

Chapter 2

Composite Indicators for Sustainability Assessment: Methodological Developments



P. Zhou and L. P. Zhang

1 Introduction

The intensive alteration to the natural environment by human beings has been posing challenges to the natural and socio-economic systems (Munda and Saisana 2011). Global warming causes a huge economic loss to agriculture sector of some countries including China in the past decade (Chen et al. 2016). Air pollution has severe negative health effects, especially for those vulnerable people such as the elderly, infant and child (He et al. 2016). Some pollution reduction mandates by central and local governments also triggered severe unintended consequences. For example, due to the effect “polluting thy neighbor”, the most downstream county of a province in China has up to 20% more water polluting activities than other counterparts since 2001 (Cai et al. 2016). Undoubtedly, problems resulted from climate change, environmental pollution, depletion of natural resources and others have been threatening the development of our society (Tilman et al. 2002). A rising concern has been voiced in scientific community and policy circles on how human beings should interact with nature, and how they are responsible for future generations in a sustainable way (Baumgärtner and Quaas 2010). Indeed, there are so many efforts and initiatives towards sustainable development in our society. However, whether those activities are adequate for pursuing sustainable development is still questionable (Sala et al. 2015). To provide a scientific basis for fighting with climate change and avoiding unanticipated consequence, the status of sustainability should be evaluated in a solid and reliable manner to assess whether the

P. Zhou (✉) · L. P. Zhang

College of Economics and Management, Nanjing University of Aeronautics and Astronautics, 29 Jiangjun Avenue, Nanjing, China
e-mail: cemzp@nuaa.edu.cn

P. Zhou

School of Economics and Management, China University of Petroleum,
No. 66 Changjiang West Road, Qingdao, China

© Springer Nature Singapore Pte Ltd. 2018

R. Pang et al. (eds.), *Energy, Environment and Transitional Green Growth in China*, https://doi.org/10.1007/978-981-10-7919-1_2

target of “meet the needs of the present without compromising the ability of future generations to meet their own needs” has gradually been realized (WCDE 1987).

Theoretically, sustainability is a multi-dimension (e.g. economic, social and ecological) concept encompassing internal relationships between different dimensions, which brings difficulty in sustainability assessment (Mayer 2008). In addition, the issues such as multi-interpretation in concept, the determination of boundaries and measurability, also cause a rising concern on the reliability of sustainability assessment (Hák et al. 2012). Different methods have been introduced into sustainability assessment, e.g. indicators, product-based assessment, and integrated assessment (Ness et al. 2007). The indicator approach, owing to some desirable properties like simplicity, quantifiability and timely identification of the trends, has acquired compelling attention in the literature of environmental and ecological economics (Díaz-Balteiro and Romero 2004). At the end of last century, the United Nations suggested to develop indicators for sustainable development to provide an analytical foundation for policy analysis and decision making at different levels (UNCED 1992). Since then, various indicators have been developed, e.g. Ecological Footprint (Wackernagel and Rees 1998), Environmental Sustainability Index (Esty et al. 2005), Human Development Index (UNDP 2014), Environmental Policy Stringency Index (Botta and Kozluk 2014), World Energy Trilemma Index (WEC 2016), Oxford Sustainability Index (OCC 2016), and Environmental Performance Index (Hsu et al. 2016). According to Zhou and Ang (2008), the existing sustainability indicators may broadly be grouped into non-composite and composite indicators. Non-composite indicators are usually in the form of a set of indicators or an integrated indicator. The approach of composite indicators aims to aggregate various indicators into a single real-valued score to represent an entity’s sustainable performance. As Nardo et al. (2008) argued, composite indicators can reduce the visible size of indicators and are easier to interpret than a set of individual indicators. Hence, composite indicators have recently gained much popularity in sustainability assessment. Hereafter we refer to composite indicators for sustainability assessment as composite sustainability indicators (CSI) for convenience purpose.

The reliability of a CSI heavily depends on the underlying methods which are used for constructing the CSI. In the past decades, scholars have contributed to developing alternative methods for constructing CSI. See, for example, van den Bergh and Veen-Groot (2001), Cherchye and Kuosmanen (2004), Díaz-Balteiro and Romero (2004), Munda (2005), Despotis (2005a, b), Zhou et al. (2007) and Zanella et al. (2015). In parallel, Ebert and Welsch (2004) showed how to construct a meaningful environmental index from the social choice perspective. Zhou et al. (2006a) proposed an information loss criterion for comparing different aggregation functions. More recently, Pollesch and Dale (2015, 2016) investigated on the application of aggregation theory and normalization methods to sustainability assessment. Zhou et al. (2017) further looked through the meaningfulness of composite environmental indices and showed that a cardinally meaningful composite indicator can be constructed by nonparametric frontier approach. Several scholars have also reviewed past CSI studies with emphasis on their theoretical and

conceptual developments, e.g. Parris and Kates (2003), Ness et al. (2007) and Mori and Christodoulou (2012). As pointed out by Mayer (2008), the identification of bias introduced by method choice plays a significant role in improving the utility of CSI for supporting policy making. The study by Böhringer and Jochem (2007) highlights the significance of scientifically sound methods for normalization, weighting and aggregation in building meaningful CSI. The purpose of this chapter is to provide a systematic literature review of the methodological developments in constructing CSI. It is expected that such a review provides not only a sketch of the mainstream methods with their strengths and weaknesses but also useful insights on the choice of an appropriate method for constructing CSI in various application scenarios.

The rest of this chapter is organized as follows. Section 2 provides a description of the framework on CSI construction. Section 3 summarizes the most popular or promising methods in constructing CSI. In Sect. 4, we discuss the influential factors and principles in method choice at the stages of normalization and aggregation. The last section concludes this study with discussions on potential future research topics.

2 Generic Procedure of Constructing CSI

The construction of CSI starts from the determination of a set of indicators for the entities whose sustainable performance is to be evaluated. The given information may be represented by a performance matrix X as shown in Eq. (1) that deals with m entities and n indicators.

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} \quad (m, n \geq 2) \quad (1)$$

$$W = [w_1 \quad \cdots \quad w_n]$$

In Eq. (1), x_{ij} refers to the performance value of entity i pertinent to indicator j , and W is a weight vector in which w_j denotes the weight assigned to indicator j . The indicators for sustainability performance evaluation are usually measured by different units. In order to aggregate individual indicators into CSI, some aggregation methods require each indicator to be dimensionless by certain transformation function, i.e. $V = v(X)$. We assume that the performance matrix after normalization is denoted by

$$V = \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{m1} & \cdots & v_{mn} \end{bmatrix} \quad (m, n \geq 2) \quad (2)$$

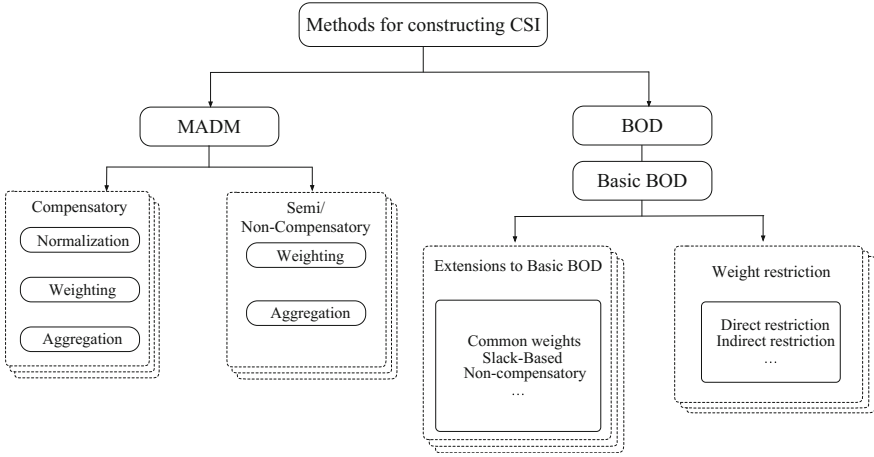


Fig. 1 Classification of the methods for constructing CSI

Once the weight vector W is determined, different aggregation techniques might be used to combine individual indicators into real-valued CSI. In general, the aggregation techniques aim to:

Seek a function $r_i = f_1(X(\text{or } V), W)$ for providing sustainable performance rankings of entities, and/or

Seek a function $u_i = f_2(X(\text{or } V), W)$ for providing sustainable performance index for each entity.

As described above, the construction of CSI from a set of pre-defined indicators mainly involves the normalization of indicators, the assignment of indicator weights and the search for an appropriate aggregation function. In literature, various methods have been used for these three steps, which might broadly be classified into two categories. One is based on multi-attribute decision making (MADM), and the other is based on benefit of the doubt (BOD) that is a data envelopment analysis (DEA)-like approach (see Fig. 1). Although BOD methods in the broad sense can be attributed to MADM, they are different from each other in several aspects. For example, by MADM, a set of common weights are shared by each entity. However, BOD methods often aim to find a different set of weights which are most favorable for each entity. Besides, CSI based on BOD methods may not require normalization.

3 MADM Methods

MADM is a well-established methodology with the aim to make choice under multiple conflict criteria be more explicit, rational and efficient (Yoon and Hwang 1995). MADM methods can be involved in all procedure of constructing CSI. In the

followings, we shall describe the steps with focus on the methods used. It starts from normalization followed by data aggregation, by which the importance of indicator weighting is highlighted and discussed.

3.1 Normalization Methods

The underlying indicators for assessing sustainability are generally in different measurable units, and different indicators have distinct ranges or scales (Mayer 2008). As such, normalization procedure is often taken for making underlying indicators comparable (Nardo et al. 2008). The commonly used normalization methods in constructing CSI may be categorized into three categories, namely, standard deviation from the mean (i.e. z-score), distance from a reference, and distance from best and worst performers (i.e. re-scaling). Table 1 provides a description of the three normalization methods.

Z-score is used to statistically measure the relationship between the value of a sustainable indicator and the mean of the values of the sustainable indicator system. It has an average of zero, indicating that it can avoid introducing aggregation distortions stemming from the differences in indicator means (Freudenberg 2003). The positive (negative) value indicates that it is above (below) the mean by how many standard deviations. Z-score transforms the original variables into a common scale. These desirable characteristics make it be frequently used in normalization, see Floridi et al. (2011).

Distance to a reference aims to normalize the underlying indicators by measuring the distance of an entity to a reference point. When the sustainable indicators are ratio-scale and the distance to a reference method is used for normalization, the CSI derived from the simple additive weighting aggregation function are found to be meaningful (Ebert and Welsch 2004). In operation, the method of distance to a reference needs to first determine the reference point, which could be the leader of the entities (Zhou et al. 2006a) or an external benchmark (Nardo et al. 2008). A popular practice in application is to use the base time as a reference so that the sustainability performance of entities could be dynamically monitored. Examples of such studies can be found in Kang (2002), Kang et al. (2002), Krajnc and Glavič (2005) and

Table 1 The commonly used normalized methods

Method	Formula	Notes
Z-score	$v_{ij} = \frac{(x_{ij} - \bar{x}_j)}{\sigma_j}$	\bar{x}_j and σ_j are the mean and standard deviation of the value of indicator x_{ij} , respectively. v_{ij} is the normalized value of indicator x_{ij}
Distance to a reference	$v_{ij} = \frac{x_{ij}}{x_{ij}^r}$	x_{ij}^r is the value of reference indicators
Re-scaling	$v_{ij} = \frac{x_{ij} - \min_i x_{ij}}{\max_i x_{ij} - \min_i x_{ij}}$	$\max_i x_{ij}$ and $\min_i x_{ij}$ are the maximum and the minimum values of indicator x_{ij} across all entities

Cherchye et al. (2007a). In the circumstance, it is also possible to make meaningful comparison over time when panel data is involved (Cherchye et al. 2007a).

Re-scaling method attempts to re-scale the original indicators to dimensionless range [0, 1] by using the global maximum and minimum. One well-known example is the Human Development Index. Other examples using the re-scaling method can be found in Neumayer (2001), Díaz-Balteiro and Romero (2004), Hajkowicz (2006), Gómez-Limón and Riesgo (2009), and Gómez-Limón and Sanchez-Fernandez (2010).

3.2 Aggregation Methods

The aggregation process implies the search for an appropriate function that can incorporate multiple indicators into a single composite indicator. In the literature, there are many aggregation functions available for use. Most of them can be represented by the following equation:

$$CI_i = \begin{cases} \left[\sum_{j=1}^n w_j (v_{ij})^\beta \right]^{\frac{1}{\beta}} & \text{for } \beta \neq 0 \\ \prod_{j=1}^n v_{ij}^{w_j} & \text{for } \beta = 0 \end{cases} \quad (3)$$

When the parameter β is assigned to different values, the aggregation function will collapse to different forms. It should be pointed out that the values of β have an impact on the trade-offs between different indicators. More discussions can be found in Decancq and Lugo (2013). Table 2 shows several aggregation functions which are often used in constructing CSI.

Table 2 Several popular aggregation methods

Function name	Formula
Simple additive weighting	$CI_i = \sum_{j=1}^n w_j v_{ij}$ when $\beta = 1$
Weighted product	$CI_i = \prod_{j=1}^n v_{ij}^{w_j}$ when $\beta = 0$
Weighted displaced ideal	$CI_i = (1 - \lambda) \cdot \min w_j v_{ij} + \lambda \sum_{j=1}^n w_j v_{ij}$ (λ is the compensatory parameter that ranges from 0 to 1)
Social multi-criterion evaluation method	$CI_i = \sum_{j=1}^m e_{jk} (e_{jk} = \sum_{i=1}^m w_i (P_{jk}) + \frac{1}{2} w_i (I_{jk}))$

3.2.1 Simple Additive Weighting

When the parameter $\beta = 1$, Eq. (3) reduces to the simple additive weighting (hereafter referred to as SAW). In the context of constructing CSI, the simple additive weighting might be the most commonly used aggregation function, e.g. Kang (2002), Kang et al. (2002), Krajnc and Glavič (2005), Esty et al. (2005), Hajkowicz (2006), Singh et al. (2009), Murillo et al. (2015), and Global Warming Potential (IPCC 2001). The SAW method is easy to understand and can visualize the relative contribution of each indicator to the CSI. Since the assumption of preferentially independent relationship between indicators may not be satisfied in practice, statistical techniques such as principal component analysis (PCA) and factor analysis (FA) are often applied before aggregation (Grupp and Schubert 2010). Additionally, the use of SAW allows for the full substitutability between the indicators so that the weights imply trade-offs, which is inconsistent with the meaning of importance coefficients quoted by many earlier studies (Munda and Nardo 2009). From a practical point of view, this characteristic of the SAW method is not desirable since it violates the spirit of sustainable development (Ayres et al. 1998).

3.2.2 Weighted Product Method

With the parameter $\beta = 0$, Eq. (3) is referred to as weighted product (WP) method. Although WP method is not widely applied in constructing CSI, it has attracted much attention owing to its several desirable characteristics, e.g. semi-compensatory property (Nardo et al. 2008), meaningfulness for ratio-scale indicators (Ebert and Welsch 2004; Böhringer and Jochem 2007), and less information loss (Zhou and Ang 2009). In application, WP method has been used for constructing HDI to replace SAW method by the United Nations Development Programme, which could be affected by these earlier studies as discussed by Tofallis (2013).

Due to the exponent property, WP method requires that all ratings are greater than one (Yoon and Hwang 1995). The relative contribution by each individual indicator to the CSI is not visualized as that in SAW. Furthermore, the results usually do not have a numerical upper bound. The former problem can be solved by multiplied by 10^l . The later problem, Yoon and Hwang (1995) suggested to compute the distance between each entity and the ideal entity as follows:

$$R_i = \frac{\prod_{j=1}^n v_{ij}^{w_j}}{\prod_{j=1}^n (v_{ij}^*)^{w_j}} \quad (i = 1, \dots, m) \quad (4)$$

where v_{ij}^* is the best value for the j th indicators. It is clear that $0 \leq R_i \leq 1$, in which 1 (0) indicates the most (least) sustainable entity.

3.2.3 Weighted Displaced Ideal

Weighted displaced ideal (WDI) method is on the basis of the ideal solution theory that aims to calculate the distance between the normalized value of each entity and the “ideal” entity (Zeleny and Cochrane 1981). This concept has further been generalized by Díaz-Balteiro and Romero (2004) which can provide solutions of “total compensability” among the sustainable indicators and “total non-compensability” of the indicators, as well as a compromise set of solutions between these two extreme cases. By setting $\beta \rightarrow +\infty$, Eq. (3) is transformed to the form $CI_i = \min w_j v_{ij}$, in which substitutability between indicators are prohibited. To reach a balanced evaluation, Díaz-Balteiro and Romero (2004) introduced a parameter λ representing the degree of substitutability between indicators. When $\lambda = 0$, non-compensatory between indicators is assumed. When $\lambda = 1$, the WDI method will be simplified as the SAW method which assumes full compensability. For $0 < \lambda < 1$, partial compensation between indicators will be allowed. This aggregation function is also attracted an increasing attention, e.g. Zhou et al. (2006a), Zhou and Ang (2009), Gómez-Limón and Riesgo (2009), Blancas et al. (2010), Gómez-Limón and Sanchez-Fernandez (2010), and Pollesch and Dale (2015).

3.2.4 Social Multi-criterion Evaluation Method

Social multi-criteria evaluation (SMCE) method, introduced by Munda (2005), is a non-compensatory technique to provide rankings of the entities based on a Condorcet-type of aggregation procedure. SMCE aims to improve the quality of composite indicators by overcoming two technical weaknesses: independence between indicators and the meaningfulness of weights (Munda and Nardo 2003). Once the weights are determined, SMCE undergoes two steps to obtain the overall sustainability rankings. At first, an outranking matrix is built by pair-wise comparison. Elements ($e_{jk} (j \neq k)$) in the matrix is the score of the sum of the weights for corresponding indicators under the condition of indicator j performing better than indicator k . A half of the weight will be added if the relationship between indicator j and k is indifference. Such process can be expressed by

$$e_{jk} = \sum_{i=1}^m (w_i (P_{jk}) + \frac{1}{2} w_i (I_{jk})) \quad (5)$$

where $P_{jk} (I_{jk})$ indicates preference (indifference) relationship. In this step, $n(n - 1)$ pair-wise combinations need to be compared. The second step is to sum up the relevant scores for a complete pre-order of entities. For instance, to rank three entities (e.g. E1, E2, and E3), all possible permutations of these entities are E1E2E3, E1E3E2, E2E1E3, E2E3E1, E3E1E2, and E3E2E1. Then the value of each permutation can be calculated. The one with the highest score is used to determine the ranking of the entities.

Compared with the aggregation methods described above, SMCE requires a large amount of computation, especially in the second step. The information including in the final results might also be limited. Nevertheless, SMCE provides a novel framework for assessing sustainability. When CSIs are derived from SMCE, their underlying subjective aspect only comes from the determination of indicator's weight. This property decreases the uncertainty in constructing CSI and also relieves some burden in sensitivity analysis. Besides, SMCE is a totally non-compensatory aggregation method which may better reflect the concept of strong sustainability.

3.3 Weighting Methods

From the previous section, it is clear that there is close relationship between the indicator weights and data aggregation. The existing weighting methods could be partitioned into three categories: exogenous (or called normative) methods, endogenous (or called data-driven) methods, and hybrid methods. The main difference between the three categories lies in the degree of value judgement of decision-makers or experts involved in determining the weights. Exogenous methods, mainly dependent on the value judgement of decision-makers or experts are determined by participatory methods. Endogenous methods on the other hand mainly rely on the data distribution, tending to let data "speak". Hybrid methods attempt to balance the exogenous and endogenous methods.

3.3.1 Exogenous Methods

Equal weighting, arbitrary weighting, and analytic hierarchy process (AHP) are three frequently used exogenous methods. See, for example, Ecological Footprint, Hope et al. (1992) and Murillo et al. (2015) for equal weighting, Kang (2002), Kang et al. (2002) and Krajnc and Glavič (2005) for AHP. Equal weights are usually applied in the circumstance of the absence of comprehensive understanding for the entity. With the improvement of data collection techniques and the extensive research on sustainability, equal weights have gradually been abandoned in constructing CSI. AHP and other methods like budget allocation processes and conjoint analysis are heavily dependent on a thorough understanding about how each entity works. The challenge in the application of these exogenous methods is the choice of appropriate experts (Decancq and Lugo 2013). Once this problem is properly handled, the reliability of the exogenous methods would significantly increase. It should be noted that exogenous methods are ex ante approaches, which makes performance comparison across time and space be feasible.

3.3.2 Endogenous Methods

Statistical weighting and BOD methods are two major endogenous families in determining the weights for constructing CSI. Statistical weighting methods are based on statistical properties of the data, e.g. PCA, FA, and regression analysis (RA). PCA is basically a multivariate statistical technique to summarize the data. FA is based on the assumption that some observed indicators rely on a certain number of unobserved factors. Although the basic assumption of these two methods is distinct, in practice, one usually does not distinguish the difference. Once the principal components are extracted, the factor loading matrix and eigenvalues of the associate principal components can be calculated. The weights of indicators then equal to the ratio of squared factor loadings to the corresponding eigenvalues. The technical details can be found in Gómez-Limón and Riesgo (2009). Despite their statistical soundness, the meaning of the weights estimated by PCA or FA fails in accordance with the original meaning, i.e. importance, since these two methods measure the overlapping information between two or more correlated indicators (Shen et al. 2013). RA approach determines weight by multiple regression or linear programming and assumes that individual indicator relies on the sum of an observed variable and an error term. Thus, the RA approach is usually defined as unobserved components model or observed derived weight method (Nardo et al. 2008).

3.3.3 Hybrid Methods

In addition to assess sustainability, one valuable characteristic of CSI is to compare among all entities so that decision makers can detect the gap and then take actions to improve sustainability performance. From this perspective, endogenous methods somewhat fail to perform this function. Exogenous methods do not suffer the issue since they do not rely on the data distribution. However, exogenous methods depend on the value judgments that might be affected by different expert panels. As mentioned by Decancq and Lugo (2013), expert groups might be underrepresented or simply uninformed resulting in a skewed weighting scheme.

Hybrid methods are proposed to combine exogenous methods with endogenous methods. Decancq and Lugo (2013) listed two hybrid methods, namely stated preference weight and hedonic weights. Stated preference weight, instead of imposing weights by expert panel, is directly based on individual opinions. Hedonic weights also rely on the individual self-reported preference. After obtaining the preference matrix, weights can be estimated by a linear regression. An example of hybrid weighting methods is the BOD model with weight restrictions that can be determined by experts.

4 BOD Methods

Two problems in constructing CSI by MADM models are the information loss caused by normalization (Zanella et al. 2015) and the subjectivity in determining the weights. Fixed weight stemming from MADM models has been controverted with argument that different cultural and social settings value individual weighting framework in different ways (Cherchye et al. 2008).

Alternatively, as suggested by Lovell et al. (1995) and Lovell (1995), linear programming models can be used to construct the ‘best practice’ frontier for the entities. The linear programming approach for constructing CSI is usually defined as benefit of the doubt (BOD). BOD roots in DEA which was originally proposed for evaluating the relative efficiency of a homogeneous set of entities which use multiple inputs to produce multiple outputs. In DEA, the weights of inputs and outputs can be endogenously determined by raw data without using price information. In methodology, CSI based on BOD borrow the idea of DEA for the purpose of weighting and aggregation.

4.1 Basic BOD Model

For each of entities, the basic BOD model explores its most favorable weights (Cherchye et al. 2007b). It can be formulated as follows:

$$\begin{aligned}
 CI_i &= \max \sum_{j=1}^n w_j x_{ij} \\
 s.t. \quad &\sum_{j=1}^m w_j x_{ik} \leq 1 \quad k = 1, \dots, n \\
 &w_j \geq 0
 \end{aligned} \tag{6}$$

Model (6) is equivalent to the input oriented DEA model with the assumption of constant returns to scale and a dummy input for all the evaluated entities. It provides the optimal aggregated performance values for all entities by solving the model n times. Different from the MADM models, the weight assignment based on BOD adheres to a posterior weighting scheme and the weights of the individual indicators weight may differ between entities. Model (6) holds several desirable properties, such as normalization-free and the invariance with respect to ratio scale transformations (Athanasoglou 2015). Normalization-free can avoid the information loss caused by data transformation. Invariance allows practitioners to aggregate individual indicators into a meaningful composite indicator (Ebert and Welsch 2004). In essence, model (6) measures how far the evaluated entity is from the best practice entity under most favorable weights (Zhou et al. 2007).

Model (6) has been used in many application contexts. The earliest literature may date back to Mahlberg and Obersteiner (2001) who introduced model (6) to

reassess HDI. Despotis (2005a, b) used an extension to model (6) to reevaluate HDI. It is worth pointing out that Cherchye and his collaborators applied the model in several backgrounds including sustainable development (Cherchye and Kuosmanen 2004), internal market (Cherchye et al. 2007a) and technology achievement (Cherchye et al. 2008).

4.2 Weight Restriction in Basic BOD Model

The BOD model brings new perspective for constructing CSI, while it also suffers from some shortcomings. For example, model (6) assumes that the weights are nonnegative. It is possible that all the weights are assigned to a single indicator which may not be expected since all the selected indicators are theoretically importance and thus need to be considered (Zhou et al. 2007). Besides, it could open up the debate on the CSI's credibility and acceptability. To overcome the problems, it is appropriate to restrict weights in certain ways. A straightforward way is to introduce non-Archimedean infinitesimal variable ε into the model, e.g. Despotis (2005a, b) and Kao (2010). With such modification, however, it is still possible to diagnose an entity well performing even if it is only superior with respect to one indicator but performs poorly with respect to the remaining indicators (Mahlberg and Obersteiner 2001). Hence, further restrictions on weights usually are considered in practice.

Broadly speaking, weight restriction could be classified into two categories, i.e. direct restriction and indirect restriction (Allen et al. 1997). Direct restriction on weights could be formulated in the forms of Eqs. (7) and (8) which are respectively termed as "Type I Assurance Regions" and "Type II Assurance Regions" by Thompson et al. (1990). The Greek letters in Eqs. (7) and (8) are specified by decision makers to reflect their preference regarding the relative importance of indicators. w' could be the combination of weights. The use of direct restriction on weights can be found in Mahlberg and Obersteiner (2001), Cherchye and Kuosmanen (2004), and Cherchye et al. (2007b). Indirect restrictions on weights could be formulated as the form of Eq. (9) which was originally proposed by Wong and Beasley (1990). ϕ and φ also indicate the preference of decision-makers. Rather than restricting actual weights, Eq. (9) places lower and upper bounds on the relative contribution of each indicator to the entity's aggregate performance value. This restriction method has been adopted in many previous studies, e.g. Zhou et al. (2007, 2010), Cherchye et al. (2008), Zanella et al. (2015) and Athanassoglou (2015).

$$\alpha_j \leq \frac{w_j}{w_{j+1}} \leq \beta_j \quad (7)$$

$$\lambda w' \leq \kappa w_j \leq \gamma w' \quad (8)$$

$$\phi_j \leq \frac{w_j x_{ij}}{\sum_{j=1}^n w_j x_{ij}} \leq \varphi_j \quad (9)$$

The above restrictions cannot only overcome the problem aforementioned but also introduce “valued judgment” to incorporate prior views or information in assessing the performance of entities. The prior information can be incorporated via the determination of boundaries by MADM method such as AHP, BAP and the social surveys, e.g. Cherchye et al. (2008). Direct restriction usually incorporates information of marginal rates of substitution between indicators which is sensitive to the units of measurement (Allen et al. 1997). Consequently, it is often difficult to specify meaningful substitution in real-life applications (Zanella et al. 2015). In contrast, indirect restriction method holds the desirable property of ratio-scale invariance (Cherchye et al. 2008; Zhou et al. 2007). This is particularly compelling in the case of constructing environmental performance index (Ebert and Welsch 2004). Furthermore, as Cherchye et al. (2008) discussed, Eq. (9) can be expressed as pie share constraints which are pure numbers and can be easily grasped by decision makers. Nevertheless, the meaning of Eq. (9) is not so straightforward since the implied restrictions on weights are entity-specific. Hence, Wong and Beasley (1990) suggested several modifications. One of the modifications, i.e. replacing x_{ij} with $\sum_{j=1}^n \frac{x_{ij}}{n}$ in Eq. (9), which represents the level of the i th indicator of the “average” entity, has also been applied in constructing composite indicators, e.g. Zanella et al. (2015).

4.3 Extensions of Basic BOD Model

Due to its striking properties, BOD model has been extended to solve various problems in constructing CSI, e.g. hierarchy problem, compensability, comparability, etc.

The basic BOD model usually treats all the indicators at the same level and thus leaves out the information of the hierarchical structure of indicators. This hierarchy problem might be unrealistic due to the fact that multiple layer indicator framework is constructed in order to evaluate the increasing complicated sustainable performance in a more comprehensive way. According to Becker (2005), frameworks are mostly hierarchical extending from broad categories of data and information to detailed measures. To overcome this limitation, Shen et al. (2013) improved the basic BOD model to fit the property of hierarchical indicator system by specifying weights in each category of each layer. More straightforwardly, in the situation of multiple hierarchical indicator framework, the basic BOD model is first used to determine the “best practice” performance of certain layer indicators. Then the aggregation of higher layer indicators can be done by MADM methods. See, for example, Kao et al. (2008).

In addition, due to the linear characteristic of its objective function, BOD also faces the compensatory issue as discussed earlier. Munda and Nardo (2009) suggested that it is compulsory to construct non-compensatory composite indicators so that weights are theoretical consistent with the meaning of importance. To relax the compensatory characteristic, Zhou et al. (2010) combined the WP aggregation method with basic BOD model to construct a multiplicative optimization approach with semi-compensatory characteristic to reach a compromise solution. Pakkar (2014) proposed a similar model for constructing Technology Achievement Index. Fusco (2015) introduced directional penalties to enhance the non-compensatory characteristic of basic BOD model to take into account the preference structure among indicators. Generally, the methods take similar perspective, i.e. imposing more penalties upon the indicators with worse performance.

The basic BOD model on the basis of conventional DEA technique distinguishes efficient and inefficient entities in the DEA terminology, and is not suitable for ranking the performance of entities (Kao 2010). The main strength of basic DEA models is to recognize the inefficient entities. Hence, many studies have been devoted to improve the comparability of basic BOD model under the framework of composite indicators. For example, from an opposite perspective of Model (6), Zhou et al. (2007) proposed a model to seek the “worst” set of weights for each entity, and use an adjusting parameter to combine the “best practice” and the “worst practice” to form composite indicators. Several studies have adopted this model to construct composite indicators in various contexts, e.g. Domínguez-Serrano and Blancas (2011), Rogge (2012), and Blancard and Hoarau (2013). Athanassoglou (2015) further improved the worst-case of basic BOD model for constructing composite indicators.

Besides, the concept of common-weight is also applied to enhance the comparability of basic BOD model. Its basic idea is that every entity need to use the same benchmark for calculating the performance score. Despotis (2005a, b) initially introduced the concept of common-weight, in which basic BOD model is firstly used to determine most favorable weights for entities and then a goal programming model is developed to discriminate entities with the same performance score. Dong et al. (2015) used similar two-stage method to measure farm sustainability. Kao et al. (2008) proposed a similar two-stage model for evaluating the national competitiveness. Kao (2010) combined the concept of common-weights with Malmquist productivity index. Built upon Zhou et al. (2007), Hatefi and Torabi (2010) also proposed a common-weights MCDA-DEA approach in which the common-weights are calculated in one step. Tofallis (2013) also used two-stage model to seek a common set of weights to apply to all entities. More recently, Hatefi and Torabi (2016) further analyzed how to improve the composite indicators of inefficient entities on the basis of a slack analysis framework.

BOD methods are flexible and systematic for constructing CSI. In recent years, conventional DEA models are also used to establish composite indicators, e.g. environmental performance index (EPI). Application of conventional DEA models to construct EPI might begin with the establishment of environmental production technology (Zhou and Ang 2008). Then an EPI can be constructed by different types

of DEA models with different properties. See, for example, Zaim et al. (2001), Zhou et al. (2006b), Zhou and Ang (2008), Blancard and Hoarau (2013), and Wang (2015). More recently, Zhou et al. (2017) evaluated previous studies and showed that the range adjusted DEA model can generate a cardinally meaningful composite index.

5 The CSI Robustness and Beyond

5.1 Selection Principle

So far, we have examined three methodological aspects pertinent to CSI construction. Besides, two additional issues have often been questioned, i.e. comparability and meaningfulness. Comparability is mainly caused by the incommensurability of indicators' measurement units. Martinez-Alier et al. (1998) theoretically showed that the incommensurability does not imply incomparability but weak comparability, which means that there is a good potential for applying multi-indicator evaluation methods (e.g. MADM and BOD) to sustainability assessment. Although the above argument provides theoretical comparability foundation, sustainability assessment still faces the difference and ambiguity caused by measurement units, which may make CSI meaningfulness.

Ebert and Welsch (2004) first discussed how to construct a meaningful environmental index, which has been used as a criterion for investigating whether an environmental or sustainable index is meaningful or not by Böhringer and Jochem (2007) and Singh et al. (2009). Meaningful CSI indicates that the preference orderings does not vary with different scale of underlying indicators. Ebert and Welsch (2004) classified different scales into four categories according to the concept of comparability (measurability) of scales: interval-scale non-comparability, interval-scale full comparability, ratio-scale non-comparability, and ratio-scale full comparability. If interval-scaled indicators are full comparable, the arithmetic mean aggregation function satisfies continuous, strongly monotone, and separable properties and thus can generate a meaningful index. If ratio-scaled indicators are non-comparable, geometric mean aggregation function is recommended. Table 3 provides a summary of different cases. It should be pointed out that it is impossible to construct meaningful CSI when there exist indicators with distinct measurement scales (Böhringer and Jochem 2007; Ebert and Welsch 2004). More recently, Pollesch and Dale (2015) investigated aggregation functions for six different scales of indicators in constructing an appropriate meaningful CSI. Zhou et al. (2017) generalized the meaningfulness concept by Ebert and Welsch (2004) and showed how to construct a cardinally meaningful index.

In addition to scales, many other factors can have impact on the selection of aggregation function in constructing CSI, e.g. interactive phenomena between indicators, the types of weight, and the assumption of sustainability. When there are interactive phenomena between indicators, some preliminary treatments should be firstly conducted to eliminate those interactive relationships. However, as Mayer

Table 3 Aggregation rules for indicators by Ebert and Welsch via Böhringer and Jochem (2007)

	Non-comparable	Full comparable
Interval scale	Dictatorial ordering	Arithmetic mean
Ratio scale	Geometric mean	Any homothetic function

(2008) stated, without a clear understanding of interactive relationship between indicators and how those relationships influence the results, it is hard for decision makers to formulate policy with the aim to increase economic equity, environmental improvement, and further increase possibilities for long-term sustainability. Hence, those aggregation methods taking interaction into consideration, e.g. Choquet integral with fuzzy measure, might be a good choice. The types of weight can also have impact on the application of aggregation function. For example, weights, no matter on which weighting methods, can be classified into two categories: ordinal and cardinal ones. Ordinal weights usually cannot be handled well by compensatory aggregation methods. In this situation, non-compensatory approach may be an appropriate choice. In addition, the assumption of sustainability theoretically determines the choice of aggregation algorithm (Munda 2005). There are usually two economic paradigms of sustainability: weak sustainability and strong sustainability (Dietz and Neumayer 2007; Neumayer 2013). From weak sustainability perspective, natural capital is considered to be substitutable. In this view, those compensatory aggregation algorithms might be suitable. From the perspective of strong sustainability, natural capital is regarded as non-substitutable. Then, the non-compensatory or semi-compensatory aggregation schemes may be more appropriate.

In general, we may summarize the procedure for selecting an appropriate approach to constructing CSI as follows. First and most importantly, economic paradigms (weak or strong sustainability) should be clearly defined, based on which either compensatory or non-compensatory aggregation scheme can be determined. The indicator framework following the definition of different paradigms can also be established. With the premise of indicator framework, practitioners can check the scales of indicators and assign weight for each indicator. Once indicator framework show the property of the same measurement scales, the procedure can continue. Otherwise, indicators with different scales should be replaced by other proxy indicators with the same scale. Additionally, there are two other factors that should be considered, namely interactive phenomena between indicators and hierarchical structure. If practitioners decide to model the interactive relationship between indicators, the way for assigning weights to indicators might be different on which the selection of aggregation scheme will directly be influenced.

5.2 Uncertainty and Sensitivity Analysis

It must be acknowledged that each method has its own merits. However, as Booyesen (2002) discussed, every element of methods used to construct composite indicators cannot escape from criticism. The disagreements originate from many

facets, and one main source is the robustness of CSI. Theoretically, different combinations of methods can be used to construct CSI which implies that it is possible to derive very different results.

Two alternative approaches are used in constructing CSI to increase their robustness. One is to ensure the transparency of the whole construction process. This requires vivid statement of the models including those important aspects, such as mathematical and descriptive properties. In addition, the way by which such models are used and integrated in a decision process still needs to be elaborated clearly. The other approach is to assess the uncertainties by sensitivity analysis. Sensitivity analysis can answer the question why those entities with similar sustainable performance get distinct rankings, and can also be used to globally analyze the variation in CSI when different aspects vary over a reasonable range of possibilities (Saisana et al. 2005; Munda and Saisana 2011). For instance, Zhou et al. (2010) compared different rankings of entities obtained by a large set of randomly chosen weighting schemes. Munda and Saisana (2011) analyzed the stability of sustainability rankings by different aggregation rule while keeping the weights of indicators unchanged.

Keeping transparency and conducting sensitivity analysis are posterior uncertainty analysis. Correspondingly, there are also a priori uncertainty analysis methods, e.g. the Shannon-Spearman measure (SSM) developed by Zhou et al. (2006a, b) and Zhou and Ang (2009). SSM is based on the concept of information loss in the process of aggregating underlying indicators into a composite index. Intuitively, methods with smaller SSM, i.e. less loss of information, may be regarded as better ones. Methods with zero SSM are deemed inheriting full information, and thus are regarded as perfect model. In this sense, SSM might be another approach for uncertainty evaluation of CSI.

5.3 *Beyond Rankings*

Although CSI intuitively provides the index values and ranking results, as emphasized by Nardo et al. (2008) and Grupp and Schubert (2010), it can be a means of initiating discussion to facilitate communication between different stakeholders. The influential CSI can draw the attention from policy makers towards the importance of sustainable development. Its intuitive construction also provides opportunity to uncover the debate for the public, instead of excluding them straightaway. Besides, CSI may help to stir policy competition about best practice in sustainable development policies and become a useful monitoring tool to avoid unintended consequence caused by unsuitable policies. The process of constructing CSI also provides the possibility of further analyzing the questions. For example, where is the strength? Which aspect of the entity should be improved? What is the real contribution of certain indicators to CSI? The information hidden in the CSI can be visually exhibited with the help of spider diagrams or radar charts, by which the strengths and weaknesses can be easily and intuitively represented. The correlation analysis

between underlying indicators and the values of CSI can illustrate the contribution of each indicator, and then help identify the priority of improvement.

6 Conclusions

This chapter provides a state-of-the-art review of CSI construction with focus on the methodological developments. We firstly introduce the general structure of CSI construction. Then, we classify the methods for constructing CSI into two groups, i.e. MADM and BOD. In MADM, methods for normalization, weighting and aggregation together with their pros and cons are respectively discussed. It is found that z-score normalization scheme, hybrid weighting methods and compensatory/semi-compensatory aggregation functions are most commonly used in application. Non-compensatory aggregation scheme has received increasing attention by some recent studies. In BOD, the basic BOD model, weight restriction and other extensions are described. A new trend is that analysts tend to incorporate various MADM methods into BOD to construct CSI. Finally, we investigate the principles for selecting appropriate aggregation methods in constructing CSI. Uncertainty and sensitivity analysis have also been discussed in order to establish CSI with robustness.

CSI has evolved as a popular tool for the purpose of monitoring sustainability performance and providing valuable information for supporting policy analysis and decision making. However, various challenges still exist, e.g. the conceptual issue of sustainability, dimensional diversity, data availability and so on. The widely accepted definition of sustainability includes the impacts on the next generation, which implies that it is important to incorporate the influence of time and geographical factor. When taking geographical factors into account, practitioners may also need to consider entities' different culture and development patterns. There is also a rising concern on how to construct a meaningful CSI from both theoretical and methodological perspectives, as the existing CSIs seldom satisfy the axiomatic requirements of the meaningfulness definition. In this sense, further efforts are still required to improving the meaningfulness and robustness of existing CSIs.

Acknowledgements We are grateful to the financial support provided by the National Natural Science Foundation of China (nos. 71273005 & 71573119) and the Funding of Jiangsu Innovation Program for Graduate Education (KYZZ16_0159).

References

- Allen, R., Athanassopoulos, A., Dyson, R.G., Thanassoulis, E., 1997. Weights restrictions and value judgements in data envelopment analysis: Evolution, development and future directions. *Annals of Operations Research* 73, 13–34.
- Athanassoglou, S., 2015. Revisiting worst-case DEA for composite Indicators. *Social Indicators Research* 128(3), 1259–1272.

- Ayres, R.U., van den Bergh, J.C.J.M., Gowdy, J., 1998. Viewpoint: Weak versus strong sustainability. Tinbergen Institute Discussion Paper (No. 98-103/3).
- Baumgärtner, S. Quaas, M., 2010. What is sustainability economics? *Ecological Economics* 69(3), pp. 445–450.
- Becker, J., 2005. Measuring progress towards sustainable development: An ecological framework for selecting indicators. *Local Environment* 10, 87–101.
- Blancard, S., Hoarau, J.-F., 2013. A new sustainable human development indicator for small island developing states: A reappraisal from data envelopment analysis. *Economic Modelling* 30, 623–635.
- Blancas, F. J., Caballero, R., González, M., Lozano-Oyola, M., Pérez, F. 2010. Goal programming synthetic indicators: An application for sustainable tourism in Andalusian coastal counties. *Ecological Economics* 69(11), 2158–2172.
- Booysen, F., 2002. An overview and evaluation of composite indices of development. *Social Indicators Research* 59, 115–151.
- Botta, E., Kozluk, T. 2014. Measuring environmental policy stringency in OECD countries: A composite index approach. OECD Economics Department Working Papers, No. 1177, OECD Publishing, Paris.
- Böhringer, C., Jochem, P.E., 2007. Measuring the immeasurable—A survey of sustainability indices. *Ecological Economics* 63, 1–8.
- Cai, H., Chen, Y., Gong, Q., 2016. Polluting thy neighbor: Unintended consequences of China's pollution reduction mandates. *Journal of Environmental Economics and Management* 76, 86–104.
- Chen, S., Chen, X., Xu, J., 2016. Impacts of climate change on agriculture: Evidence from China. *Journal of Environmental Economics and Management* 76, 105–124.
- Cherchye, L., Kuosmanen, T., 2004. Benchmarking sustainable development: A synthetic meta-index approach. Research Paper, UNU-WIDER, United Nations University.
- Cherchye, L., Lovell, C.K., Moesen, W., Van Puyenbroeck, T., 2007a. One market, one number? A composite indicator assessment of EU internal market dynamics. *European Economic Review* 51, 749–779.
- Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T., 2007b. An introduction to 'benefit of the doubt' composite indicators. *Social Indicators Research* 82, 111–145.
- Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T., Saisana, M., Saltelli, A., Liska, R., Tarantola, S., 2008. Creating composite indicators with DEA and robustness analysis: The case of the Technology Achievement Index. *Journal of the Operational Research Society* 59, 239–251.
- Decancq, K., Lugo, M.A., 2013. Weights in multidimensional indices of wellbeing: An overview. *Econometric Reviews* 32, 7–34.
- Despotis, D., 2005a. Measuring human development via Data Envelopment Analysis: The case of Asia and the Pacific. *Omega* 33, 385–390.
- Despotis, D., 2005b. A reassessment of the Human Development Index via Data Envelopment Analysis. *Journal of the Operational Research Society* 56, 969–980.
- Díaz-Balteiro, L., Romero, C., 2004. In search of a natural systems sustainability index. *Ecological Economics* 49, 401–405.
- Dietz, S., Neumayer, E., 2007. Weak and strong sustainability in the SEEA: Concepts and measurement. *Ecological Economics* 61, 617–626.
- Domínguez-Serrano, M., Blancas, F.J., 2011. A gender wellbeing composite indicator: The best-worst global evaluation approach. *Social Indicators Research* 102, 477–496.
- Dong, F., Mitchell, P.D., Colquhoun, J., 2015. Measuring farm sustainability using data envelope analysis with principal components: The case of Wisconsin cranberry. *Journal of Environmental Management* 147, 175–183.
- Ebert, U., Welsch, H., 2004. Meaningful environmental indices: A social choice approach. *Journal of Environmental Economics and Management* 47, 270–283.

- Esty, D.C., Levy, M., Srebotnjak, T., De Sherbinin, A., 2005. Environmental sustainability index: benchmarking national environmental stewardship. New Haven: Yale Center for Environmental Law & Policy.
- Floridi, M., Pagni, S., Falorni, S., Luzzati, T., 2011. An exercise in composite indicators construction: Assessing the sustainability of Italian regions. *Ecological Economics* 70, 1440–1447.
- Freudenberg, M. 2003. Composite indicators of country performance: A critical assessment, OECD Science, Technology and Industry Working Papers, No. 2003/16, OECD Publishing, Paris.
- Fusco, E., 2015. Enhancing non-compensatory composite indicators: A directional proposal. *European Journal of Operational Research* 242, 620–630.
- Gómez-Limón, J.A., Riesgo, L., 2009. Alternative approaches to the construction of a composite indicator of agricultural sustainability: An application to irrigated agriculture in the Duero basin in Spain. *Journal of Environmental Management* 90, 3345–3362.
- Gómez-Limón, J.A., Sanchez-Fernandez, G., 2010. Empirical evaluation of agricultural sustainability using composite indicators. *Ecological Economics* 69, 1062–1075.
- Grupp, H., Schubert, T., 2010. Review and new evidence on composite innovation indicators for evaluating national performance. *Research Policy* 39, 67–78.
- Hák, T., Moldan, B., Dahl, A.L., 2012. Sustainability indicators: A scientific assessment. Island Press.
- Hajkowicz, S., 2006. Multi-attributed environmental index construction. *Ecological Economics* 57, 122–139.
- Hatefi, S., Torabi, S., 2010. A common weight MCDA–DEA approach to construct composite indicators. *Ecological Economics* 70, 114–120.
- Hatefi, S.M. and Torabi, S.A., 2016. A slack analysis framework for improving composite indicators with applications to human development and sustainable energy indices. *Econometric Reviews*.
- He, G., Fan, M., Zhou, M., 2016. The effect of air pollution on mortality in China: Evidence from the 2008 Beijing Olympic Games. *Journal of Environmental Economics and Management* 79, 18–39.
- Hope, C., Parker, J., Peake, S., 1992. A pilot environmental index for the UK in the 1980s. *Energy Policy* 20, 335–343.
- Hsu, A., de Sherbinin, A. C., Esty, D., Levy, M., 2016. Environmental Performance Index. New Haven, CT: Yale University. Available: www.epi.yale.edu.
- Intergovernmental Panel on Climate Change (IPCC), 2001. Climate Change 2001: The Scientific Basis. Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, UK.
- Kang, S.M., 2002. A sensitivity analysis of the Korean composite environmental index. *Ecological Economics* 43, 159–174.
- Kang, S.M., Kim, M.S., Lee, M., 2002. The trends of composite environmental indices in Korea. *Journal of Environmental Management* 64, 199–206.
- Kao, C., 2010. Malmquist productivity index based on common-weights DEA: The case of Taiwan forests after reorganization. *Omega* 38, 484–491.
- Kao, C., Wu, W.-Y., Hsieh, W.-J., Wang, T.-Y., Lin, C., Chen, L.-H., 2008. Measuring the national competitiveness of Southeast Asian countries. *European Journal of Operational Research* 187, 613–628.
- Krajnc, D., Glavič, P., 2005. A model for integrated assessment of sustainable development. *Resources, Conservation and Recycling* 43, 189–208.
- Lovell, C.A.K., 1995. Measuring the macroeconomic performance of the Taiwanese economy. *International Journal of Production Economics* 39, 165–178.
- Lovell, C.A.K., Pastor, J.T., Turner, J.A., 1995. Measuring macroeconomic performance in the OECD: A comparison of European and non-European countries. *European Journal of Operational Research* 87, 507–518.

- Mahlberg, B., Obersteiner, M., 2001. Remeasuring the HDI by Data Envelopment Analysis. Available at SSRN 1999372.
- Martinez-Alier, J., Munda, G., O'Neill, J., 1998. Weak comparability of values as a foundation for ecological economics. *Ecological Economics* 26, 277–286.
- Mayer, A.L., 2008. Strengths and weaknesses of common sustainability indices for multidimensional systems. *Environment International* 34, 277–291.
- Mori, K., Christodoulou, A., 2012. Review of sustainability indices and indicators: Towards a new City Sustainability Index (CSI). *Environmental Impact Assessment Review* 32, 94–106.
- Munda, G., 2005. Measuring sustainability: A multi-criterion framework. *Environment, Development and Sustainability* 7, 117–134.
- Munda, G., Nardo, M., 2003. On the methodological foundations of composite indicators used for ranking countries. *Ispira, Italy: Joint Research Centre of the European Communities*.
- Munda, G., Nardo, M., 2009. Noncompensatory/nonlinear composite indicators for ranking countries: A defensible setting. *Applied Economics* 41, 1513–1523.
- Munda, G., Saisana, M., 2011. Methodological considerations on regional sustainability assessment based on multicriteria and sensitivity analysis. *Regional Studies* 45, 261–276.
- Murillo, J., Román, J., Suriñach, J., 2015. The business excellence attraction composite index (BEACI) in small areas. Design and application to the municipalities of the Barcelona province. *Applied Economics* 47, 161–179.
- Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Giovannini, E., Hoffmann, A., 2008. Handbook on constructing composite indicators: Methodology and User guide. OECD publishing.
- Ness, B., Urbel-Piirsalu, E., Anderberg, S., & Olsson, L. 2007. Categorising tools for sustainability assessment. *Ecological Economics* 60(3), 498–508.
- Neumayer, E., 2001. The human development index and sustainability—A constructive proposal. *Ecological Economics* 39, 101–114.
- Neumayer, E., 2013. Weak versus strong sustainability: Exploring the limits of two opposing paradigms. Edward Elgar Publishing.
- Oxford City Council (OCC), 2016. Oxford Sustainability Index Report 2016. https://www.oxford.gov.uk/downloads/file/2655/oxford_sustainability_index_2016.
- Pakkar, M.S., 2014. Using Data Envelopment Analysis and Analytic Hierarchy Process to construct composite indicators. *Journal of Applied Operational Research* 6, 174–187.
- Parris, T.M., Kates, R.W., 2003. Characterizing and measuring sustainable development. *Annual Review of Environment and Resources* 28, 559–586.
- Pollesch, N., Dale, V., 2015. Applications of aggregation theory to sustainability assessment. *Ecological Economics* 114, 117–127.
- Pollesch, N., Dale, V., 2016. Normalization in sustainability assessment: Methods and implications. *Ecological Economics* 130, 195–208.
- Rogge, N., 2012. Undesirable specialization in the construction of composite policy indicators: The Environmental Performance Index. *Ecological Indicators* 23, 143–154.
- Saisana, M., Tarantola, S., Saltelli, A., 2005. Uncertainty and sensitivity techniques as tools for the analysis and validation of composite indicators. *Journal of the Royal Statistical Society A* 168, 1–17.
- Sala, S., Ciuffo, B., Nijkamp, P., 2015. A systemic framework for sustainability assessment. *Ecological Economics* 119, 314–325.
- Shen, Y., Hermans, E., Brijs, T., Wets, G., 2013. Data Envelopment Analysis for composite indicators: A multiple layer model. *Social Indicators Research* 114, 739–756.
- Singh, R.K., Murty, H., Gupta, S., Dikshit, A., 2009. An overview of sustainability assessment methodologies. *Ecological Indicators* 9, 189–212.
- Thompson, R.G., Langemeier, L.N., Lee, C.-T., Lee, E., Thrall, R.M., 1990. The role of multiplier bounds in efficiency analysis with application to Kansas farming. *Journal of Econometrics* 46, 93–108.
- Tilman, D., Cassman, K.G., Matson, P.A., Naylor, R., Polasky, S., 2002. Agricultural sustainability and intensive production practices. *Nature* 418, 671–677.

- Tofallis, C., 2013. An automatic-democratic approach to weight setting for the new Human Development Index. *Journal of Population Economics* 26, 1325–1345.
- UNCED, 1992. Rio declaration on environment and development. Report of the United Nations Conference on Environment and Development.
- UNDP (United Nations Development Programme). 2014. Human development report - Sustaining human progress: Reducing vulnerabilities and building resilience. New York.
- van den Bergh, J.C.J.M., van Veen-Groot, D.B., 2001. Constructing aggregate environmental-economic indicators: A comparison of 12 OECD countries. *Environmental Economics and Policy Studies* 4(1), 1–16.
- Wackernagel, M., & Rees, W. 1998. Our ecological footprint: reducing human impact on the earth (No. 9). New Society Publishers.
- Wang, H., 2015. A generalized MCDA–DEA (Multi-Criterion Decision Analysis–Data Envelopment Analysis) approach to construct slacks-based composite indicator. *Energy* 80, 114–122.
- Wong, Y.-H., Beasley, J., 1990. Restricting weight flexibility in Data Envelopment Analysis. *Journal of the Operational Research Society* 41(9), 829–835.
- Yoon, K.P., Hwang, C.-L., 1995. Multiple attribute decision making: An introduction. Sage publications.
- World Commission on Environment Development (WCED), 1987. Our common future. Oxford University Press.
- World Energy Council (WEC), 2016. World Energy Trilemma Index: Benchmarking the sustainability of national energy systems. <https://trilemma.worldenergy.org/reports/main/2016/2016%20Energy%20Trilemma%20Index.pdf>.
- Zaim, O., Färe, R., Grosskopf, S., 2001. An economic approach to achievement and improvement indexes. *Social Indicators Research* 56(1), 91–118.
- Zanella, A., Camanho, A.S., Dias, T.G., 2015. Undesirable outputs and weighting schemes in composite indicators based on Data Envelopment Analysis. *European Journal of Operational Research* 245, 517–530.
- Zeleny, M., Cochrane, J.L., 1981. Multiple criteria decision making. McGraw-Hill.
- Zhou, P., Ang, B. W. 2008. Indicators for assessing sustainability performance. In: *Handbook of Performability Engineering* (905–918). Springer London.
- Zhou, P., Ang, B., 2009. Comparing MCDA aggregation methods in constructing composite indicators using the Shannon-Spearman measure. *Social Indicators Research* 94, 83–96.
- Zhou, P., Ang, B., Poh, K., 2006a. Comparing aggregating methods for constructing the composite environmental index: An objective measure. *Ecological Economics* 59, 305–311.
- Zhou, P., Ang, B. W., Poh, K. L. 2006b. Slacks-based efficiency measures for modeling environmental performance. *Ecological Economics* 60(1), 111–118.
- Zhou, P., Ang, B., Poh, K., 2007. A mathematical programming approach to constructing composite indicators. *Ecological Economics* 62, 291–297.
- Zhou, P., Ang, B., Zhou, D., 2010. Weighting and aggregation in composite indicator construction: a multiplicative optimization approach. *Social Indicators Research* 96, 169–181.
- Zhou, P., Delmas, M.A., Kohli, A., 2017. Constructing meaningful environmental indices: A nonparametric frontier approach. *Journal of Environmental Economics and Management* 85, 21–34.