Socio-Economic Planning Sciences 83 (2022) 101339

Contents lists available at ScienceDirect

Socio-Economic Planning Sciences

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

Dealing with weighting scheme in composite indicators: An unsupervised distance-machine learning proposal for quantitative data

Eduardo Jiménez-Fernández [a], Angeles Sánchez [b,*], Mario Ortega-Pérez [c]

[a] Departamento de Teoría e Historia Económica, Campus Cartuja s/n. Universidad de Granada, 18071, Granada, Spain
[b] Departamento de Economía Aplicada, Campus Cartuja s/n. Universidad de Granada, 18071, Granada, Spain
[c] Institute for Research and Innovation in Bioengineering, Camino de Vera, s/n. Universitat Politècnica de València, 46022, València, Spain

ARTICLE INFO

Keywords:
Composite indicator
P2 distance
Unsupervised machine learning
Benchmarking
Weighting scheme
MARS
PACS
C02
C15
C44
C43

ABSTRACT

There is increasing interest in the construction of composite indicators to benchmark units. However, the mathematical approach on which the most commonly used techniques are based does not allow benchmarking in a reliable way. Additionally, the determination of the weighting scheme in the composite indicators remains one of the most troubling issues. Using the vector space formed by all the observations, we propose a new method for building composite indicators: a distance or metric that considers the concept of proximity among units. This approach enables comparisons between the units being studied, which are always quantitative. To this end, we take the P2 Distance method of Pena Trapero as a starting point and improve its limitations. The proposed methodology eliminates the linear dependence on the model and seeks functional relationships that enable constructing the most efficient model. This approach reduces researcher subjectivity by assigning the weighting scheme with unsupervised machine learning techniques. Monte Carlo simulations confirm that the proposed methodology is robust.

1. Introduction

Composite indicators have clear advantages that justify their increasing use for summarising complex and multidimensional realities that are not directly measurable. For instance, they are used to support decision makers, make comparisons and assess the progress of units (companies, countries, regions, etc.) over time or facilitate communication with the general public [21,23,31].

Composite indices developed by international organizations and institutions choose simplicity as the best methodological option. The most widely used aggregation method is the arithmetic mean. Some examples of indices that use the arithmetic mean include the United Nations Human Development Index from 1990 until 2010 when it was substituted by the geometric mean [35]; the Ease of Doing Business ranking, which studies business regulation at country level [36]; the Better Life Index developed by the Organisation for Economic Cooperation and Development [27] to visualise well-being; the Canadian Index of Wellbeing developed by the Canadian Research Advisory Group University of Waterloo [5]; or the Sustainable Development Goals (SDG) Index supported by Cambridge University Press to assess where each country stands in achieving the SDGs [30]. Other institutions have combined the arithmetic mean and principal component analysis (PCA) in their indices, such as the World Economic Forum's Global Competitiveness Index since 2008 [33] and the European Commission's European Regional Competitiveness Index from 2009 to 2019 [2]. PCA is applied to verify whether the indicators within each dimension are internally consistent and then aggregate them by an arithmetic mean in a second step. In addition to the composite indices developed by international organizations, it is worth highlighting empirical applications that use data envelopment analysis (DEA) in academia. DEA is a methodological approach to evaluate the performance of a set of observations referred to as decision making units (DMUs) that subsequently transform multiple inputs into multiple outputs. DEA methodology has the advantage that it does not depend on the method chosen to normalise the data or on the weights used for aggregation. Recent studies in this line have constructed composite indicators to assess the level of competitiveness of Costa Rican counties [19], evaluate the provision of local municipal services in Flanders [9] or to assess the performance of public hospitals in Portugal from the perspective of users and providers [1].

These indices and rankings, some of which are more influential than others, are taken as a reference to make comparisons between

* Corresponding author.
E-mail address: sancheza@ugr.es (A. Sánchez).

https://doi.org/10.1016/j.seps.2022.101339
Received 4 October 2021; Received in revised form 11 May 2022; Accepted 11 May 2022
Available online 21 May 2022

companies, countries or regions and guide decision-making on public policies and in private companies. However, despite the popularity of these methods, none of them enables benchmarking because they do not provide a mathematical structure to analyse the results through a metric that permits comparisons between units. Benchmarking is one of the required characteristics in a composite indicator and is broadly defined as the capability to interpret results according to a specific frame [21]; pp. 106–107). Within the setting of the construction of composite indicators, a metric is the natural way to establish the proximity or distance between the analysed observations and therefore perform benchmarking in a rigorous and reliable way. Additionally, the weighting schemes of these methods have serious disadvantages that call into question the veracity of their results. Neither the arithmetic mean nor the geometric mean avoid the redundant or overlapping information of single indicators, and PCA only removes the linear redundant information [4,13,15,24,26].

This study presents a new method for building composite indicators that can be applied in formative measurement models [6,10], using quantitative data and a partially compensatory aggregation method based on the mathematical concept of distance or metric. To this end, we draw on the P2 Distance method (hereafter DP2) developed by Pena Trapero [28] that has been widely used in very recent applications (see, for instance, Refs. [8,22,31,32]. Taking into account the available tools when Professor Pena Trapero developed the DP2 method, this method constituted a great methodological advance, particularly due to the introduction of a metric in the construction of composite indicators. Nonetheless, DP2 has some limitations. In this study we address these weaknesses using machine learning (ML) techniques.

More specifically, our composite indicator is the outcome of a weighted ℓ^2 metric, where the weights are computed using unsupervised ML algorithms. Our proposal makes several notable contributions. Firstly, our composite indicator is able to measure distances to perform benchmarking between the units studied in a rigorous way. Secondly, it efficiently eliminates the redundant information provided by the single indicators, so that the weights of the single indicators properly reflect their relative importance. Thirdly, it satisfies a sufficiently large number of mathematical properties to be considered a reliable method. Finally, our composite indicator has passed a robustness analysis.

To achieve our goal, the rest of the paper is structured as follows. In section 2, the DP2 method is reviewed and its positive aspects and limitations are highlighted. Section 3 introduces the DL2, the methodology proposed in this study. Specifically, we analyse the formula to calculate DL2 composite indicators, select the best set of disjoint polynomials between the composite indicator and the set of single indicators, estimate the weights of the single indicators in the calculation of the composite indicator, remove redundant information, and the iterative method or algorithm for calculating the values of the DL2 composite indicators in each unit. Section 4 analyses the mathematical properties the DL2 methodology satisfies and its goodness of fit. Section 5 examines three strategies for checking the robustness of the proposed composite indicator. Section 6 compares the properties that the DP2 and DL2 methods satisfy. Finally, section 7 concludes.

2. Drawing on P2 distance

As in any empirical analysis in which data are used to test a theory or estimate a relationship between variables, the construction of composite indicators requires performing a series of stages to ensure a reliable result. This is a complex task that involves, at least, the following steps: (1) defining the phenomenon to be measured (latent construct), which in turn requires identifying the nature and direction of the structural relationships between the latent construct and the observed variables; (2) selecting a group of variables or single indicators that represent the phenomenon to be studied according to the conceptual framework; (3) normalising the single indicators; (4) weighting and aggregating the normalised indicators using a mathematical method (compensatory, partially compensatory or non-compensatory) and (5) validating the composite index [15,21,23,26]. Likewise, to maximise the robustness and validity of a composite indicator, the most appropriate methodological choices must be made in each of the previous steps.

In this section, we review how the DP2 method responds to these stages, which will allow us to highlight its strengths and identify some weaknesses. Thus, our methodological proposal for constructing composite indicators focuses on overcoming the drawbacks of the DP2 method while taking advantage of its strengths.

Let us first introduce some formal technical concepts and definitions regarding the metric or distance. Let Λ be a nonempty set and let $\mathbb{R}$ be the set of real numbers. A function $d : \Lambda \times \Lambda \rightarrow \mathbb{R}^+$ is said to be a metric or distance if for all $A, B, C \in \Lambda$ the following statements are satisfied:

1 $d(A, B) \geq 0$; $d(A, B) = 0$ if and only if $A = B$,
2 $d(A, B) = d(B, A)$,
3 $d(A, B) \leq d(A, C) + d(C, B)$ (triangular inequality).

Let X be an $n \times m$-dimension matrix where columns $X_{\cdot 1}, \ldots, X_{\cdot m}$ represent the single indicators and the rows of X refer to the i observations or units (regions, countries, etc.). Let $X_{i\cdot} = (x_{i1}, \ldots, x_{im})$ be an m-dimension row vector associated to the i-observation and let $X_{*\cdot} = (x_{*1}, \ldots, x_{*m})$ be a hypothetical unit we call the target vector or baseline. For instance, the vector $X_{*\cdot}$ may represent the best or worst case scenario for each of the single indicators depending on the phenomenon to be measured [1]. Let $d_{ij} = |x_{ij} - x_{*j}|$ be the distance from the i-observation $i \in \{1, \ldots, n\}$ to the j-coordinate of the target vector. For each $j \in \{1, \ldots, m\}$, $R^2_{j,\ldots,1}$ represents the coefficient of determination in the multiple linear regression of $X_{\cdot j}$ over the preceding indicators $X_{\cdot j-1}, \ldots, X_{\cdot 1}$ assuming $R_1^2 = 0$. Let $\omega_j = 1 - R^2_{j,\ldots,1}$ be the weights computed following an iterative process explained below.

Distance *DP*2 can be defined as follows:

$$DP2(X_{i\cdot}, X_{*\cdot}) = \sum_{j=1}^{m} \frac{d_{ij}}{\sigma_j}\omega_j \tag{1}$$

where σ_j is the standard deviation of the j-single indicator $j \in \{1, \ldots, m\}$ subject to the standard deviation $\sigma_j \neq 0$.

Next, we review how the DP2 solves the main issues in the construction of composite indicators and the drawbacks detected in the DP2 method.

1. Under the DP2 framework, the $\frac{|x_{ij} - x_{*j}|}{\sigma_j}$ term transforms the single indicators into dimensionless numbers because the units of measure in the numerator and denominator are cancelled. However, this transformation does not ensure that the scale of measurement is the same

[1] Some single indicators may be positively correlated with the latent variable (positive polarity), whereas others may be negatively correlated with it (negative polarity). For instance, investment in R&D would be positively associated with sustainable development (latent variable), whereas CO2 emissions would be negatively associated. In this case, the target vector would be formed by the worst case scenario in all single indicators: the minimum value of the indicators with positive polarity and the maximum value of the single indicators with negative polarity. Thus, the greater the distance of one unit from the target vector, the higher the value of the composite indicator (i.e. the higher the level of sustainable development).

for all the indicators, since the transformed indicators have a minimum value of zero but the maximum value is not limited. The maximum value depends on the specific distributions of the indicators. To correct this drawback, the indicators must be normalised.

2. Generally, the single indicators that make up the composite indicators are correlated, as they provide information from the same constructor. Hence, to perform a composite indicator of a latent phenomenon, a partially compensatory or non-compensatory aggregation technique of the single indicators should be chosen. In the case of DP2, several studies have tested the results reached with DP2 using other composite indicators developed under a multi-criterion approach (with a double reference point) and concluded that the results are very similar to those of weak and mixed indices (see Refs. [8,20]. This indicates that DP2 can be considered a partially compensatory approach. Regarding this, the method should provide an adequate treatment to avoid the overlapping of information. Unlike other methods (see Ref. [15], DP2 introduces the coefficient factor $1 - R^2$ for this purpose. Nevertheless, the coefficient of determination R^2 only detects the linear correlations between single indicators. This is one of the limitations of the DP2 method that we try to overcome with our proposal.
3. DP2 provides an iterative method to objectively assign weights to the single indicators in the composite indicator. To do so, the Fréchet distance (FD) is taken as a starting point. The FD corresponding to the i-observation is defined as follows

$$FD_i(X_{i\cdot}, X_{*\cdot}) = \sum\nolimits_{j=1}^{m} \frac{d_{ij}}{\sigma_j} \tag{2}$$

All the single indicators in the FD have the same weight or importance. In a first step, DP2 computes the pairwise correlation coefficients between each single indicator and the FD and then sorts the indicators from highest to lowest according to the absolute values of the pairwise correlation coefficient. The indicators are introduced in the model following the previous rank and the weights are calculated according to this criterion. The process continues iteratively until the difference between two average adjacent $DP2$s is less than a fixed threshold. This criterion does not guarantee convergence in the order of the value of the composite indicator of the units or observations. In other words, the DP2 criterion can choose some values of the composite indicator so that the average of the difference between these values and the previous ones is very small but with large differences in the ranking of the units. This is why the criterion to reach the final value of the composite indicator in each unit should take into account the order convergence of units rather than the convergence in mean.

4. DP2 allows comparing observations and provides a mathematical structure to analyse the results through a metric, except when there are collinearity problems among single indicators. More specifically, the procedure will not provide a satisfactory result when a single indicator is a linear combination of other indicators. In this case, for some $j \in \{1, \ldots, m\}$ the weight ω_j is equal to zero and, therefore, the formula $DP2$ is not a distance or metric (statement (1) of a metric is not satisfied) because the defined weights are not always strictly positive. This is an essential aspect in order to inherit the rich properties of a metric. For example, let us assume that function d defines a composite indicator to measure socio-economic status. Let T be the vector of the worst observations of a given set of regions, that is, the hypothetical region with the lowest socio-economic status. Let us assume that A, B denote two different regions belonging to the same set such that $d(T, A) = d(T, B) = 1$ and $d(A, B) = 3$. Fig. 1 shows the measures of the distance between observations where the

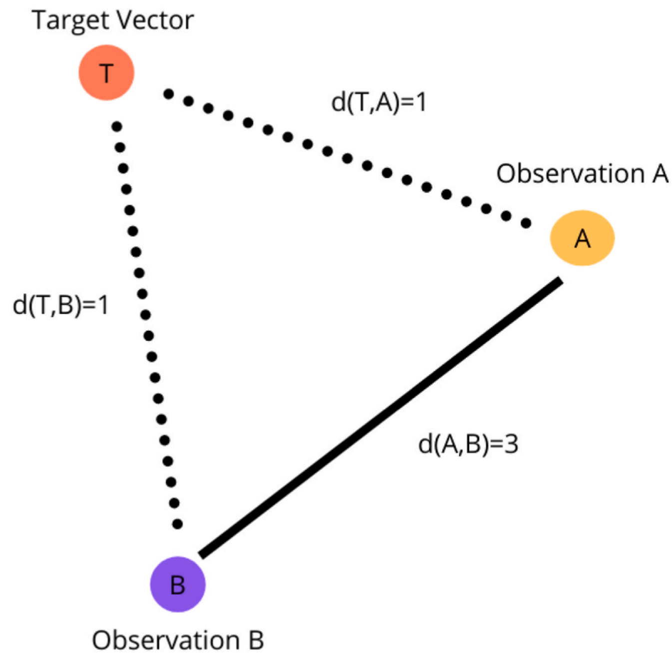

Fig. 1. Triangle inequality is not satisfied.

triangle inequality is not satisfied. Accordingly, $3 = d(A, B) > d(T, A) + d(T, b) = 1 + 1$. A metric space is a set with an associated distance function. This function allows us to establish the concept of proximity, so that for any pair of points of the set we can know the distance between them and therefore perform a range according to this function. In relation to this, the function defined in Fig. 1 is not a distance since it does not satisfy the triangular inequality. Therefore, we cannot know which units have a higher socio-economic status than others.

3. A new proposal: the DL2 composite indicator

In this section we present our proposal for constructing a composite indicator, Distance-Learning or DL2. This composite indicator is the outcome of a weighted ℓ^2 metric[2] in which the weights are computed using iterative ML algorithms. The technique is based on the following concepts. Firstly, the measurement model is formative and works with quantitative data. Secondly, it is based on benchmarking and, thirdly, it is partially compensatory.

Like the review of the DP2 method in the previous section, this section is divided into six parts where we describe how to apply the (DL2) method to determine the values of the composite indicator and overcome the drawbacks of DP2.

To clarify the proposed methodology, the pseudo-code of the proposed algorithm is described in what follows.

Algorithm 1. Computation of the composite indicator $D(Z_{s\cdot}, Z_{*\cdot})$ withrespect to a reference vector $Z_{*\cdot}$.

[2] The metric induced by the norm $\| X_i \| = \sqrt{\sum_{i=1}^{\infty} |x_i|^2}$ associated to ℓ^2 spaces is the generalised way of expressing the classical Euclidean distance.

Algorithm 1: Computation of the composite indicator $D(Z_{s\cdot}, Z_{*\cdot})$ with respect to a reference vector $Z_{*\cdot}$.

Data: (Inputs)

$\ell = 1$,

$error = \alpha$,

$\max\_iterations > 1$,

$p - value_1 = 1$,

τ,

$weights = rep(1, n)$,

Result: (0utput)

Composite indicator $D(Z_{s\cdot}, Z_{t\cdot})^{(\ell)}$

Initialisation: Compute $D(Z_{s\cdot}, Z_{t\cdot})^{(0)}$ and the weights $\{\omega_1^{(0)}, ..., \omega_m^{(0)}\}$;

repeat

  Compute $D(Z_{s\cdot}, Z_{t\cdot})^{(\ell)}$ using the weights $\{\omega_1^{(\ell-1)}, ..., \omega_m^{(\ell-1)}\}$

  Compute the next weights $\{\omega_1^{(\ell)}, ..., \omega_m^{(\ell)}\}$ of the composite indicator $D(Z_{s\cdot}, Z_{t\cdot})^{(\ell)}$

  Apply Kendall correlation to $D(Z_{s\cdot}, Z_{t\cdot})^{(\ell)}$ $D(Z_{s\cdot}, Z_{t\cdot})^{(\ell-1)}$

  Compute $p - value^{\ell}$ (null hypothesis of no association)

  Compute Kendall $\hat{\tau}^{(\ell)}$

  $\ell = \ell + 1$;

until $\ell \leq$ max_iterations or $(\hat{\tau}_\ell \leq \tau$ and $p - value_\ell < error)$;

- **Normalisation**

In a first stage, we pre-process the set of data X. The quantitative variables are normalised by re-scaling (or Min-Max) that converts single indicators into a common scale, namely into the interval [0, 1]. This prevents some indicators from weighing more than others in the composite indicator (DL2). Let $M_j = max(x_{ij})$, $m_j = min(x_{ij})$ denote the maximum and minimum corresponding to each j-single indicator $j \in \{1, …, m\}$, then

$$z_{ij} = \frac{x_{ij} - m_j}{M_j - m_j} \tag{3}$$

if the single indicator has positive polarity and

$$z_{ij} = \frac{M_j - x_{ij}}{M_j - m_j} \tag{4}$$

if the single indicator has negative polarity. Let Z denote the $n \times m$ matrix whose columns are the single standardised indicators.

- **DL2 formula**

Definition 1 Let $D : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}^+ \cup \{0\}$ be a map and $\omega_j \in \mathbb{R}^+$ for all $j \in \{1, …, m\}$. We define $DL2$ as follows:

$$D(Z_{s\cdot}, Z_{t\cdot}) = \left(\sum_{j=1}^{m} |z_{sj} - z_{tj}|^2 \omega_j\right)^{1/2} \tag{5}$$

where s and t are two compared units or observations[3].

- **Fréchet distance.** In a third stage, we compute the FD with respect to the vector reference $Z_{*\cdot} = (z_{*1}, …, z_{*m})$ as follows:

$$FD(Z_{s\cdot}, Z_{*\cdot}) = \left(\sum_{j=1}^{m} |z_{sj} - z_{*j}|^2\right)^{1/2} \tag{6}$$

The FD does not take into account the overlap of information that may exist between the single indicators; however, it provides a first approximation to the final composite indicator. The FD depicts the initial seed of our supervised algorithm. Note that if the single indicators were independent of each other, the information they contribute to the composite indicators would not overlap in conceptual terms. In such a case, the methodology presented in this article would simply consist of calculating the FD.

- **Selecting the best set of disjoint polynomials.** In a fourth stage, when the iteration is equal to one, the algorithm searches functional relationships between the set of single indicators and the FD. To perform this task, we use multivariate adaptive regression splines (MARS) [12].MARS estimates different regression slopes at different

[3] This formula provides the distance between two observations for any normalisation of data. Note that with the normalisation chosen in this article, z_{*j} is the null vector.

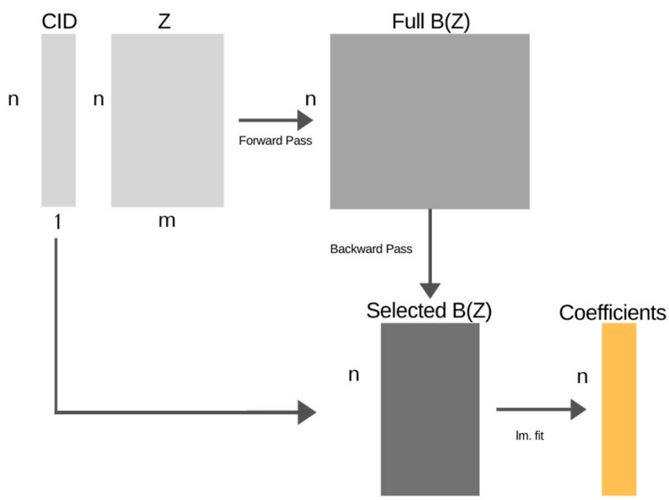

Fig. 2. Overview of EARTH steps.

intervals for each predictor and selects the best set of disjoint polynomials between the composite indicator and the set of single indicators.

Unlike *DP*2, which selects weights using ordinary linear regression (OLS) as a hinge to establish relationships between the composite indicator and single indicators, MARS is a non-parametric method that extends the model by looking for non-linear interactions between the single indicators and the composite indicator. Moreover, the algorithm is insensitive to the basic assumptions of linear regression, which enables it to detect irrelevant indicators in the model [18]. To perform this task, the model explicitly includes polynomial parameters or step functions.

Let us shorten $D_i^{\ell-1} = D(Z_{i\cdot}, Z_{*\cdot})$ as the composite indicator DL2 computed in the $\ell - 1$ iteration $\ell \in \{1, \dots, m\}$ (for more details, see Algorithm to determine DL2 below) associated to the i observation or unit and the target vector. The set of disjoint polynomials $B(z_{ih})$ are functions that depend on the respective variables z_{ih}, where each $B(z_{ih})$ $h \in \{1, \dots, H\}$ can be written as $B(z_{ih}) = \max(0, z_{ih} - c)$ or $B(z_{ih}) = max(0, c - z_{ih})$, where c is a threshold value and H represents the number of explanatory indicators, which includes interactions of the predictor variables. The final model is a combination of the generated set of disjoint polynomials including possible interactions between predictors. The MARS model can be written as follows:

$$D_i^{\ell-1} = \beta_0 + \beta_1 B(z_{i1}) + \dots + \beta_m B(z_{iH}) + \epsilon_i \tag{7}$$

where the coefficients β_j are estimated by minimising the sum of squared errors and the error term ϵ follows a normal distribution $N(0, \sigma^2)$.

To select the best model, an ML algorithm is used. A first step is to start with a model containing only the β_0 intercept and iteratively add disjoint polynomials to the model. During the training process, MARS selects new disjoint polynomials that minimise the sum of squared (residual) errors (SSE) using OLS. The forward step continues until a matrix of predictors $B(Z)$ is completed. In general, at the end of this process, $B(Z)$ has a much larger number of columns than the original single indicators $H > m$ (Fig. 2). The second phase of this algorithm uses the backward stepwise process. The functions that contribute least to the fit are removed through 10-fold cross-validation (CV) [7,11] until the best sub-model is found. The entire procedure is executed using the R EARTH package ("Notes on the EARTH package", Stephen Milborrow, personal notes, September 15, 2020). The steps explained above are illustrated in Fig. 2.

- **Computing the weights of DL2.** In a fifth stage, we compute the weights of the DL2 (ω_j in Equation (5)) corresponding to the variable importance (VI) function using the simple feature importance ranking measure (FIRM) of the VIP package [14]. This tool provides a standardised model-based approach for measuring a single indicator's importance across the growing spectrum of supervised learning algorithms. This function allows us to classify the single indicators in terms of their relative influence on the predicted DL2. Roughly speaking, VI provides a measure of the strength of the relationship between single indicators. Thus, VI quantifies the relative "flatness" of the effect of each feature z_j with respect to the other indicators $\{z_1, \dots, z_{j-1}, z_{j+1}, z_{j+1}, \dots, z_k\}$, considering that the estimation is evaluated on the functional relationship $\widehat{B}$ obtained in the previous step. The normalised scores provided by VI have a similar role as the correction factor $1 - R^2_{j,\dots,1}$ in the DP2 methodology.

For example, if VI assigns a value close to 0 to the first single indicator, the indicator will contribute very little information to the model. When a score is equal to zero, we assign the value $min(\omega_j)/m$ to the corresponding weight, where $min(\omega_j)$ is the minimum nonzero weight and m is the number of single indicators [4]. Conversely, if VI assigns a value of 1 to a single indicator, then this indicator will be the most relevant in the model and the assigned weight will be 1.

- **Algorithm to determine DL2.** We now analyse the iterative method of calculation. To compute the composite indicator, we begin by calculating the FD, for which all single indicators have the same relevance, i.e. $\omega_1 = \dots = \omega_m = 1$. The FD will be the first composite indicator for iteration one and is called $D^{(0)}$. Assuming $D^{(0)}$ as a response variable, we calculate the set of disjoint polynomials $\widehat{B}(Z)$ that best approximates the single standardised indicators. We then compute the variable importance with respect to $D^{(0)}$ and obtain the first weights $\{\omega_1^{(1)}, \dots, \omega_m^{(1)}\}$. The metric or distance (Equation (5)) allows us to calculate a new composite indicator with the previous weights we call $D^{(1)}$. This iterative process generates a succession of weights $\{\omega_1^{(\ell)}, \dots, \omega_m^{(\ell)}\}_{\ell=1}^m$ and composite indicators $\{D^{(\ell)}\}_{\ell=1}^m$.

Note that the weights of iteration ℓ have been calculated with respect to the composite indicator of iteration $\ell - 1$. Therefore, it is necessary to decide when the algorithm should be stopped. Each composite indicator induces a rank into the observations. We use the non-parametric hypothesis test (the Kendall rank correlation coefficient or Kendall's τ coefficient) as a measure to compute the ordinal association between $D^{(\ell-1)}$ and $D^{(\ell)}$. This tool provides the rank similarity of the two composite indicators. In the case that there are no ties between the indicator $D^{(\ell-1)}$ and $D^{(\ell)}$, the correlation coefficient is expressed as follows

$$\tau = \frac{\sum_{i<j} (sign(D_j^{(\ell)} - D_i^{(\ell)}) * sign(D_j^{(\ell-1)} - D_i^{(\ell-1)}))}{D}$$

where $D = n(n-1)/2$. In the case of ties, the expression D is somewhat more complex (see Ref. [16]; chapter 3). Thus, under the null hypothesis of no association, the τ statistic provides. Consequently, a τ in the interval [0.9, 1] indicates strong agreement between two consecutive composite indicators. Moreover, the algorithm stops when we reject the null hypothesis and the τ statistic is greater than 0.9. Intuitively, this result will confirm that the weights added to the model have not changed the ranking of the composite indicator.

[4] Note that the weigths are not completely eliminated to ensure our DL2 continues to be a metric or distance.

4. Properties of the DL2

The methodology presented in this article is based on a weighted distance metric. The mathematical properties of this type of structure hold and are listed below.The proofs of all statements can be found in the Appendix.

i Map D defines a metric or distance. Firstly, the distance between two observations is positive or zero. In the latter case, the two units must be the same. Secondly, the distance from observation A to observation B is the same as the distance from B to A. Thirdly, according to Fig. 1, the distance from B to T plus T to A must be greater or equal than the distance from B to A.

ii Map D is well defined. The composite indicator assigns a unique interpretation or value for each unit or observation.

iii Monotonicity. If a single indicator (with positive polarity) increases (decreases) while keeping the others constant, the computed index should increase (decrease).

iv Invariance by origin and scale changes. The standardisation is invariant by origin and scale changes.

v Transitivity. The composite indicator satisfies that if A is at least as great as B, and B is at least as great as C, then A is at least as great as C.

vi Homogeneity. A proportional increase (decrease) in all the single indicators generates a proportional increase (decrease) of equal magnitude in the composite indicator.

vii Symmetry. Permutations of the simple indexers lead to the same result.

Theorem 2 Let $D : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}$ be a map. The composite indicator distance defined as Equation (5) satisfies the properties listed above.

5. Robustness of the DL2

This section focuses on the robustness assessment of the DL2 in terms of its capacity to produce correct and stable measures. We develop three strategies to test whether the composite indicators built with the DL2 method are able to deal with adversities that may arise due to the selection of single indicators, the way single indicators are introduced in the model, and changes in the number of units analysed over time. To this end, we use Monte Carlo simulations. We perform a data set Z where 10 single indicators $\{Z_{\cdot 1}, \ldots, Z_{\cdot 10}\}$ are analysed and 400 random uniform observations are generated for each single indicator ($i \in \{1, \ldots, 400\}$).

The first strategy to validate the DL2 is to examine the selection of single indicators. The selection of single indicators is a fundamental step that is closely related to the concept to be modelled and the choice must be supported by the theoretical and empirical literature [21]. Nevertheless, some of the single indicators could be a perfect or almost perfect linear combination of the rest. We want to know how this situation could affect the values of the DL2 composite indicator. Let $\{Z_{\cdot 1}, \ldots, Z_{\cdot 10}\}$ be the single indicators generated through uniform random variables and $Z_{\cdot Added} = \sum_{j=1}^{10} \alpha_j Z_{\cdot j}$ a linear convex combination. We built a Monte Carlo procedure on 100 random convex linear combinations. For each linear combination, we calculated the corresponding weights using the proposed methodology. Fig. 3 shows that the *Zadded* single indicator is irrelevant for the computed iterations.

However, we observe that if we perform the same computation, but this time for the distance P2, the variable Zadded is not eliminated and is also the most relevant (Fig. 4).

The second strategy to validate the DL2 involves checking whether the way single indicators are introduced in the model alters the values of the composite indicator in the units. Given that our proposal relies on VI to estimate the functional relationships between the composite indicator (DL2) and the set of single indicators, it is reasonable to test whether permutations of the single indicators change their weights or importance in the composite indicator. Therefore, we calculated the DL2 corresponding to 100 different permutations according to the rank of the single indicators $Z_{\cdot j}$ $j \in \{1, \ldots, 10\}$. The results provided the same weighted scheme, producing equal values of DL2. Therefore, the algorithm is invariant to how single indicators are introduced in the model and the calculated weights are not altered by these permutations (results are available upon request).

Finally, the third strategy checks to what extent an uncertain number of units (in the future) would affect the units' rank. This is a situation that can arise when companies that are part of a database in one year leave the market in other years (e.g. due to closure), when there are administrative changes that affect the number of units (e.g. recognition of new municipalities) or simply when for one year or several years the information of the single indicators is not available for some of the units analysed. When studying the robustness of a method for constructing stable composite indicators, it is essential to check if the method will still be able to provide a reliable and comparable measurement over time in the event any of these situations occur.

To check whether composite indicators built with the DL2 method are able to deal with such situations, we analyse the variability in the range of scores of the DL2 when some units are removed at random. To perform this test, we again use a Monte Carlo procedure. Firstly, 10 random observations or units are deleted from the database and the composite indicator DL2 is calculated with the remaining 390 observations. Secondly, we use the original database with the 400 units to calculate the DL2. Once computed, we remove the DL2 values corresponding to the same 10 observations that were eliminated in the previous step. As a result of the two steps, two composite indicators are obtained and compared using Spearman's, Kendall's and robust range correlation statistics. Assuming that the probability of a type I error is $\alpha = 0.05$, a 100 data set was analysed, for which 10 observations were

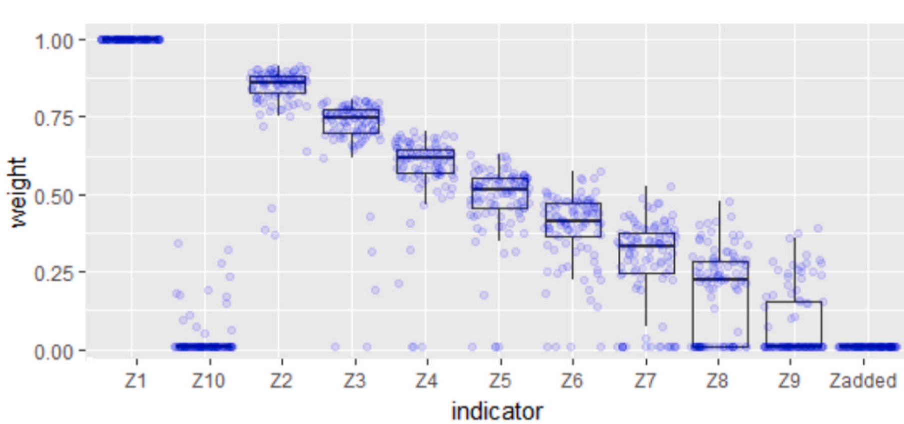

Fig. 3. Weights of single indicators in 100 iterations for DL2 ($n = 400$).

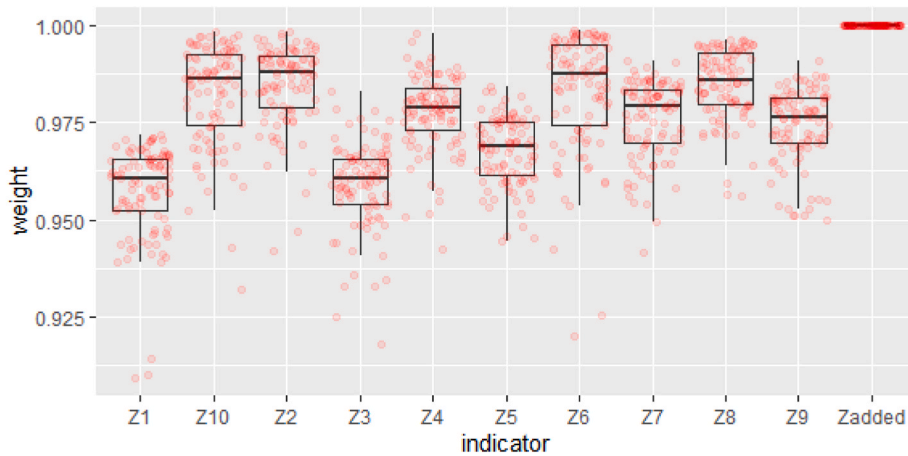

Fig. 4. Weights of single indicators in 100 iterations for DP2 ($n = 400$).

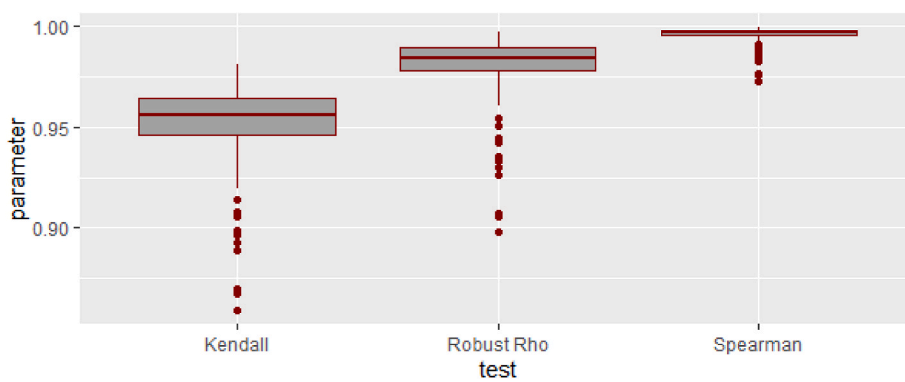

Fig. 5. Results of simulations of DL2 when 10 observations are randomly deleted ($n = 390$).

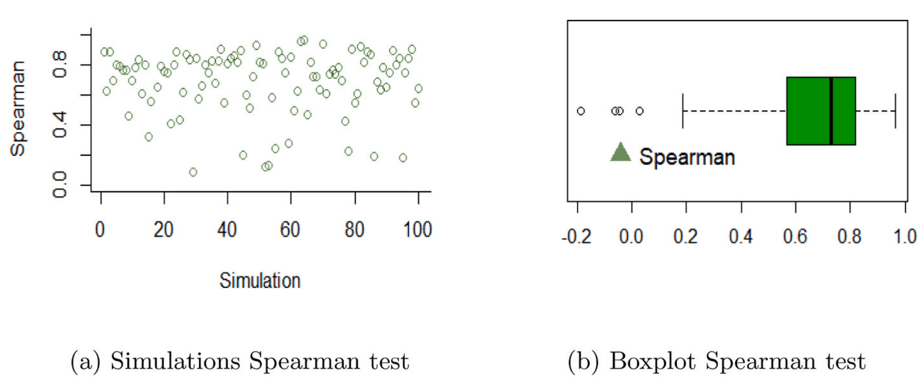

Fig. 6. Montecarlo simulation to compare DL2 and DP2 composite indicators
(a) Simulations Spearman test
(b) Boxplot Spearman test.

randomly deleted. Fig. 5 shows the three correlation tests computed. All the tests show evidence of a strong correlation between the two composite indicators. Thus, the random elimination of observations does not produce significant differences in the units' ranking in the composite indicator. Consequently, the composite indicators constructed with the DL2 method are stable over time even if there are changes in the units analysed.

6. Comparison between DL2 and DP2 methods

In this section, we compare the DL2 and DP2 methods in terms of the results that would be achieved in an empirical application and the mathematical properties that verify the composite indicators constructed with both methodologies.

For the first comparison, a Monte Carlo algorithm was designed to analyse the ranking of the units in terms of the composite indicator values using both methods. One hundred databases with 400 observations or units and 10 single indicators were generated. The single indicators follow a normal distribution with randomly chosen mean and variance parameters. In addition, each database was designed to have strong, intermediate and weak correlations between the single indicators to ensure that the algorithm presents extreme cases and to

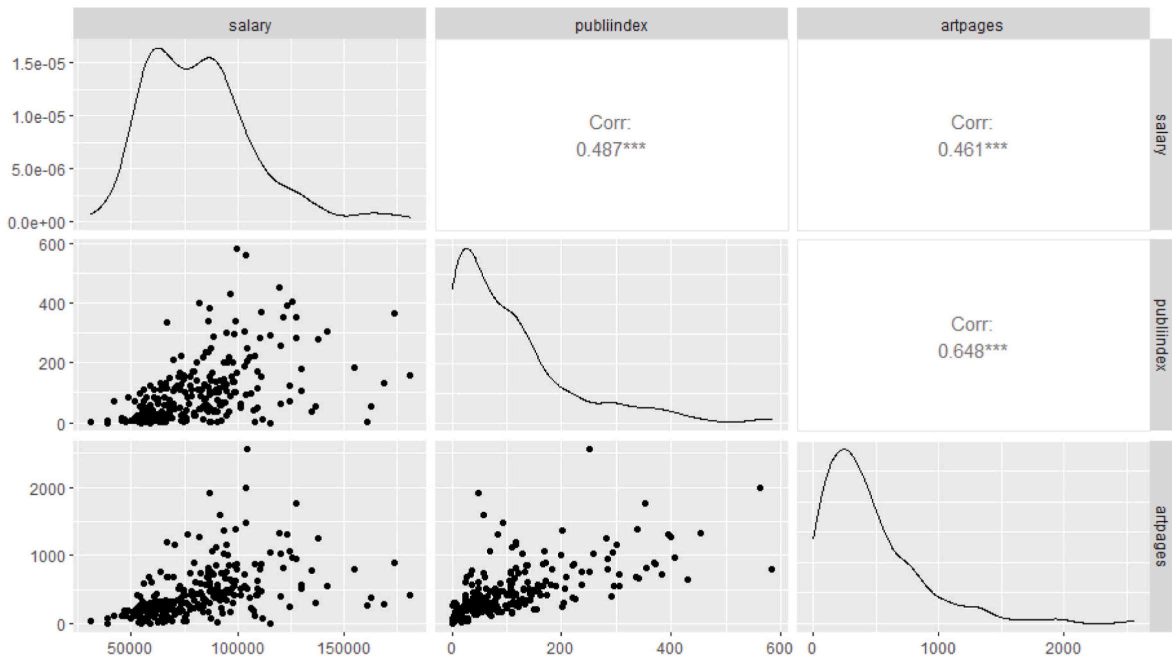

Fig. 7. Correlogram with ggpairs ($n = 246$).

Table 1
Weights of aggregation methods.

	Salary	Pubiindex	artpages
DL2	0.3684520	0.2296465	0.4019016
DP2	0.7685526	0.8976993	0.8111156

highlight the differences in the methodologies. Fig. 6 shows the results of the comparison.

The boxplot of Spearman's test (part b of Fig. 6) indicates that 75% of the simulations have a Spearman's correlation between the values of both composite indicators (DL2 and DP2) greater than 0.6, but only 25% of them showed correlations above 0.85. Two plausible explanations can be given for these divergences in the ranking of both composite indicators, which is why Spearman's correlations appear so low. Firstly, as analysed in a previous section, the methodology presented in this paper (DL2) introduces improvements that allow solving the collinearity problems that may occur in some databases, whereas DP2 is unable to detect single indicators with multicollinearity problems. Secondly, some simulations in both methods can reach the maximum number of iterations introduced to stop the algorithm and, therefore, the optimal solution is not reached. In these cases of non-convergence, the rankings of the two methods show a larger difference.

Additionally, for the second comparison between the DL2 and DP2 methods, we consider the bigsalary data set [3] available in the Wooldridge R software package. The total number of observations is 246. The variables considered for this comparison are the identifier of each faculty member (id), the annual salary measured in dollars (salary), an indicator of publications (publiindex), and the standardised total article pages (artpages). Fig. 7 shows high correlations among the variables analysed. These correlations make it necessary to eliminate redundant information to avoid overestimates in the calculated composite indicator. We compute DL2 and DP2. It is worth highlighting that DP2 calculates weights that are not a convex linear combination. If this were the case, the weights would correspond to 0.3102296, 0.3623602, 0.3274103, that is, they would be very similar to those provided by DL2.

We also tested for associations between the DL2 and DP2 indicators using Pearson's product-moment correlation coefficient, Kendall's τ or Spearman's ρ (see Table 1). The null hypothesis for which the parameter corresponding to each test is zero is rejected, obtaining correlations of 0.9544713, 0.8417787 and 0.9659782, respectively. Table 2 shows the first 10 observations where *dell*2 denotes the DL2 composite indicator,

Table 3
Comparison of aggregation methods.

	Distance	Existence	Monotonicity	Invariance
DP2	×	×	✓	✓
DL2	✓	✓	✓	✓
	Transitivity	Homogeneity	Symmetry	Exhaustiveness
DP2	✓	×	✓	✓
DL2	✓	✓	✓	✓

Table 2
First 10 observations ranked by DL2

id	salary	publiindex	artpages	dell2	dell2_rank	dp2	dp2_rank
154	30813.67	2.24	29.00	0.05	1	0.06	1
17	38770.00	0.00	0.00	0.19	2	0.23	2
139	38934.50	4.86	65.00	0.22	3	0.38	3
182	45533.33	6.87	105.00	0.39	4	0.64	4
146	45319.00	15.19	110.50	0.40	5	0.72	8
54	47113.00	4.29	105.00	0.43	6	0.66	5
148	41748.33	71.44	112.00	0.45	7	1.14	35
164	49890.00	7.51	71.00	0.47	8	0.72	9
48	46492.00	10.67	193.50	0.49	9	0.84	15
122	50971.00	9.38	129.00	0.53	10	0.86	17

*dp*2 denotes the *DP*2 composite indicator and *dell*2*_rank* and *dp*2*_rank* denote the rank obtained from the worst numbering in each of the methods.

The third comparison between the DP2 and the proposed DL2 methodologies focuses on the mathematical properties a composite indicator should fulfil to assess its goodness of fit. Table 3 summarises the comparison between DL2 and DP2 in terms of these properties.

In a previous section we showed that since the DL2 is a weighted metric distance, it satisfies the following seven properties: distance, well-defined, monotonicity, invariance by origin and scale changes, transitivity, homogeneity and symmetry. To these seven properties, we add the concept of exhaustiveness. Exhaustiveness refers to the fact that a composite indicator should take full advantage, and in a useful way, of the information provided by the single indicators. In this vein, a composite indicator is better than another one if it provides more useful information about the phenomenon being studied, but it must also be able to eliminate duplicate information [28,29,37]. Composite indicators built with DL2 are exhaustive because the weights of the single indicators are computed according to their relevance through VI scores, and the model is able to avoid overlapping information.

As regards the composite indicators built with the DP2 method, it should be noted that they fulfil all the properties except for distance (in all cases), existence and homogeneity. These weaknesses have been analysed in a previous section. Composite indicators constructed with DP2 would not be a metric or distance if there were a collinearity problem. In such a case, $R^2_{j,\ldots,1} = 1$ for some $j \in \{1, \ldots, m\}$, thus indicating that statement (i) in section 2 is not satisfied. If the standard deviation of a single indicator were equal to zero (i.e. the single indicator has the same value in all the units), problems of existence could occur since Equation (1) would not be well defined. Let α be a positive real number. The non-homogeneity would be due to the fact that $DP2(\alpha X_{i\cdot}) = DP2(X_{i\cdot})$. Lastly, it is worth noting that DP2 would be considered exhaustive, although it only detects relationships between single and composite indicators of a linear nature.

7. Conclusions

The increasing use of composite indicators in economics and the social sciences is warranted because they allow comparisons to be made between units (companies, territories, etc.) and assess the progress or evolution of units over time [21,23,31]. Consequently, for composite indicators to be effective, they must be developed with robust methods that ensure that the two objectives of benchmarking and stability over time are achieved. However, the mathematical approach in which the most widely used techniques to build composite indicators are grounded (i.e. arithmetic and geometric means, PCA and DEA) does not enable addressing these issues in a reliable way. Indeed, the weighting and aggregation aspects of these techniques have received much criticism in the most recent literature (see Refs. [4,13,15,17].

In addition to weighting scheme and aggregation, in this paper we showed that none of these techniques provide a mathematical structure for analysing results through a metric or distance. Consequently, it is unfeasible to establish proximity or distance between the units analysed, so that both quantitative and ordinal comparisons (unit rankings) lack a solid basis. This paper has proposed a new method for building composite indicators called DL2. This method is based on the mathematical concept of distance or metric that enables comparisons between the units being studied. Its application is intended for quantitative data in formative measurement models in the context of compensatory aggregation.

Our proposal took as a starting point the Distance P2 or DP2 method developed by Pena Trapero [28] given its remarkable advantages studied in a previous section: it provides a mathematical structure (except in extreme cases) that enables the units to be ranked according to distance, it avoids linear overlapping information between single indicators and the composite indicator, it fulfils most of the mathematical properties required to assess goodness of fit, and it is quite versatile, thus allows the analysis of a wide range of multidimensional phenomena (see, for instance, Refs. [8,22,25,31,32].

Likewise, we identified the limitations of the DP2 method and improved them by taking advantage of ML techniques, as well as the growing computational capacity. Our improvements included, firstly, correcting cases in which DP2 is not a metric since the defined weights are not always strictly positive. Secondly, because the DP2 method inherently relies on linear models, it does not behave efficiently when single indicators are poorly correlated with the composite indicator. The proposed DL2 method corrects this weakness by means of unsupervised ML algorithms. From the composite indicator generated by the unweighted metrics, the algorithm optimizes the best functional relationship (not necessarily linear) between the composite indicator and the single indicators. By means of ML, DL2 ranks the single indicators in order of importance by assigning weights to the metric based on this relationship. The algorithm stops when the order of the units (in terms of composite indicator values) remains unchanged.

We also analysed the mathematical properties of our proposed method to study its goodness of fit and concluded that it is a distance, it is well-defined, and it satisfies the properties of monotonicity, invariance by origin and scale changes, transitivity, homogeneity and symmetry. To the best of our knowledge, this kind of analysis has been scarcely addressed in the literature. Furthermore, we compared the DP2 to our method and identified the properties that DP2 might not fulfil in some cases. In this regard, the method we have proposed overcomes these weaknesses of the DP.

Finally, the Monte Carlo simulations and real data set confirm that the proposed methodology DL2 is robust for building composite indicators. The results of the composite indicator or unit rankings remain stable over time, even when the number of analysed units changes. The method detects and eliminates multicollinearity problems among the single indicators. Likewise, the weighting scheme is not altered by permutations in the order in which the single indicators are computed. The requirement of robustness should be a mandatory step in any composite indicator proposal, since the results can guide public decisions regarding the allocation of economic resources, which can be scarce and susceptible to alternative uses. Nevertheless, little attention has been paid to this step in empirical applications [13].

Appendix

Proofs of the mathematical properties listed in 8.

iMap D defines a metric or distance. Firstly, the distance between two observations is positive or zero. In the latter case, the two units must be the same. Secondly, the distance from observation A to B is the same as the distance from B to A. Thirdly, according to Fig. 1, the distance from B to T plus T to A must be greater or equal than the distance from B to A.

Verification is immediate for all statements except for triangular inequality (statement 3).

Proof i

$$\begin{aligned}D(Z_{s\cdot},Z_{p\cdot}) &= \left(\sum_{j=1}^{m}|z_{sj}-z_{pj}|^2\omega_j\right)^{1/2} = \left(\sum_{j=1}^{m}|z_{sj}\omega_j^{1/2}-z_{pj}\omega_j^{1/2}|^2\right)^{1/2}\\ &= \left(\sum_{j=1}^{m}|z_{sj}\omega_j^{1/2}-z_{tj}\omega_j^{1/2}+z_{tj}\omega_j^{1/2}-z_{pj}\omega_j^{1/2}|^2\right)^{1/2}\\ &\le^{*} \left(\sum_{j=1}^{m}|z_{sj}-z_{tj}|^2\omega_j\right)^{1/2} + \left(\sum_{j=1}^{m}|z_{tj}-z_{pj}|^2\omega_j\right)^{1/2}\\ &= D(Z_{s\cdot},Z_{t\cdot})+D(Z_{t\cdot},Z_{p\cdot}).\end{aligned}$$

in which the inequality (*) is obtained through Minkowski and Holder inequality.

iiMap D is well defined. The composite indicator assigns a unique interpretation or value for each unit or observation.

Proof ii

For any $Z_{i\cdot}$, $i \in \{1, \ldots, n\}$, map D exists and belongs to $\mathbb{R}^+ \cup 0$.

iiiMonotonicity. If a single indicator (with positive polarity) increases (decreases) while keeping the others constant, the computed index should increase (decrease). DL2 is monotone.

Proof iii

Let $D(Z_{i\cdot}) = (z_{i1}, \ldots, z_{im})$ be the m-vector of an observation corresponding to the i observation and let $D(Z_{*\cdot})$ be the vector reference. Without loss of generality, let us assume that for $j = 1$ $z_{*1} < z_{i1} < z_{i1} + \epsilon$, where z_{*1} is the best scenario and *varepsilon* is some positive constant, then $|z_{i1} - z_{*1}| < |z_{i1} + \epsilon - z_{*1}|$.

$$\begin{aligned}D(Z_{i\cdot},Z_{*\cdot}) &= \left(\sum_{j=1}^{m}|z_{ij}-z_{*j}|^2\omega_j\right)^{1/2}\\ &= \left(|z_{i1}-z_{*1}|^2\omega_1+|z_{i2}-z_{*2}|^2\omega_2+\cdots+|z_{im}-z_{*m}|^2\omega_m\right)^{1/2}\\ &< \left(|z_{i1}+\epsilon-z_{*1}|^2\omega_1+|z_{i2}-z_{*2}|^2\omega_2+\cdots+|z_{im}-z_{*m}|^2\omega_m\right)^{1/2}\\ &= D(Z'_{i\cdot},Z_{*\cdot})\end{aligned}$$

where $Z'_{i\cdot}$ is the m-dimension vector whose first coordinate corresponds to the increment $z_{i1} + \epsilon$.

Conversely, assuming that the Z_{1*} single indicator has negative polarity, $z_{i1} < z_{i1} + \epsilon < z_{*1}$, where z_{*j} is the worst scenario is some positive constant, then $|z_{i1} - z_{*1}| > |z_{i1} + \epsilon - z_{*1}|$.

$$\begin{aligned}D(Z_{i\cdot},Z_{*\cdot}) &= \left(\sum_{j=1}^{m}|z_{ij}-z_{*j}|^2\omega_j\right)^{1/2}\\ &= \left(|z_{i1}-z_{*1}|^2\omega_1+|z_{i2}-z_{*2}|^2\omega_2+\cdots+|z_{im}-z_{*m}|^2\omega_m\right)^{1/2}\\ &> \left(|z_{i1}+\epsilon-z_{*1}|^2\omega_1+|z_{i2}-z_{*2}|^2\omega_2+\cdots+|z_{im}-z_{*m}|^2\omega_m\right)^{1/2}\\ &= D(Z'_{i\cdot},Z_{*\cdot})\end{aligned}$$

On the other hand, an increase in a single indicator with negative polarity must generate a decrease in the composite indicator, responding positively to a positive change in any indicators and negatively to a negative change. The proof is symmetric with respect to the positive polarity.

ivInvariance by origin and scale changes. The standardisation tool is invariant by origin and scale changes.

Proof iv

Let $M_j = max(x_{ij})$, $m_j = min(x_{ij})$ denote the maximum and minimum corresponding to each j-single indicator, then

$$z_{ij} = \frac{x_{ij-m_j}}{M_j - m_j} \tag{8}$$

for each $j \in \{1, \ldots, m\}$. It is sufficient to check that the standardisation is invariant by change of scale for each single indicator. Let $\nu_{ij} = \alpha x_{ij} + \beta$ denote a origin and scale changes, where α is a positive real number and β any real number. For all $j \in \{1, \ldots, m\}$

$$\begin{aligned}\min\{\nu_{ij}\} &= \min\{\alpha x_{ij}+\beta\} = \alpha\min\{x_{ij}\}+\beta.\\ \max\{\nu_{ij}\} &= \max\{\alpha x_{ij}+\beta\} = \alpha\max\{x_{ij}\}+\beta.\end{aligned}$$

hence,

$$\frac{\nu_{ij}-\min\{\nu_j\}}{\max\{\nu_j\}-\min\{\nu_j\}} = \frac{\alpha x_{ij}+\beta-\min\{\alpha x_{ij}+\beta\}}{\max\{\alpha x_{ij}+\beta\}-\min\{\alpha x_{ij}+\beta\}} = \frac{x_{ij-m_j}}{M_j-m_j} = z_{ij} \tag{9}$$

Therefore, it is immediate to check that the standardisation is invariant by origin and scale changes.

vTransitivity.

Proof v

Let $Z_{s\cdot}, Z_{t\cdot}, Z_{\ell\cdot} \in \mathbb{R}^+$ be three observations and let $Z_{*\cdot} \in \mathbb{R}^+$ be the vector reference. We assume that $D(Z_{s\cdot}, Z_{*\cdot}) < D(Z_{t\cdot}, Z_{*\cdot})$ and $D(Z_{t\cdot}, Z_{*\cdot}) < D(Z_{\ell\cdot}, Z_{*\cdot})$ then $D(Z_{s\cdot}, Z_{*\cdot}) < D(Z_{\ell\cdot}, Z_{*\cdot})$.

viHomogeneity. A proportional increase (decrease) in all the single indicators generates a proportional increase (decrease) of equal magnitude in the composite indicator.

Proof vi

Let α be a real positive constant. Then

$$D(\alpha Z_{s\cdot},\alpha Z_{t\cdot}) = \left(\sum_{j=1}^{m}|\alpha z_{sj}-\alpha z_{tj}|^2\omega_j\right)^{1/2} = \alpha D(Z_{s\cdot},Z_{t\cdot}). \tag{10}$$

viiSymmetry.

Proof vii

In the proposed method, the value of the composite indicator does not depend on the rank of the indicators introduced in the n × m X matrix data.

Funding sources

European Commission, project 813234. ERDF-Universidad de Granada, project B-SEJ-242-UGR20. Ministerio de Ciencia e Innovación (España), project PID2019-105708RB.

Funding for open access charge: Universidad de Granada / CBUA.

Declaration of competing interest

The authors declare no conflicts of interest.

References

[1] Alves Pereira M, Santos Camanho A, Figueira JR, Cunha Marques R. Incorporating preference information in a range directional composite indicator: the case of Portuguese public hospitals. Eur J Oper Res 2021;294(2):633–50. https://doi.org/10.1016/j.ejor.2021.01.045.

[2] Annoni P, Dijkstr L. The EU regional competitiveness index 2019. Luxembourg: Publications Office of the European Union; 2019.

[3] Baser O, Pema E. The return of publications for economics faculty. Econ Bull 2003;1:1–13.

[4] Becker W, Saisana M, Paruolo P, Vandecasteele I. Weights and importance in composite indicators: closing the gap. Ecol Indicat 2017;80:12–22. https://doi.org/10.1016/j.ecolind.2017.03.056.

[5] Canadian Index of Wellbeing. How are Canadians really doing? The 2016 CIW National report. Waterloo, ON: Canadian Index of Wellbeing and University of Waterloo; 2016.

[6] Coltman T, Devinney TM, Midgley DF, Venaik S. Formative versus reflective measurement models: two applications of formative measurement. J Bus Res 2008;61(12):1250–62.

[7] Craven P, Wahba G. Estimating the correct degree of smoothing by the method of generalized cross-validation. Numer Math 1979;31:377–403.
[8] Cuenca-García E, Sánchez A, Navarro-Pabsdorf M. Assessing the performance of the least developed countries in terms of the Millennium Development Goals. Eval Progr Plann 2019;72:54–66. https://doi.org/10.1016/j.evalprogplan.2018.09.009.
[9] D'Inverno G, De Witte K. Service level provision in municipalities: a flexible directional distance composite indicator. Eur J Oper Res 2020;286(3). https://doi.org/10.1016/j.ejor.2020.04.012.
[10] Diamantopoulos A, Riefler P, Roth KP. Advancing formative measurement models. J Bus Res 2008;61:1203–18.
[11] Friedman JH, Silverman BW. Flexible parsimonious smoothing and additive modeling. Technometrics 1989;31(1):3–21.
[12] Friedman JH. Multivariate adaptive regression splines. Ann Stat 1991;19(1). 1-141.
[13] Greco S, Ishizaka A, Tasiou M, Torrisi G. On the methodological framework of composite indices: a review of the issues of weighting, aggregation, and robustness. Soc Indicat Res 2019;141:61–94. https://doi.org/10.1007/s11205-017-1832-9.
[14] Greenwell BM, Boehmke BC, McCarthy AJ. A simple and effective model-based variable importance measure. 2018. arXiv:1805.04755v1 [stat.ML].
[15] Jiménez-Fernández E, Ruiz-Martos MJ. Review of some statistical methods for constructing composite indicators. Estudios de Economía Aplicada-Studies of Applied Economy 2020;38(1):1–15. https://doi.org/10.25115/eea.v38i1.3002.
[16] Kendall MG. Rank correlation methods. fourth ed. 1976. Griffin.
[17] Keogh S, O'Neill S, Walsh K. Composite measures for assessing multidimensional social exclusion in later life: conceptual and methodological challenges. Soc Indicat Res 2021. https://doi.org/10.1007/s11205-021-02617-7.
[18] Kuhn M. Building predictive models in R using the caret package. J Stat Software 2008;28(5):1–26. https://doi.org/10.18637/jss.v028.i05.
[19] Lafuente E, Araya M, Leiva JC. Assessment of local competitiveness: a composite indicator analysis of Costa Rican counties using the "Benefit of the Doubt" model. Soc Econ Plann Sci 2020:100864. https://doi.org/10.1016/j.seps.2020.100864.
[20] Luque M, Pérez-Moreno S, Robles JA, Rodríguez B. Measuring child and maternal health in developing countries: a proposal of new hybrid MDG composite indices. Appl. Res. Quality Life 2017;12(3):737–58.
[21] Maggino F. Developing indicators and managing the complexity (chapter 4). In: Maggino F, editor. Complexity in society: from indicators construction to their synthesis. Springer International Publishing; 2017. p. 87–114. https://doi.org/10.1007/978-3-319-60595-1.
[22] Martín JM, Salinas Fernández JA, Rodríguez Martín JA, Ostos Rey S. Analysis of tourism seasonality as a factor limiting the sustainable development of rural areas. J Hospit Tourism Res 2020;44(1). https://doi.org/10.1177/1096348019876688.
[23] Mazziotta M, Pareto A. Synthesis of indicators: the composite indicators approach (chapter 7). In: Maggino F, editor. Complexity in society: from indicators construction to their synthesis; 2017. p. 159–92.
[24] Mazziotta M, Pareto A. Use and misuse of PCA for measuring well-being. Soc Indicat Res 2019;142(2):451–76. https://doi.org/10.1007/s11205-018-1933-0.
[25] Montero J-M, Chasco C, Larraz B. Building an environmental quality index for a big city: a spatial interpolation approach combined with a distance indicator. J Geogr Syst 2010;12(4):435–59.
[26] OECD. Handbook on constructing composite indicators. Methodology and user guide. Paris: OECD publishing; 2008. https://doi.org/10.1787/9789264043466-en.
[27] OECD. How's Life? 2017: measuring well-being. OECD Publishing; 2017. https://doi.org/10.1787/how-life-2017-en.
[28] Pena Trapero JB. Problemas de la medición del bienestar y conceptos afines. Una aplicación al caso español (Problems of welfare measurement and related concepts. An application to the Spanish case). Madrid, Spain: INE; 1977.
[29] Pena Trapero JB. La medición del bienestar social: una revisión crítica(Measuring social welfare: a critical review). Estudios de Economía Aplicada/Studies of Applied Economy 2009;27(2):299–324.
[30] Sachs J, Schmidt-Traub G, Kroll C, Lafortune G, Fuller G. SDG indicator and dashboards report 2018. New York: Bertelsmann Stiftung and Sustainable Development Solutions Network (SDSN); 2018.
[31] Sánchez A, Chica-Olmo J, Jiménez-Aguilera JD. A space-Time study for mapping quality of life in Andalusia during the crisis. Soc Indicat Res 2018;135(2):699–728. https://doi.org/10.1007/s11205-016-1497-9.
[32] Sánchez A, Ruiz-Martos M. Europe 2020 strategy and citizens' life satisfaction. J Happiness Stud 2018;19:2315–38. https://doi.org/10.1007/s10902-017-9928-0.
[33] Schwab K, Porter M. The global competitiveness report 2008–2009. Geneva, Switzerland: World Economic Forum; 2008.
[35] UNDP. Human development indices and indicators. New York: United Nations Development Programme; 2018.
[36] World Bank. Doing business 2020. Comparing business regulation in 190 economies. Washington, DC: International Bank for Reconstruction and Development/The World Bank; 2020.
[37] Zarzosa Espina P. Aproximación a la medición del bienestar social (Approach to the measurement of social welfare). University of Valladolid; 1996.