

## REVIEW OF SOME STATISTICAL METHODS FOR CONSTRUCTING COMPOSITE INDICATORS

E. JIMENEZ-FERNANDEZ

Departament d'Economia, UNIVERSITAT JAUME I, SPAIN

E-mail: jimeneze@uji.es

MARIA J. RUIZ-MARTOS

Departamento de Teoría e Historia Económica, UNIVERSIDAD DE GRANADA, SPAIN

E-mail: ruizmartosmaria@ugr.es

---

Recibido: 31 de Diciembre de 2019

Aceptado: 26 de Enero de 2020

### ABSTRACT

The methodology for the construction process of composite indicators is reviewed in a step-by-step approach ranging from the ex-ante definition of the latent variable that is intended to be measured through the aggregation process. We focus on comparing four statistical aggregations methods in terms of their weighting and aggregation approaches: Distance  $P_2$ , Principal Component Analysis, Data Envelopment Analysis and Mazziotta-Pareto Index. An empirical comparison among them is provided and the composite indicators divergences are discussed.

**Keywords:** Composite indicators, Weighting, Aggregation, DEA-BoD, PCA, Distance  $P_2$ , Mazziotta Pareto Index.

### RESUMEN

La metodología para el proceso de construcción de indicadores compuestos se ha examinado a través de un enfoque gradual que va desde la definición de la variable latente que se pretende medir hasta el proceso de agregación. En particular, nos centramos en la comparación de cuatro métodos de agregación estadística respecto de sus enfoques de ponderación y agregación: Distancia  $P_2$ , Análisis de Componentes Principales, Análisis de Envoltorio de Datos e Índice de Mazziotta-Pareto. Adicionalmente, se proporciona una comparación empírica entre ellos y se examinan las divergencias de los indicadores compuestos.

**Palabras clave:** Indicadores compuestos, Ponderación, Agregación, DEA-BoD, PCA, Distancia  $P_2$ , Índice de Pareto de Mazziotta.

JEL Classification Codes: I30, C13, C51, C55

## 1 INTRODUCTION

In social sciences, the use of indicators is ever spreading. Indicators, single and composite, aim to measure some concept or latent variable. Most socioeconomic phenomena are multidimensional, which renders a single indicator unable to capture the inherent complexity in, for example, development, poverty, well-being (Maggino, 2017; Greco et al. 2019), and favors a multi-indicator approach. Composite indicators, which synthesize the information conveyed by a wide range of indicators, constitute a popular alternative. The most well-known example of composite indicators of human well-being is the Human Development Index (HDI) by the United Nations Development Programme (UNDP 1990, 2001, 2010).

Constructing a composite indicator, however, goes beyond the purely mathematical operation involved in reducing data dimensionality (Mazziotta and Pareto, 2018). The construction of composite indicators should follow a respectful methodological approach to ensure that the *big picture* fundamentally captures what it is meant to (OECD, 2008). The methodological process to construct a composite indicator starts with the precise definition of the conceptual framework (a *defined process* of measurement, Maggino, 2017, p.87), which conditions the selection of single indicators that (attempt to) measure the various dimensions of the concept and the aggregation method - differential weighting allowed- of the resulting system of indicators, and finishes with the robustness analysis of the composite indicator. This measurement process inevitably involves some subjective choices whose consequences should be clearly stated by the researcher (Maggino, 2017, p.89).

There are different aggregation approaches for constructing composite indicators. We can distinguish between compensatory and non-compensatory methods. This refers to the possibility that low values in a single indicator may or may not be compensated by high values in another indicator. The appropriateness of the (degree of) compensability of the aggregation technique depends on the conceptual framework (Section 3). Examples of compensatory methods are linear and geometric aggregation (e.g. Saisana and Tarantola, 2002; Bandura, 2008, 2011; Greco et al. 2019). Examples of non-compensatory techniques are ELECTRE and PROMETHEE methods. The downside of non-compensatory approaches is their computational complexity, which minimizes their popularity (Greco et al., 2019).

This paper reviews, first, the methodological steps<sup>1</sup> in the construction of a composite indicator. Secondly, it discusses some popular aggregation methods to construct composite indicators of human well-being. These methods are characterized by eliciting weights based on statistical methods (*data-driven techniques*, Decanq and Lugo, 2013, p. 19 in Greco et al., 2019): Data Envelopment Analysis (DEA), Distance  $P_2$  ( $DP_2$ ), Mazziotta-Pareto Index (MPI) and Principal Components Analysis (PCA)<sup>2</sup>. We focus on these methodologies because, first, they are widely used (for instance: Saisana and Tarantola, 2002; Bandura, 2008; Somarriba and Pena, 2009; Greyling and Tregenna, 2016; Yang et al. 2017; Sanchez and Ruiz-Martos, 2018; for a more thorough survey see Greco et al. 2019). Secondly, their approaches to the computation of weights are intrinsically different, which results in severely dissimilar measures and makes each one of them appropriate for a specific measurement exercise. We review the desired properties of an aggregation method and the properties verified by the four methodologies. Finally, we compare these methods with respect to their weighting schemes; and the consequences of eliminating observations and adding noise (introducing an indicator which is a lineal combination of the other indicators).

Main conclusion is that the selection among these aggregation methods requires a refinement of the conceptual framework so as to define the ultimate purpose of the measurement exercise. That is, it does not suffice to state the targeted multidimensional concept, e.g., well-being. It is necessary to establish how exactly we aim to measure it. If the research goal is to produce a ranking of observations (countries, regions, etc.) regarding, eg, well-being then PCA and  $DP_2$  should be applied, with a preference for the latter (see Mazziotta and Pareto, 2019 and below). If the research goal is, however, to determine which dimension/s (or individual indicator/s) is/are more efficient to maximize well-being for each observation (e.g. in which dimensions of well-being each country is more efficient so as to address public policies), then DEA type methodologies and MPI should be applied.

Next section defines the four methodologies. Section 3 describes the methodological steps required to construct a composite indicator and why human well-being is commonly viewed as following a particular model of measurement. Section 4 reviews the desired properties of an aggregation method. Section 5 compares the methodologies and Conclusions follow.

## 2. SOME AGGREGATION METHODS

Last decades have witnessed a development of the measurement of multidimensional socio-economic phenomena. In this paper, we focus on several methodologies to analyze their weaknesses and strengths: Data

<sup>1</sup> For a more thorough methodological discussion see, for instance, OCDE (2008) and Maggino (2017).

<sup>2</sup> Greco et al. (2019) do not discuss the  $DP_2$  methodology.

Envelopment Analysis (DEA), Distance  $P_2$  ( $DP_2$ ), Principal Components Analysis (PCA) and Mazziotta-Pareto Index (MPI). There is no universal method for constructing composite indices, therefore, depending on the targeted phenomenon and on how it is measured, a methodology is more suitable than other. To carry out this analysis, we check a set of properties for each of the above methods. Throughout the section, sub-index  $i$  will correspond to an observation (region, country, etc.) and sub-index  $j$  to a single indicator. We begin by providing a brief description of each of them.

### 2.1. Data Envelopment Analysis (DEA)

This methodology, originally related to Management Science, is used to analyze the technical efficiency of public-sector decision-making units (DMUs). This methodology optimizes for each individual observation a discrete piecewise frontier through the set of Pareto-efficient Decision Making Units (DMU) (Charnes, A. et. al. 1979). With respect to other parametric approaches, one of its virtues is that it does not require specific assumptions on the distribution of the error terms. Another crucial feature is that the weights assigned to well-being domains are endogenously generated at the observation level.

DEA gathers a set of methodologies for evaluating performance. We focus on the so-called DEA-BoD approach, where BoD is the abbreviation of the Benefit-of-the-Doubt (BoD) principle (Cherchye et al. 2007). The basic idea of the DEA-BoD approach is to impose on each observation the optimal set of weights such that the observation achieves the best relative position with respect to the remaining observations. For example, Mariano et al. (2015) review the literature using DEA to measure human development. Other approaches combine DEA and Multi-Criteria-Decision-Making (MCDM) techniques to improve DEA-BoD while retaining a structure scheme of weightings for well-being domains across observations (Despotis, 2005; Peiro & Picazo 2018).

We concentrate on the additive model DEA-BOD, which has been used to compute well-being and quality of life composite indicators, for instance, González et al. (2010), Reig-Martínez (2012), Mizobuchi (2014). We suppose that  $X$  is a  $n \times m$  dimension matrix, where each column represents a single indicator. Let  $y_i$  be the composite index associated to the  $i$ -observation. For each  $i \in \{1, \dots, n\}$ , the Additive model DEA consists on maximising  $n$  DMU problems as follows:

$$\begin{aligned} \text{Maximise}_{\alpha_{ij}} y_i &= \sum_{j=1}^m \alpha_{ij} x_{ij} \\ \text{Subject to: } \sum_{j=1}^m \alpha_{ij} x_{lj} &\leq 1 \text{ for all } l \in \{1, \dots, n\} \\ \alpha_{ij} &\geq 0 \text{ for all } j \in \{1, \dots, m\}. \end{aligned}$$

The DEA computations maximize the relative efficiency score for each DMU, where the constraint condition is that the set of weights so obtained for each DMU must also be feasible for all the other DMUs. Then, the main difference with respect to other parametric approaches is that the analysis is based on individual observations, and not on population estimations. Such a strategy provides a single aggregate measure for each observation (DMU) through the input factors (single indicators), i.e., produces its respective composite indicator. A priori, the optimization procedure does not require a suitable specification of weights for each single indicator. As stated above, it is not necessary to impose the functional relationship between the composite index and the set of single indicators that define it (Charnes et al. 1997).

### 2.2. Distance $P_2$

Distance  $P_2$  (Pena Trapero, 1977) is an iterative method to obtain a metric (the composite indicator) by aggregating various single indicators as a weighted sum. Let  $X$  be a  $n \times m$ -dimension matrix, in which  $n$  is the number of observations and  $m$  is the number of single indicators. Let  $X_j$  be the  $j$ -single indicator  $j \in \{1, \dots, m\}$ . We define the reference vector  $X_* = (x_{*1}, \dots, x_{*m})$  as a fictitious vector whose coordinates belong to a theoretical observation with the best-worst possible scenario for all the single indicators. For each observation, we define  $d_{ij} = |x_{ij} - x_{i*}|$  as the distance from the  $i$ -observation  $i \in \{1, \dots, n\}$  to the  $j$ -coordinate of reference vector  $j \in \{1, \dots, m\}$ . For instance, if the composite indicator is measuring regional development, the composite indicator measures the distance between each region and a fictitious reference. Thus, the reference vector summarizes the results of a fictitious region with the worst possible scenario for all the indicators, for more details see Sánchez & Ruiz (2018).

Initially it is calculated the Frechet Distance (DF) corresponding to the  $i$ -observation as follows

$$DF_i = \sum_{j=1}^m \frac{d_{ij}}{\sigma_j} = \sum_{j=1}^m \frac{|x_{ij} - x_{i*}|}{\sigma_j} \quad (1)$$

where  $\sigma_j$  is the standard deviation of the  $j$ -single indicator  $j \in \{1, \dots, m\}$ . Concerning the weights, this method

allows several options. First, the researcher can assign to each element of the sum, i.e. to  $|x_{ij} - x_{i^*j}|/\sigma_j$ , a weight according to the relative stated importance of the indicator (e.g., experts opinion, political agreement). Secondly, we can assume the DF distance so that, implicitly, all indicators have the same weight. These options are to be considered in those cases where statistical relationships among single indicators do not represent the actual influence among them (Saisana and Tarantola 2002). However, when there is no information about nor agreement on the importance of the indicators, Distance  $P_2$  provides an iterative method to assign weights based on the linear correlations between indicators. This strength is also its weakness, as it depends solely on the linear relationships that may exist. The Distance  $P_2$  is defined as follow:

$$DP_2 = \sum_{j=1}^m \frac{d_{ij}}{\sigma_j} (1 - R_j^2) = \sum_{j=1}^m \frac{|x_{ij} - x_{i^*j}|}{\sigma_j} (1 - R_{j,\dots,1}^2) \quad (2)$$

where for each  $j \in \{1, \dots, m\}$ ,  $R_j^2$  represents the coefficient of determination in the multiple linear regression of  $X_j$  over  $X_{j-1}, \dots, X_1$  assuming  $R_1^2 = 0$ . The weight  $(1 - R_{j,\dots,1}^2)$  deletes the information conveyed by the preceding indicators and, thus, avoids double counting. This property will be called “completeness”. This method is sensitive to the order in which the indicators are introduced. To avoid subjectivity of choice, the pairwise correlation coefficients between each indicator and the DF are assessed, and then the indicators are sorted from highest to lowest according to the absolute values of these pairwise correlation coefficients. This property is called “neutrality” by Zarzosa Espina (1996). The indicators are then introduced following the previous order and the weights calculated accordingly. The process continues iteratively until the difference between two averages adjacent  $DP_2$ s is zero. As an example of this methodology see Sánchez & Ruiz (2018).

### 2.3. Principal Components Analysis (PCA)

PCA is a mathematical tool whereby an orthogonal transformation of the reference system of the set of observations transforms the set of indicators of possibly correlated variables into a set of linearly uncorrelated variables. Its origin can be imputed to Pearson (1901) or even Cauchy (1829). Let  $X_j$  be the  $j$ -single indicator  $j \in \{1, \dots, m\}$  and let  $p \leq m$  be the number of principal components  $Y_1, \dots, Y_p$  that are obtained as linear combinations of the original data  $X_1, \dots, X_m$ .

$$\begin{aligned} Y_1 &= a_{11} X_1 + \dots + a_{1m} X_m \\ &\vdots \\ &= \vdots \\ &\vdots \\ Y_p &= a_{p1} X_1 + \dots + a_{pm} X_m \end{aligned}$$

where the factor loadings  $A_l = (a_{l1}, \dots, a_{lm})$  satisfy that

$$\sum_{j=1}^m a_{lj} = 1, \text{ for all } l \in \{1, \dots, p\} \quad (3)$$

and  $Y_1, \dots, Y_p$  is a orthogonal set of vectors (i.e., uncorrelated). The goal of this method is to maximize the variance

$$\begin{aligned} \text{Var}(Y_1) &= \text{Var}(A_1 X) = A_1' \Sigma A_1 \\ \text{Subject to:} \\ A_1' A_1 &= 1 \end{aligned}$$

where  $\Sigma$  is the covariance matrix of the data set  $X_1, \dots, X_m$ . Let  $\lambda_1 \geq \dots \geq \lambda_m$  denote the set of eigenvalues of the covariance matrix  $\Sigma$ . Using Lagrange Multiplier approach and Roché-Frobenius theorem,  $\text{Var}(Y_1) = \lambda_1$ . Therefore, the maximum eigenvalue provides the maximum variance, and the corresponding eigenvector whose coordinates are the factor loadings. The second component  $Y_2$  is computed solving the previous optimization problem but also by imposing  $Y_1' Y_2 = 0$ . The process continues iteratively until all the components are computed. Karamizadeh et al. (2009) provide a recent description of this methodology. A detailed review of the literature and applicability of this approach is in Greco et al. (2019).

### 2.4. Mazziotta-Pareto Index

Let  $x_{ij}$  be the  $i$ th-observation corresponding on the  $j$ th-indicator. The Min-max method is used as normalization method through the following formulation:

- If the polarity of the  $j$ -indicator is positive:

$$y_{ij} = \frac{x_{ij} - \min(X_j)}{\max(X_j) - \min(X_j)} \quad (4)$$

- If the polarity of the  $j$ -indicator is negative:

$$y_{ij} = \frac{\max(X_j) - x_{ij}}{\max(X_j) - \min(X_j)} \quad (5)$$

The Mazziotta-Pareto Index (MPI) is a non-linear function composite index (Maggino, F. 2017b). After Min-Max normalization, we define  $z_{ij} = 100 + 10(y_{ij} - M(Y_j)/\sigma(Y_j))$ , where the  $M(Y_j)$  is the mean of the j-Min-Max normalized indicator and  $\sigma(Y_j)$  is the standard deviation of the j-Min-Max normalized indicator. The MPI index is defined as follows:

$$MPI_{z_i}^{\pm} = M_{z_i} \pm \sigma_{z_i} cv_{z_i} = \frac{1}{m} \sum_{j=1}^m z_{ij} \pm \sigma_z cv_{z_i} \quad \text{for } i \in \{1, \dots, n\} \quad (6)$$

Where  $M_{z_i}$ ,  $\sigma_{z_i}$  and  $cv_{z_i} = \sigma_{z_i}^2/M_{z_i}$  denote the mean, standard deviation and the coefficient variation of each i-th observation. The sign  $\pm$  is related to the targeted phenomenon. The penalty is positive when an upward adjustment is required. When a downward adjustment proceeds, the penalty is negative.

### 3. METHODOLOGICAL STEPS FOR CONSTRUCTING COMPOSITE INDICATORS

Measuring in social sciences requires a robust conceptual definition of the target, a consistent collection of observations and a subsequent analysis of the relationship between observations and defined concepts (Maggino, 2017, p. 87). The relationship between target and indicators determines the model of measurement and conditions the construction process of the composite indicator, especially the aggregation method (Maggino, 2017, p.97).

#### 3.1. Model of measurement

The model of measurement may be reflective or formative (Maggino, 2017; Mazziotta and Pareto, 2018). In a reflective model, indicators are functions of the latent variable, which is the independent variable (changes in latent variable trigger changes in indicators). Conversely, in a formative model, the latent variable depends on the indicators. In the following, let  $R$  represent the multidimensional latent variable we aim to measure and  $X_j$ , individual indicator  $j$ .

- **Reflective model**

Indicators are manifestations of the latent variable  $R$ . Hence, causality is from the concept  $R$  to the indicators  $X_j$  that reflect the concept. Each indicator will be a linear function of the underlying variable  $R$  plus a measurement error:

$$X_j = \lambda_j R + \varepsilon_j$$

where  $\lambda_j$  is the coefficient or loading that captures the effect of  $R$  on  $X_j$  and  $\varepsilon_j$  is the measurement error of that indicator.

Measurement errors are assumed independent and unrelated to the latent variable. Individual indicators are: interchangeable (removing one of them does not affect essentially the latent variable); intercorrelated (two uncorrelated indicators cannot share a common cause); and, moreover, positively correlated if they share equal polarities (i.e., equally related to the latent variable), conversely, negatively correlated (Maggino, 2017; Mazziotta and Pareto, 2018).

The key is highly correlated single indicators (Maggino, 2017, p.121). The right approach is to reduce dimensionality by a factor or scaling model, i.e., factor analysis (main goal is to test a reflective approach and allows to synthesize indicators belonging to the same dimension) and PCA (Maggino, 2017, Mazziotta and Pareto, 2019). For example, measuring intelligence through a questionnaire, more intelligence more correct answers in all dimensions (Simonetto, 2012 in Mazziotta and Pareto, 2019, p. 454).

- **Formative model**

As indicators cause the latent variable, a change in the latent variable does not necessarily imply changes in all its measures (Mazziotta and Pareto, 2018). The concept is defined by the indicators.

$$R = \sum_j \lambda_j X_j + \zeta$$

where  $\lambda_j$  captures the effect of indicator  $X_j$  on the latent variable  $R$ , and  $\zeta$  is the error term.

Indicators are not interchangeable (omitting one of them implies missing a part of the latent variable), and intercorrelations are not explained by the measurement model (high correlations are possible but not generally expected, causing a multicollinearity problem). In fact, correlated indicators may be redundant and make the conceptual component measured by both to get over weighted in the composite indicator (Maggino, 2017). Moreover, correlations and polarities are independent. Since such a model does not assume correlated indicators, the correlation structure of the data cannot be used to determine the latent variable. Instead, the latent variable is estimated

by taking a weighted average of the indicators conforming the concept (Shwartz et al. 2015 in Mazziota and Pareto, 2019, p. 4). DEA-BoD, Distance  $P_2$  and MPI are examples of formative models.

A common view is that human well-being follows a formative model (Diamantopolus et al. 2008; Mazziota and Pareto, 2017; Mazziota and Pareto, 2019). Well-being depends on health, income, occupation, services, safety, environment, etc. So the improvement of any one of these indicators would imply an improvement in well-being, even if the other indicators remain invariant. Subsequently, an improvement of well-being does not necessarily imply and improvement in all its indicators. When the composite indicator follows a formative model, the following issues are critical (Diamantopolus and Winklhofer, 2001 in Maggino, 2017, p. 120):

- Content specification -domain that the composite indicator aims to capture- is inextricably linked to the specification of the indicators.
- Indicator specification: indicators must cover the entire latent variable domain (nor too many or too few).
- Indicator collinearity: excessive collinearity among indicators makes difficult to disentangle particular influences of individual indicators. Multicollinear indicators may be redundant.
- External validity. The composite indicator should be related to other measures by means of individual indicators being related to external indicators.

### 3.2 Indicators: selection and comparison

The selection of indicators must address the complexity of the targeted phenomenon (multidimensionality, nature -objective versus subjective, quantitative versus qualitative-, distinct level of observations -micro and macro-, dynamics -internal and external conditions, trends and relationships between phenomena); allow for relativity and comparability (e.g., same concept may be measured by different indicators in different areas and cautious interpretation of the results), and avoid over reductionism (system of indicators may be simplified by reducing the number of indicators following the conceptual model or by synthesizing into a composite indicator) (Maggino, 2017).

Comparison of indicators is mandatory but they have diverse measurement units and ranges. Hence, before aggregation, indicators should be made comparable. This is achieved by normalization, i.e., by transforming indicators into pure, dimensionless, numbers. Several normalization techniques are available: ranking, standardization, re-scaling or indicization. Each of these techniques has pros and cons mainly concerning interval level of information, sensitivity to outliers and implicit weighting (Mazziota and Pareto, 2017, p.170). The selection of a normalization technique should be guided by the conceptual framework, indicators variability among observations and the ultimate purpose of the measurement exercise. Moreover, polarity of indicators (i.e., the positive or negative sign of the relationship between them and latent variable) must be defined. Indicators with negative polarity must be “inverted” by a linear or non-linear transformation (Mazziota and Pareto, 2017). The aim of normalizing and dealing with polarity is that a rise in the normalized indicator should imply an increase in the composite index (Salman, 2003 in Mazziota and Pareto, 2017, p.166). Furthermore, as socio-economic data are mainly ordinal and discrete, it should be clearly stated how ordinal indicators are dealt with, i.e., how metric analysis is carried out of non-metric data while maintaining consistency with the true nature of the phenomenon (Maggino, 2017, p.127).

### 3.3 Synthesizing indicators: weighting criteria and aggregative-compensative approach

Before aggregation, we should consider, if convenient, the relative importance of the single indicators, i.e., the weighting scheme, which basically implies values judgement (OCDE, 2008, p.31). One can contemplate, on the one hand, no-weights and equal weights, and, on the other hand, participatory and statistical weighting techniques. No-weights encounters two main problems: double counting (OECD, 2008) and discarding indicators information (Saisana and Tarantola, 2002). Most frequent is to assign equal weights. Unlike the no-weights approach, equal weighting may imply differential weighting when single indicators are grouped in dimensions, weights are attached to dimensions and there is a distinct number of indicators per dimension (Greco et al. 2019). Equal weights are favored when there is: no consensus on the distribution of weights; lack of theoretical basis; insufficient statistical knowledge; or when simplicity or objectivity are called upon. Its main inconveniences are that it does not permit to differentiate between crucial and non-crucial indicators and that, precisely for this, can be considered as subjective as differential weighting (Greco et al., 2019).

Participatory methods (e.g., budget allocation processes, analytic hierarchy processes, conjoint analysis) incorporate various stakeholders (experts, politicians, citizens) to agree upon a weighting scheme ideally through a clearly stated process. However, this approach is only functional when there is consensus about the national policy (OCDE, 2008; Maggino, 2017; Greco et al. 2019). Statistical weighting or data-driven techniques (e.g. factor analysis, principal components analysis, data envelopment analysis, unobserved components models, Distance  $P_2$ ) determine the weights based only on statistical techniques. The methodologies discussed here are all data-driven techniques.

The interpretation of the weights is essential. A weight can be viewed as the relative importance of an indicator

or as the trade-off between pairs of indicators (OCDE, 2008, p.31-33; Greco et al. 2019, p.75). The proper interpretation depends on the compensatory versus non-compensatory nature of the aggregation technique. The aggregative-compensative approach assumes that only one latent variable is being measured. Serious issue is the correlation among the indicators to be aggregated; the interpretation of which depends, as discussed, on the model of measurement. A reflective model requires very high correlation among indicators, because they all are manifestations of the same, commonly multidimensional, latent variable. Hence, indicators referring to the same dimension may be aggregated. A formative model typically encompasses indicators that measure independent dimensions of the multidimensional latent variable. Correlation among indicators suggests overlapping and may induce discarding one of them, which should be done while preserving comparability among observations and over time (Maggino, 2017, p.122).

The main criticism of the aggregative approach charges against its main strength, uni-dimensionality. The argument is that conveying into an uni-dimensional measure a multidimensional, complex and dynamic concept, such as well-being, rises critical conceptual, methodological and technical issues. For instance, aggregating may result in two distinctive observations being assigned the same score. Hence, it is fundamental to identify the befitting aggregation technique. This step should consider the issues of comparability and measurement homogeneity, which both refer to normalization and polarity (above), and compensability (Maggino, 2017).

A compensatory technique allows for low values in some indicators to be compensated by high values in other indicators. Compensation among indicators determines the interpretation of the weights. The compensatory approach implies that weights should be interpreted as trade-offs, and not as importance coefficients (OCDE, 2008; Maggino, 2017; Greco et al. 2019). Both the linear (composite indicator is the sum of the weighted indicators) and geometric (composite indicator is the product of indicators, each of them raised to the power of its weight) aggregation schemes are compensatory. In linear aggregation, compensability is constant; in geometric aggregation (used, for instance, by the HDI since 2010), compensability is lower for those indicators with worse values. All of these compensability issues should be taken into account.

A multidimensional phenomenon such as well-being, where each dimension may be represented by several indicators, may require to build a composite indicator for each dimension and, then, obtain the overall index by aggregating the partial composite indicators. In which case, a compensatory approach could be followed within dimensions and a non or partially compensatory approach<sup>3</sup> among dimensions (Mazziotta and Pareto, 2017). However, non-compensatory approaches, such as multi-criteria, are computationally costly with a high number of observations (Munda and Nardo, 2007 in OCDE, 2008 p. 33).

#### 4. PROPERTIES OF THE AGGREGATION METHOD

In table 1 we summarize the mathematical properties of the aggregation methods described in section 2. Following the properties pointed out by Pena Trapero (1977, 2009) and Zarzosa Espina (1996), as well as the desired properties for Indicators Construction by Maggino (2017), we indicate the performance of the methods discussed on these mathematical properties.

#### 5. ANALYSIS OF THE WEIGHTING SCHEME

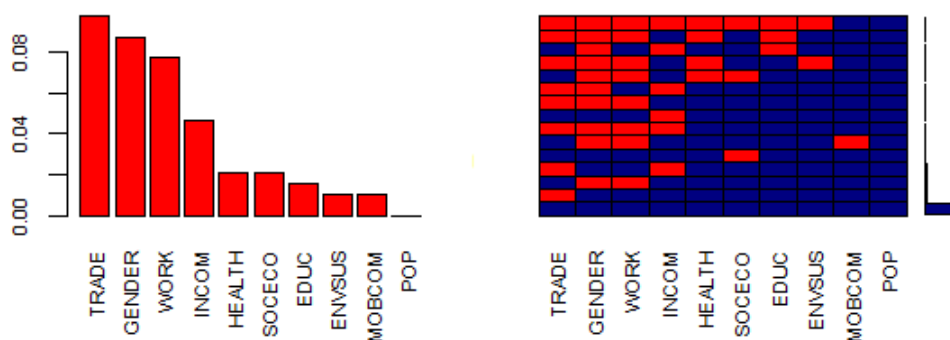
Here the aim is to highlight some important aspects of the analyzed methods. We use the Human Development database, where ten single indicators were chosen from the Human Development Report Office (HDRO) for 2017. Table 2 depicts the definition of the selected single indicators and additional information. To use PCA, all indicators have been chosen with the same polarity. Following OECD's Regional Well-Being Dataset, we normalize using Equation 7 for each single indicator  $X_j, j \in \{1, \dots, 10\}$ . This normalization has been used for DEA, MPI, and PCA, such that the data set will express the distance each observation has from the worst scenario in each indicator. For the Distance  $P_2$  calculation, the reference vector's coordinates are the minimum of each single indicator. The greater the distance to the minimum, the greater the calculated distance  $P_2$ .

$$\hat{x}_{ij} = \frac{x_{ij} - \min(X_j)}{\max(X_j) - \min(X_j)} \quad (7)$$

<sup>3</sup> MPI is an example that summarizes partially non-compensatory indicators.

**Table 1:** Summary of the mathematical properties

<b>1. Existence</b>		
For any observation (country, region, etc) the composite indicator defined from each method is well defined.	<b>Distance P2</b>	True
	<b>PCA</b>	True
	<b>DEA</b>	True
	<b>MPI</b>	True ( if for each observation if $M_{zi}$ is different to 0)
<b>2. Monotony</b>		
An increase(decrease) in one single indicator with positive(negative) polarity while keeping the other indicators constant produces an increase in the composite indicator.	<b>Distance P2</b>	True
	<b>PCA</b>	True (if all single indicators has the same polarity)
	<b>DEA</b>	True
	<b>MPI</b>	True
<b>1.3 Symmetry</b>		
The composite indicator does not depend on the order of the single indicators.	<b>Distance P2</b>	False (The methodology impose an order).
	<b>PCA</b>	True
	<b>DEA</b>	True
	<b>MPI</b>	True
<b>1.4 Invariance</b>		
	<b>Distance P2</b>	True
	<b>PCA</b>	False (depends on the normalization has been chosen)
	<b>DEA</b>	False (depends on the normalization has been chosen)
	<b>MPI</b>	True
<b>1.5 Completeness</b>		
The weigs of the single indicators are introduced according to their relevance avoiding duplication of information.	<b>Distance P2</b>	True
	<b>PCA</b>	False (The weighting scheme is not related to the relevance of each indicator)
	<b>DEA</b>	False (The information provided by each indicator is particular to each observation. Therefore, we cannot know what information is provided by each indicator as a whole)
	<b>MPI</b>	False (The information provided by each indicator is particular to each observation)
<b>1.6 Objectivity</b>		
The ranking or the weights are not arbitrarily determined.	<b>Distance P2</b>	True
	<b>PCA</b>	True
	<b>DEA</b>	True
	<b>MPI</b>	True



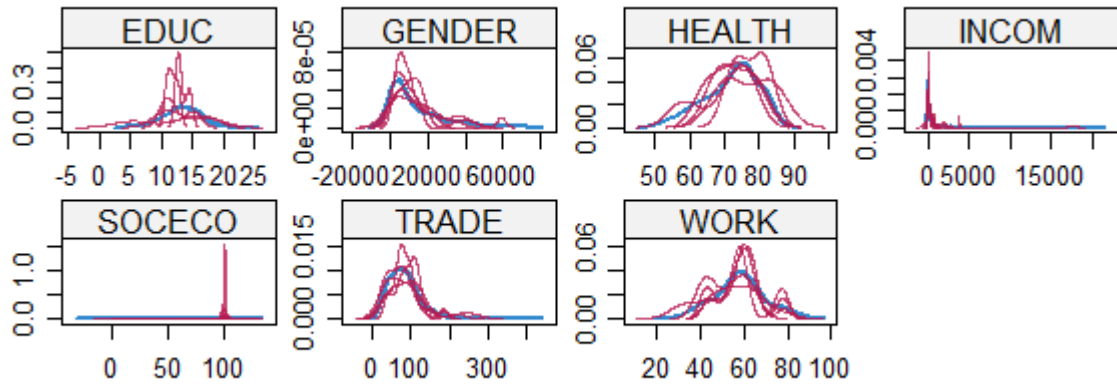
**Figure 1:** Percentage and patterns of missing values for each indicator

The database contains missing values. To balance it, the missing values have been replaced with the imputed values using predictive mean matching (PMM). Figure 1 shows that almost 89% of the samples are not missing any information. The worse single indicator contains 8% of missing values and the best none.



**Table 2:** Single indicators at country level

<b>Indicator</b>	<b>Definition</b>	<b>Source</b>	<b>Polarity</b>
Population ( <b>POP</b> )	Urban population %	UNDESA (2018a). World Urbanization Prospects: The 2018 Revision. New York. <a href="https://esa.un.org/unpd/wup/">https://esa.un.org/unpd/wup/</a> . Accessed 17 May 2018.	Positive
Education ( <b>EDUC</b> )	Expected years of schooling (years)	UNESCO Institute for Statistics (2018), ICF Macro Demographic and Health Surveys, UNICEF Multiple Indicator Cluster Surveys and OECD (2017a).	Positive
Environmental sustainability ( <b>ENVSUS</b> )	Renewable energy consumption (% of total final energy consumption)	World Bank (2018a). World Development Indicators database. Washington, DC. <a href="http://data.worldbank.org">http://data.worldbank.org</a> . Accessed 6 July 2018.	Positive
Gender ( <b>GENDER</b> )	Estimated gross national income per capita, female (2011 PPP \$)	HDRO calculations based on ILO (2018a), UNDESA (2017a), World Bank (2018b) and IMF (2018).	Positive
Health ( <b>HEALTH</b> )	Life expectancy at birth (years)	UNDESA (2017a). World Population Prospects: The 2017 Revision. New York. <a href="http://esa.un.org/unpd/wpp/">http://esa.un.org/unpd/wpp/</a> . Accessed 10 May 2018.	Positive
Income/composition of resources ( <b>INCOM</b> )	Gross domestic product (GDP), total (2011 PPP \$ billions)	World Bank (2018a). World Development Indicators database. Washington, DC. <a href="http://data.worldbank.org">http://data.worldbank.org</a> . Accessed 6 July 2018.	Positive
Mobility and communication ( <b>MOBCOM</b> )	Mobile phone subscriptions (per 100 people)	ITU (International Telecommunication Union) (2018). ICT Facts and Figures 2018. <a href="http://www.itu.int/en/ITU-D/Statistics/Pages/stat/">www.itu.int/en/ITU-D/Statistics/Pages/stat/</a> . Accessed 18 July 2018.	Positive
Socio-economic sustainability ( <b>SOCECO</b> )	Rural population with access to electricity (%)	World Bank (2018a). World Development Indicators database. Washington, DC. <a href="http://data.worldbank.org">http://data.worldbank.org</a> . Accessed 6 July 2018.	Positive
Trade and financial flows ( <b>TRADE</b> )	Exports and imports (% of GDP)	World Bank (2018a). World Development Indicators database. Washington, DC. <a href="http://data.worldbank.org">http://data.worldbank.org</a> . Accessed 6 July 2018.	Positive
Work, employment and vulnerability ( <b>WORK</b> )	Employment to population ratio (% ages 15 and older)	ILO (International Labour Organization) (2018a). ILOSTAT database. <a href="http://www.ilo.org/ilostat">www.ilo.org/ilostat</a> . Accessed 13 April 2018.	Positive



**Figure 2:** Empirical density function of variables (with and without missing)

We observe that the density functions have not been altered by the imputation of missing values. For instance, Figure 2 shows magenta slope in the density of the imputed data of the 5th worse single indicators, while blue slope density shows the observed data.

The use of PCA and Distance  $P_2$  is not suitable when the correlations between the indicators is very weak. Table 3 shows correlations among single indicators. The single indicators chosen have high and medium correlation to avoid this problem.

**Table 3.** Correlation among single indicators

	POP	EDUC	ENVSUS	GENDER	HEALTH	INCOM	MOBCOM	SOCECO	TRADE	WORK
POP	1.00	0.58	-0.51	0.60	0.59	0.13	0.42	0.51	0.23	-0.15
EDUC	0.58	1.00	-0.48	0.67	0.80	0.16	0.57	0.70	0.22	-0.12
ENVSUS	-0.51	0.00	1.00	-0.38	-0.60	-0.14	-0.44	-0.73	-0.27	0.34
GENDER	0.60	0.67	-0.38	1.00	0.65	0.16	0.45	0.48	0.43	0.06
HEALTH	0.59	0.80	-0.60	0.65	1.00	0.16	0.54	0.83	0.27	-0.12
INCOM	0.13	0.16	-0.14	0.16	0.16	1.00	0.05	0.15	-0.19	0.01
MOBCOM	0.42	0.57	-0.44	0.45	0.54	0.05	1.00	0.56	0.35	-0.09
SOCECO	0.51	0.70	-0.73	0.48	0.83	0.15	0.56	1.00	0.20	-0.29
TRADE	0.23	0.22	-0.27	0.43	0.27	-0.19	0.35	0.20	1.00	-0.01
WORK	-0.15	-0.12	0.34	0.06	-0.12	0.01	-0.09	-0.29	-0.01	1.00

Please note that our choice of single indicators is irrelevant, it is simply a reference framework to observe the possible divergences among the discussed methods. Then, we calculate the composite index with the selected single indices using the four procedures analyzed. Kendall rank correlation test is used to measure the ordinal association between two composite indicators (Kendall, 1938). This non-parametric statistic test assesses whether two variables may be regarded as statistically dependent by determining if there exist a monotonic relationship between them, namely, preserving the rank ( $\tau$  statistic close to one) under the null hypothesis of independence.

Intuitively, Kendall tau measures the difference between the probability that data are in the same order for the two computed indicators and the probability that data are in different orders. If close to 0, there is no evidence that the ranks are equal. Table 4 shows the close order-relationship among the composite indicators, from which it follows that some of these procedures are unreliable for constructing composite indices if the goal is to rank observations. The null hypothesis of independence was not rejected between the composite indicators constructed by Distance  $P_2$  versus PCA; DEA versus PCA. However, independence was rejected between Distance  $P_2$  versus DEA; Distance  $P_2$  versus MPI; DEA versus PCA, and DEA versus MPI, as the concordance pairs are very low. These divergences are caused because each of the methods befits a specific measurement goal, as we discuss next.

**Table 4.** Results of Kendall's rank correlation test (p-value)

	DEA	PCA	MPI
<b>DP2</b>	0,4535 (<0,0001***)	0,0181 (0,7066)	0,3441 (<0,0001***)
<b>DEA</b>		0,2825 (0,5928)	0,2358 (<0,0001***)
<b>PCA</b>			0,0988 ( 0,0402*)
p-value significance	*** p<0.001	** p<0,01	* p<0,05

### 5.1 Weighting schemes: indicators ranking

To understand divergences between the methods, we consider the previous database. The weighting scheme of each method provides part of the answer. As described in section 2, the weights are distinctively obtained.

DEA-BoD is a good tool for designing good strategies or policies per observation. It provides a collection of weights (different for each observation), endogenously determined. However, the differential weighting inherent in the process prevents comparison among observations (Greco et al. 2018). To overcome this shortcoming, Peiro et al. (2018) propose to combine DEA and Multi-Criteria-Decision-Making (MCDM) techniques to achieve both a common set of weights and allow for comparisons in the ranking of observations. The latter method gives a unique structure of weights that allows a ranking, although it is not possible to compare the results of the compound index between pairs, since the result obtained through DEA-BoD-MCDM is not a metric. Also, Spearman's rank correlations between the well-being scores were obtained between the DEA-BoD and different DEA-BoD-MCDM approaches. As a result, a high rank and statistically significant at the 1% confidence level correlation was found. We take the DEA-BoD rank as reference, because there are no significant differences between the ranks of the latter approaches. As stated in section 2, DEA-BoD approach imposes on each observation under evaluation the optimal set of weights such that the observation achieves the best relative position with respect to remaining observations. Thus, for each observation, a whole weighting scheme is computed. Hence, we have not a unique weight associated to each indicator but a collection of weights. To provide an approximation to these collections, Table 5 depicts the summary statistics (Max., 3rd Qu., Median, Mean, 1rd Qu., Min) for each indicator<sup>4</sup>. This analysis shows that SOCECO is the most important indicator, being HEALTH and GENDER the less relevant.

**Table 5:** Summary Statistics of DEA Weights

Stats.	POP	EDUC	ENVSUS	GENDER	HEALTH	INCOM	MOBCOM	SOCECO	TRADE	WORK
Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1 <sup>st</sup> Qu.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Median	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
Mean	0.037	0.026	0.193	0.016	0.013	0.080	0.082	0.608	0.106	0.101
3 <sup>rd</sup> Qu.	0.000	0.000	0.302	0.000	0.000	0.000	0.000	1.000	0.000	0.047
Max	1.000	0.758	1.000	0.998	0.938	1.000	1.000	1.000	1.000	1.000

As described in MPI Equation 6, for each observation (country or region) and after normalization, the arithmetic mean is calculated. Hence, we can assume that all indicators have the same weight, to which a function of indicator variability called *penalty* is added to minimize duplication of information. In this way, the penalty could restore the unbalance produced by a poor performance in some indicators. A variant of the previous is the Adjusted Mazziotta-Pareto Index (AMPI). The AMPI summarizes a set of indicators that are assumed to be non-substitutable, Mazziotta & Pareto (2018). However, as with DEA-BoD, this aggregation technique does not allow to compare observations with each other, and consequently the ranking provided has a difficult interpretation.

PCA uses the factor loadings as indicators weights, for instance, to make HDI (Noorbakhsh, 1996). Since the number of components/factors to be retained must be chosen by the decision maker, subjectivity is introduced to a certain degree. In the context of composite indices, the first component is commonly chosen for the composite index (Greyling & Tregenna 2016). Extensive PCA-related literature is provided by Greco et al. (2019). Although the first component alone only explains a portion of the variance of the indicators, we take the factor loadings of the first component as the weights for the computed composite indicator. Table 6 shows the coordinates of the eigenvector<sup>5</sup> (weights) corresponding to the high eigenvalue  $\lambda = 4.70347$  that explains the 47% of variance of the composite indicator. With this methodology, TRADE is the most relevant indicator, followed by HEALTH and SOCECO. Notice that PCA assumes a linear relationship between the indicators. Thus, in the current case, that TRADE, WORK and INCOM have low correlations with the rest of indicators can lead to unwanted results. In addition, though the reductionism of this approach is useful to avoid duplicity of information, the weights so endogenously obtained could not necessarily correspond to the real links among the indicators (Saisana & Tarantola 2002).

<sup>4</sup> This difficulty in understanding the relevance of weights and the contribution of each to the composite indicator is one of the weaknesses of this approach.

<sup>5</sup> Eigenvector provides the direction for which the variance of the first component is maximized.

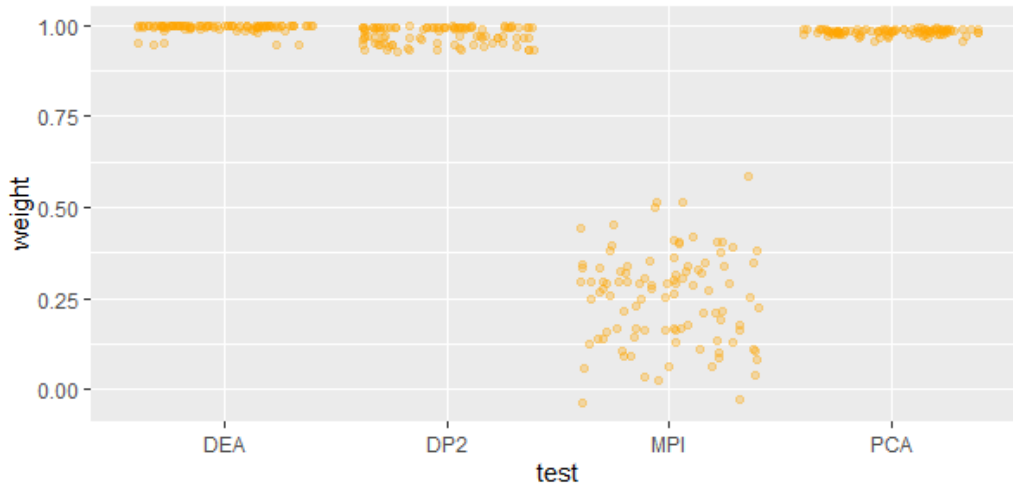
Distance  $P_2$  provides another rank distribution of indicators. After ordering the indicators through the correlation between the Frechet's distance and the indicators, we compute the coefficient of determination in the multiple linear regression of  $x_j$  over  $x_{j-1}, \dots, x_1$  assuming  $R_1 = 0$ , then the weights  $1 - R_{j,\dots,1}$  are obtained. Notice that Distance  $P_2$  measures a distance. Namely, let  $X_i = (x_{i1}, \dots, x_{im})$  denote the vector whose coordinates correspond to the indicators values for each observation  $i \in \{1, \dots, n\}$ . Then Distance  $P_2$  is symmetric  $DP_2(X_i, X_j) = DP_2(X_j, X_i)$ , positive  $DP_2(X_i, X_j) > 0$  if  $i \neq j$  and  $DP_2(X_i, X_j) = 0$  if  $i = j$ , and satisfies the triangular inequality  $DP_2(X_i, X_j) + DP_2(X_j, X_k) > DP_2(X_i, X_k)$ . The latter property allows for comparisons between observations, and this is the only one out of the analyzed methods that obeys this property. However, the method has a strong dependence on the linearity of the model, for which when the indicators have very weak correlations the results can be wrong. Table 6 shows GENDER as the most relevant indicator followed by INCOM and WORK, being SOCECO the least relevant indicator.

**Table 6:** Indicators weights in DEA,  $DP_2$ , PCA

Indicator	POP	EDU	ENVS	GEND	HEAL	INCO	MOBC	SOCE	TRA	WOR
DEA	0.036	0.025	0.1929	0.0160	0.0126	0.080	0.0822	0.6082	0.1058	0.100
DP2	0.562	0.325	0.4183	1	0.5812	0.876	0.6470	0.2857	0.7529	0.803
PCA	0.116	0.154	0.1187	0.1219	0.1703	0.0754	0.1055	0.1576	0.3577	0.108

### 5.2 Sensitivity analysis

Sometimes, the observations may be chosen for further efficiency or deleted if they are contaminated by data errors (Wilson, 1995). In this way, an approach within sensitivity analysis studies responses to manipulations - addition or subtraction- of the number of observations. The objective is to analyze if the ranking of the composite index is statistically altered when several observations are, e.g., deleted. Starting from the original sample, we eliminate ten random observations in each iteration, with reposition, using Monte Carlo procedure and compute the composite index. The Kendall rank correlation test is calculated for the resulting composite indices. Figure 3 shows that MPI is sensitive to deletion of observations, while the random elimination of observations leaves invariant the ranks in the other methods.



**Figure 3:** Tau outputs from Kendall rank correlation test for sensitivity analysis

High correlation of indicators can generate duplication of information in the composite index. Aggregation methods should be able to overcome this overlapping of information. Here, the weighting scheme plays an essential role. To test the behavior of the studied methods, we introduce an indicator that is a lineal combination of the other indicators (noise) so as to determine the extent to which they are capable of eliminating this redundant information. Through 100 random combinations of the 10 indicators, we compute the weights for Distance  $P_2$ , PCA and DEA-BoD<sup>6</sup>. Figure 4 shows the results of the Monte Carlo method used. DEA reasonably discriminates the noise introduced. However, PCA is not able to discriminate noise through the weights. The weights are actually the coordinates of the eigenvector corresponding to the highest eigenvalue of the covariance matrix, i.e. the value that maximizes the variance of the first component, or rather, of the composite index, and are not strictly speaking weights that balance the model according to the relevance of the indicators. On the other hand, Distance  $P_2$  computes the

<sup>6</sup> MPI has not been included in this analysis because the weights are all the same for this procedure.

correlations among indicators but is unable to detect if any of the indicators is a linear combination of the others.

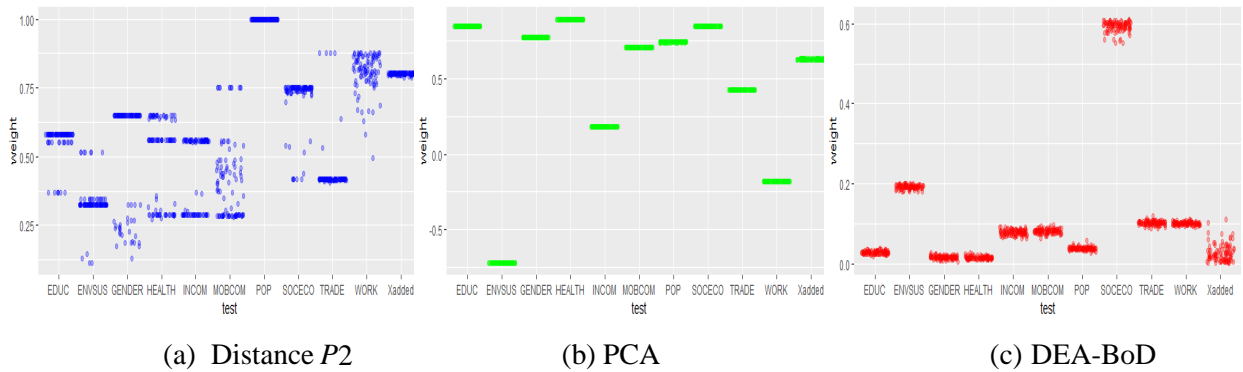


Figure 4: Indicators weights

## 6 CONCLUSIONS

Most socioeconomic phenomena are multidimensional, which renders a single indicator unable to capture the inherent complexity in, for example, development, poverty, well-being. The construction of composite indicators should follow a respectful methodological approach to ensure that the *big picture* fundamentally captures what it is meant to (OECD, 2008). The methodological process starts with the precise definition of the conceptual framework (Maggino, 2017, p.87), which may be reflective or formative, and conditions the selection of single indicators that (attempt to) measure the various dimensions of the concept. Indicators should be selected befitting the targeted phenomenon. Following the conceptual framework, individual indicators must be normalized and, then, aggregated taken into consideration compensability issues and weighting schemes. Finally, the robustness of the composite index should be assessed.

We focus on four aggregation methods to analyze their weighting and aggregation approaches. The choice of these statistical methods is based upon their differences in aggregation and weighting schemes. First, a method that imposes on each observation the optimal set of weights so that it is rated in the best relative position with respect to remaining observations. The DEA-BoD optimizes for each individual observation a discrete piecewise frontier through the set of Pareto-efficient Decision Making Units (DMU). Secondly, PCA, a method whose fundamental virtue is reducing the dimensionality of the dataset when there are high correlations among indicators. Thirdly, a method that builds a metric that inherits analytical properties allowing observations to be compared, Distance  $P_2$ . Finally, a method that produces a composite index that penalizes substitutability among indicators, Mazziotta–Pareto Index (MPI).

From the 2017 Human Development database, ten individual indicators are selected to construct a composite indicator and study the divergences of the four methods. For comparison purposes, all indicators have the same polarity. Likewise, the same normalization approach has been applied, except for Distance  $P_2$  for which a reference vector has been chosen in accordance with the above approaches.

The core of PCA's philosophy is to optimize the variation of the new components that reduce the dimensionality of the indicators, as long as they are highly correlated and have the same polarity. PCA is more suitable for reflective models. The use of this methodology for the construction of composite indices outside this context can lead to important errors. The procedure is stable when eliminating observations, maintaining the ranking of the calculated composite index. Additionally, when an indicator (linear combination of the principals) is added, PCA does not discriminate against it.

DEA-BoD is befitting to study the efficiency of each observation separately. However, the composite index constructed does not allow comparisons, which complicates the interpretation of the results as a whole. The procedure is stable when eliminating observations, maintaining the ranking of the calculated composite index. Additionally, when an indicator (linear combination of the principals) is added, DEA assigns a weight that makes this new indicator irrelevant in the model.

Distance  $P_2$  computes a composite index based on a metric. Thus, unlike DEA-BoD, it allows comparison between observations and provides a mathematical structure for the analysis of the results. However, given the dependence of the model on its linearity, when indicators are poorly correlated, Distance  $P_2$  does not behave efficiently, as PCA. The procedure is stable when eliminating observations, maintaining the ranking of the calculated composite index. In addition, when adding a linear combination of the main indicators, Distance  $P_2$  does not provide a satisfactory result.

MPI partially avoids compensability. Unlike Distance  $P_2$  or PCA, the indicators can be poorly correlated.

However, the MPI penalty will not always catalyze imbalances between indicators. The penalty, calculated for each observation, can also complicate the understanding of the composite index, as in the DEA-BoD approach. The procedure is not stable when observations are eliminated, generating significant alterations in the ranking of the calculated composite index.

## REFERENCES

- BANDURA, R. (2008). *A survey of composite indices measuring country performance: 2008 update*. Technical report, Office of Development Studies, United Nations Development Programme (UNDP), New York.
- BODENHOFER, U. & KLAWONN, F. (2008). *Robust rank correlation coefficients on the basis of fuzzy orderings: initial steps*. 15.1, 5-20.
- CAUCHY AL. (1829) *Sur l'équation à l'aide de laquelle on détermine les inégalités séculaires des mouvements des planètes*. vol. 9. Oeuvres Complètes (11ème Série); Paris: Blanchard; 1829.
- CHARNES, A., COOPER, W., LEWINA, Y. & SEIFORD, L. M. (1997) *Data Envelopment Analysis Theory, Methodology and Applications*, Journal of the Operational Research Society, 48:3, 332-333.
- CHERCHYE, L., MOESEN, W., ROGGE, N., & VAN PUYENBROECK, T. (2007). *An introduction to "benefit of the doubt" composite indicator*. Social Indicators Research, 82(1), 111-145.
- DECANCO, K., & LUGO, M. A. (2013). *Weights in multidimensional indices of wellbeing: An overview*. Econometric Reviews, 32(1), 7-34.
- DESPOTIS, D. K. (2005a). *A reassessment of the human development index via data envelopment analysis*. Journal of the Operational Research Society, 56(8), 969-980.
- DIAMANTOPOULOS, A., & WINKLHOFER, H. M. (2001). *Index construction with formative indicators: An alternative to scale development*. Journal of Marketing Research, 38, 269-277.
- DIAMANTOPOULOS, A., RIEFLER, P., & ROTH, K. P. (2008). *Advancing formative measurement models*. Journal of Business Research, 61, 1203-1218.
- FATTORE, M. (2016). *Partially ordered sets and the measurement of multidimensional ordinal deprivation*. Social Indicators Research, 128(2), 835-858.
- GONZALEZ, E., CARCABA, A., VENTURA, J. (2010) *The Importance of the Geographic Level of Analysis in the Assessment of the Quality of Life: The Case of Spain*. Social Indicators Research. 102, 209-228.
- GRABISCH, M., MARICHAL, J.L., MESIAR, R., PAP, E. (2011) *Aggregation functions: Means*. Information Sciences, Elsevier, 181 (1), 1-22.
- GRECO, S., ISHIZABA, A., TASIOUR, M., TORRISI, G. (2019). *On the Methodological Framework of Composite Indices: A Review of the Issues of Weighting, Aggregation, and Robustness*. Social Indicators Research, 141, 61-94.
- GREYLING, T., & TREGENNA, F. (2016). *Construction and analysis of a composite quality of life index for a region of South Africa*. Social Indicators Research, 131(3), 88-930.
- KARAMIZADEH, S., ABDULLAH, S.M., MANAF, A.A., ZAMANI, M., & HOOMAN, A. (2009). *Principal Component Analysis*. Encyclopedia of Biometrics.
- KENDALL, M.; GIBBONS, J. D. (1990) [First published 1948]. *Rank Correlation Methods*. Charles Griffin Book Series (5th ed.). Oxford: Oxford University Press. ISBN 978-0195208375.
- MAZZIOTTA, M., PARETO, A. (2012). *A non-compensatory approach for the measurement of the quality of life*. In F. Maggino & G. Nuvolati (Eds.), *Quality of life in Italy: Research and reflections* (pp. 27-40). New York: Springer.
- MAZZIOTTA, M., & PARETO, A. (2013). *Methods for constructing composite indices: One for all or all for one?* Rivista Italiana di Economia Demografia e Statistica, LXVII(2), 67-80.
- MAZZIOTTA, M. & PARETO, A. (2017). *Synthesis of Indicators: The Composite Indicators Approach* (Chapter 7). In F. Maggino (ed.) *Complexity in Society: From Indicators Construction to their Synthesis* (pp. 115-138). Cham, Switzerland: Springer International Publishing.
- MAZZIOTTA, M., & PARETO, A. (2018). *Measuring Well-Being Over Time: The Adjusted Mazziotta-Pareto Index Versus Other Non-compensatory Indices*. Social Indicators Research, 136 (3), 967-976.
- MAZZIOTTA, M., & PARETO, A. (2019). *Use and Misuse of PCA for Measuring Well-Being*. Social Indicators Research, 142, 451-476.
- MAGGINO, F. (2017). *Complexity in Society: From Indicators Construction to their Synthesis*. Cham, Switzerland: Springer International Publishing.
- MAGGINO, F. (2017a). *Dealing with Syntheses in a System of Indicators* (Chapter 5). In F. Maggino (ed.) *Complexity in Society: From Indicators Construction to their Synthesis* (pp. 115-138). Cham, Switzerland: Springer International Publishing.
- MAGGINO, F. (2017b). *Developing Indicators and Managing the Complexity* (Chapter 4). In F. Maggino (ed.) *Complexity in Society: From Indicators Construction to their Synthesis* (pp. 87-114). Cham, Switzerland: Springer International Publishing.
- MARIANO, E. B., SOBREIRO, V. A., & REBELATTO, D. A. N. (2015). *Human development and data envelopment analysis: A structured literature review*. Omega, 54, 33-49.
- MIZOBUCHI, H. (2014). *Measuring world better life frontier: a composite indicator for OECD better life index*. Social Indicators Research, 118, 987-1007.
- MUNDA, G. & NARDO, M. *Non-Compensatory/Non-Linear Composite Indicators for Ranking Countries: A Defensible Setting*. Applied Economics, Taylor & Francis (Routledge), 2009, 41 (12), 1513-1523.
- NOORBAKHSH, F. (1996). *The human development indices: Are they redundant?* Occasional Papers No. 20. Centre for Development Studies, University of Glasgow, Glasgow.

- OECD & JOINT RESEARCH CENTRE (2008). Handbook on constructing composite indicators: methodology and user guide. Paris: OECD.
- PEARSON, K. (1901). LIII. *On lines and planes of closest fit to systems of points in space*. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2:11, 559-572.
- PEIRO-PALOMINO, J., PICAZO-TADEO, A. J. (2018). *OECD: One or Many? Ranking Countries a Composite Well-Being Indicator*. Social Indicators Research. 139-847.
- PENA TRAPERO, J.B. (1977) *Problemas de la medición del bienestar y conceptos afines (Una aplicación al caso español)*, INE, Madrid.
- PENA TRAPERO, J. B. (2009). La medición del bienestar social: una revisión crítica. *Estudios de Economía Aplicada*, 27(2), 299-324.
- REIG-MARTINEZ, E. *Social and Economic Wellbeing in Europe and the Mediterranean Basin: Building an Enlarged Human Development Indicator*. Social Indicators Research (2013) 111:527–547.
- SAISANA, M., & TARANTOLA, S. (2002). *State-of-the-art report on current methodologies and practices for composite indicator development*. European Commission, Joint Research Centre, Institute for the Protection and the Security of the Citizen, Technological and Economic Risk Management Unit, Ispra, Italy.
- SANCHEZ, A., & RUIZ-MARTOS, M. (2018). Europe 2020 Strategy and Citizens' Life Satisfaction. *Journal of Happiness Studies*.
- SANCHEZ, A., CHICA-OLMO, J., AND JIMENEZ-AGUILERA, J.D. (2018). A Space–Time Study for Mapping Quality of Life in Andalusia During the Crisis. *Social Indicators Research*, 135(2), 699-728.
- SOMARRIBA, N., & PENA, B. (2009). Synthetic indicators of quality of life in Europe. *Social Indicators Research*, 94(1), 115-133.
- SHWARTZ, M., RESTUCCIA, J. D., & ROSEN, A. K. (2015). *Composite measures of health care provider performance: A description of approaches*. *The Milbank Quarterly*, 93, 788–825.
- UNDP. (1990). *Human development report 1990: Concept and measurement of human development*. Oxford: Oxford University Press.
- UNDP. (2001). *Human development report 2001*. New York: Oxford University Press.
- UNDP. (2010). *Human development report 2010*. New York: Palgrave MacMillan.
- WILSON, P.W. (1995). *Detecting influential observations in data envelopment analysis*. *J Prod Anal* 6, 27-45.
- YANG, F.-C., KAO, R.-H., CHEN, Y.-T., HO, Y.-F., CHO, C.-C., & HUANG, S.-W. (2017). *A common weight approach to construct composite indicators: The evaluation of fourteen emerging markets*. *Social Indicators Research*. <http://doi.org/10.1007/s11205-017-1603-7>. (advance online publication).
- ZARZOSA ESPINA, P. (1996). *Aproximación a la medición del bienestar social*. Valladolid: University of Valladolid.