Social Indicators Research (2019) 145:1125–1160
https://doi.org/10.1007/s11205-019-02125-9

The Importance of Variables in Composite Indices: A Contribution to the Methodology and Application to Development Indices

Martin Schlossarek[1] · **Miroslav Syrovátka**[1] · **Ondřej Vencálek**[2]

Accepted: 19 April 2019 / Published online: 18 May 2019
© Springer Nature B.V. 2019

Abstract
The paper examines the issue of weights and importance in composite indices of development. Building a composite index involves several steps, one of them being the weighting of variables. The nominal weight assigned to a variable often differs from the degree to which the variable affects the scores of the overall index. The newly suggested notion of *importance* is based on the idea that an important indicator, if omitted from the index, causes large changes in countries' results. We propose a method of measuring the importance and apply it to inequality variables in composite indices of development. The results show a low importance for most inequality variables, and for some of them, a large discrepancy between the nominal weights and the importance. We argue that the importance of variables should be considered in the process of index construction. This may imply a modification of the index when there is a large discrepancy between the nominal weights and importance and when the importance of some variables is extremely low. Whether any such modification is justified must be decided within the context of the particular index.

Keywords Composite indicator · Composite index · Importance · Weighting · Development indicator · Inequality

Abbreviations

CDI	Commitment to Development Index
EPI	Environmental Performance Index
HDI	Human Development Index
IAH	Index of Inequality-Adjusted Happiness
IEWB	Index of Economic Well-being
IHDI	Inequality-adjusted Human Development Index
MAI	Measure of absolute importance
MCI	Modified composite index

✉ Martin Schlossarek
martin.schlossarek@upol.cz

[1] Department of Development and Environmental Studies, Palacký University Olomouc, 17. listopadu 12, 771 46 Olomouc, Czech Republic

[2] Department of Mathematical Analysis and Applications of Mathematics, Palacký University Olomouc, 17. listopadu 12, 771 46 Olomouc, Czech Republic

MRI	Measure of relative importance
OCI	Original composite index
PCA	Principal component analysis
OECD	Organisation for Economic Co-operation and Development
SSI/HW	Sustainable Society Index/Human Well-being component
UNDP	United Nations Development Programme
WHI/QL	World Happiness Index/Quality of Life component

1 Introduction

The paper examines the issue of weights in composite indices of development and related phenomena (such as quality of life, happiness, and well-being). Composite indices allow for the aggregating of complex phenomena into a single number as well as the comparing of the results for other objects, which are often countries. Composite indices measuring country performance are increasingly popular (for a survey see Bandura 2008; Yang 2014), but their reliability depends critically on their construction. Building a composite index involves several steps; in each step a constructor must choose a certain approach and each decision made has direct implications on the scores and rankings. It is therefore essential that a constructor is aware of these implications, provides a complete methodology and data, and substantiates all construction steps. One of these steps is the weighting of variables from which the index is composed of.[1] In order to aggregate the variables into an index, each variable is assigned a weight which is often assumed to reflect its *empirical importance* (hereinafter referred to as *importance*) with respect to the phenomenon being measured.

The starting point of our research is the fact that the weights assigned to individual variables (i.e. their nominal weights) generally differ from the degree of how much the variables affect the scores of an overall index. In the paper, we (a) propose a new method of measuring the importance of variables in composite indices and compare it with alternative method introduced by Paruolo et al. (2013), (b) analyze the importance of inequality variables in composite indices of development, and (c) provide recommendations for the construction of composite indices. With this, the paper contributes to the methodology of composite indices.

The role of the nominal weight of indicators is often understood by users as a measure of importance, i.e. how much indicator influences (affects) the results of an overall index. It should be noted that constructors of indices add to the confusion of index users. Indices are often presented with user-friendly schemes that only show variables included in the index and their nominal weights. However, the nominal weights do not provide credible information on the importance of variables if it is not accompanied by other methodological procedures, e.g. the method of data normalization. The communication of nominal weights alone implicitly creates a perception that the nominal weight is a measure of the importance of the variable. This misleading notion of the nominal weight is even more critical for the users of those composite indices where the users are called on to decide about the

[1] Many composite indices have a two-layer hierarchical structure of variables composed of several components (dimensions) which include several indicators. We use the term "indicator" for the lower hierarchical level and "component" for the higher hierarchical level. We call the resulting structure "index," "composite index," and "overall index."

weights of the variables. We argue that for sake of transparency, it is necessary to communicate to the users of indices the difference between the nominal weights and the importance. Additionally, we insist that the information about the empirical importance is crucial for the constructors of indices who may take it into account when deciding about the inclusion of variables into an index and about the final weights of variables.

In this paper, we propose a new method of how to measure the importance. This method can be used for increasing the transparency of composite indices as well as for their construction. In the empirical part, we test the method on five selected development indices with a special focus on their inequality variables.[2] The paper answers three research questions. The theoretical question is (1) "How to measure the importance of indicators and components in composite development indices?" This is followed by the two empirical research questions (2) "What is the importance of inequality variable(s) in composite indices?" and (3) "How do the nominal weights of inequality variables compare with their importance?"

The paper is structured as follows. In Sect. 2, we describe the selected issues of setting weights and analyze the factors that affect the importance of variables. The core of the paper lies in the following three sections. In Sect. 3, we introduce the notion of importance and the method of measuring it. In Sect. 4, we apply this method to five indices of development; more specifically on their inequality variables. In Sect. 5, we discuss the results and draw insights that can be used for the construction of composite indices.

2 Weights and Importance of Composite Indices

Setting weights is perhaps the most frequently challenged issue in the construction of composite indices. The Organisation for Economic Co-operation and Development (OECD 2008) list various methods for setting weights. Weighting methods can be classified into two broad categories. Weights are either determined by statistical models or by expert/participatory assessment. Composite indices are often structured in two levels; namely components (higher level) composed of indicators (lower level). Each component and indicator have a certain weight, i.e. the *nominal* weight assigned explicitly or implicitly in case of equal weights.

Statistical weighting methods are based purely on the characteristics of the data. Decancq and Lugo (2010) distinguish between descriptive models and explanatory models. One of the most common descriptive models, the principal components analysis (PCA), groups indicators according to the degree of correlation in a composite index comprising of the least number of factors (principal components) with the minimum loss of information (i.e. with the highest share of explained variability) contained in the original set of variables. Principal components can also be used to assign the weights of variables (for detailed discussion of PCA see Jolliffe 1986). Another example of a descriptive model is a cluster analysis (see Anderberg 1973). Explanatory models work with the notion that observed variables are related to unobserved latent variables. A well-known representative of explanatory models is factor analysis (see Gorsuch 1974), however, other (more complex) methods based on structural equation modeling (e.g. Multiple Indicator Multiple

[2] The motivation behind the focus on inequality indicators stems from the growing interest in inequality in academia and public discourse (e.g. see Wilkinson and Pickett 2009; Piketty 2014) as well as in global policy (the Sustainable Development Goals include Goal 10 which aims to reduce inequalities).

Causes) are also common (see Bollen 1989; Krishnakumar and Nagar 2008). It should be noted that statistical weighting methods do not assess the importance of variables from a theoretical perspective. Additionally, they do not allow people´s preferences to be considered (Ginsberg et al. 1986; de Kruijk and Rutten 2007). As highlighted by the OECD (2008), statistical methods tend to overcome double counting problem when indicators measure overlapping phenomena.[3]

There is a philosophical dimension in the question of whether high correlation necessarily means double counting. As Saisana et al. (2005, p. 314) put it, "[d]epending on a school of thought, one may see a high correlation between subindicators as something to correct for … [or] consider the existence of correlations as a feature of the problem, not to be corrected for, as correlated subindicators may indeed reflect non-compensational different features of the problem." We can also approach the question from a pragmatic point of view. Let us assume a development index that is constituted from per capita income, income inequality, and headcount index measuring absolute income poverty. Let us also assume that the poverty indicator highly positively correlates with the sub-index aggregated from two former indicators. Perhaps, the poverty indicator may be eliminated—not (only) because of the high correlation, but because the sub-index aggregated from income and inequality is a good way of poverty operationalization. Therefore, the sub-index effectively substitutes the poverty indicator. On the other hand, if it happens that, for example, a combination of health and education highly positively correlates with income, the case for dropping income on the ground of parsimony is much weaker as the two former indicators (as well as the combination of them) measure different phenomena. Therefore, one can argue that because of the theoretical importance of every single indicator out of the three, they may all be retained even if the results empirically overlap.

Participatory weighting methods use people's expertise for setting weights (Saisana and Tarantola 2002). For example, in a budget allocation process, experts are asked to allocate a certain amount of points (usually 100) among the set of variables according to their relative importance. Alternatively, experts may set the weight of each variable on a fixed scale (e.g. from 0 to 10). The final weights are derived from the average weights assigned by these experts. In some cases, the public can be asked to assign weights, but this is only relevant where people at large can serve as "experts" on the issue. This may be relevant, for example, in setting the weights for a quality-of-life index as people may be considered to have "expertise" here, while for setting the weights for a competitiveness index, people's answers may be generally less useful. Expert weights may better reflect the theoretical importance of variables, but their credibility depends on the selection of experts, their involvement in the task, and—as argued by Nord (2013)—their understanding of the process.

Another example of participatory weighting method is the analytic hierarchy process. In this method, people compare pairwise various phenomena. The final weights can be derived, for example, by regression analysis based on obtained data (Salomon et al. 2012). In fact, various alternatives of the exact procedure exist (for a detailed review, see Gan et al. 2017). A particular case of participatory weighting methods occurs when the creators of an index use their own expertise to set the weights. In reality, this is a common method though the degree of participation is limited.

[3] Examples of other data-driven methods are frequency-based approaches and benefit-of-the-doubt approaches described and analyzed by Decancq and Lugo (2010). For a comprehensive discussion and list of literature on statistical weighting methods, see Greco et al. (2018) or Ding et al. (2017).

Decancq and Lugo (2010, p. 4) note a specific problem that arises when setting weights for indices of well-being. They write that "setting the weights in a multidimensional analysis seems to become a problem of choosing between Scylla and Charybdis, between Hume's guillotine and paternalism." The former represents a challenge for methods that are data-driven as it is impossible to justify the weights from the factual distributions as "ought" cannot be derived from "is" (Hume 1740). When the setting of weights relies on value judgment on "good life" (well-being), the obvious question emerges, i.e. whose values count? This approach therefore cannot escape from some form of paternalism as the well-being of some people will be inevitably judged using value judgements which are not their own. Based on these arguments, it may seem that choosing a statistical method is free of value judgement, but this is not completely true. As the OECD (2008, p. 31) put it, "regardless of which method is used, weights are essentially value judgements." By choosing the weighting method, the creator reveals their preference to certain inherent characteristics of the method. As none of the procedure leads to an "objective" weighting, alternative weighting methods may be employed to test the robustness of weights suggested by one method. To increase the credibility and transparency, the choice of the method and the entire procedure of setting weights should be described and properly substantiated.

An interesting clash between theory and practice emerges. Despite several sophisticated weighting methods available, the weights are assigned by their creator(s) in most indices, though in some cases, they are assigned after informal discussion with other experts. Additionally, weights determined by non-statistical methods often result in "equal weights," a statistically surprising phenomenon (OECD 2008; Bandura 2008).[4] As argued by Hagerty and Land (2006), equal weighting minimizes the disagreement among all possible individuals' weights in most situations. Whether the reason is simplicity or justification, in reality most composite indices use equal weights at some level of their hierarchical structures. This is the case with the most well-known development index—the Human Development Index (HDI)—but also with other indices such as the Multidimensional Poverty Index and Commitment to Development Index (CDI).[5]

The methodology of setting weights is often not adequately substantiated and the tradeoffs are not transparent (Chowdhury and Squire 2006; Greco et al. 2017). Ravallion (1997, p. 633) challenged the construction of the HDI that led to an implicit monetary valuation of life that was much higher in rich countries than in poor countries. He writes that "the value judgements underlying these tradeoffs built into the HDI are not made explicit, and they are questionable."[6] Taking a different example, the authors of the Environmental Performance

[4] Equal weights can be applied on various levels in the structure of composite index. For example, if the composite index is composed from three components, with five indicators in each, it is possible to assign equal weights to each of the three components and/or to each of the five indicators. If the components have a different number of indicators, the equal weights of components do necessarily mean different weights of indicators and the equal weights of indicators necessarily imply different weights of components.

[5] For the HDI see United National Development Programme (UNDP 2016), for the Multidimensional Poverty Index see Alkire and Robles (2017), and for the CDI see Käppeli et al. (2017). Interestingly, equal weights in the HDI and CDI were later corroborated by other studies. Chowdhury and Squire (2006) asked a group of experts around the world to set the weights for components of the two indices; the weights derived from experts' opinion were not substantially different from equal weights, especially for the HDI. Nguefack-Tsague et al. (2011) showed that principal component analysis based on the correlation matrix of the HDI components leads to practically the same weights.

[6] It should be noted that the tradeoffs are affected not just by "equal" weights, but also by other methodological choices. While this specific critique applies to the HDI, Ravallion (2011) also made a more general case against composite indices of development.

Index (EPI) published a guide on the methodology of the composite indices illustrated in the EPI where they provide a discussion on the issue of weighting. They describe that the "weights for the EPI were established after considering expert recommendations including perceived quality of data, importance of the indicators and categories for policymaking, and the degree to which the indicators provide direct measurement of environmental performance" (Hsu et al. 2013, p. 62). Yet, the report neither describes the process of expert consultations nor does it explain the final weights of the indicators, although the authors employed several methods for the robustness check of the weighting. We argue that the weighting scheme, i.e. weights assigned to each variable, should be substantiated (especially if they are set by the creator(s) of the index) and the resulting tradeoffs discussed and defended.

Going further, similarly to Decancq and Lugo (2010), Paruolo et al. (2013) and Becker et al. (2017), we argue that distinguishing between the weights and importance is a crucial and neglected issue in the process of the construction of indices. By importance, we mean the degree to which an indicator or component affects the higher hierarchical level (a component or an overall index). Consider the example of the OECD Better Life Index. Users are challenged to rate different topics according to their *importance* and the application automatically recalculates a country's result. However, the users' choices are reflected only in the nominal weights. Without information on other characteristics of the index´s construction and data included, the nominal weights bring limited information about the empirical importance of an indicator.

There are plenty of factors which influence the importance of a variable in the overall index. Among them include various features of the structure of the data and the method of normalization and aggregation. When a change is made in one of these factors, it is difficult (if not impossible) to make general predictions of what will happen with the importance of a variable (i.e. we are unable to establish the general rule which says whether transforming the data of a variable into a more linear shape increases or decreases its importance of within the index). We have identified three key factors which allow for the predictions of what will happen in case the characteristics are changed: (a) the nominal weight, (b) the variability of the original data (i.e. the degree to which data of a variable vary from a mean) and how it is affected by normalization, and (c) the correlation between the variable and the other variables of the overall index.

A variable with a higher nominal weight (e.g. 20%) has a higher importance in the overall index than a variable with a lower nominal weight (e.g. 10%), after controlling for other factors. A variable with a higher variability has a higher importance in the overall index than a variable with a lower variability, after controlling for other factors. However, normalization methods usually modify the variability of the original data and therefore affect the importance of the indicators.[7] The two most common methods of normalization are z-scores and min–max. Z-scores transform the original values to standardized values with a mean 0 and standard deviation 1. For variables with normal distribution, about 95% of standardized values lie within the range from -1.96 to $+1.96$, i.e. approximately two standard deviations below and above the mean. Min-max

[7] The modification of variability is common but not a necessary feature of normalization. For example, the former methodology of the CDI normalized indicators linearly to a set of transformed values with a fixed average, preserving the variability of the original data. The consequence was that indicators (components) with a low variability had a low influence on the scores of the component (index), while highly variable indicators (components) had a large influence (see Syrovátka and Hák 2015).

Table 1 Index Alfa

	Indicator 1 (x_1)	Indicator 2 (x_2)	Indicator 3 (x_3)	Index Alfa
Country A	0.20	0.70	0.80	0.52
Country B	0.20	0.55	0.90	0.48
Country C	0.30	0.80	0.10	0.46

Table 2 Index Beta

	Indicator 1 (y_1)	Indicator 2 (y_2)	Indicator 3 (y_3)	Index Beta
Country A	0.00	0.60	0.88	0.53
Country B	0.00	0.00	1.00	0.40
Country C	1.00	1.00	0.00	0.60

transforms the original values to the fixed range 0–1, with minimum and maximum values often set to actual minimum and maximum of the dataset.

The main reason for normalization is the invariance of the entire process under a change of scale. Changing the scale by multiplying certain indicators by a factor greater than 1 (e.g. using *cm* instead of *m* corresponds to multiplying by 100) increases most of the measures of variability of the indicator such as its standard deviation or range and thereby, if the normalization is omitted, increases the influence of the indicator on the overall index. Normalization avoids this undesirable effect.

The third factor affecting the importance of a variable is its association with other variables included in the index. The level of association is usually measured by the Pearson correlation coefficient ranging from -1 (perfect negative correlation) to $+1$ (perfect positive correlation). The closer the correlation coefficient between the given variable and other variables is to 1, the lower the importance of the given variable (the lower the potential of the variable to affect the score of the composite index). The following illustration with index Alfa and index Beta shows how the weight can differ from the importance in composite indices. First, let us consider a composite index $\text{Alfa} = w_1x_1 + w_2x_2 + w_3x_3$ where x_1, x_2, and x_3 are indicators and w_1, w_2, and w_3 are the nominal weights, where $(w_1, w_2, w_3) = (2, 2, 1)/5$. The values of x_1, x_2, and x_3 for three fictitious countries as well as their respective scores of index Alfa can be seen in Table 1. We can see that country C has the lowest score of index Alfa even though it has the highest score in x_1 and x_2 and has the lowest score only in x_3, which bears the lowest nominal weight (20% in comparison with 40% for each x_1 and x_2).

Second, let us consider a composite index $\text{Beta} = w_1y_1 + w_2y_2 + w_3y_3$ where y_1, y_2, and y_3 are indicators and w_1, w_2, and w_3 are the nominal weights, where $(w_1, w_2, w_3) = (3, 3, 4)/10$. The values of y_1, y_2, and y_3 for three fictitious countries as well as respective scores of index Beta can be seen in Table 2. Be aware that the values of y_1, y_2, and y_3 were calculated according to the formula $y_i = (x_i - \min(x_i))/(\max(x_i) - \min(x_i))$, i.e. values of y_1, y_2, and y_3 are normalized values of x_1, x_2, and x_3 used for the calculation of index Alfa. Therefore, the differences of results obtained by using index Alfa and index Beta are driven by the different weights as well as by normalization.

As we can see in Table 2, even though the nominal weight of y_3 is the highest (40%, while both y_1 and y_2 have only 30%), country C has the highest score of index Beta. The reason behind these counter-intuitive results is the fact that the effect of the higher weight of the third indicator (y_3 and x_3, respectively) in index Beta compared to index Alfa was outweighed by the effects of min-max normalization used before the calculation of index Beta. The example clearly shows that the nominal weights are only very crude approximations of the importance as in this case the third indicator (y_3 and x_3, respectively) influences the results less when having a higher nominal weight (and vice versa).

Recent papers by Paruolo et al. (2013) and Becker et al. (2017) deal with the issue of disparity between the weights and importance. They build on previous studies mostly related to the sensitivity analysis (see Saltelli and Tarantola 2002; Saisana et al. 2005; Xu and Gertner 2008a, b). The methodology of Paruolo et al. (2013) is discussed in Sect. 3.5, their measure of importance is calculated in Sect. 4.2, and the results are discussed in Sect. 5.2. Similarly, we argue that after the initial calculation of the scores of an index, an analysis of the importance of the variables should be performed. However, we show that our notion of importance differs from their definition.

Based on the analysis of importance, the author should reconsider the weighting scheme and other relevant steps of index construction. This is especially relevant for composite indices where users are called on to decide about the weights of indicators. The constructors of these indices should assure that the users are aware of the difference between nominal weights and importance.

3 Methods

In this section, we introduce a new method of measuring the importance of variables in an index. We apply it to the level of indicators within an index (i.e. importance of indicators in an index), but it can be also used on the level of indicators within a component as well as the components within an index. Our concept of importance of a given indicator is based on the correlation between

- the original composite index (OCI) which includes the examined indicator and
- the modified composite index (MCI) which excludes the examined indicator.

The stronger the correlation is between the original and modified composite index, the lower the importance of the examined indicator. The correlation coefficient equal to 1 means that the exclusion of the indicator does not affect (up to a linear transformation) the composite index and therefore has no importance.

3.1 Procedure

We start by describing the procedure of assessing the importance of individual indicators in the composite index. The entire procedure can be divided into three steps: (1) the computation of the original and modified composite indices, (2) the computation of the absolute importance for individual indicators, and (3) the computation of the relative importance for individual indicators. The detailed description of the three steps follows.

Step 1 In the first step, we calculate the original index and p modified versions of the composite index, where p is the number of indicators from which the overall index is composed of. We assume that the original composite index can be expressed as a weighted average of indicators $x_1, \dots, x_p$:

$$OCI = \sum_{i=1}^{p} w_i x_i,$$

where $w_1, \dots, w_p$ denote the nominal the weights of individual indicators. Note that in practice, we use the previous formula for several statistical units (countries) distinguished by index j:

$$OCI_j = \sum_{i=1}^{p} w_i x_{ij}, \quad \text{for } j = 1, \dots, N,$$

where N denotes the number of statistical units. In what follows, we omit the index j for simplicity.

The computation of MCI is based on the same formula, but one of the indicators is omitted each time:

$$MCI_k = \sum_{i \neq k} w_i x_i = OCI - w_k x_k, \quad \text{for } k = 1, \dots, p.$$

The subscript k identifies the excluded indicator. Let us take note that the weights for the computation of the modified composite index do not sum up to one, but to $1 - w_k$. This could be remedied easily by using the weights $w_i / \left(1 - w_k\right)$ in the formula above. However, this transformation makes no difference in what follows and therefore it is omitted to make the process clearer.

Step 2 In the second step, we calculate the correlations between the original composite index and all of its modified versions. We denote these correlations by the letter r indexed by the identifier of indicator:

$$r_k = corr\left(OCI, MCI_k\right), \quad \text{for } k = 1, \dots, p.$$

The correlation coefficients show the degree of influence of each indicator on the composite index. A low or even negative correlation between the original index and the modified index (i.e. the original index where a given indicator was omitted) implies a high influence of the given indicator. To make the interpretation easier, we define a *measure of absolute importance* (MAI) of the k-th indicator as:

$$MAI_k = \left(1 - r_k\right) * 50, \quad \text{for } k = 1, \dots, p. \tag{1}$$

The outcome is expressed on a 0–100 percent scale. MAI = 0% means that the indicator has no influence within an index: without the indicator, ranking, and relative distances[8] between countries (or any other units) remain to be the same. If MAI is between 0 and 50%, the ranking and relative distances tend to be similar for both the original and modified index (i.e. there is a positive correlation between both indices). For MAI = 50%, there is no correlation between the original and modified index, indicating a very strong

[8] By the relative distance, we mean a gap in scores. For example, the gap between countries on the first and the second position in comparison with the gap between the second and the third place according to the overall index.

importance of the examined indicator. MAI between 50 and 100% indicates an extremely strong importance of the indicator. In this case, the results based on a modified index tend to be the opposite of the results based on the original index (i.e. countries being among the best in the original index tend to be among the worst in the modified index and vice versa). For MAI = 100%, the ranking and relative distances for a modified index are inverse to the original index: this is theoretically the highest possible importance that can be assigned to the indicator.

MAI can be used for the comparison of the absolute influence of similar indicators across composite indices. In Sect. 4, we calculate MAI of inequality indicators in several composite indices. If a certain inequality indicator has MAI 4% in the composite index A and the same inequality indicator has MAI 7% in another composite index, the degree to which the inequality indicator is important for the overall index is higher in the latter case.

Step 3 In the third step, we rescale MAI to a *measure of relative importance* (MRI) of indicators in the composite index. We use the following formula:

$$MRI_k = \frac{MAI_k}{\sum_{i=1}^{p} MAI_i} \times 100, \text{ for } k = 1, \dots, p. \tag{2}$$

MRI allows for (a) the comparison of the relative importance of various indicators within one index and (b) the comparison of the relative importance and nominal weights, as both use the same scale (the total sum is always 100 percent). For example, if the nominal weight of an indicator is 20% and MRI is 13%, the indicator has a lower relative importance for the composite index than the nominal weight may suggest.

The absolute importance shows how far the linear association (measured by the Pearson correlation coefficient) between the composite index is with and without the examined indicator from being perfect. Note that if the association is not linear, the Pearson correlation coefficient is not an appropriate measure of association. Instead, the Spearman correlation coefficient measuring the monotone association should be used; a non-monotonic relationship can rarely occur in practice and it may reveal serious conceptual problems in the index construction.

The suggested measure serves as a useful tool for the comparison of the absolute importance of similar indicators across composite indices (in this paper we compare the absolute importance of inequality indicators within five indices of development). However, it is not the most suitable tool for the comparison of the relative importance of various indicators within one index and it does not allow for the comparison of the importance and nominal weights. For this purpose, we rescale the values of the absolute importance so that the sum of all values is 100%.

Theoretically, it can happen that all MAIs are equal to zero. In that case, the ratio in expression (2) is not defined. As this situation occurs when all correlation coefficients are equal to one ($r_k = 1$, for $k = 1, \dots, p$), any of the indicators can be omitted with no effect on the composite index. Therefore, they are all equally (un)important.

3.2 Some Theoretical Considerations

In this section, we discuss some mathematical properties of the proposed measure of importance. Since the measure of the absolute importance is based on the correlation between the original and modified composite index, it is useful to study the properties of this correlation coefficient.

It can be shown that the correlation coefficient r_k can be expressed in terms of

(1) nominal weight w_k;
(2) correlation coefficient $\rho_k = corr(x_k, OCI)$ measuring the correlation between the k-th indicator and the original composite index (note that its value is influenced by the correlation between all indicators included in the composite index);
(3) variability of the k-th indicator and the original composite index; more specifically on the ratio of their standard deviations which is denoted by RV_k here, i.e. $RV_k = SD(x_k)/SD(OCI)$, where SD stands for standard deviation.

Note that these are exactly the three key factors influencing the importance of an indicator listed in Sect. 2.

Using the notation above, one can show that

$$r_k = \frac{1 - w_k\rho_k RV_k}{\sqrt{1 - 2w_k\rho_k RV_k + (w_k RV_k)^2}}. \tag{3}$$

Let us briefly discuss a special case of the uncorrelated indicators $x_1, \dots, x_p$. Although this case is rare in practice, it may be worthwhile to investigate. In the case of uncorrelated indicators, the formula above simplifies to

$$r_k = \sqrt{1 - \frac{\left(w_k SD(x_k)\right)^2}{\sum_{i=1}^{p}\left(w_i SD(x_i)\right)^2}}. \tag{4}$$

From this expression, one should set the nominal weights w_k proportional to the multiplicative inverse (reciprocal) of the standard deviations $SD(x_k)$ to make all (independent) indicators equally important. Therefore, if the data are standardized (in this case $SD(x_1) = \dots = SD(x_p) = 1$) one should use equal weights ($w_1 = \dots = w_p = 1/p$). If the data are not standardized, the highest nominal weight should be assigned to the indicator with the lowest variance. For the derivation of expressions 3 and 4, see the attachment.

3.3 The First Series of Illustrative Examples

In this section we illustrate the good properties of the suggested measures of components' importance in several simple situations.

Let us consider a composite index $\mathrm{I} = \mathrm{w}_1\mathrm{x}_1 + \mathrm{w}_2\mathrm{x}_2 + \mathrm{w}_3\mathrm{x}_3$ where x_1, x_2, and x_3 are mutually independent variables with uniform distributions on the interval [0, 1].

(A) Let us first consider equal weights (equal to 1/3). In this case, all three components are equivalent since they have the same distribution and the same nominal weights.

The correlation coefficients needed for evaluation of components' importance can be computed using the formula (3):

$$\mathrm{r}_1 = \mathrm{r}_2 = \mathrm{r}_3 = \sqrt{1 - \frac{1}{3}} \approx 0.8165.$$

Therefore, the measures of absolute importance are all equal to $(1-0.8165)\times 50 = 9.2$ (%). This means that omitting any of the three components would change the information provided by the composite coefficient only very slightly. Moreover, the equality of MAIs implies equality of relative importance of all three components. Therefore, $MRI_1 = MRI_2 = MRI_3 = 33.3\%$. Indeed, all three components are equally important in this case.

(B) Let us now consider unequal nominal weights, for example (w1, w2, w3) = (1, 2, 3)/6. It seems that the second component is twice more important than the first since its nominal weight is twice as high. The third component seems to be three times more important than the first one.

The correlation coefficients can be again computed using the formula (3). They are not equal any longer. Their values are

$$r_1 = \sqrt{1 - \frac{1}{1+4+9}} \cong 0.964, r_2 = \sqrt{1 - \frac{4}{1+4+9}} \cong 0.845, r_3 = \sqrt{1 - \frac{9}{1+4+9}} \cong 0.598.$$

The measures of absolute importance are then evaluated using the formula (1): $MAI_1 = 1{,}8\%$, $MAI_2 = 7{,}7\%$, $MAI_3 = 20{,}1\%$. Omitting the first components can be done with only negligible change in the composite index. On the other hand, omitting the third component causes a much more dramatic change. Comparing the effect sizes by MRIs, we obtain the following values: $MRI_1 = 6.1\%$, $MRI_2 = 26.1\%$, $MRI_3 = 67.8\%$. Recall that the nominal weights are equal to 16.7%, 33.3%, and 50.0% respectively. The newly suggested measures of importance show bigger differences among the importance of individual components.

(C) Similar results can be obtained if the weights are equal but the (uniform) distributions differ in dispersion. Consider $x_1 \sim Unif(0,1)$, $x_2 \sim Unif(0,2)$, and $x_3 \sim Unif(0,3)$. The standard deviations of these distributions are $1/\sqrt{12}, 2/\sqrt{12}$, and $3/\sqrt{12}$. Using the formula (3), we obtain exactly the same correlation coefficients as in the previous case (B). Consequently, we obtain the same MAIs as well as MRIs. The third component is therefore the most important one. This is due to its higher dispersion (if compared to the other two components). This example illustrates the importance of standardization in the process of an index definition. If the components are standardized before entering the computation of the index, they have the same support—they all have uniform distribution on the interval $(-\sqrt{3}, \sqrt{3})$. Since their nominal weights are equal, their importance is then equal as well.

3.4 The Second Series of Illustrative Examples: The Problem of Nonlinearity

Let us consider a composite index $I = w_1x_1 + w_2x_2 + w_3x_3$ where x_1, x_2, and x_3 are deterministic nonlinear functions of each other, more specifically $x_1 = x$, $x_2 = x^2$, and $x_3 = x^3$. Consider equal nominal weights. Which of the three components is the most influential? The answer depends on the values of x, as will be illustrated further.

(A) Let x range from 1 to 10, i.e. x = 1, 2, 3, …, 9, 10. In this case, the component x^3 plays a crucial role in the composite index, as illustrated in Table 3.

Table 3 Illustrative example A

x_1w_1	x_2w_2	x_3w_3	Index
0.33	0.33	0.33	1.00
0.67	1.33	2.67	4.67
1.00	3.00	9.00	13.00
1.33	5.33	21.33	28.00
1.67	8.33	41.67	51.67
2.00	12.00	72.00	86.00
2.33	16.33	114.33	133.00
2.67	21.33	170.67	194.67
3.00	27.00	243.00	273.00
3.33	33.33	333.33	370.00

Table 4 Illustrative example B

x_1w_1	x_2w_2	x_3w_3	Index
0.003333	0.000033	0.000000	0.003367
0.006667	0.000133	0.000003	0.006803
0.010000	0.000300	0.000009	0.010309
0.013333	0.000533	0.000021	0.013888
0.016667	0.000833	0.000042	0.017542
0.020000	0.001200	0.000072	0.021272
0.023333	0.001633	0.000114	0.025081
0.026667	0.002133	0.000171	0.028971
0.030000	0.002700	0.000243	0.032943
0.033333	0.003333	0.000333	0.037000

All three values of MAI are less than 1% indicating minimal effect to the omission of any component. This is not surprising since the components are functions of each other and therefore only a minute piece of information is lost by omitting one of them. The relative measures of importance correctly reflect the dominance of the third component in the composite index. Their values are $MRI_1 = 0.03\%$, $MRI_2 = 0.\ 76\%$, $MRI_3 = 99.20\%$.

(B) Consider now the situation in which the values of x are close to zero. More specifically, consider $x = 0.01, 0.02, \ldots, 0.10$, i.e. 100 times smaller than in the previous case (A). In this case, the higher powers of x are negligible compared to the values x themselves, see Table 4.

In this case, $MAI_1 = 1.1\%$ while the other two measures of absolute importance are much smaller. Therefore, similarly as in the previous case, no large effect on the composite index can be gained by the omission of any component. The relative measures of importance correctly identify x_1 to be the most influential component: $MRI_1 = 98.90\%$, $MRI_2 = 1.07\%$, $MRI_3 = 0.03\%$.

(C) The previous cases differ only in scale—the values of x in the example B were obtained by dividing the values of x from the example A by 100. The different scales have led to a different importance of the individual components. However, these interpretation

Table 5 Illustrative example C

x_1w_1	x_2w_2	x_3w_3	index
−0.495	−0.366	−0.292	−1.154
−0.385	−0.337	−0.286	−1.007
−0.275	−0.288	−0.267	−0.830
−0.165	−0.219	−0.231	−0.616
−0.055	−0.132	−0.172	−0.359
0.055	−0.024	−0.084	−0.053
0.165	0.102	0.039	0.307
0.275	0.249	0.203	0.727
0.385	0.415	0.414	1.213
0.495	0.600	0.676	1.772

problems can be easily avoided by standardization. It can be easily shown that both previously considered cases have the same standardized version. Table 5 illustrates the creation of the composite index from the standardized components.

The measures of absolute importance are again less than 1%. What is more interesting now is the comparison of the relative importance: $MRI_1 = 54.91\%$, $MRI_2 = 1.34\%$, $MRI_3 = 43.75\%$. This result can be interpreted as a relative smaller importance of the x^2 in presence of x and x^3 regardless of the considered scale. In fact, the first and the third component play the crucial role.

The examples presented in this section show that the suggested methodology can be applied even in situations where the relationship between the composite index and an index obtained by omission of one its component is not linear. In this paper, we assumed the composite index to be a linear combination of its components. However, the suggested measures of importance can be applied even if the composite index is a nonlinear function of its components, e.g. geometrical mean. They are meaningful as far as the relationships between the original composite index and its modifications can be well described by the Pearson's correlation coefficients, i.e. as far as they are linear or at least approximately linear. We show that this assumption is fulfilled in the case of all five indices inspected in Sect. 4 (see Figs. 1, 2, 3, 4 and 5). Figure 5 proves that it can be fulfilled even for nonlinear transformations as a geometric mean. Before applying this method, it is advised to check the form of dependence by visual inspection of the corresponding graphs.

Note that if we used Spearman correlation coefficient instead of Pearson correlation coefficient, all indicators would be equally (un)important, as discussed in Sect. 3.1.

3.5 Comparison to Pearson's Correlation Ratio

Our notion of the importance measured by *absolute importance* and *relative importance* differs from the importance or *main effect* used by Paruolo et al. (2013). They define the importance (or the main effect) of an indicator as "the expected fractional reduction in variance of the composite indicator that would be obtained if that variable could be fixed" (Paruolo et al. 2013, p. 8), i.e. they use the statistics proposed by Pearson (1905), which he called the "correlation ratio" (also known as the *main effect index*, or the *first order sensitivity index*; see Becker and Saltelli 2015).

We assume that the rank of observations (countries) and relative distances between them matter. An indicator has a low importance if it brings little additional information to the combination of all other indicators forming the overall index. If such an indicator is omitted from the index, the rank of countries as well as the relative distances between them remains similar. On the contrary, an indicator with a high importance changes the rank of countries and the relative distances between them to a substantial extent. Paruolo et al. (2013) assume that an indicator has a low importance if the loss of variability after fixing the indicator[9] is minor (and vice versa). Their formula was further developed and adjusted by Becker et al. (2017). However, the main principle—the focus on the loss of variability after fixing certain indicators—remains to be in the center of their approach.

Nevertheless, we do not see any considerable value in observing changes in variability after the fixing of an indicator, especially in the context of dimensionless indices. If the variability drops significantly, but the country positions (the ranks and relative distances) remain similar, such an indicator is not important for the interpretation of the overall score of the country. If the variability drops negligibly but the country positions (the ranks and relative distances) change significantly, such an indicator is important. We note that the scores of all social indicators, especially dimensionless composite indices, are necessarily interpreted in the context of other observations (countries, people). Therefore, changes in the variability do not bear relevance for the interpretation of the result.

4 Results

In this section, we apply suggested methodology on the five composite indices of development that include inequality. We analyze the absolute importance, the relative importance, and the nominal weights of inequality variables. Our aim is to answer two research questions. Q1: What is the degree of importance of inequality variable(s) in composite indices? Q2: How do the nominal weights of inequality variables compare with their relative importance?

We reviewed composite development indices and identified those which include inequality and for which recent data are available. These are: the World Happiness Index (WHI), the Sustainable Society Index (SSI), the Index of Inequality-Adjusted Happiness (IAH), the Index of Economic Well-being (IEWB), and the Inequality-adjusted Human Development Index (IHDI).[10] We included only "overall development" indices, excluding those that focus only on a certain development dimension (e.g. education and health). After a brief introduction of each index, we calculate the measures of absolute and relative importance. In Tables 6, 7, 8, 9, 10 and 11, we present MRI and MAI for the respective indicators and the three key factors affecting the importance of indicators (nominal weight, correlation with other indicators, and variability of the indicator). We also present ranks

[9] Fixing values of a certain indicator decreases the variability of the composite index not just because one source of variability disappears, but also because other sources of variability (other indicators) are in general correlated with this indicator and thus the reduction of variability in one indicator brings at least a partial reduction of variability from other sources.

[10] In the case of the SSI and WHI, inequality is an indicator included in only one component. We investigate the importance of inequality in this component. In the case of the IEWB, inequality itself is one component of the index. In other cases (the IHDI and IAH), we assess the importance of indicators in a composite index.

within each factor: the top rank (1) is assigned to the highest nominal weight, the lowest multiple correlation coefficient, and the highest variance.[11] An indicator which ranks well (closest to 1) in all three factors tends to have high importance, while an indicator which ranks poorly (farthest from 1) tends to have low importance. The sum of ranks for all three factors is presented in the penultimate column of each table. This sum can be used to measure the strength of the three factors affecting the importance of the indicator and can therefore serve as a crude approximation of its importance. In the last column of each table, we present the correlation ratio which was suggested as measure of importance by Paruolo et al. (2013).

The aim of this section is to review the differences between the weights and importance of inequality indicators and to identify the factors which drive the difference. This analysis allows us to describe general patterns in terms of the weights and importance of indicators in all five indices and to suggest recommendations for their construction in Sect. 5. Additionally, we show the differences between MAI and the correlation ratio: roots of these differences and their implications are also discussed in Sect. 5.

4.1 Relative Importance of Inequality Indicators Within the Respective Indices

(a) Quality of Life component of the World Happiness Index

The World Happiness Index (WHI) was introduced by Globeco (2001) and the most recent edition is Globeco (2016), which includes data from 60 countries. The WHI for the country level is composed of four components; the one which includes inequality is the Quality of Life (QL). The component consists of GDP per capita (GDP), the Gini coefficient (GINI), life expectancy (LIFE), the incidence of suicides (SUIC), and air quality (AIR). The component´s score is the sum of ranks of the indicators:

$$QL = \text{GDP rank} + \text{GINI rank} + \text{LIFE rank} + \text{SUIC rank} + \text{AIR rank}.$$

We created five modified versions of QL, calculated their correlations with the original QL, subtracted their correlation coefficients from 1 to get MAI of individual indicators, and standardized them to MRI. The results are presented in Table 6.

MAI varies between 0.96 and 2.28. While nominal weights are 20% for each indicator, MRI varies between 11% and 26%. The two indicators with the highest relative importance are air quality and incidence of suicides (each nearly 26%). Inequality (represented by the Gini coefficient) has a slightly higher MRI (22%) than its nominal weight (20%). On the other side, life expectancy and GDP per capita have an MRI lower than 20%. The authors use normalization by ranks to standardize the indicators which leads to roughly the same variance of all the indicators. That explains why MAI/MRI of life expectancy is lower than MAI/MRI of inequality despite the column Sum of ranks predicts the opposite: in variance (in this case unimportant), life expectancy ranks first while inequality ranks fourth. The differences between the importance and nominal

[11] As all examined indices (components) assign the same nominal weights to all indicators, the factor of nominal weight does not contribute to differences in relative importance between indicators within an index. For this reason, we do not comment on the factor of nominal weights in this section.

weights are driven mainly by the different correlations among the indicators (life expectancy correlates much less than inequality).

According to correlation ratio (see column Paruolo's importance), the most important indicators are GDP per capita and life expectancy (the least important according to MAI/MRI) and the least important is air quality (the most important according to MAI/MRI).

(b) Human Well-being component of the Sustainable Society Index

The Sustainable Society Index (SSI) was introduced by van de Kerk and Manuel (2008) and is updated every 2 years; the most recent version is from 2016 (Sustainable Society Foundation 2016a, b). The analyzed data includes records from 154 countries. The index has three components that are not aggregated into a single index. The component relevant for our research is Human Well-being (HW). It is calculated as:

$$HW = (FOOD*DRINK*SANIT*EDUC*HEALTH*GENDER*DIST*POP*GOV)^{1/9}.$$

There are nine indicators in HW component: FOOD—Sufficient Food; DRINK—Sufficient Drink; SANIT—Safe Sanitation; EDUC—Education; HEALTH—Healthy Life; GENDER—Gender Equality; DIST—Income Distribution; POP—Population Growth; GOV—Good Governance. Each indicator has a nominal weight of approximately 11% (1/9). The two indicators representing inequality—Gender Equality and Income Distribution—have together a nominal weight of 22%. The results of inspection of importance are shown in Table 7.

MAI varies substantially, from 0.03 to 2.37. Five out of nine indicators have MAI equal or lower than 0.10, only two indicators have MAI higher than 1. Income inequality has the highest relative importance, more than 45%. On the other hand, MRI of gender (in)equality is close to zero and is the least influential. When summed together, income distribution and gender equality together have a relative importance of nearly 46%, more than twice their combined nominal weight of 22%. The results seem to be driven mainly by the variance as ranks for MRI and for variance are very similar. On the other hand, a wide gap between MAI/MRI of income inequality and other indicators is driven by multiple correlation. While all the other indicators correlate strongly with the rest of indicators (0.65–0.91), the correlation of income inequality with the other indicators is much lower (0.37). In general, we can say that the column Sum of Ranks predicts the results of MAI/MRI very well: ranks are switched only in case of indicators where the differences of their respective MAI/MRI values were very low.

Vast differences in the importance seem to be driven by the inconsistencies in the application of min-max normalization (the combination of real minimum and maximum with the theoretical minimum and maximum). The income inequality measured by the ratio of income of the richest 10% to the income of poorest 10% of people in a country has no theoretical maximum; the authors therefore use the real minimum and maximum values. Gender (in)equality is operationalized by the Gender Gap Index (GGI) with a theoretical minimum of 0 and theoretical maximum of 1 and the authors work with these values when setting the benchmarks. However, in practice, GGI varies much less: in the 2016 data, the actual minimum is 0.48 and the maximum is 0.88. This leads to a much lower variability of the normalized indicator of gender inequality than of the income inequality (see column Variance); as an effect, income inequality has more than 70 times higher relative importance than gender inequality (which is the least important indicator in the index). We suspect that such a vast discrepancy was not the intention of the authors.

According to correlation ratio (see the column Paruolo's importance), the most important indicator is healthy life followed by safe sanitation. The least important indicator is gender equality, followed by the second least important indicator, income distribution. There seems to be no relationship between MAI/MRI and the correlation ratio.

(c) Index of Inequality-Adjusted Happiness

The Index of Inequality-Adjusted Happiness (IAH) evolved from a paper by Veenhoven (2003) and Veenhoven and Kalmijn (2005), while its most recent formula was introduced by Kalmijn and Veenhoven (2014). The most recent edition is Veenhoven (2017), which includes data from 143 countries. The index consists of two indicators: the average level (i.e. mean) of life satisfaction (m) and the standard deviation of life satisfaction (s). The formula for calculation is:

$$\text{IAH} = 8.28(\text{m} - \text{s}) + 17.2.$$

As can be seen from Table 8, MAI of the level of satisfaction is very high (almost 24), while MAI of the inequality indicator (the standard deviation of life satisfaction) is 1.14. It is the variability which causes the importance to differ dramatically (logically, with only two indicators in the index, the correlation is same for the both of them). The relative importance of the inequality indicator is less than 5%, more than 20 times lower than the influence of level of life satisfaction. The column Sum of ranks correctly predicts a higher importance of level of satisfaction compared to the inequality indicator.

Similar to MAI/MRI, the correlation ratio (see column Paruolo's importance) shows that the importance of the level of satisfaction is much higher than the importance of the inequality indicator.

A very low importance of inequality on the results is startling considering that the index has "inequality" in its name. One may argue that inequality is an added specification to what is primarily being measured, i.e. happiness. But it seems to us that the index was developed with the aim to show that inequality in happiness is something that should be considered, rather than ending up as a negligible factor.

(d) Index of Economic Well-being

The Index of Economic Well-being (IEWB) was created under the Centre for the Study of Living Standards (Osberg and Sharpe 1998) and later revised (Salzman 2003). It is composed of four components: Consumption Flows (CONSUM), Wealth Stocks (WEALTH), Equality (EQUAL), and Economic Security (SECUR). The components are aggregated by the arithmetic mean with default nominal weights of 25% each (the weights may be changed by the users). The formula is:

$$\text{IEWB} = (\text{CONSUM} + \text{WEALTH} + \text{EQUAL} + \text{SECUR})/4.$$

The results in Table 9 show that for OECD countries and Alberta, equality is the most influential component with MAI of 5.80 and a relative importance of 41%. Equality is followed by wealth stocks with MRI of 35%, consumption flows with a relative importance of 18%, and economic security (MAI of 0.84, MRI 6%). Given that the variance of equality and wealth is approximately the same and that at the same time, multiple correlation with other components is lower in the case of wealth, one can be surprised that its MAI/MRI is lower. When analyzing the correlation matrix, we found that equality correlates negatively with consumption while wealth correlates positively with all other components.

Table 6 Results for the quality of life component (2016 edition). *Source*: The authors, based on the data from Globeco (2016)

Indicator	Importance		Nominal weight	Multiple correlation (rank)	Variance (rank)	Sum of ranks (rank)	Paruolo's importance
	MAI	MRI					
AIR	2.28	25.63	20 (1–5)	0.437 (2)	313.47 (3)	8 (2–3)	0.28
SUIC	2.26	25.43	20 (1–5)	0.284 (1)	314.15 (2)	6 (1)	0.31
GINI	1.97	22.20	20 (1–5)	0.496 (3)	280.54 (4)	10 (4)	0.31
LIFE	1.42	15.96	20 (1–5)	0.831 (4)	325.78 (1)	8 (2–3)	0.68
GDP	0.96	10.78	20 (1–5)	0.853 (5)	281.26 (5)	13 (5)	0.72

Table 7 Results for the human well-being component (2016 edition). *Source*: The authors, based on data from Sustainable Society Foundation (2016a, b)

Indicator	Importance		Nominal weight	Multiple correlation (rank)	Variance (rank)	Sum of ranks (rank)	Paruolo's importance
	MAI	MRI					
DIST	2.37	45.12	11.11 (1–9)	0.366 (1)	6.587 (2)	8 (1)	0.28
POP	1.31	24.93	11.11 (1–9)	0.715 (4)	5.975 (3)	12 (2)	0.58
SANIT	0.64	12.19	11.11 (1–9)	0.890 (8)	8.738 (1)	14 (3–4)	0.73
GOV	0.64	12.18	11.11 (1–9)	0.822 (5)	3.406 (4)	14 (3–4)	0.54
EDUC	0.10	1.84	11.11 (1–9)	0.845 (7)	2.660 (5)	16 (6–7)	0.67
DRINK	0.07	1.27	11.11 (1–9)	0.839 (6)	2.160 (6)	16 (6–7)	0.74
FOOD	0.06	1.05	11.11 (1–9)	0.652 (2)	1.440 (8)	15 (5)	0.43
HEALTH	0.04	0.78	11.11 (1–9)	0.905 (9)	1.491 (7)	21 (9)	0.76
GENDER	0.03	0.64	11.11 (1–9)	0.688 (3)	0.377 (9)	17 (6–7)	0.27

Table 8 Results for Index of Inequality-Adjusted Happiness (2005–2014 edition). *Source*: The authors, based on data from Veenhoven (2017)

Indicator	Importance		Nominal weight	Multiple correlation (rank)	Variance (rank)	Sum of ranks (rank)	Paruolo's importance
	MAI	MRI					
m	23.89	95.45	50 (1–2)	0.33 (1–2)	1.68 (1)	4 (1)	0.96
s	1.14	4.55	50 (1–2)	0.33 (1–2)	0.10 (2)	5 (2)	0.27

The negative pairwise correlation is transformed into a positive multiple correlation as the multiple correlation can only be positive.[12] The results in Table 10 show the different properties of the IEWB calculated for Canada and its provinces lead to different results—consumption has the highest importance (36%), while equality is the second most important

[12] A negative correlation among indicators is a rare phenomenon within development indices. We checked pairwise correlations among examined variables in all indices presented in Tables 6, 7, 8, 9, 10 and 11 and found that out of 74 correlations, only four were negative.

Table 9 Results for the Index of Economic Well-being for selected OECD countries and Alberta (edition 2016). *Source*: The authors, based on data from Osberg and Sharpe (2016a)

Indicator	Importance		Nominal weight	Multiple correlation (rank)	Variance (rank)	Sum of ranks (rank)	Paruolo's importance
	MAI	MRI					
EQUAL	5.80	40.98	25 (1–4)	0.874 (3)	0.041 (1)	6.5 (1–3)	0.51
WEALTH	4.95	34.93	25 (1–4)	0.685 (2)	0.040 (2)	6.5 (1–3)	0.58
CONSUM	2.57	18.15	25 (1–4)	0.671 (1)	0.018 (3)	6.5 (1–3)	0.43
SECUR	0.84	5.99	25 (1–4)	0.899 (4)	0.008 (4)	10.5 (4)	0.53

Table 10 Results for the Index of Economic Well-being for Canada and its provinces (edition 2016). *Source*: The authors, based on data from Osberg and Sharpe (2016b)

Indicator	Importance		Nominal weight	Multiple correlation (rank)	Variance (rank)	Sum of ranks (rank)	Paruolo's importance
	MAI	MRI					
CONSUM	1.95	36.35	25 (1–4)	0.673 (1)	0.007 (2–4)	6.5 (1)	0.37
EQUAL	1.54	28.71	25 (1–4)	0.679 (2)	0.007 (2–4)	7.5 (2–3)	0.82
WEALTH	0.96	17.77	25 (1–4)	0.970 (4)	0.014 (1)	7.5 (2–3)	0.92
SECUR	0.92	17.16	25 (1–4)	0.959 (3)	0.007 (2–4)	8 (4)	0.80

component (29%). Wealth and economic security have approximately the same importance (over 17%). In both tables, we can see that column Sum of ranks predicts ranking of indicators in (MRI and MAI) relatively well.

The IEWB clearly shows that the importance of variables could change when the index is applied to structurally different sets of observations. While the relative importance of all four components varies around the level of their nominal weight (from 17.16 to 36.35) when using data from Canada and its 10 provinces, it varies more strongly when using the dataset of 14 OECD countries and Alberta (MRI varies from 5.99 to 40.98). Economic equality is measured by the Sen–Shorrocks–Thon index (with the nominal weight within the inequality component of 75%) and the Gini index (25%). All indicators are standardized by the min-max method; benchmarks are consistently based on real minimum and maximum values for all the indicators. However, these minimum and maximum values are not taken from a specific year, but from the entire observed period. This means that the importance is not only influenced by the selection of countries (regions) considered, but also by the length of the analyzed period (both datasets start at the beginning of the 1980s).

Note that the importance measured by the correlation ratio (see the column Paruolo's importance) differs dramatically in case of both datasets from the importance measured by MAI/MRI.

(e) Inequality-adjusted Human Development Index

The Human Development Index is arguably the most popular composite indicator of development. It was introduced in the first Human Development Report of the United Nations Development Programme (UNDP 1990; for the most recent methodology, see UNDP

Table 11 Results for Inequality-adjusted Human Development Index (2015 edition). *Source*: The authors, based on data from UNDP (2016)

Indicator	Importance		Nominal weight	Multiple correlation (rank)	Variance (rank)	Sum of ranks (rank)	Paruolo's importance
	MAI	MRI					
INC	0.39	26.11	16.66 (1–6)	0.911 (3)	0.031 (2)	8.5 (1–2)	0.86
IINC	0.35	23.53	16.66 (1–6)	0.369 (1)	0.011 (6)	10.5 (4)	0.22
SCH	0.27	17.71	16.66 (1–6)	0.956 (5)	0.032 (1)	9.5 (3)	0.91
ISCH	0.25	16.82	16.66 (1–6)	0.890 (2)	0.021 (3)	8.5 (1–2)	0.76
LIFE	0.17	11.05	16.66 (1–6)	0.950 (4)	0.017 (4)	11.5 (5)	0.82
ILIFE	0.07	4.78	16.66 (1–6)	0.966 (6)	0.015 (5)	14.5 (6)	0.92

2010). The Inequality-adjusted Human Development Index (IHDI) is a modified HDI adjusted by inequality developed by Alkire and Foster (2010). The index includes six indicators. The first three indicators represent the three dimensions of human development: (1) living standard—income per capita (INC); (2) education—mean years of schooling (SCH), and (3) health—life expectancy (LIFE). The other three indicators measure the distribution in the respective dimensions. They are based on the reversed Atkinson index calculated for (4) income per capita (IINC), (5) mean years of schooling (ISCH), and (6) life expectancy (ILIFE). IHDI is aggregated by a geometric mean of the three-dimension indicators adjusted for inequality:

$$\mathrm{IHDI} = ((\mathrm{IINC} * \mathrm{INC}) * (\mathrm{ISCH} * \mathrm{SCH}) * (\mathrm{ILIFE} * \mathrm{LIFE}))^{1/3}.$$

Each indicator has a nominal weight of 16.67% (1/6). Therefore, the three indicators measuring inequality have a joint nominal weight of 50%. Our analysis is based on data from UNDP (2016), which includes 151 countries. Table 11 shows the results for the six indicators. MAI of all indicators is very low, ranging from 0.07 to 0.39. Income has the highest relative importance (26%), followed by income inequality (24%), and mean years of schooling (18%). The sum of the MRI scores of the three inequality indicators is almost 45%, slightly lower than the sum of their nominal weights (50%). However, the relative importance of inequality in life expectancy (ILIFE) is very low, less than a quarter of its nominal weight (4% versus 17%).

The high MRI of inequality in income (IINC) is driven by a correlation with the other indicators (0.369), which is by far the lowest multiple correlation among the IHDI indicators (ranging between 0.890 and 0.966; see the respective column of Table 11). This high difference in correlation of IINC and other IHDI indicators explains why the column Sum of ranks underestimates the importance of IINC. On the other hand, inequality of income (IINC) also has the lowest variability among all indicators: that is the reason why its importance is lower than the importance of income (INC).

It can be argued that the IHDI is balanced in the sense that it gives roughly equal weight to the dimensions and their distributions. On the other hand, income-related indicators have a much higher relative importance than the indicators of education and of health. It is also true that each of the six indicators (three dimensional and three distributional) has an absolute importance below 1%; this is mainly caused by a very high correlation among them. For all indicators, the min-max method with benchmarks based on theoretical arguments is used. The Atkinson index is used to measure inequality. The advantage of the Atkinson index is that it allows systematic integration to the level (mean value) of the measured phenomenon.

Note that according to the correlation ratio (see the column Paruolo's importance), the most important indicator is inequality in life expectancy (the least important indicator if measured by MAI/MRI). Measured by correlation ratio, the least important is income inequality, which is the second most important indicator according to MAI/MRI. Also in case of the IHDI, there seems to be no relationship between MAI/MRI and the correlation ratio.

4.2 Absolute Importance of Inequality Indicators Within the Respective Indices

In this section, we use MAI to compare the absolute importance of inequality indicators across the five composite indices (see Table 12). Out of the eight indicators of inequality, four income inequality indicators rank within the first five positions. Therefore, compared to other dimensions of inequality (happiness, education, health, and gender), income tends to have a higher absolute importance across the five indices. This tendency is driven mainly by the relatively low correlation with the other indicators in each index (except for the IEWB) and by the relatively higher variance (except for the IHDI).

5 Discussion

In the preceding sections, we proposed the method of measuring the importance of variables within a composite index. Then, we applied the method on a set of development indices that include inequality indicators. The importance of an indicator is driven by three factors: its nominal weight, its correlation with other indicators, and its relative variance. This implies that there is a difference between the importance and nominal weight of an indicator and this difference is driven by the two remaining factors.

In the following text, we discuss the results of MAI and MRI calculated for the five development indices with special attention being paid to inequality indicators. Based on the theory laid down in Sect. 3 and results presented in Sect. 4, we contribute to the theory of constructing composite indices by providing some recommendations related mainly to the issues of weighting, normalization, and transparency.

5.1 Discussion of Empirical Results

First, the importance of inequality indicators in composite indices of development is generally low: the average MAI for the eight examined indicators is 1.5%. However, this is not specific for inequality indicators, but a general characteristic of indices composed from more than a few indicators (the average MAI for the examined non-inequality indicators is 2.2%). The results show substantial differences in the absolute importance among inequality indicators, ranging from 0.04 to 5.81%. While it is not surprising that the absolute importance is relatively low given that the indices are composed of more than a few correlated indicators, the absolute importance of some indicators in the overall index is negligible (the extreme cases among inequality indicators are gender inequality in the SSI/HW with 0.04% and health inequality in the IHDI with 0.06%). Again, the issue of very low importance is not an exclusive problem of inequality indicators (see results for the SSI/HW in Table 7).

Second, out of the eight inequality indicators (within the five composite indices), four have a lower relative importance than their nominal weight and four have a higher relative importance than their nominal weight. Income inequality and non-income inequality show a different pattern: the four indicators with higher importance than their nominal weights are the four income inequality indicators. On the opposite side, all non-income inequality indicators have a lower real importance than their nominal weights. It seems that the main reason for these opposing tendencies is a lower correlation of the income inequality indicators with other indicators in comparison with the correlation for the non-income inequality indicators. While income distribution is unbounded, many non-income indicators have natural upper-bounds to which many countries are gradually approaching; this automatically increases the level of correlation between the level (mean value) and inequality in the same dimension.[13] However, a (large) difference in the importance of income inequality and non-income inequality may not apply in all cases and may be balanced by the index constructor (see Sect. 5.2).

Third, the importance is partly derived from the dataset on which the index is based. We illustrate this on the results of the IEWB. This index was calculated separately for (a) 14 countries and Alberta, a province of Canada, and for (b) Canada and its ten provinces. The difference in the importance of the components reflects the different datasets.[14] If an index is calculated for a specific group of countries, but the original concept is universal (i.e. the index was aimed to be applied to all countries), the importance of the indicators will change each time new countries are added to the dataset. The degree of change depends on the number of original countries, the number of new countries, and on the structural differences between the new and old countries in the underlying indicators (i.e. on the differences in the variability of the indicators and the differences in the relationship between them).[15]

Fourth, the correlation coefficient between the nominal weight and the importance of indicators in our sample is 0.52, a moderately strong relationship.[16] The interpretation of the strength of the relationship depends not just on the correlation coefficient, but also on what relationship we expect. The fact that the relationship is positive is unsurprising and the fact that it is not a perfect correlation is in accordance with theory which shows that

[13] The average income can be increased by increasing the income at the lower and upper parts of the income distribution (e.g. for a person who has 1 USD per day and for a person with 1000 USD per day), while it is not possible to increase life expectancy by significantly raising the length of life of those who already have reached a very high age (it is practically impossible to raise the length of life of 90-year-old person by 50 years; the same is possible for infants). Therefore, to increase the life expectancy significantly, inequality in length of life must be reduced.

[14] The IEWB is inconsistent in that it mixes data for countries and for provinces in two datasets, but we want to raise a more general issue that is practically relevant for indicators calculated at a country level.

[15] For example, if a few new countries are now added to the IHDI (currently calculated for 151 countries) the importance of indicators will probably change negligibly; adding new countries to the IEWB (calculated for 14 selected OECD countries) will probably have a higher effect on the importance as the original country base is small and new countries will increasingly be structurally different (especially if new countries are outside the OECD). We suppose that indices that do not have large universal coverage at their inception will be first calculated for richer countries (with available data) and the newer countries will be increasingly poorer countries, likely more different than the original set of countries.

[16] In some cases, the difference between the nominal weight and the importance is striking: for example, the inequality indicator in the IAH has s nominal weight of 50%, while its MRI is lower than 5%.

nominal weight is only one of the three factors affecting the importance of an indicator. In a way, given how nominal weights are generally interpreted, the relationship may be understood as surprisingly weak.

5.2 Discussion of Methodology, Comparison of Results with Pearson's Correlation Ratio

We define importance as the degree how much the variables affect the scores of the overall index. Therefore, a measure of importance should indicate how much a variable influences (affects) value of an overall index. A variable affects values of an index through three channels: (1) shifting the distribution; (2) narrowing/widening the distribution; and (3) changing the ranking and relative distances between the units.

Only in some cases *all* three ways of changes matter for the interpretation of the results. We show this with an example of intelligence. Factor A causes that all people become more (or less) intelligent without changing their ranking and differences between them (channel 1). Factor B increases the difference between the most and least intelligent people (channel 2). Factor C reverses the distribution of intelligence, causing people who were more intelligent to become less intelligent and vice versa (channel 3). In this case, all these factors are important for the overall assessment of the situation.

How does MAI/MRI capture the effects of the variables on the index through the three channels? We show this by using a case of the IEWB (the version for OECD countries and Alberta) as it has only a few indicators and observations whose changes are easier to interpret. The biggest shift of the distribution is caused when we omit economic security: the mean value of the index drops from 0.61 to 0.57 (channel 1). Omitting economic security also causes the highest change in the narrowing/widening of the distribution: the difference between the best and the worst country increases from 0.41 to 0.47 (channel 2). Despite this, economic security is deemed by MAI/MRI as being the least important indicator in the IEWB. This is because MAI/MRI measures only channel 3, i.e. changing the ranking and relative distances between units. The fact that MAI/MRI does not capture the first two channels of importance follows from the characteristics of the Pearson´s correlation coefficient and the procedure of the MAI/MRI calculation described in Sect. 3 of this paper.

To what extent does it matter that MAI/MRI does not capture two out of the three channels? We think that the majority of users of (development) indices interpret the results rather in their relative sense. For example, one country's IHDI value of 0.850 will be interpreted as very good less because it is close to a maximum of 1, but because only a few countries reach such a high value (channel 1). Similarly, one will probably interpret a difference of 0.100 in the IHDI of two countries as being relatively high less because it is 10% of the theoretical range (0–1), but because such a difference usually means that the ranks of those countries differ by 20 or more positions.

When we compare the results based on MAI/MRI with Paruolo´s measure of importance, we find out that these results differ considerably and, often, are nearly reversed (see the results for the Quality of Life component of the World Happiness Index in Table 6). The data in Tables 6, 7, 8, 9, 10 and 11 suggest that when the multiple correlation of an indicator with other indicators is high, then also the Paruolo´s measure of importance is high (and vice versa). This seems surprising as this goes directly against the patterns which we can observe in MAI/MRI. This is explained by the different notions of importance. For us, an important indicator is the one which brings additional information beyond that

Table 12 Inequality-related indicators and their absolute importance. *Source*: The authors, based on data from Globeco (2016), Sustainable Society Foundation (2016a, b), Veenhoven (2017), Osberg and Sharpe (2016a), and UNDP (2016)

Indicator	Index	MAI	Inequality dimension	Multiple correlation[a] (rank)	Variance ratio[b] (rank)	Nominal weight (rank)	Sum of ranks
EQUAL	IEWB[c]	5.80	Income	0.877 (6)	1.55 (2)	25.00 (2)	10
DIST	SSI/HW	2.37	Income	0.366 (2)	1.81 (1)	11.11 (7–8)	10.5
GINI	WHI/QL	1.97	Income	0.496 (4)	0.93 (4)	20.00 (3)	11
s	IAH	1.14	Happiness	0.329 (1)	0.12 (7)	50.00 (1)	9
IINC	IHDI	0.35	Income	0.369 (3)	0.53 (6)	16.66 (4–6)	14
ISCH	IHDI	0.25	Education	0.890 (7)	1.00 (3)	16.66 (4–6)	15
ILIFE	IHDI	0.07	Health	0.966 (8)	0.72 (5)	16.66 (4–6)	18
GENDER	SSI/HW	0.03	Gender	0.688 (5)	0.10 (8)	11.11 (7–8)	20.5

[a]Multiple correlation is multiple correlation with other indicators in the index

[b]Variance ratio is the variance of a given indicator divided by the average variance of all indicators in the index

[c]We use results for the IEWB calculated for OECD countries (Table 9) rather than for the provinces of Canada (Table 10) as all other indices are calculated on country level

involved in the combination of the other indicators in the index. In other words, the most important indicator is the one which changes the results of the index more than the other indicators.

For Paruolo et al. (2013), the most important indicator is the one which most accurately depicts the phenomenon measured by the index. In other words, the *most important* indicator is the one which would be best to use if the task is to measure some complex phenomenon by only one indicator; the indicator which would be worst for this task is the *least important*. Our approach towards importance starts with the *least important* indicator. When searching for the least important indicator, our question is, "Which indicator we can drop with minimal loss of the information captured by the original index?" An indicator whose drop would cause the highest loss of the information provided by the index is the most important. We show this with the case of the IHDI. For Paruolo et al. (2013), inequality in life expectancy is the most important indicator of the IHDI: if you were tasked to measure development by just one indicator, this would be a good choice (while life expectancy indirectly captures many aspects of development, its inequality does it even better as it de facto contains the usual length of life span and premature deaths). For us, it is the least important: when combined with other indicators, it does not bring much additional information (*because* it indirectly captures many aspects of development). However, measuring development just by inequality in incomes would be unreasonable as it does not capture well the other dimensions of development. That is why for Paruolo et al. (2013) this indicator is the least important in the IHDI. For us, it is the second most important indicator: dropping it would cause a high loss of information (it is an inclusive part of development and other dimensions of development do not capture it).

We conclude that there are two approaches to importance and neither of them is superior; different contexts require different approaches. For example, if we want to find a single indicator which captures the most information from a composite index, the approach of Paruolo et al. (2013) is useful. If we want to reduce the number of indicators in an index, the newly suggested approach is more appropriate. In the following section we also show other uses of our notion of importance.

5.3 Recommendations for Constructing Composite Indices

In the preceding text, we noted the discrepancy between the nominal weights and relative importance of variables. Does this discrepancy matter? It matters to the extent that these concepts are misunderstood, which we believe are, at least, among users. We argue that the creators of indices should be aware of the discrepancy and take this matter into account when constructing an index. They should check if a discrepancy exists and if it does, there are three possible ways how to respond:

- to decrease the discrepancy to a reasonable level by changing the way of normalization, aggregation, or by changing/omitting/adding variables;
- to adjust the nominal weights in a way which allows a creator to achieve the intended level of importance of the variables; this does not decrease the discrepancy between the nominal weight and the importance, and therefore this should be made transparent for users;
- if the discrepancy is not perceived as a problem by the index creator, the only step needed is to make the issue transparent for users.

Usually, indices are structured in two hierarchical levels, i.e. components and indicators. We argue that after the initial decision about the structure of the index (including the initial calculation of the scores), creators should include importance checks into the final step of index construction (see the robustness and sensitivity check in OECD 2008). We propose the following procedure. First, the creator checks for potential discrepancies between the nominal weights and relative importance of the indicators within individual components. In the case of high discrepancy, the creator decides on the appropriate response out of the three options listed above. Second, the same is done on the level of components within the index. Naturally, this discrepancy is not the only issue that the creators of composite indices must deal with and therefore should be considered in the given context. For example, if there are good reasons to stay with certain (perhaps consistent) method of normalization, changing the normalization for one specific indicator may not be the best way how to sort out the discrepancy; in other cases, this may be appropriate.

Two out of three factors affecting importance (variance, correlation) are always dependent on data.[17] As shown above on the example of the IEWB, the importance of variables changes once we change the set of the analyzed subjects (countries). Even with the same set of analyzed countries, the importance of variables changes in time; see the example of the IHDI indicators in Table 13.[18] While some indicators (the SCH, LIFE, ILIFE, IINC) have quite stable levels of importance in all six editions, others (the INC, ISCH) fluctuate

[17] In the case when the nominal weights are determined by statistical methods, they are also dependent on data. One can even argue that nominal weights set by participatory methods are also indirectly dependent on data as the experts do not assign weights without considerations of the realities (captured in the actual data structure).

[18] We note that the number of countries included in different IHDI editions slightly fluctuates (ranging from 132 countries in 2012 up to 152 countries in 2014). Year-by-year differences in importance presented in Table 13 are, therefore, partly driven also by changes in the number and structure of countries included.

more. Generally, the analysis of MRI values of the IHDI indicators shows that the importance of indicators remains relatively stable in time. The only three changes in MRI ranks are related with the IINC which dropped from the second to the third position in 2012, then climbed to the first a 1 year later and finally fell back to the second in 2015. We note that the creators of indices should not pay much attention to slight differences in the importance of variables; it is sufficient to reach *roughly* the same MRI scores if the aim is to have an index with equal importance of all indicators (components).

There are various methods how to adjust the importance of variables. Adjusting the nominal weights and changing the normalization of variables are methods how to change the importance of a variable with predictable effects. Increasing (decreasing) the nominal weight leads to an increase (decrease) of the importance. Z-scores normalization increases the importance of variables with low variability, while decreasing the importance of variables with high variability. Normalization by ranking has the same effect. Min-max normalization can be also used to change the importance. For example, the IHDI uses income, schooling, and life expectancy (i.e. their average levels and distribution), with life expectancy (again both the level and the distribution) having the lowest real importance (MRI). Current, to an extent arbitrary, minimum and maximum benchmarks of 20 and 85 years lead to a limited normalized range of 0.445–0.980. Using real minimum and maximum benchmarks of 48.9 and 83.7 years (data from 2015 edition) would lead to a maximum normalized range of 0–1, with higher variance (0.061 rather than 0.017) and higher importance of this indicator within the index.

Omitting the indicator (component) from the index should be considered by the creators if its importance is extremely low (and it is not possible to find a reasonable way how to increase it). The omission has a predictable effect on the other indicators (components): their importance increases. And vice versa, adding a new variable leads to an average decrease of MAI of all other variables. However, omitting/adding variables affects the importance of *all* variables in the index. Therefore, it is not as useful as adjusting the nominal weights or changing the method of normalization if we want to change the importance of one specific variable. The method of aggregation can also somehow influence the importance of variables. However, it is difficult to predict the direction of change and again, it cannot be used to influence one specific variable.

Another method how to adjust the importance of variables is changing the operationalization of a measured phenomenon. For example, there are several indicators of income inequality. Among others, it can be measured by the Gini index (see Gini 1912), quantile ratios (the idea introduced by Kuznets 1955) including the recently proposed Palma ratio (see Cobham et al. 2016), the Atkinson index (see Atkinson 1970), and the (adjusted) standard deviation. The different measures of inequality bring different results including different variances. Interestingly, the Atkinson index includes the parameter ε representing aversion to inequality. Increasing (decreasing) this parameter shall automatically lead to an increase (decrease) of the importance of the indicator in an index.

It is possible to adjust inequality indicators to narrow the difference between the nominal weights and their relative importance in some of our examined indices. For example, measuring the income inequality by the Gini coefficient instead of the ratio of income of the richest 10% to the poorest 10% people in the Quality of Life component of the WHI would allow authors to use the theoretical minimum (0) and maximum (1) when normalizing the indicator; this would lead to a drop of its importance and, therefore, to a lower discrepancy between the nominal weight and MRI. Adjusting the parameter ε of the Atkinson indices in the IHDI can be used to narrow the differences between their nominal weights

Table 13 MRI for indicators of the IHDI in its different editions. *Source*: The authors, based on UNDP (2010, 2011, 2013, 2014, 2015, 2016)

Indicator	2010	2011	2012	2013	2014	2015	2010–2015
INC	36.7	27.0	27.3	23.1	24.4	26.1	23.1–36.7
IINC	20.2	26.1	22.3	25.1	26.4	23.5	20.2–26.4
SCH	19.9	21.7	23.8	21.3	19.6	17.7	17.7–23.8
ISCH	10.5	13.6	14.7	17.0	16.2	16.8	10.5–17.0
LIFE	8.9	8.4	8.5	10.2	9.4	11.1	8.4–11.1
ILIFE	3.8	3.2	3.4	3.2	4.0	4.8	3.2–4.8

and relative importance. In such a case, the lower parameter should be assigned for income inequality and the higher for inequality in health.

6 Conclusion

In this paper, we attempted to examine two issues related to the weighting of variables in composite indices. We analyzed the factors affecting the importance of variables and proposed the method of measurement of the importance. Our notion of importance measured by absolute importance and relative importance differs from the approach by Paruolo et al. (2013) as we see less value in observing changes in variability after the fixing of an indicator, especially in the context of dimensionless indices. Instead, we argue that it is changing countries' positions (the ranks and relative distances) after omitting the examined variable that matters more when assessing its importance for the overall index. Based on this approach, we proposed the method for measuring of the importance.

We then applied the method to a specific group of composite development indices, namely those that include inequality. Our analysis brought three conclusions. First, the absolute importance of inequality variables in composite development indices is generally low (ranging from 0.04 to 5.81%) which is not surprising for indices that are composed of more than a few correlated indicators. Second, for some composite indices the nominal weights and relative importance of inequality variables differ significantly. Third, the relative importance tends to be higher than the nominal weight for those inequality variables that measure income inequality as opposed to those measuring other types of inequality.

The role of the nominal weight of indicators is often incorrectly understood by users as a measure of importance (i.e. how much an indicator affects results of an overall index). We argue that the importance of indicators should be examined in the process of index construction. For those composite indices where experts are called on to decide on weights of the indicators, they should be aware of the difference between nominal weights and importance. It is possible to reduce a large discrepancy between the two through various instruments, such as normalization and a change of operationalization of the measured phenomenon. Alternatively, adjusting the nominal weights can be used to target a specific level of importance; however, this does not decrease the discrepancy between the nominal weight and importance. If a large discrepancy persists, the users of the index should be appropriately informed that nominal weights do not represent importance. On the ground of parsimony, if the importance of a variable is extremely low its omission should be considered.

Finally, a word of caution. In the paper we have empirically showed how real importance can be measured and how this often significantly differs from the nominal weights. This is an important perspective that may help to improve the construction of composite indices. Yet, this is at the same time only one (statistical) perspective that should not override all other concerns. We argue that the discrepancy between nominal weights and importance should be examined in the process of index construction; whether this needs to be undertaken (and how) should be decided within the context of a particular index.

Funding Funding was provided by Univerzita Palackého v Olomouci (Grant No. IGA_PrF_2019_025).

Appendix

See Figs. 1, 2, 3, 4 and 5.

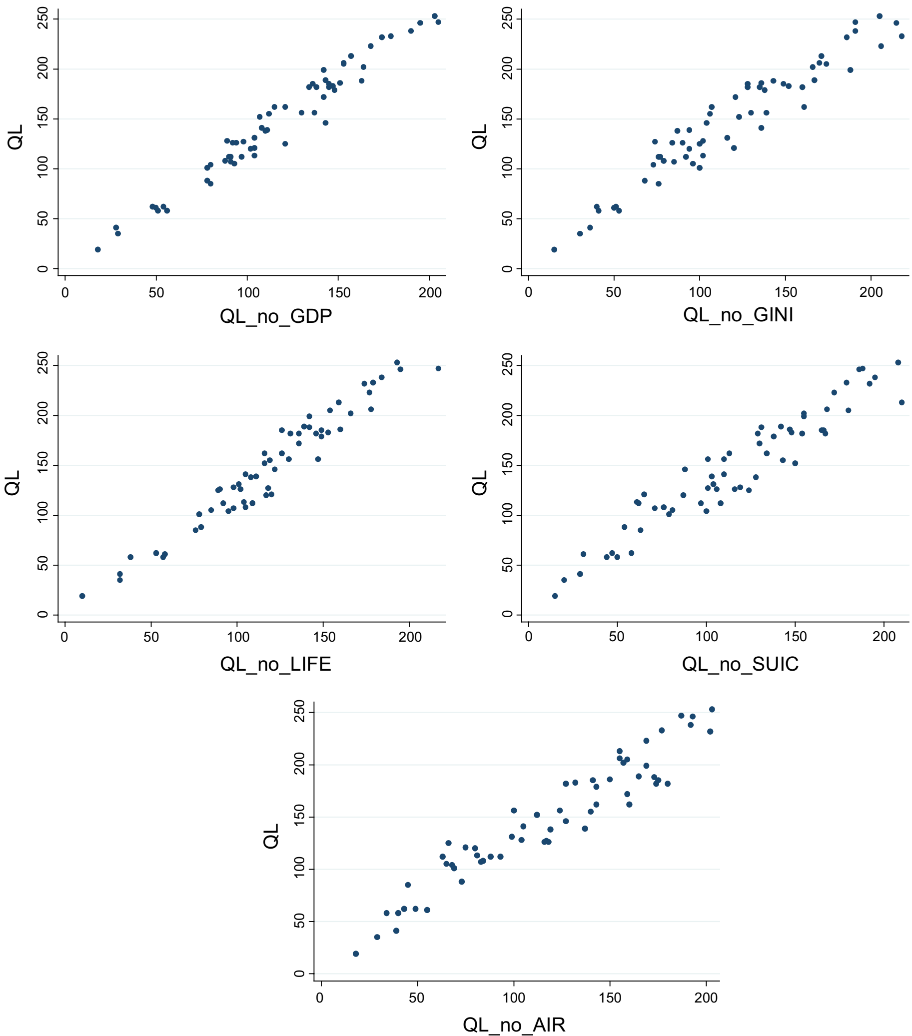

Fig. 1 Association of Quality of Life component and its modified versions. *Source*: The authors based on the data from Globeco (2016)

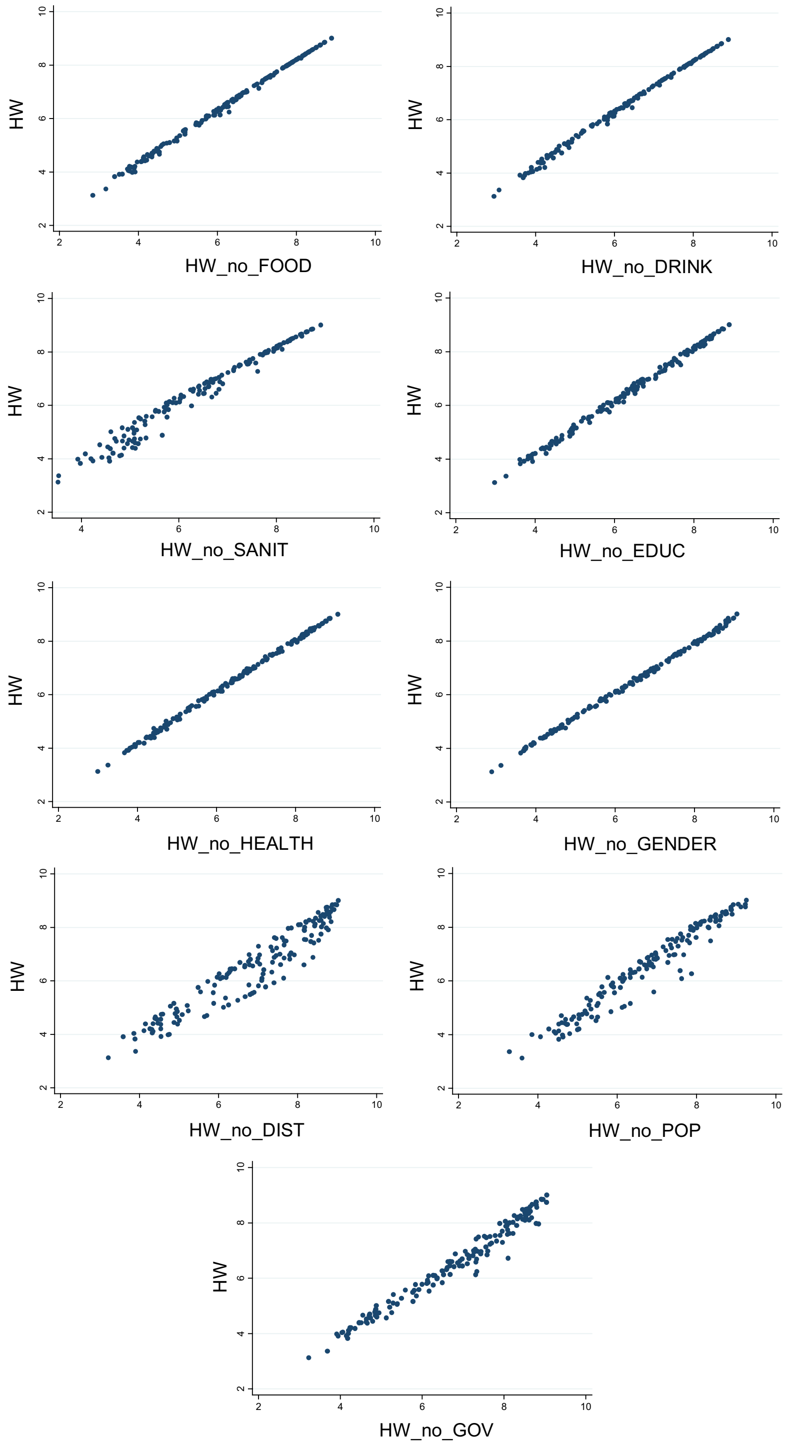

Fig. 2 Association of Human Well-being component and its modified versions. *Source*: The authors based on data from Sustainable Society Foundation (2016a, b)

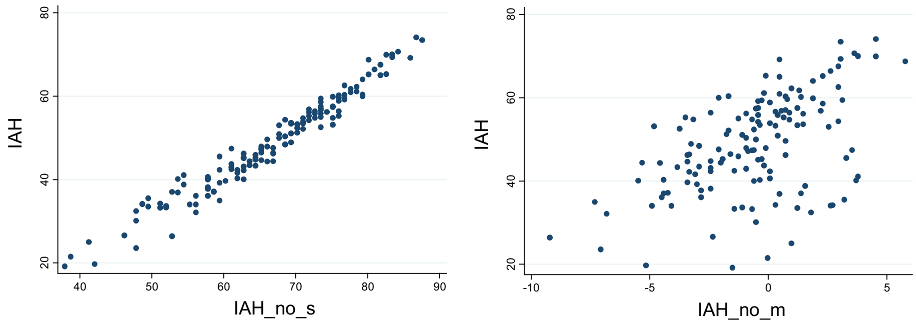

Fig. 3 Association of Index of Inequality-Adjusted Happiness and its modified versions. *Source*: The authors based on data from Veenhoven (2017)

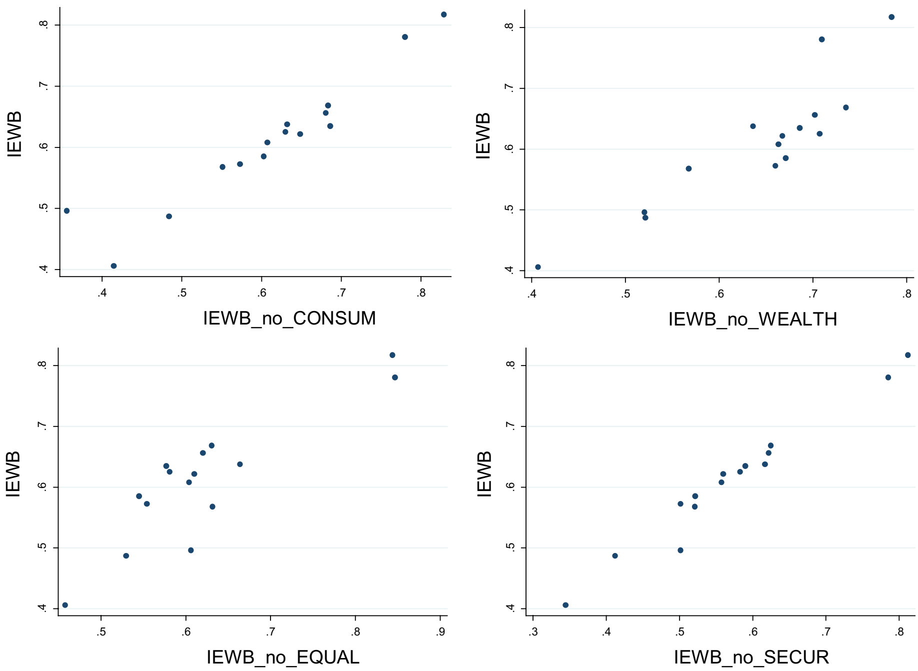

Fig. 4 Index of Economic Well-being for selected OECD countries and Alberta and its modified versions. *Source*: The authors based on data from Osberg and Sharpe (2016a)

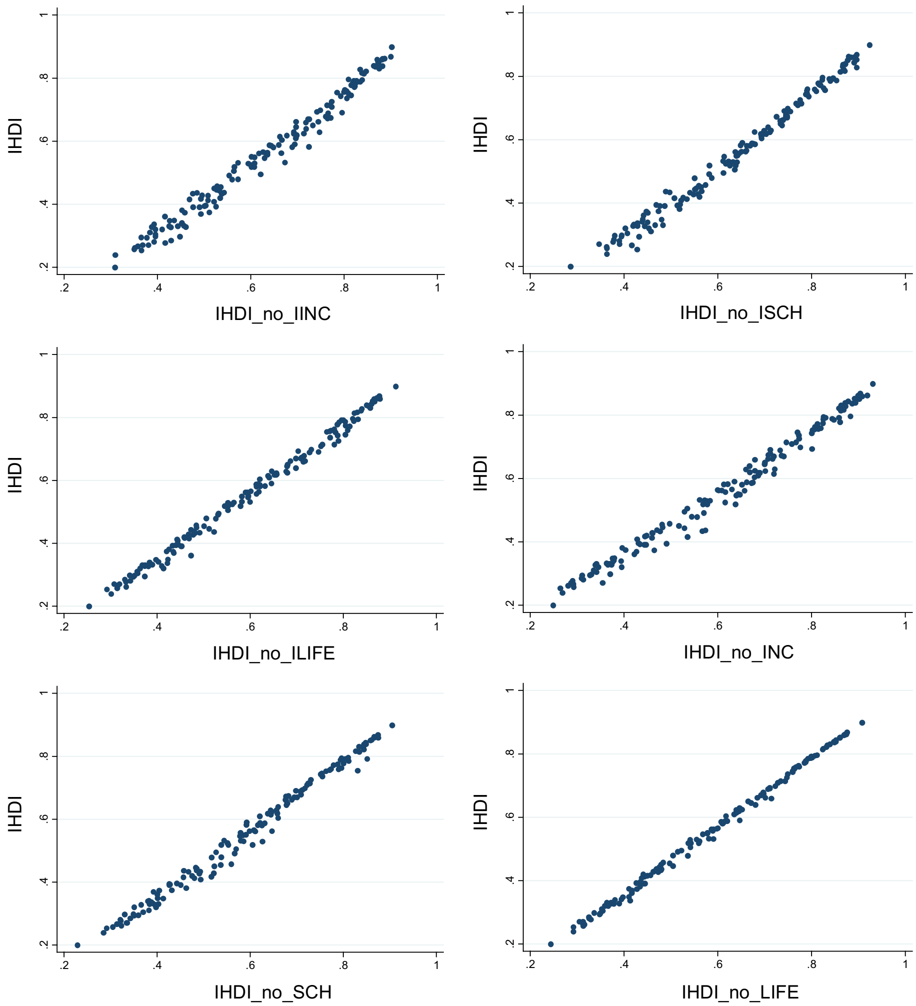

Fig. 5 Inequality-adjusted Human Development Index and its modified versions. *Source*: The authors based on data from UNDP (2016)

Derivation of Formula (3)

Let us show that the correlation coefficient r_k used in the definition of measure of absolute importance (*MAI*) of the *k*-th indicator can be expressed in form (3).

$$
\begin{aligned}
r_k &= corr\left(OCI, MCI_k\right) = corr\left(OCI, OCI - w_k x_k\right) \\
&= \frac{cov\left(OCI, OCI - w_k x_k\right)}{\sqrt{var(OCI)}\sqrt{var\left(OCI - w_k x_k\right)}} = \frac{var(OCI) - w_k cov\left(OCI, x_k\right)}{\sqrt{var(OCI)}\sqrt{var(OCI) - 2w_k cov\left(OCI, x_k\right) + w_k^2 var\left(x_k\right)}}.
\end{aligned}
$$

The equalities above are based on the definition of correlation coefficient and properties of variance and covariance of a sum.

Now, we express $cov\left(OCI, x_k\right)$ by means of correlation coefficient $\rho_k = corr\left(x_k, OCI\right)$. Since $\rho_k = \frac{cov\left(OCI, x_k\right)}{SD(OCI)SD\left(x_k\right)}$, it holds $cov\left(OCI, x_k\right) = \rho_k SD(OCI)SD\left(x_k\right)$.

Using this expression, the formula for r_k can be rewritten in the following form:

$$
\begin{aligned}
r_k &= \frac{var(OCI) - w_k\rho_k SD(OCI)SD\left(x_k\right)}{\sqrt{var(OCI)}\sqrt{var(OCI) - 2w_k\rho_k SD(OCI)SD\left(x_k\right) + w_k^2 var\left(x_k\right)}} \\
&= \frac{var(OCI) - w_k\rho_k SD(OCI)SD\left(x_k\right)}{\sqrt{var(OCI)}\sqrt{var(OCI)\left[1 - 2w_k\rho_k SD\left(x_k\right)/SD(OCI) + w_k^2 var\left(x_k\right)/var(OCI)\right]}}
\end{aligned}
$$

The final step consists in dividing both nominator and denominator by $var(OCI) = SD(OCI)^2$. Denoting $RV_k = SD\left(x_k\right)/SD(OCI)$, it immediately follows that

$$
r_k = \frac{1 - w_k\rho_k RV_k}{\sqrt{1 - 2w_k\rho_k RV_k + \left(w_k RV_k\right)^2}}.
$$

Derivation of Formula (4)

In the case of uncorrelated indicators $x_1, \dots, x_p$, it holds

$$
var\left(\sum_i w_i x_i\right) = \sum_i w_i^2 var(x_i), \quad cov\left(x_k, \sum_i w_i x_i\right) = w_k var\left(x_k\right),
$$

and therefore

$$
\rho_k = \frac{cov\left(\sum_i w_i x_i, x_k\right)}{SD(OCI)SD\left(x_k\right)} = w_k \frac{SD\left(x_k\right)}{SD(OCI)},
$$

By substituting this term into expression (3), one obtain the following equality

$$
r_k = \frac{1 - \frac{\left(w_k SD\left(x_k\right)\right)^2}{\sum_{i=1}^p \left(w_i SD\left(x_i\right)\right)^2}}{\sqrt{1 - \frac{\left(w_k SD\left(x_k\right)\right)^2}{\sum_{i=1}^p \left(w_i SD\left(x_i\right)\right)^2}}} = \sqrt{1 - \frac{\left(w_k SD\left(x_k\right)\right)^2}{\sum_{i=1}^p \left(w_i SD\left(x_i\right)\right)^2}}
$$

References

Alkire, S., & Foster, J. (2010). *Designing the inequality-adjusted human development index*. OPHI working paper 37. Oxford: Oxford Poverty & Human Development Initiative.

Alkire, S., & Robles, G. (2017). *Multidimensional poverty index summer 2017: Brief methodological note and results. MPI methodological notes 44*. Oxford: Oxford Poverty & Human Development Initiative.

Anderberg, M. R. (1973). *Cluster analysis for applications*. New York: Academic Press Inc.

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory, 2*(3), 244–263.

Bandura, R. (2008). *A survey of composite indices measuring country performance: 2008 update*. UNDP/ODS working paper. New York, NY: Office of Development Studies, United Nations Development Programme.

Becker, W., Saisana, M., Paruolo, P., & Vandecasteele, I. (2017). Weights and importance in composite indicators: Closing the gap. *Ecological Indicators, 80,* 12–22.

Becker, W., & Saltelli, A. (2015). Design for sensitivity analysis. In A. Dean, M. Morris, J. Stufken, & D. Bingham (Eds.), *Handbook of design and analysis of experiments* (pp. 627–674). Boca Raton: Chapman & Hall/CRC Press.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Chowdhury, S., & Squire, L. (2006). Setting weights for aggregate indices: An application to the commitment to development index and human development index. *The Journal of Development Studies, 42*(5), 761–771.

Cobham, A., Schlogl, L., & Sumner, A. (2016). Inequality and the tails: The Palma proposition and ratio. *Global Policy, 7*(1), 25–36.

de Kruijk, H., & Rutten, M. (2007). *Weighting dimensions of poverty based on people's priorities: Constructing a composite poverty index for the maldives*. IIDE discussion paper 2007 08-01. Rotterdam: Institute for International & Development Economics.

Decancq, K., & Lugo, M. A. (2010). *Weights in multidimensional indices of well-being: An overview*. https://lirias.kuleuven.be/bitstream/123456789/262767/1/DPS1006.pdf. Accessed January 2, 2018.

Ding, Y., Fu, Y., Lai, K. K., & Leung, W. K. J. (2017). Using ranked weights and acceptability analysis to construct composite indicators: A case study of regional sustainable society index. *Social Indicators Research*. https://doi.org/10.1007/s11205-017-1765-3.

Gan, X., Fernandez, I. C., Guo, J., Wilson, M., Zhao, Y., Zhou, B., et al. (2017). When to use what: Methods for weighting and aggregating sustainability indicators. *Ecological Indicators*, *81*, 491–502.

Gini, C. (1912). *Variabilità e Mutuabilità. Contributo allo Studio delle Distribuzioni e delle Relazioni Statistiche*. Bologna: C. Cuppini.

Ginsberg, N., Osborn, J., & Blank, G. (1986). *Geographic perspectives on the wealth of nations*. Department of Geography research paper no. 220. Chicago: University of Chicago.

Globeco. (2001). *L´indice Globeco du Bonheur Mondial*. http://www.globeco.fr/images/PDF/BMB/bonheur-mondial-2001.pdf. Accessed February 21, 2017.

Globeco. (2016). *Bonheur mondial. Edition 2016*. http://globeco.fr/images/PDF/BMB/bonheur-mondial-2016.pdf. Accessed January 22, 2017.

Gorsuch, R. L. (1974). *Factor analysis*. Philadelphia: W.B Saunders.

Greco, S., Ishizaka, A., Matarazzo, B., & Torrisi, G. (2017). Stochastic multi-attribute acceptability analysis (SMAA): An application to the ranking of Italian regions. *Regional Studies*. https://doi.org/10.1080/00343404.2017.1347612.

Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2018). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research*. https://doi.org/10.1007/s11205-017-1832-9.

Hagerty, M. R., & Land, K. C. (2006). Constructing summary indices of quality of life: A model for the effect of heterogenous importance weights. *Sociological Methods & Research, 35*(4), 455–496.

Hsu, A., Johnson, L. A., & Lloyd, A. (2013). *Measuring progress: A practical guide from the developers of the environmental performance index (EPI)*. New Haven: Yale Center for Environmental Law & Policy.

Hume, D. (1740). *A treatise of human nature*. London: Penguin Books. **(reprinted in 1985)**.

Jolliffe, I. T. (1986). *Principal component analysis*. New York: Springer.

Kalmijn, W., & Veenhoven, R. (2014). Index of inequality-adjusted happiness (IAH) improved: A research note. *Journal of Happiness Studies, 15*(6), 1259–1265.

Käppeli, A., Hillebrandt, H., & Mitchell, I. (2017). *The commitment to development index: 2017 edition. Methodological overview paper*. Washington, D.C.: Center for Global Development.

Krishnakumar, J., & Nagar, A. L. (2008). On exact statistical properties of multidimensional indices based on principal components, factor analysis, MIMIC and structural equation models. *Social Indicators Research, 86*(3), 481–496.

Kuznets, S. (1955). Economic growth and income inequality. *American Economic Review, XLV*(1), 1–28.

Nguefack-Tsague, G., Klasen, S., & Zucchini, W. (2011). On weighting the components of the Human Development Index: A statistical justification. *Journal of Human Development and Capabilities, 12*(2), 183–202.

Nord, E. (2013). Disability weights in the Global Burden of Disease 2010: Unclear meaning and overstatement of international agreement. *Health Policy, 111*(1), 99–104.

OECD. (2008). *Handbook on constructing composite indicators: Methodology and user guide*. Paris: OECD Publications.

Osberg, L., & Sharpe, A. (1998). *An index of economic well-being for Canada.* Research report. Human Resources Development Canada. http://www.csls.ca/iwb/articles.asp. Accessed January 2, 2017.

Osberg, L., & Sharpe, A. (2016a). *Database of the index of economic well-being for selected OECD countries and Alberta, 1980–2014*. Centre for the Study of Living Standards. http://www.csls.ca/iwb/oecd.asp. Accessed February 21, 2017.

Osberg, L., & Sharpe, A. (2016b). *Database of the index of economic well-being for Canada and the provinces, 1981–2014*. Centre for the Study of Living Standards. http://www.csls.ca/iwb/prov.asp. Accessed February 21, 2017.

Paruolo, P., Saisana, M., & Saltelli, A. (2013). Ratings and rankings: Voodoo or science? *Journal of the Royal Statistical Society, 176*(3), 609–634.

Pearson, K. (1905). *On the general theory of skew correlation and non-linear regression*. London: Dulau & Co.

Piketty, T. (2014). *Capital in the twenty-first century*. Cambridge: The Belknap Press of Harvard University Press.

Ravallion, M. (1997). Good and bad growth: The human development reports. *World Development, 25*(5), 631–638.

Ravallion, M. (2011). Mashup indices of development. *World Bank Research Observer, 27*(1), 1–32.

Saisana, M., Saltelli, A., & Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society, 168*(2), 307–323.

Saisana, M., & Tarantola, S. (2002). *State-of-the-art report on current methodologies and practices for composite indicator development*. Ispra: Joint Research Centre.

Salomon, J. A., Vos, T., Hogan, D. R., Gagnon, M., Naghavi, M., Mokdad, A., et al. (2012). Common values in assessing health outcomes from disease and injury: Disability weights measurement study for the Global Burden of Disease Study 2010. *Lancet, 380*(9859), 2129–2143.

Saltelli, A., & Tarantola, S. (2002). On the relative importance of input factors in mathematical models: Safety assessment for nuclear waste disposal. *Journal of the American Statistical Association, 97*(459), 702–709.

Salzman, J. (2003). *Methodological choices encountered in the construction of composite indices of economic and social well-Being.* Research paper. http://www.csls.ca/events/cea2003/salzman-typol-cea2003.pdf. Accessed February 23, 2017.

Sustainable Society Foundation. (2016a). *Calculation formulas of the Sustainable Society Index, SSI-2016*. http://www.ssfindex.com/ssi2014/wp-content/uploads/pdf/calculation_formulas-2016.pdf. Accessed February 23, 2017.

Sustainable Society Foundation. (2016b). *SSI 2016*. http://www.ssfindex.com/ssi2016/wp-content/uploads/pdf/ssi2016.pdf. Accessed February 23, 2017.

Syrovátka, M., & Hák, T. (2015). Measuring rich-country policies toward the global environment: A critical analysis of the environmental component of the Commitment to Development Index. *Journal of Environment & Development, 24*(1), 105–128.

UNDP (United Nations Development Programme). (1990). *Human development report 1990*. New York, NY: Oxford University Press.

UNDP (United Nations Development Programme). (2010). *Human development report 2010. The real wealth of nations: Pathways to human development*. New York, NY: Palgrave Macmillan.

UNDP (United Nations Development Programme). (2011). *Human development report 2011. Sustainability and equity: A better future for all*. New York, NY: Palgrave Macmillan.

UNDP (United Nations Development Programme). (2013). *Human development report 2013. The rise of the south: Human progress in a diverse world*. Washington DC: Communications Development Incorporated.

UNDP (United Nations Development Programme). (2014). *Human development report 2014. Sustaining human progress: Reducing vulnerabilities and building resilience*. Washington DC: Communications Development Incorporated.

UNDP (United Nations Development Programme). (2015). *Human development report 2015: Work for human development*. Washington DC: Communications Development Incorporated.

UNDP (United Nations Development Programme). (2016). *Human development report 2016: Human development for everyone*. Washington DC: Communications Development Incorporated.

van de Kerk, G., & Manuel, A. (2008). A comprehensive index for a sustainable society: The SSI—the Sustainable Society Index. *Ecological Economics, 66*(2–3), 228–242.

Veenhoven, R. (2003). Equality-adjusted happiness in nations. In *Paper presented at the conference of the International Society for Quality of Life Studies (ISQOLS)*, Frankfurt (DE).
Veenhoven, R. (2017). *Inequality-adjusted happiness in 159 nations 2005–2014*. World Database of Happiness. https://worlddatabaseofhappiness.eur.nl/hap_nat/findingreports/RankReport_InequalityHappiness.php. Accessed March 23, 2017.
Veenhoven, R., & Kalmijn, W. M. (2005). Inequality-adjusted happiness in nations: Egalitarianism and utilitarianism married in a new index of societal performance. *Journal of Happiness Studies, 6,* 421–455.
Wilkinson, R., & Pickett, K. (2009). *The spirit level: Why greater equality makes societies stronger*. New York, NY: Bloomsbury.
Xu, C., & Gertner, G. Z. (2008a). A general first-order global sensitivity analysis method. *Reliability Engineering and System Safety, 93*(7), 1060–1071.
Xu, C., & Gertner, G. Z. (2008b). Uncertainty and sensitivity analysis for models with correlated parameters. *Reliability Engineering and System Safety, 93*(10), 1563–1573.
Yang, L. (2014). *An inventory of composite measures of human progress*. United Nations Development Programme, Human Development Report Office. http://hdr.undp.org/sites/default/files/inventory_report_working_paper.pdf. Accessed 23 Mar 2017.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.