed, the need to exhaustively represent one country's wealth can lead to define in the same indicator "stock" variables—to characterize wealth—and "flow" variables like the economic growth. On the other hand, the complexity of certain phenomena makes CIs' constructors simplify variables and only keep relevant and representative ones. The human capital represented by enrolment and literacy rates in HDI illustrates these simplifications of social realities. One bad definition of CI at the beginning has evidently impacts on the other steps of CI construction and in particular, co-linearity analysis, normalization and interpretation of the CI.

In sum, without neglecting other steps, the definition of CI is an important prerequisite for its success since a poorly constructed theoretical framework results in biased and hard to interpret findings and consequently to inadequate policies. However, it does not mean that we should only start with available and easily accessible data to elaborate a CI; the definition of relevant variables ought to guide statistical data mobilization.

2.2 Sources of Data and Imputation of Missing Data (Step 3)

After defining relevant variables according to the theoretical framework, the next step deals with data mobilization. Definition of a CI by identification of its sub-indicators and component variables should allow the determination of data types necessary for the construction of the final indicator. We conventionally distinguish two types of data: primary data and secondary data. Primary data are directly collected via surveys, observations or experiments done by researchers for a specific problem. Secondary data are available before the study is done and can come from statistical institutes, administrative sources or polling organizations.

In practice, needed data are not always fully available. To deal with this difficulty, researchers use many statistical tips. The missing patterns can be of three types depending on their links with the variable of interest on one hand and the other observed variables on the other hand.

They could be "*Missing Completely At Random*", thus "do not depend on the variable of interest (Y) or on any other observed variable (X_i) in the data set or on any other observed variable in the data set" (OECD and JRC 2008). Formally, $X_i \perp (X_j; Y) \quad \forall i \neq j$. In this case, it is possible to omit records from the analysis (*case deletion*) without producing a biased indicator. One example of this treatment is the removal of some countries from HDI ranking when some data are missing. However, this removal reduces the quality of information revealed by the CI, especially when the variable represents an important element. Moreover, it is not possible to make a comparative study between the original sample and the reduced one. In this case, a substitution of the variable for which data are not available could be considered.

Missing data could directly depend on variables of interest (*Non Missing At Random*). Formally, $X = f(Y)$ with $X_i \perp X_j \quad \forall i \neq j$.

Missing data could also be conditional on other variables in the data set but do not depend on variables of interest (*Missing At Random*). $X_i = f(X_j) \quad \forall i \neq j$ and $X_i \perp Y$. In these last two cases, missing data can be imputed with statistical tools (use of central tendency indicators such as means, medians or modes) or econometric ones (such as linear regression). These approximations help to deal with one difficulty but raise another issue regarding the reliability of the CI because of the uncertainties they could produce. Indeed, the imputed values are considered as equivalent to observed data. Yet, one unique imputed value cannot represent the whole uncertainty. Regarding these variables as equivalent to

observed data is an underestimation of this uncertainty; thus tends to reduce the variance of the sample and the confidence interval of the indicator (Donzé 2001). Similarly, Saisana and Saltelli (2010) show that the extent of the consideration of the uncertainty in collected data can lead to a significant variation of the final indicator's value. The quality of the indicator depends strongly on the quality of data used and the latter, in case of imputation, depends on the robustness of mobilized tools.

In our example, raw data come from the World Bank database, World Development Indicators (WDI). We consider for each country the most recent year for which data are available for the three variables. Unfortunately, in order to have a full panel, we were not able to work with a more recent year than 2008. The choice of our sample (high, intermediate and low income countries) is based on World Bank classification according to the level of GNI per capita, whereas the selected countries are done randomly in order to have a representative sample of countries. Obviously, other selection criteria, such as the level of human development (based on HDI), would lead to different choices as shown in Table 9. According to our criteria, the set of selected countries is composed by five high income countries (Australia, Germany, France, USA & Qatar); five intermediate income countries (Brazil, Russia, India, China, Bulgaria) and five low income countries (Algeria, Burkina Faso, Burundi, Cambodia, Vietnam).

2.3 Multivariate Analysis (Step 4)

The multivariate analysis aims to analyze the general structure of data in order to find an eventual correlation between sub-indicators (in the case of SDI, it relates to relations between used variables). The advantage of this analysis is that it allows early identification of inconsistencies in the indicator's formulation and corrects it when it is needed—for example with the inverse weighting of correlated sub-indicators. Indeed, if the analysis reveals a negative correlation between two sub-indicators, both of them should not be components of the final indicator since their effects will neutralize each other and thus constitute a bias in some aggregation functions, such as arithmetic mean. Different weightings should be made if these indicators represent important and district criteria. In practice, variables can be correlated with each other (see Table 1) and not considering the endogeneity of these variables could result in biased estimators. In the case of correlation between variables, Principal Component Analysis (PCA) gives weights allowing for the taking into account these interactions between variables. Weights are determined following three steps.

In the first step, we verify that correlations exist between variables;

At the second step, we select components called factors that explain the most the variance of the sample. PCA proceeds to a linear combination of all variables related to each other. Principal components are identified, if the next three conditions are met:

Table 1 Correlation matrix between variables of SDI and pairwise correlation test

	GNI per capita	Gini index	CO ₂ per capita
GNI per capita	1.0000		
Gini index	-0.0023	1.0000	
CO ₂ per capita	-0.9086***	0.1110	1.0000

*** Correlation coefficients significant at the 1 % level

- (1) The eigenvalue associated with the variable should be ≥ 1 ;
- (2) The individual contribution of the variable to the total variance should be $\geq 10\%$;
- (3) The cumulative variance of the variables in a decreasing order should be $\geq 60\%$.

The third step consist in obtaining weights from a rotation matrix which gives coefficients related to interactions between variables called loading factors (OECD and JRC 2008). With components chosen from the second step, weights are calculated by dividing the square of loading factors by the respective variance of each component.

Visibly, Table 1 shows negative correlations between the GNI per capita and CO₂ per capita and Gini index. However, because of the normalization method used below i.e. $I = \frac{Value_{max} - Value_{country}}{Value_{max} - Value_{min}}$, these negative correlations must be interpreted as positive coefficients. In other words, as GNI and CO₂ emissions are positively linked (seen as negative in terms of sustainable development-SD), the complementary value of CO₂ given by the normalization method is negatively correlated to GNI. One has also to note that correlations are weak between variables. The correlation coefficients between GNI per capita and Gini index is $< 1\%$, those between CO₂ emissions per capita and Gini index is somewhat more than 10% , whereas correlation between CO₂ emissions and GNI per capita is much greater at more than 90% , meaning that industrialized and emergent countries emit a much larger quantity of CO₂ because of the importance of their total production. Togtokh (2011) also highlighted these weak correlations between economic and social dimensions of SD whereas emissions are positively and strongly correlated with income.

Strictly speaking, the Gini coefficient is not representative of the social dimension, as it is weakly correlated with both the GNI and the CO₂ per capita. It should be replaced by a more relevant variable. But as said above, the challenge is not to have an ideal SDI.

2.4 Normalization of Data (Step 5)

This step aims at unifying different measurement units when data for all variables can have a common or equivalent measurement. Depending on the indicator’s type—warning indicator (existence of a critical level for a given phenomenon) or indicator for comparing performances (international indicators), different methods exist and suggest reference scales. One could cite *Ranking*, *Standardization* (or z-scores), *Denominator-Based Weight*. In this article, the two most used approaches are presented namely *Min–Max* and *Benchmark scale-ratio*.

2.4.1 Min–Max

In practice, it is the most used method especially to normalize international indicators such as HDI. Algebraically, $SI_{ij}^t = \frac{I^t - \min(i')(I^t)}{\max(I)(I^t) - \min(i')(I^t)}$ where $\min(i')(I^t)$ is the weakest score performed by one of the entities. Entity i' could be different from I , which means that the weakest score could belong to one entity other than the one for which the indicator j is normalised (i). t denotes year; $\max(I)(I^t)$ is the highest score performed by one of the entities. I could be different from i and should be different from i' except when all entities are both best and worst. By definition, the then normalized sub-indicator ranges from 0 to 1 and rankings of all entities are made with reference to relative positions of the indicator in this range. The min–max method is very sensitive to extreme values.

2.4.2 Benchmark Scale-Ratio

This method associates scores with performances made in a field with reference to a threshold more or less arbitrarily chosen. This threshold could be the performance of the reference country at the initial year: $SI'_{ij} = \frac{I'_{ij}}{I'_{ij=J}^0}$. Two other approaches are also used: the threshold could be $I'_{ij=J}$, i.e. performance of the reference country at the current year; or it could be I'_{ij} , i.e. performance of the considered country at the initial year.

The normalized indexes of SDI are summarized in Table 2. Major clarifications have to be made in the normalization of sub-indicators of SDI. The indexes corresponding to the “social” and “environmental” dimensions are “warning indicators”, which means that the SDI’s score is improved when the values of variables decline (Gini index and CO₂ emissions per capita). In other words, the warning indicators refer to indicators built to warn of the existence of a threshold for a given phenomenon. The existence of these types of indicators allows policy makers to take action at the right time to avoid exceeding critical thresholds. Thus, the normalization formula in this case is: $I = \frac{Value_{max} - Value_{country}}{Value_{max} - Value_{min}}$.

The index corresponding to the “economic” dimension is a “prosperity indicator”. A “prosperity indicator” is an indicator for which its growth improves the composite indicator positively. Example: the level of national revenue (Gross National Income) is a prosperity indicator for a Sustainable Development Indicator (SDI) or for Human Development Index (HDI). So the normalization method respects the traditional formula: $I = \frac{Value_{country} - Value_{min}}{Value_{max} - Value_{min}}$. Both “warning” and “prosperity” indicators are named by Areal and Riesgo (2015) as “less is better” indicators and “more is better” indicators respectively.

Considering the “economic” dimension, Qatar has the best performance (1.00) whereas the persistent poverty in Burundi is reflected by a zero score for this country. Also, most of the countries in the sample have a lower than 0.50, even industrialized economies such as Australia and France. Next to Qatar, only the USA and Germany manage to get a score bigger than 0.50.

Table 2 Normalization by the Min–Max method

Country	GNI per capita	Gini index	CO ₂ per capita
Australia	0.480	0.740	0.622
Algeria	0.100	0.734	0.935
Germany	0.503	0.996	0.805
Brazil	0.130	0.000	0.959
Bulgaria	0.172	1.000	0.865
Burkina Faso	0.008	0.576	0.998
Burundi	0.000	0.811	1.000
Cambodia	0.019	0.641	0.994
China	0.077	0.468	0.892
France	0.463	0.831	0.881
India	0.031	0.807	0.971
Qatar	1.000	0.520	0.000
Russia	0.262	0.476	0.755
USA	0.635	0.531	0.634
Vietnam	0.030	0.725	0.970

Source: Authors’ calculations

For the “social” aspect, there is no “best” nor “worst” performance thanks to which we could evaluate the other countries when we refer to the Gini coefficient (social policies are different from one country to another). Nonetheless, it must be noted that some countries get better scores than others and that in this sample, Bulgaria tends toward a more egalitarian distribution of income than the rest of the sample while Brazil stays quite inequalitarian. The developed countries such as Germany and France, two pillars of European Union, as well as Australia, get high scores in this field, probably thanks to the effectiveness of their social protection policies.

With regards to the “environmental” aspect, regularly highlighted in discussions related to sustainable development, it is interesting to note that Qatar, the leader in economic matters, gets the weakest score for environmental issues, whereas countries with the most limited production capacities and thus low CO₂ emissions, have high scores (Burkina Faso, Burundi). In the group of high income countries, there is one notable distinction: while European ones manage to get good scores, USA and Australia are only better than Qatar. Also, the gap between Qatar and the other 14 countries in terms of CO₂ emissions is very large, since none of the latter has a score <0.50. These results are not surprising since, as highlighted by Table 1, a high correlation is found between CO₂ emission per capita and GNI per capita.

It would be more interesting to compare the results of the two main used normalization methods namely min–max and scale-ratio normalization. Unfortunately, the second one is not adapted to our topic because it requires the need of a benchmark.⁷ If it is reasonable to consider the 1992 pollution level (the 1st Earth Summit) as a reference in the environmental variable standardization, the choice of a baseline for economic and social dimensions is subject to debate among researchers (see Klugman et al. 2011). What baseline to choose for all countries in the study? This choice is it legitimate and accepted by all? At the individual level, each country can set its reference level according its development priorities.

2.5 Weighting and Aggregation (Steps 6 and 7)

These two steps are closely linked and difficult to dissociate in practice because the chosen weighting method implicitly imposes the aggregation method. Nevertheless, some methods allow to explicitly distinguish these two steps.

2.5.1 Budget Allocation Process

This method consists in asking each expert (or stakeholder) to allocate a budget of an X amount between different fields of a phenomenon. The mean of allocated scores allows calculating weights of the indicators and the composite score is their weighted sum. Although the optimal allocation of this budget coming from experts in the field and so gives to the CI a professional legitimacy, choices strongly depend on the perception of the phenomenon by the experts. So, this method tends to be founded on an implicit subjectivity, the risk being that the expert opinions could differ from both the target audience’s opinion and reality which will be likely to occur if the number of experts is not sufficiently large and representative. In such a case, this too limited number of experts can produced biased weights. Nevertheless, even if the list of experts is large, it is advised to verify the

⁷ Nevertheless, we introduced both methods in Sect. 2.4 in order to highlight the difference between them and, therefore, the risks of uncertainty induced.

logic of the value judgment of the expert or any other stakeholder by calculating a coherence index⁸ of value judgments (Saaty 1987; Saaty 1990). When the value of this index is greater than 10 %, then there is an incoherence in the value judgment and thus in the budget allocation of the player which has to be identified and corrected.

Furthermore, when the phenomenon is multidimensional and the budget has to be allocated between these dimensions, this method could give weights make no sense—taking into account a bigger number of variables in the construction of a CI doesn't necessarily lead to a high quality indicator which is representative of the phenomenon—the reasonable number of sub-indicators has to be around twelve (Nardo et al. 2005).

To illustrate this method, we asked 21 experts⁹ to allocate 100 points between the three dimensions of sustainable development.

Regarding budget allocation done by the 21 experts, Table 3 shows that the three variables chosen are all crucial in Sustainable Development (SD) issues; no dimension has received zero. The minimum weight is given to social dimension (0.1) whereas economic dimension received the maximum weight (0.6); the environmental dimension is an intermediary position. However, on average, experts give more importance to social issues (0.355), followed by environmental issues (0.335). These results are well distributed to the extent that the differences between the average values and median values are negligible. We can therefore conclude that globally, experts have converging views on issues of sustainable development.

2.5.2 Maximization of Scores

This method is directly derived from the Benefit of the doubt (BOD) method, itself an application of the DEA (*Data Envelopment Analysis*) approach (OECD and JRC 2008; Blancard and Hoarau 2013). The DEA approach consists in constructing from best performances an efficiency frontier and then, determines other participants' performances relative to this. Thus, BOD gives a relative weight to an individual i considering best performances. By associating score 1 to the best performance, the least effective individuals' scores will logically be inferior to 1 but still be positive according to this formula: $w_i = \frac{\text{Performance of } i}{\text{Benchmark}} \leq 1$. Defined that way, the relative weight of each individual i depends on their performance compared with the "ideal" situation—the benchmark.

By definition, aggregation by this method maximizes scores given by BOD while probability constraint is respected, i.e. non negativity of scores or sub-indicators weights, the sum of weighted sub-indicators should be inferior or equal to 1 ($\sum_{j=1}^J I_j w_j \leq 1$). For a set of sub-indicators, this method gives the benefit of the doubt to the individual whose global performance might be evaluated by only keeping dimensions for which they are most effective (OECD and JRC 2008; Blancard and Hoarau 2013). The underlying idea is that each country has political priorities and seeks to maximize its actions in dimensions judged essential. Formally, CI results from the following maximization program:

⁸ $I = x \frac{w_j}{w_j'}$; x being the ratio of the budget sums allocated to sub-indicator SI_j and SI_j' ; w_j and w_j' relative weights of sub-indicators SI_j and SI_j' obtained from the allocation of budget X .

⁹ We asked 21 researchers from the University of Nantes and professionals working on issues of sustainable development to allocate 100 points to the three main dimensions of sustainable development namely economy, social and environment. The 21 experts were randomly chosen by emailing.

Table 3 BAP weighting results from 21 experts

	Economic	Social	Environmental
Min	0.200	0.100	0.150
Mean	0.310	0.355	0.335
median	0.300	0.350	0.333
Mode	0.200	0.400	0.300
Max	0.600	0.500	0.500

Full details on BAP weighting and experts' list are given in appendix in Table 10

Source: Authors' calculations

$$CI_i = \arg \max_{w_{ij}} \sum_{j=1}^J w_{ij}SI_{ij}$$

$$sc. \begin{cases} 0 < \sum_{j=1}^J w_{ij}SI_{ij} \leq 1 \\ w_{ij} > 0 \end{cases}$$

with i the individual or country ($i = 1, \dots, I$); j the index of sub-indicator (SI) representing one dimension ($j = 1, \dots, J$) and w_{ij} relative weight associated with sub-indicator j in the CI of individual i .

Just as other methods, this one can present weaknesses when dimensions are not substitutable. For example if they are of equal importance (case of sustainable development) or complementary, the maximization method only keeps scores for dimensions for which the individual makes the most efforts and thus does not treat the phenomenon in its totality. There might be from this moment a risk of imbalance in the phenomenon apprehension. The second limit is inherent to the weighting method. Indeed, by limiting scores in the range $[0,1]$, this method does not allow one to observe the worst performances and the individuals who perform outside of the predefined efficiency frontiers are confined to this pre-established range. Some individuals might present poor scores lower than the imposed limit inferior. However, these scores could be a warning when analyzing the global performance of the entity. By restricting the weights of the sub-indicators to 1, the method excludes the possibility that better situations exist. Entities located on this frontier do not have incentives to improve their performances for the latter are considered references to which other entities should tend. Moreover, this method becomes very complex, inoperable even, when it comes to representing this possibility frontier from a large number of sub-indicators. The same difficulty occurs when there are among entities participating in the study "many best performers" for one dimension. In principle, the method of BOD is inappropriate in the case of the SDI as it makes the implicit assumption that the dimensions of the measured phenomenon are perfectly substitutable. In other words, if we use the method of BOD, we assume implicitly that countries can choose to pollute more to further develop their economy, for example. If perfect substitution is not allowed, one should discard the BOD method.

All things considered, BOD presents two essential advantages. Firstly, the BOD gives to the entity an opportunity to be excellent in at least one of many dimensions of the studied phenomenon, generally where the entity focuses its efforts. Secondly, the weights of the sub-indicators are endogenously determined, thus keep the CI free from any criticism about subjectivity in the weighting process.

In order to better illustrate the overall view of these weighting methods, we present in the same table (see Table 4) weighting results from BAP, PCA and BOD approaches. Before commenting them, let examine some methodological points:

Table 4 The weights of the dimensions according to the BAP, PCA and BOD methods

Country	BAP			PCA			BOD		
	GNI/capita	GINI	CO ₂ /capita	GNI/capita	GINI	CO ₂ /capita	GNI/capita	GINI	CO ₂ /capita
	Australia	0.310	0.355	0.335	0.493	0.008	0.500	0.963	0.070
Algeria	0.310	0.355	0.335	0.493	0.008	0.500	0.257	0.001	0.999
Germany	0.310	0.355	0.335	0.493	0.008	0.500	0.648	0.677	0.000
Brazil	0.310	0.355	0.335	0.493	0.008	0.500	0.258	0.000	1.000
Bulgaria	0.310	0.355	0.335	0.493	0.008	0.500	0.000	1.000	0.000
Burkina Faso	0.310	0.355	0.335	0.493	0.008	0.500	0.235	0.000	1.000
Burundi	0.310	0.355	0.335	0.493	0.008	0.500	0.000	0.452	0.634
Cambodia	0.310	0.355	0.335	0.493	0.008	0.500	0.257	0.001	0.999
China	0.310	0.355	0.335	0.493	0.008	0.500	0.258	0.000	1.000
France	0.310	0.355	0.335	0.493	0.008	0.500	0.939	0.047	0.597
India	0.310	0.355	0.335	0.493	0.008	0.500	0.181	0.303	0.754
Qatar	0.310	0.355	0.335	0.493	0.008	0.500	1.000	0.000	0.000
Russia	0.310	0.355	0.335	0.493	0.008	0.500	0.258	0.000	1.000
USA	0.310	0.355	0.335	0.493	0.008	0.500	0.939	0.047	0.597
Vietnam	0.310	0.355	0.335	0.493	0.008	0.500	0.257	0.001	0.999

Source: Authors' calculations

Weights in BAP are the average allocation of 100 points between the three dimensions of SD by the experts. PCA considers the weight of each variable as its relative contribution to the variations of the total variance of the CI. In BOD, a small difference in the way of computing weights exists between the “economic” dimension and the other two dimensions. For “social” and “environmental”, the lower the value of Gini index or CO₂ emissions of a country relative to the rest, the better their situations relative to these dimensions and the relative weight is equal to the value of the benchmark countries divided by the value of the entity. The benchmark countries have a weight equal to 1 for this dimension. The best economic performances are rewarded. Relative weights are determined by reporting the GNI per capita of the country in question to the best value of this variable in the sample. To give meaning to comparisons, the weights are normalized to 1.

The BAP method results come from Table 3. Weights from this method are very different from PCA ones. With the PCA method, the relative weight of the “economic” dimension is 0.493, the “environmental” one is 0.500 whereas the “social” one is 0.008. The low weight combined with the social dimension is not surprising given the weak correlation between the Gini coefficient and the other variables identified in Table 1. At the same time, BOD results are analyzed dimension by dimension. In each one, the best performers are respectively Qatar (leader in the “economic” dimension), Bulgaria in “social” matters and Brazil, Burkina Faso, Burundi, China and Russia in the “environmental” dimension (low emission of CO₂ per capita). We calculated the average weight of each dimension over the whole sample and got the following results (in the same order): 0.430; 0.173; 0.0676. So most countries have a good performance in the social dimension, to which the weights given are beneficial. On the contrary, the environmental dimension is generally given a low weight, for most of countries got poor results in this field relative to best performer.

Although PCA can provide weights that clearly take in account the correlations between the variables, the PCA results should be taken with caution since a robust PCA needs a relatively high number of variables and the correlations between variables must be higher or equal to 0.30. Indeed, even if PCA weighting system allows for “objective weights” which are generated following the endogenous structure of data, the constructor may be in an uncomfortable situation when all variables are not well correlated (Table 1). For the OECD and JRC Handbook (2008), the PCA weighting system is not suitable when variables are uncorrelated.

2.5.3 Indicators Average

It can be an additive, multiplicative or harmonic mean. We present hereby the arithmetical and the geometrical ones. Indeed, the arithmetical mean is the most used weighting-aggregation method in practice, probably because of its simplicity in being understood by a large public and its transparency. However, equal weight which seems “neutral” (as it gives the same importance to different dimensions of the treated phenomenon), can be a source of discriminations. In other words, it is very sensitive to extreme values, which can give biased results when data contain outliers. Furthermore, this method is based on an implicit assumption that a perfect substitutability between different dimensions and that the latter are of equal importance. Thus dimensions for which values are relatively low would be over-estimated. The geometric mean takes into account the lack of perfect substitution between sub-indicators and rewards entities that perform evenly in all fields. We can see how weighting-aggregation methods influence the calculation results of the CI as well as the ranking results by referring to Tables 5, 6 and to Bravo (2015); Areal and Riesgo (2015).

Table 5 Scores and ranking according to different weighting methods and the additive^a aggregation method

Country	EQUAL		BAP		PCA		BOD	
	Score	Rank	Score	Rank	Score	Rank	Score	Rank
Australia	0.614	4	0.620	6	0.553	4	0.859	14
Algeria	0.590	8	0.605	7	0.523	7	0.960	12
Germany	0.768	1	0.779	1	0.658	2	1.000	1
Brazil	0.363	15	0.362	15	0.544	5	0.992	9
Bulgaria	0.679	3	0.698	3	0.525	6	1.000	1
Burkina Faso	0.527	11	0.541	11	0.508	10	1.000	1
Burundi	0.604	5	0.623	4	0.506	12	1.000	1
Cambodia	0.551	10	0.566	10	0.511	8	0.998	8
China	0.479	14	0.489	14	0.488	15	0.912	13
France	0.725	2	0.734	2	0.675	1	1.000	1
India	0.603	6	0.621	5	0.507	11	0.982	10
Qatar	0.507	12	0.494	13	0.497	14	1.000	1
Russia	0.498	13	0.503	12	0.510	9	0.822	15
USA	0.600	7	0.598	8	0.634	3	1.000	1
Vietnam	0.575	9	0.592	9	0.505	13	0.977	11

Source: Authors' calculations

^a The additive aggregation method involves to calculate $\sum_j w_j I_j$ where w_j is the weight of the sub-index j and I_j is the value of the normalized sub-index j

Table 6 Scores and ranking according to different weighting methods and the geometric^a aggregation method

Country	EQUAL		BAP		PCA	
	Score	Rank	Score	Rank	Score	Rank
Australia	0.604	3	0.853	6	0.548	4
Algeria	0.410	7	0.846	7	0.310	7
Germany	0.739	1	0.920	1	0.639	2
Brazil	0.000	13	0.713	15	0.00	13
Bulgaria	0.530	5	0.887	3	0.391	6
Burkina Faso	0.169	12	0.815	11	0.094	12
Burundi	0.000	13	0.854	4	0.000	13
Cambodia	0.230	11	0.827	10	0.141	11
China	0.318	8	0.788	14	0.265	8
France	0.697	2	0.902	2	0.641	1
India	0.289	9	0.853	5	0.177	9
Qatar	NA ^b	NA	NA	NA	NA	NA
Russia	0.455	6	0.795	12	0.446	5
USA	0.598	4	0.842	8	0.633	3
Vietnam	0.275	10	0.839	9	0.173	10

Source: Authors' calculations

^a The geometric aggregation method involves calculating $\prod_j I_j^{w_j}$

^b Not allowed because of aggregation function

With the additive aggregation method, results do not vary much from one weighting method to another except the case of BOD. Germany always occupies the first place, with equal weighting, BAP, or BOD excepted PCA method in where Germany is dethroned by France. France, Bulgaria and Australia always have good scores whereas Russia, China and Brazil are in the last three places regardless of the weighting method used. BOD method provides seven leaders namely Germany, Bulgaria, Burkina Faso, Burundi, France, Qatar and USA.

With the geometric method, we veer BOD method because it is inconsistent with nonlinear programs. The analysis is done considering the three equal BAP and PCA weighting methods. Germany stays the leader in equal and BAP weighting methods whereas Qatar becomes the last of the group. France ranking is also stable even improved in the case of PCA. The emerging and developing countries positions are mixed. As we can see, emerging countries such as China, India Russia are in the intermediate position while developing countries (e.g.: Vietnam, Burundi, Burkina Faso) hold the last places. Table 6 also highlights the inconsistency between some aggregation methods and some types of data. In the present case, Qatar's score calculation is not possible because negative values are not allowed in geometric aggregation method.

3 Discussion of Results and Sensitivity Analysis

In this section we discuss results (ranks and scores) of different combinations of approaches and their implications in terms of sustainable development. We also analyze the sensitivity of the indicator and scores related to changes of methods and then, see its validity by conducting a pair wise correlation test with some known CIs.

3.1 Discussion of the Sustainable Development Index

As a reminder, three dimensions of sustainable development—Economic, Social, Environmental—are measured via three respective corresponding variables—GNI per capita based on purchasing power parity (PPP constant 2005 \$US); Gini index (base 100); Emissions of CO₂ per capita (in metric tons).

The additive aggregation method gave similar results between the first three equal, BAP and PCA weighting methods used except the case of PCA method. Scores of countries in the sample do not vary much from one method to another and ranking is mostly the same (except for Burundi and the USA). Germany is always at the top whereas new economic powers such as Russia, China and Brazil occupy the last places, probably because of the imbalance between efforts spent on economic growth and those used on social and environmental matters. BOD induces some different results, as Qatar and Burundi are rewarded for their respective performances in economic and environmental matters. These figures illustrate how this method promotes aspects where the entity has the most advantage, whereas in sustainable development all aspects are assumed to have equal importance. Also, with BOD, 7 countries share the best score of 1.000, because of the restriction of scores in a [0, 1] range, even if they don't really have the same performance.

The geometric aggregation method does not change ranking results related to equal with the PCA methods. The BAP combined with multiplicative method provide higher scores for all countries. However Qatar, leader in the economic dimension, gets the last place with

both equal and PCA methods. Indeed, its ranking suffers from too much CO₂ emitted per capita, (0 point for the environmental aspect so 0 for the total grade). We observe no compensation between sub-indicators, which is different from the additive aggregation.

Table 7 Sensitivity of the indicator and scores to weighting methods (additive aggregation)

Country	EQUAL-BAP		EQUAL-PCA		EQUAL-BOD	
	Score variation (%)	Rank variation	Score variation (%)	Rank variation	Score variation (%)	Rank variation
Australia	0.9	2	10	0	-34.5	-2
Algeria	2.6	-1	11	13	-2.29	-1
Germany	1.4	0	14	-100	-30.25	0
Brazil	-0.3	0	-50	67	92.17	0
Bulgaria	2.8	0	23	-100	-47.26	2
Burkina Faso	2.7	0	4	9	-13.52	1
Burundi	3.2	-1	16	-140	-65.65	4
Cambodia	2.7	0	7	20	1.08	-2
China	2.1	0	-2	-7	32.99	0
France	1.2	0	7	50	-28.95	-3
India	3.0	-1	16	-83	-15.73	-2
Qatar	-2.4	1	2	-17	-97.41	11
Russia	1.1	-1	-3	31	21.83	0
USA	-0.4	1	-6	57	-28.7	0
Vietnam	2.9	0	12	-44	-2.73	-2

Country	BAP-PCA		BAP-BOD		PCA-BOD	
	Score variation (%)	Rank variation (%)	Score variation (%)	Rank variation (%)	Score variation (%)	Rank variation (%)
Australia	-50	9	-62.4	0	-49	-50
Algeria	-13	10	-66.4	1	-15	-29
Germany	0	14	-53.9	2	-52	50
Brazil	0	-50	-96.5	0	95	-200
Bulgaria	67	22	-49.2	1	-90	83
Burkina Faso	9	2	-63.2	-2	-18	0
Burundi	80	15	-9.7	-3	-97	92
Cambodia	-20	6	-69.0	1	-7	-50
China	0	-3	-75.5	0	34	7
France	-150	6	-65.1	3	-38	-400
India	-33	15	-62.6	2	-38	27
Qatar	92	3	-8.5	-11	-101	93
Russia	0	-3	-73.9	1	24	-44
USA	0	-6	-71.3	4	-22	-133
Vietnam	-22	11	-66.9	1	-17	15

Source: Authors' calculations

Based on these results, industrialized nations seem to be closer of SD objectives than the developing countries. Obviously, the addition of other dimensions, such as intergenerational equity and good governance, will deteriorate the scores of the last countries group.

3.2 Sensitivity and Uncertainty Analysis

From the above results, we conduct tests on the sensitivity of the indicator and of the scores to weighting methods, with the additive aggregation. Table 7 presents cumulative gaps attributable to weighting methods (see “Appendix 2” for theoretical explanations).

Equal and BAP weighting methods give similar results in terms of ranking. Differences are slightly more perceptible when it comes to scores. If the scores’ variations lie around 3 % between equal and BAP methods, they are very extensive in the case of PCA and BOD methods, reaching -400 %. These results highlight how countries scores and their underlying ranks vary following the method used (Table 8).

After the sensitivity test, we also evaluate the uncertainty associated to the SDI. We use additive aggregation results. Since we do not have functional specifications linking different combinations used (normalizations, weightings and aggregations), only the uncertainty related to weighting method changes are evaluated. Monte Carlo simulations could have been the alternative solutions but they require a much larger sampling which is an unsatisfied condition in our case (see “Appendix 1”). Thus we use the next relative uncertainty formula: $\Delta X/\bar{X} = t \cdot \frac{s}{\sqrt{n}}/\bar{X}$ where \bar{X} is SDI score, $t = 3.18$ Student t-value at 3 degrees of freedom ($dof = n - 1$), s standard deviation of the sample (here the standard deviation of SDI obtained with 4 weighting methods) and $n = 4$ number of variables (in SDI, $n = 4$ represents the number of weighting methods used). The absolute uncertainty ΔX measures the maximal error in the evaluation of indicator \bar{X} . The relative uncertainty $\Delta X/\bar{X}$ measures the importance of the maximal error compared with the calculated value of the indicator at a certain degree of confidence (in our application, the degree of confidence is 95 %).

Table 8 Relative uncertainty associated with SDI’s scores

Country	EQUAL Score	BAP Score	PCA Score	BOD Score	Uncertainty (%)
Australia	0.614	0.620	0.553	0.826	31.37
Algeria	0.590	0.605	0.523	0.603	37.37
Germany	0.768	0.779	0.658	1.000	25.63
Brazil	0.363	0.362	0.544	0.028	432.81
Bulgaria	0.679	0.698	0.525	1.000	27.97
Burkina Faso	0.527	0.541	0.508	0.599	39.27
Burundi	0.604	0.623	0.506	1.000	29.63
Cambodia	0.551	0.566	0.511	0.545	40.48
China	0.479	0.489	0.488	0.321	57.64
France	0.725	0.734	0.675	0.935	26.96
India	0.603	0.621	0.507	0.698	34.64
Qatar	0.507	0.494	0.497	1.000	32.03
Russia	0.498	0.503	0.510	0.389	50.47
USA	0.600	0.598	0.634	0.772	32.04
Vietnam	0.575	0.592	0.505	0.591	38.30

Source: Authors’ calculations

The uncertainty associated with the countries scores are around 30 % except that associated with the score of Brazil that reaches 432 %. The results analyzed in the previous section suggest that both BOD and PCA methods are at the origin of important variations in the calculation results for SDI of certain countries especially for Brazil. This illustrates the caution we have to keep in conclusions from analyses of CIs related to different weighting methods. The same principle applies to other approaches in the CI construction, for example, the choice of normalization, aggregation method.

Table 9 Comparing SDI with some well-known indexes and correlation test

Country	SDI (a) Score	Rank	HDI (b) score	Rank	GNI (c) Score	Rank	HSDI (d) score	Rank	EFI (e) Score	Rank
<i>(A)</i>										
Australia	0.614	4	0.938	1	0.862	5	0.884	3	0.461	13
Algeria	0.590	8	0.715	9	0.728	9	0.764	8	0.926	6
Germany	0.768	1	0.92	3	0.867	4	0.904	1	0.656	11
Brazil	0.363	15	0.742	8	0.747	8	0.783	7	0.807	8
Bulgaria	0.679	3	0.776	7	0.759	7	0.812	5	0.749	9
Burkina Faso	0.527	11	0.343	15	0.367	14	0.448	15	0.937	5
Burundi	0.604	5	0.355	14	0.25	15	0.459	14	1.000	1
Cambodia	0.551	10	0.579	13	0.495	13	0.631	13	0.968	3
China	0.479	14	0.715	9	0.706	10	0.746	9	0.881	7
France	0.725	2	0.884	4	0.892	3	0.897	2	0.625	12
India	0.603	6	0.583	12	0.591	11	0.638	12	0.998	2
Qatar	0.507	12	0.85	5	1.000	1	0.647	11	0.000	15
Russia	0.498	13	0.777	6	0.817	6	0.796	6	0.672	10
USA	0.600	7	0.937	2	0.897	2	0.879	4	0.414	14
Vietnam	0.575	9	0.635	11	0.581	12	0.691	10	0.949	4
	SDI		HDI		GNI		HSDI		EFI	
<i>(B)</i>										
SDI	1.000									
HDI	0.271		1.000							
GNI	0.146		0.957***		1.000					
HSDI	0.342		0.929***		0.829***		1.000			
EFI	-0.053		-0.704***		-0.789***		-0.414		1.000	

Panel A: (a) Use linear aggregation and equal weight system

(b) Values from <http://hdr.undp.org/fr/content/table-1-human-development-index-and-its-components>

(c) HDRO calculations based on GNI per capita (2005 PPP International \$, using natural logarithm) expressed as an index using a minimum value of \$100 and a maximum value \$75,000; GNI is then normalized to 1 using min-max method

(d) G. Bravo, The Human Sustainable Development Index: new calculations and a first critical analysis; data and sources available at doi: [10.1016/j.ecolind.2013.10.020](https://doi.org/10.1016/j.ecolind.2013.10.020)

(e) Authors' calculations based on data from http://www.footprintnetwork.org/ar/index.php/GFN/page/footprint_data_and_results/. To give sense to comparison, we normalize Ecological Footprint Index (EFI) values using min-max method and then, take the complementary value to one i.e. 1-EFI normalized to 1

Panel B: *, **, *** correlation coefficients significant at the 10, 5 and 1 % level respectively

3.3 Validity and Robustness Analysis

Comparing SDI to some well-known composite indexes such as HDI, HSDI, GNI and EFI, major changes appear in the countries' ranking. While old indexes reflect the income level of the country, the new SDI produces scores that allow a nuanced reading of development. This is typically the case of USA (see Table 9A). Results also show that countries like Bulgaria, Burkina Faso or Cambodia are more sustainable than the historical great countries like USA and Australia. Emerging countries are not well ranked in the SDI comparing to the other indexes except the case of very strong sustainability index namely EFI (see Table 9A). Regarding only sustainable development indices, Table 9B shows that most of them remain strongly correlated to the income indices except the new SDI and the EFI. According to the SDI, results seem to show that income effects are reduced in the composite score when income is divided by the population size. Relative to the strong sustainability index, results show that the EFI is a partial measure of the concept of SD. As it can be seen, there is no correlation between this index and the SDI or between the latter and the HSDI. In addition, the strong correlation between the EFI and the GNI shows the pressure of economic activity on the biological capacity of the Earth. Note that the negative correlations must be interpreted as positive coefficients because of the normalization method used.

One of the most interesting results is that our SDI is correlated with any of the existing indicators at the statistical 1 % level. The result indicates that the indicator is not redundant in addition to those already built. Despite the adjustments made by Bravo (2015), the results in Table 9 challenge the HSDI as a measure of sustainable development in view of its strong correlation with the GNI and HDI, and then almost identical rankings. The lack of significant correlation between SDI and EFI is explained by the difference of approaches of the theoretical framework of their construction. Indeed, while the ecological footprint is based on a strict and strong vision of sustainable development, the SDI accepts compensation between the three dimensions of sustainable development and integrates within the social aspects absent in the EFI.

4 Conclusion

In this paper, we review the steps of the construction of a composite indicator and illustrate them by constructing a simple sustainable development index-SDI, which includes three main dimensions: economic, social and environmental. Each step is illustrated by the most used approaches in practice. Results are discussed, as well as the sensitivity of the indicator and the scores and uncertainty related to a change in method. Regardless of the method used, the indicator remains subject to uncertainties, making scores and ranking results fluctuate. It also appears that the legitimate need to make the CIs' construction more popular and the need to develop robust tools to reduce uncertainty and sensitivity of composite scores appear as two irreconcilable goals (Aguna and Kovacevic 2010; Blancard and Hoarau 2013; Areal and Riesgo 2015). From this fact and since CIs define, guide and evaluate public action, the choice of a method in the construction of a CI requires a coherent theoretical justification without which the indicator would lack legitimacy. Regarding the validity of the index, correlation tests with some lights and well known indicators, reveal very consistent results. Finally, given its prominent role in the definition of what a country, a territory or a city would be, research in this field should be encouraged in order to develop methods making CIs more robust.

Acknowledgments The authors gratefully acknowledge Thomas Vallée for his help in technical calculations and three anonymous reviewers for their relevant comments in the previous manuscript.

Appendix 1: Uncertainty Analysis

Monte Carlo Method

Considering that every step and different methods of constructing a CI generate uncertainties that have effects on the resulting variable (here the rank attributed to a country by the value of the CI), the uncertainty analysis consists in determining a probabilistic distribution function relying inputs (sub-indicators) to the output (rank) via a random combination of different methods and steps.

Different methods exist to estimate the uncertainty of the resulting variable. Nardo et al. (2005) present the estimation of this uncertainty by the Monte Carlo method in the following way:

The first step is relative to the method used to impute missing data in the CI construction. The authors note $X_i (i = 1, \dots, k)$ the random variable corresponding to different steps and methods. The random variable X_1 characterizes the imputation of missing data and takes two distinct values: 1 when the used method consists in replacing the variable with missing data by another one strongly correlated with it. For example, the variable “Investment” could be replaced by “Savings” and vice versa. X_1 takes the value 2 when the zero value is given to the variable with missing data.

$$X_1 = \begin{cases} 1, & \text{if replacing variable} \\ 2, & \text{if zero value given to missing values} \end{cases}$$

The second random variable is relative to the normalization method of initial variables.

$$X_2 = \begin{cases} 1 & \text{if } I = [I - \min(I)]/\text{benchmark} \\ 2 & \text{if } I = (I - \bar{I})/\sigma \\ 3 & \text{if raw data} \end{cases}$$

The authors also assume that two discrete random variables X_1 and X_2 are evenly distributed on $[0; 1]$. By assuming the random number ζ , $X_1 = 1$ if $\zeta \in [0; 0,5)$ and $X_1 = 2$ if $\zeta \in [0,5; 1]$.

In an analogous way, X_3 is defined as the random variable representing the event “number of sub-indicators isolated for the analysis” knowing that the CI contains J sub-indicators. So:

$$X_3 = \begin{cases} 0 & \text{if } \zeta \in \left[0; \frac{1}{J+1}\right) \text{ all sub-indicators are used in the analysis} \\ 1 & \text{if } \zeta \in \left[\frac{1}{J+1}; \frac{2}{J+1}\right) \\ \dots & \\ J & \text{if } \zeta \in \left[\frac{J}{J+1}; 1\right] \end{cases}$$

$\frac{1}{J+1}$ is the probability that no sub-indicator is excluded from the analysis whereas $1 - \frac{1}{J+1}$ is the probability that at least one sub-indicator is excluded.

The exclusion of a sub-indicator refers to the hypothesis that some sub-indicators cannot be considered in certain methods. For example, when the aggregation method used is the

geometric one, all sub-indicators with negative values should be excluded from this aggregation. Moreover, by excluding the sub-indicator j from the simulation analysis, we isolate its contribution to the CI creation and highlight, all others thing equal, the relative importance of dimension j in the explanation of the phenomenon.

The random variable X_4 is used to capture the uncertainty related to the aggregation method. Three aggregation methods are retained: the linear method (LIN), the geometric one (GEM) and multi-criteria analysis (MCA) discussed in Sect. 4.

$$X_4 = \begin{cases} 1 & \text{if LIN}(IC = \sum w_j * SI_{ij}) \\ 2 & \text{if GEM}(IC = \prod_{j=1}^J (SI_{ij})^{w_j}) \\ 3 & \text{if MCA, CI scores are directly generated by the method} \end{cases}$$

The random variable X_5 is generated to take into account the uncertainty related to the chosen weighting system. Three weighting systems have been retained by the authors: Benefit of the Doubt (BOD), Budget Allocation Process (BAP) and Analytic Hierarchy Process (AHP). The latter is not developed in this paper.

$$X_5 = \begin{cases} 1 & \text{if BAP} \\ 2 & \text{if AHP} \\ 3 & \text{if BOD} \end{cases}$$

The last random variable generated is X_6 . It allows the capture of the uncertainty related to the judgement of the expert especially when there is incoherence in their value judgement such as an illogical allocation of points between different dimensions. X_6 takes values $0, 1, \dots, N$ where N is the number of experts participating in the study. As experts are chosen randomly, each expert selected is associated to the weight that he/she gives dimensions.

However, in the analysis if $X_5 = 3$, $X_6 = 0$ because the weighting method chosen randomly ($X_5 = 3$ corresponds with BOD) does not involve the expert's point of view. The weights in this case are endogenously determined.

Give six random variables generated above, the Monte Carlo analysis consists in defining a probabilistic function combining these six variables. It is then possible to generate N combinations from X_i^l ($i = 1, \dots, k$ with $k = 6$ in our case and $l = 1, 2, \dots, N$) random variables, then analyze the impact of each combination on the value of the final CI or on the induced ranking. The samples X_i^l could be obtained from many randomization methods such as simple random sampling, quasi-random sampling, stratification sampling, etc. (Saltelli et al. 2000). The result variable (CI or rank of country) is related to random variables by the probabilistic density function mentioned above. From an arbitrarily set threshold, it is possible to determine the characteristics of this density function from the number of simulations N obtained. By applying this Monte Carlo method to the Technology Achievement Index, Nardo et al. (2005) find that the ranking of countries varies when all uncertainties related to different steps of the CI construction are taken into account.

Appendix 2: Sensitivity Analysis by Variance Decomposition

This method aims at evaluating the output (CI) robustness since the variance is a measurement of imprecision. The analysis evaluates the contribution of each sub-indicators to the CI total variance and finds the part attributable to interactions between different inputs

(co-linearity, endogeneity, etc.). This decomposition allows the construction of CI sensitivity indexes.

With CI being considered the variable of interest and sub-indicators inputs, the first step of the method consists in specifying a function linking the output (here the CI) to explanatory variables (sub-indicators). When the functional specification linking the output variable Y to input variables X —supposedly independent—is linear ($Y = \beta_0 + \sum_{i=1}^p \beta_i X_i$) a first sensitivity index can be built (Jacques 2011). The index SRC_i (*Standardized Regression Coefficient*) expresses the part of the CI variance imputable to the variance of variable X_i . $SRC_i = \frac{\beta_i V(X_i)}{V(Y)}$ where $\beta_i V(X_i)$ is the variance of X_i .

Nardo et al. (2005) note that given uncertainties related to different levels of CI construction, the functional form of the model cannot be linear nor additive. They support a non-linear model with an undetermined specification. Although these models are not known in advance, they should verify the following properties.

- For n sub-indicators, models make possible an estimation of the total variance explained by these n factors;
- Models allow a sensitivity analysis in which inputs containing uncertainties are considered in groups rather than individually in order to estimate the part of the variance attributable to interactions between variables (cumulative effects of uncertainties, endogeneity biases of variables etc.).
- Variances are quantifiable and allow a decomposition in main variance and in residual variance (variance related to interactions between different explanatory variables);
- Variances are easy to interpret and explain;
- Finally, they allow the discussion of the CI robustness.

Given X_i inputs (sub-indicators), the relative contribution of variable X_i to the total variance of output Y (CI) is given by the variance of the conditional expectations of Y . $V_i = V_i(E(X_{-i}(Y|X_i)))$. For a precise value of $X_i = x_i^*$, it is possible to calculate the conditional mean of Y . In particular, when X_i does not influence the principal variance of Y , $V_i = 0$ and when the output variance is totally explained by this factor X_i , $V_i = V(Y)$, the other factors having no effect on the total variance.

The decomposition of the total variance into main variance and residual variance is given by: $V(Y) = V_{X_i}(E_{X_{-i}}(Y|X_i)) + E_{X_i}(V_{X_{-i}}(Y|X_i))$. Thus, when a factor X is important in the composition of the variance of Y , residual variance $E_{X_i}(V_{X_{-i}}(Y|X_i))$ is small and vice versa.

By dividing the conditional variance by the total variance, we get a first indicator of the sensitivity of CI to X_i : $S_i = \frac{V_{X_i}(E_{X_{-i}}(Y|X_i))}{V(Y)} = \frac{V_i}{V(Y)}$. S_i is the relative contribution of the i th variable to the total variance. When the variable explains the quasi-totality of variations of the output, the sensitivity indicator tends towards 1 ($S_i \rightsquigarrow 1$) as uncertainties and interactions are negligible.

Analogously, it is possible to calculate relative contributions to the total variance.

For two given factors X_j and X_i , the conditional variance in relation to two factors is written: $V_{X_i X_j}(E_{X_{-ij}}(Y|X_i, X_j))$. The residual variance (resulting from the interaction between X_i and X_j) is given by: $V_{ij} = V_{X_i X_j}(E_{X_{-ij}}(Y|X_i, X_j)) - V_{X_i}(E_{X_{-i}}(Y|X_i)) - V_{X_j}(E_{X_{-j}}(Y|X_j))$. It allows us to detect relations between different explanatory variables. In the absence of any interaction between X_i and X_j in the CI construction model, V_{ij} equals zero. In other words, all explanatory variables are independent and no-collinear.

For k explanatory variables independent from one another, the decomposition of the total variance is given by the following formula:

$$V(Y) = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + \sum_i \sum_{j>i} \sum_{l>j} V_{ijl} + \dots + V_{12\dots k}.$$

In the hypothesis of total independence between inputs, the model of decomposition of total variance is the sum of the marginal contributions of each factor.

$$\sum_{i=1}^k V_i = V(Y) \quad \text{and} \quad \sum_{i=1}^k S_i = 1.$$

From the decomposition into residual variances, it is possible to calculate the indexes of sensitivity related to interactions between the X_i explanatory variables. However, Nardo et al. (2005) show that the number of these indexes n gets larger with the number of variables k ; $n = 2^k - 1$. In practice we would rather calculate a condensed index of interactions between variables. It gives the marginal effect of factor i in explaining the total variance of Y , given the effects attributable to interactions between other variables i.e. residual variances of $k - i$ factors.

For a CI with three factors, the marginal sensitivity index is: $S_{T1} = \frac{V(Y) - V_{X_2, X_3}(E_{X_1}(Y|X_2, X_3))}{V(Y)} = S_1 + S_{12} + S_{13} + S_{123}$. S_{T1} is the ration between the sum of variances in which the indicator 1 intervenes individually (S_1) or in interaction with the other indicators (S_{12}, S_{13} et S_{123}) and the total variance of the CI.

The residual index of factor 2 is: $S_{T2} = S_2 + S_{12} + S_{23} + S_{123}$ and the residual index of factor 3 is $S_{T3} = S_3 + S_{13} + S_{23} + S_{123}$.

Homma and Saltelli (1996) show that $V_{X_2, X_3}(E_{X_1}(Y|X_2, X_3))$ could be generalised like this: $V_{X_{-i}}(E_{X_i}(Y|X_{-i}))$. It gives the contribution of $k - i$ variables to the explanation of the total variance.

So

$$S_{Ti} = \frac{V(Y) - V_{X_{-i}}(E_{X_i}(Y|X_{-i}))}{V(Y)} = \frac{E_{X_{-i}}(V_{X_i}(Y|X_{-i}))}{V(Y)} \quad \text{and} \quad \sum_{i=1}^k S_{Ti} \geq 1.$$

Finally, two sensitivity indicators (S_i and S_{Ti}) allow the appreciation of the degree of global appropriateness of the model and its robustness. In particular, when there is a significant difference between indexes S_i and S_{Ti} , this result shows the effects of endogeneity and multi-collinearity between factors X_i which are to be corrected (by deconstruction, by change of weighting or by substitution of some variables by other ones which are non-collinear). If there is no correction, the final CI would be very biased.

Appendix 3: Budget Allocation Process (BAP) Results

The question asked to experts is: Considering the three main dimensions of sustainable development namely economic, social and environment, you are asked to distribute 100 points among these three dimensions according to the importance you give to each of them knowing that the total points awarded must be equal to 100 (Table 10).

Table 10 BAP weighting results from 21 experts

Expert	Economic	Social	Environment
Expert 1	0.330	0.370	0.300
Expert 2	0.200	0.500	0.300
Expert 3	0.200	0.400	0.400
Expert 4	0.600	0.100	0.300
Expert 5	0.250	0.500	0.250
Expert 6	0.333	0.333	0.333
Expert 7	0.333	0.333	0.333
Expert 8	0.300	0.200	0.500
Expert 9	0.250	0.500	0.250
Expert 10	0.200	0.400	0.400
Expert 11	0.350	0.400	0.250
Expert 12	0.333	0.333	0.333
Expert 13	0.200	0.400	0.400
Expert 14	0.200	0.300	0.500
Expert 15	0.250	0.450	0.300
Expert 16	0.500	0.300	0.200
Expert 17	0.330	0.330	0.340
Expert 18	0.200	0.400	0.400
Expert 19	0.400	0.300	0.300
Expert 20	0.250	0.250	0.500
Expert 21	0.500	0.350	0.150
Min	0.200	0.100	0.150
Mean	0.310	0.355	0.335
Mode	0.200	0.400	0.300
Max	0.600	0.500	0.500
Median	0.300	0.350	0.333

References

- Aguna, C., & Kovacevic, M. (2010). Uncertainty and sensitivity analysis of the human development index. *Human Development Research Paper*, 11.
- Areal, F. J., & Riesgo, L. (2015). Probability functions to build composite indicators: A methodology to measure environmental impacts of genetically modified crops. *Ecological Indicators*, 52, 498–516.
- Bandura, R. (2008). *A survey of composite indices measuring country performance: 2008 update*. New York: United Nations Development Programme, Office of Development Studies (UNDP/ODS Working Paper).
- Blancard, S., & Hoarau, J.-F. (2013). A new sustainable human development indicator for small island developing states: A reappraisal from data envelopment analysis. *Economic Modelling*, 30, 623–635.
- Bornand, T., Caruso, F., Charlier, J., Colicis, O., Guio, A.-C., Juprelle, J., et al. (2011). Développement d'indicateurs complémentaires au PIB. Partie 1: Revue harmonisée d'indicateurs.
- Bravo, G. (2015). The human sustainable development index: The 2014 update. *Ecological Indicators*, 50, 258–259.
- Chiappini, R. (2012). *Les indices composites sont-ils de bonnes mesures de la compétitivité des pays?* hal.archives-ouvertes.
- Council of Europe. (2005). *Concerted development of social cohesion indicators—Methodological guide*. Council of Europe Publishing.
- Dialga, I. (2015). Du boom minier au Burkina Faso, opportunité de développement ou risques de péril pour des générations futures? *Revue Cedres Etudes Sciences Economiques*, 59, 27–47.

- Donzé, L. (2001). *L'imputation des données manquantes, la technique de l'imputation multiple, les conséquences sur l'analyse des données: l'enquête 1999 KOF/ETHZ sur l'innovation*. Ecole polytechnique fédérale de Zurich, Centre de recherches conjoncturelles.
- Gadrey, J. (2002). *De la croissance au développement. A la recherche d'indicateurs*. cippa.paris-sorbonne.
- Homma, T., & Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, 52(1), 1–17.
- Jacques, J. (2011). *Pratique de l'analyse de sensibilité: comment évaluer l'impact des entrées aléatoires sur la sortie d'un modèle mathématique*. Lille: sn.
- Klugman, J., Rodríguez, F., & Choi, H.-J. (2011). The HDI 2010: New controversies, old critiques. *The Journal of Economic Inequality*, 9(2), 249–288.
- Kutin, N., Perraudau, Y., & Vallée, T. (2015). *Sustainable fisheries management index, part 1, methodological proposal*, NUM research series (vol. 3).
- Nardo, M., Saisana, M., Saltelli, A., & Tarantola, S. (2005). Tools for composite indicators building. In *European Commission, EUR 21682 EN*, Institute for the Protection and Security of the Citizen, JRC Ispra, Italy, 131.
- OECD & JRC. (2008). *Handbook on constructing composite indicators: Methodology and user guide*. OECD Publishing.
- Ruta, G., Silva, P., Hamilton, K., Lange, G.-M., Markandya, A., Saeed Ordoubadi, M., et al. (2005). *Where is the wealth of nations? Measuring capital for the 21st century*. 34855. The World Bank.
- Saaty, R. W. (1987). The analytic hierarchy process—what it is and how it is used. *Mathematical Modelling*, 9(3–5), 161–176.
- Saaty, T. L. (1990). How to make a decision: The analytic hierarchy process. *European Journal of Operational Research*, 48(1), 9–26.
- Saisana, M., & Saltelli, A. (2010). *Uncertainty and sensitivity analysis of the 2010 environmental performance index*. OPOCE.
- Saltelli, A., Chan, K., & Scott, M. (2000). *Sensitivity analysis, probability and statistics series*. New York: Wiley.
- Stiglitz, J., Sen, A., & Fitoussi, J.-P. (2009). Report of the commission on the measurement of economic performance and social progress.
- Talberth, J., Cobb, C., & Slattery, N. (2006). *The genuine progress indicator 2006*. Oakland: A Tool for Sustainable Development.
- Thiry, G. (2010). Indicateurs alternatifs au PIB: Au-delà des nombres. L'Épargne nette ajustée en question. *Émulations*, 8, 39–57.
- Togtokh, C. (2011). Time to stop celebrating the polluters. *Nature*, 479(7373), 269.
- UNU-IHDP. (2015). *Inclusive wealth report 2014 measuring progress towards sustainability*. Cambridge: Cambridge University Press.