



A Functional approach for constructing dynamic Composite Indicators

Annalina Sarra¹ · Eugenia Nissi² · Adelia Evangelista¹ · Tonio Di Battista¹ 

Accepted: 30 September 2023
© The Author(s) 2023

Abstract

This paper contributes to the research on the development of comparable composite indicators by introducing a Functional Weighted Malmquist Productive Index that allows for comparative trend analysis. In analogy with entropy-based weighted methods, this novel dynamic indicator is derived by measuring the degree of diversification of the single method through a family of diversity indices. The paper has the merit of proposing a new dynamic composite indicator that supplements the analysis with Functional Data Analysis (FDA) tools that provide us with useful information about the order and dynamics of the composite index trajectories. The simulation study set up in this paper raises doubts about the robustness of the entropy-based weighted methods while the application of the new index to well-being dataset highlights its practical appeal.

Keywords Composite Indicators · Benefit of Doubt weighting · Time-varying efficiency · Novel dynamic index

✉ Tonio Di Battista
tonio.dibattista@unich.it

Annalina Sarra
annalina.sarra@unich.it

Eugenia Nissi
eugenia.nissi@unich.it

Adelia Evangelista
adelia.evangelista@unich.it

¹ Department of Philosophical, Pedagogical and Economic-Quantitative Sciences, University "G. d'Annunzio", Viale Pindaro, 42, Pescara 65120, Italy

² Department of Economics, University "G. d'Annunzio", Viale Pindaro, 42, Pescara 65120, Italy

1 Introduction

In recent decades, with the rapid growth in available information, there has been a parallel growth of building synthetic indicators. The deep complexity of real world makes difficult to measure and evaluate the relevant aspects of the most phenomena, like well-being, human development, environmental sustainability, industrial competitiveness, and hard to capture them by using a single perspective. The scientific community has echoed this interest; accordingly, many scholars have focused their efforts on the development and improvement of a methodology known as Composite Indicator (CI). Constructing a CI (or index) is a popular approach to achieve a simplified numerical representation of a complex phenomenon. A CI is defined as a mathematical combination of individual indicators representing different dimensions into a single index, based on an underlying model of the multidimensional concept that is being measured (Saisana et al. 2005). Greco et al. (2019) make a huge effort in providing a wide collection of numerous publications on techniques and applications related to this field of research. Albeit CIs have gained much attention, they remain the subject of controversy. Their construction involves a series of advantages and disadvantages, some of them are mentioned below. In particular, as pointed out in Smith (2002), arguments for developing CIs include, among others, the possibility to present complex or multidimensional issues in one aggregate value, much easier to interpret than trying to find a trend in many separate indicators. Also, they enable to place performance at the centre of the policy arena. In fact, CIs are being increasingly recognised as a useful tool for policy analysis, benchmarking comparison, performance monitoring, public communication and decisions in various fields, including economy, environment, technology development and society, by many national and international organizations (Badea et al. 2011; Costa 2015; Filippetti and Peyrache 2011). Points against CIs construction, on the other hand, emphasise how they can send misleading or ineffective policy messages if they are poorly constructed or misinterpreted. Opponents specifically question the credibility of CIs, stressing the subjectivity that surrounds the various steps involved in their construction. According to the recommendations of the OECD document (OECD 2008), construction decisions range from the selection of a set of sub-indicators to the standardization method, weighting, and choice of the aggregation function.

The subject of this paper is an essential aspect of index aggregation, specifically the weighting of indicators, which has been subject to considerable scrutiny in recent scholarly works (Becker et al. 2017; Greco et al. 2019; Keogh et al. 2021). The methodological proposal put forward in this study is based on a data-oriented weighted method, known as Data Envelopment Analysis (DEA) (Charnes et al. 1978). When exact knowledge of the weights is not available, the DEA approach facilitates the aggregation of a number of quantitative sub-indicators. We address the issue of unit performance change over time within this framework by proposing a Functional Weighted Malmquist Productivity Index (FWMPI) that allows for comparative trend analysis. The novel aspect of

the proposal is that it achieves units ranking by supplementing the analysis with Functional Data Analysis (FDA) tools. There are several practical reasons for considering functional data. Ramsay and Silverman (2005) described important characteristics of FDA and used a strong argument for that approach. In recent years, FDA methods have been used in a variety of fields, including medicine, economics, meteorology, and many others. In contrast, the use of a functional approach for constructing CIs is relatively new (see, for example, Fortuna et al. (2022)). In any case, the implementation of a functional approach in the construction of CIs offers various benefits. Firstly, FDA allows for a more flexible and nuanced representation of data by treating them as continuous functions, which can capture important features missed by other methods. For example, FDA enables the observation of indicator behavior over time, providing insights into its evolution. Secondly, the functional approach is particularly useful when data are not sampled at equally spaced time points, as it can handle irregular intervals. This is important in constructing CIs that are updated at different intervals depending on data availability. Thirdly, FDA allows for the introduction of new analytical tools that can complement the original data with valuable information. It is possible to incorporate external covariates or predictors into the functional model, which can help to explain the variation in the CI. This can lead to more accurate and robust CIs, as well as a better understanding of the relationships between different variables. These advantages are essential for guiding policy decisions and measuring progress towards development goals.

The rest of the paper is structured as follows. Section 2 describes the methodological context, with a focus on the proposed methodology. Section 3 presents the results of the simulated data analysis, while in Sect. 4 we summarise the application of FWMPI to well-being data. Some concluding remarks are given in Sect. 5.

2 Methodological background

In this section, we provide a detailed overview of the methodological approaches employed in our study. Specifically, we begin by examining the use of DEA and Benefit of the Doubt (BoD) weighting to construct a CI. Next, we delve into the use of an entropy-based Malmquist Productive Index to measure changes in productivity over time for the units of analysis. In the third subsection, we introduce the FWMPI as a new method for assessing productivity changes over time.

2.1 Data envelopment analysis and “Benefit of the Doubt”-weighting

When constructing CIs, there are several approaches to aggregating individual indicators. Dimensionality reduction methods, such as Principal Component Analysis (PCA) or Factor Analysis (FA), are commonly used to reduce the number of indicators to a smaller set of uncorrelated factors. These multivariate statistical techniques account for the highest variation in the dataset, replacing the original variables with the smallest possible number of factors that reflect the

underlying structure of the data. However, these methods do not explicitly take into account the relative importance of each indicator in the final index and may result in a loss of information. An alternative approach is DEA, a non-parametric mathematical programming method that assesses the relative efficiency of homogeneous Decision Making Units (DMUs) in combining inputs to produce outputs. The first DEA model was proposed by Charnes et al. (1978) in management science to measure the production efficiencies of DMUs in combining inputs to produce outputs of goods or services. Since the mentioned pioneering work, various methods and models for ranking the originally efficient units have been put forward over the years. Cook and Seiford (2009) developed a classification of DEA models.

DEA models are based on the fundamental concept of evaluating the relative efficiency of homogeneous DMUs, such as companies, banks, countries, universities, and so on. The definition of DMUs has been deliberately kept broad to enable the use of DEA in a variety of applications. DEA is concerned with calculating an efficiency score ranging from 0 to 1. DEA assigns an efficiency score less than 1 to "inefficient" units while DMUs with a score equal to 1 are deemed "efficient". In general, among a group of DMUs, a unit with higher outputs but lower inputs has a better chance of achieving a high efficiency rank.

Due to its numerous advantages, the DEA non-parametric technique has recently been adopted as an appropriate method for CIs construction. The application of DEA to the field of CIs has been dubbed the Benefit of Doubt (BoD), which was first proposed by Melyn and Moesen (1991) and has since been used by Cherchye et al. (2007, 2008); Mahlberg and Obersteiner (2001), OECD (OECD 2008), Sahoo et al. (2017); Staessens et al. (2019), and Färe et al. (2019). The main advantage of using BoD in the construction of CIs is its flexibility in retrieving weights from the data itself, allowing for the aggregation of quantitative sub-indicators when exact weight knowledge is not available. The BoD model assumes that a unit good relative performance in a particular aspect or dimension of performance indicates that this unit considers that dimension to be relatively important when estimating these weights. Similarly, poor performance indicates that a unit places less importance on that dimension. As a result, BoD generates a set of endogenous weights that are most advantageous for every unit: each DMU has been placed in its most advantageous position. Thus, if a unit underperform in comparison to others, it cannot be attributed to the unfair weighting scheme since any other set of weights would have worsened the evaluated unit ranking position. The BoD method can be seen as a tool to combine performance sub-indicators without making explicit reference to the input(s).

In order to clearly convey the underlying idea, the BoD formulation can be presented in a step-wise fashion. Let us denote with CI_j the composite index for unit j , y_{ji} the value (possibly normalised) for the unit j on indicator i ($i = 1, \dots, m$) and w_i the weight assigned to indicator i . In *Step 1*, the composite index score of a unit j is calculated as the ratio of the weighted sum of its sub-indicators to the sum of the benchmark sub-indicators y_i^B :

$$CI_j = \frac{\sum_{i=1}^m w_{ji}y_{ji}}{\sum_{i=1}^m w_{ji}y_i^B} \tag{1}$$

Step 2 involves identifying the benchmark performance, which is determined endogenously:

$$CI_j = \frac{\sum_{i=1}^m w_{ji}y_{ji}}{\max_{y_{zi} \in \{studied\ units\}} \sum_{i=1}^m w_{ji}y_{zi}} \tag{2}$$

Formally speaking, the presence of the max operator and its associated argument in the denominator of Eq. 2 indicates that the benchmark observation is derived from an optimization problem.

Step 3 specifies appropriate weights. Because the weights are endogenously selected in such a way that they can be inferred from looking at the relative strengths and weaknesses of each unit, this stage implies the BoD concept. Taking a normalisation constraint into account, emphasising that the most favorable weights are always applied to all observations, and ensuring that weights are not-negative, we have:

$$\begin{aligned}
 CI_j = \max_{w_{ji}} & \frac{\sum_{i=1}^m w_{ji}y_{ji}}{\max_{y_{zi} \in \{studied\ units\}} \sum_{i=1}^m w_{ji}y_{zi}} \\
 \text{s.t.} & \sum_{i=1}^m w_{ji}y_{zi} \leq 1 \\
 & w_{ji} \geq 0.
 \end{aligned} \tag{3}$$

Formally, the BoD model in Eq. 3 is equivalent to the Charnes, Cooper, and Rhodes input-oriented constant-returns-to-scale (CRS) model (Charnes et al. 1978), with all indicators treated as outputs and a “dummy input” equal to one for all units. From a theoretically point of view, the departure to consistently derive aggregate CI using the BoD model is Koopmans’ theorem (Koopmans 1951) and revenue corollaries, as shown in Färe and Grosskopf (2004).

2.2 Entropy-based Malmquist productivity index

DEA literature offers the possibility to determine whether different DMUs are grown up, are regressed or remain unchanged in terms of their performance over time. The first contribution to measure the productivity change of a DMU over time is the Malmquist Productivity Index (MPI). The original so-called Malmquist index is a quantity index, introduced by Malmquist (1953) for analysing the consumption of inputs in a consumer theory context. Later, Färe et al. (1994) constructed a MPI index directly from input and output data using DEA. Färe et al. (1994) created a DEA-based MPI by combining Farrell’s efficiency measurement (Farrell 1957) with Caves’ productivity measurement (Caves et al.

1982), and decomposing it into two components to analyse productivity change due to technical efficiency and change due to technology. The DEA-based MPI relies on first constructing an efficiency frontier over the whole sample realised by DEA and then computing the distance of individual observations from the frontier.

Therefore, MPI is constructed by utilising distance functions that can serve as both input and output. More in detail, an input distance describes the production technology by observing the comparative reduction of the input vector, given an output vector (Coelli and Prasada Rao 2005); conversely, the output function distance mirrors a maximum proportional expansion of the output vector, given an input vector.

MPI can be formalised using either an output-oriented or an input-oriented method.

The output-oriented MPI is the focus of this research. Let us suppose to have a set of n DMUs, each with s unitary inputs denoted by a vector \mathbf{x}_j and m outputs denoted by a vector \mathbf{y}_j , for $j = 1, \dots, n$, over the periods t and $t + 1$. According to Färe et al. (1994), the DEA-MPI output for a given DMU_0 at time $t + 1$ and t can be expressed mathematically as:

$$MPI_0 = \left[\frac{D_0^t(x_0^{t+1}, y_0^{t+1}) D_0^{t+1}(x_0^t, y_0^t)}{D_0^t(x_0^t, y_0^t) D_0^{t+1}(x_0^{t+1}, y_0^{t+1})} \right]^{\frac{1}{2}} \quad (4)$$

where (x_0^{t+1}, y_0^{t+1}) and (x_0^t, y_0^t) represent the input and output vector of the period $t + 1$ and t respectively, and the notation $D_0^t(x_0^t, y_0^t)$ represents the distance from the period t_0 to period t_1 technology. The index in Eq. 4 is the geometric mean of two output-based Malmquist indices. A magnitude of MPI greater than 1 indicates progress (positive growth) from period t_0 to period t_1 , whereas magnitudes equal to 1 and less than 1 indicate the status quo and productivity decay, respectively. The overall tendency in DMU productivity changes over time periods is traditionally obtained by taking the average of sequential productivity indices, implicitly assuming that all sectional indices are equally affecting the productivity level.

In his work, Fallahnejad (2017) suggested the use of Shannon's entropy to derive more objective weights for aggregating MPIs, thereby eliminating the problem of equal weighting in the process.

Notably, the Shannon's entropy technique is one of the most important methods for determining the relative weights of indicators in multi-criteria decision making contexts (Peykani et al. 2022). In information theory, the entropy weight method primarily uses the magnitude of the entropy value to measure the indicator weight contained in known data. The lower the entropy value, the greater the degree of differentiation and the more information that can be derived, and a higher weight should be given to the indicator on the target in the overall evaluation.

In the Fallahnejad's approach the following basic steps can be enucleated:

Step 1, first the MPI matrix (Table 1) is created taking into account the n DMUs and MPI measures over times.

Table 1 MPI matrix

	MPI1	MPI2	MPIk
DMU_1	MPI_{11}	MPI_{12}	MPI_{1k}
DMU_2	MPI_{21}	MPI_{22}	MPI_{2k}
DMU_3	MPI_{31}	MPI_{32}	MPI_{3k}
...
...
DMU_n	MPI_{n1}	MPI_{n2}	MPI_{nk}

Step 2, the MPI matrix is normalised by dividing the value of each column by the sum of its column. Thus, the normalised value of the $i - th$ MPI in the $t - th$ sample is denoted by p_{jt} and its calculation is as follows:

$$p_{jt} = \frac{MPI_{jt}}{\sum_{j=1}^n MPI_{jt}}, \quad \forall i, t. \tag{5}$$

The above normalisation allows to eliminate anomalies due to different measurement units and scales.

In *Step 3*, the entropy h_t for all normalised MPIs is calculated as:

$$h_t = -h_0 \sum_{j=1}^n p_{jt} \ln p_{jt}, \tag{6}$$

where $h_0 = \frac{1}{\ln(n)}$. For the convenience of calculation, $p_{jt} = 0$ is generally set when $p_{jt} \ln p_{jt} = 0$.

Step 4 involves the computation of the degree of diversification, defined as:

$$d_t = 1 - h_t, \quad t = 1, \dots, k. \tag{7}$$

As previously stated, the degree of diversification indicates the amount of useful information provided by the relevant MPI measures to the overall aggregated index. It follows that if the DMU productivity values are close, the weight of a given year can be considered weak in the aggregating process.

In *Step 5* the degree of importance of MPI at time t is obtained by setting

$$w_t = \frac{d_t}{\sum_{s=1}^k d_s}, \quad t = 1, \dots, k, \tag{8}$$

where $\sum_{t=1}^k w_t = 1$.

In *Step (6)*, the weighted MPI is calculated as:

$$WMPI_j = \sum_{t=1}^k w_t MPI_{jt}, \quad j = 1, \dots, n. \tag{9}$$

It is apparent that a model which generates scores that are nearly identical for all units would have little impact on the ranking and should, therefore, be deemed relatively unimportant.

2.3 The proposed approach: FWMPI

Although the traditional literature has demonstrated that the results of entropy-weighted methods are reliable and effective, some studies, primarily based on engineering practice, have raised concerns about their rationality in decision making. For example, Zhu et al. (2020) point out some flaws in the entropy-weighted method that result in distorted decision-making outcomes. The authors pointed out that when there are too many zeros in the measured values, the standardised results of the entropy-weighted method are prone to distortion. Therefore, the index with the lowest actual differentiation degree will be given too much weight. Secondly, in multi-index decision-making involving categorisation, the classification degree can accurately reflect the information content of the index. The entropy-weighted methods, on the other hand, only consider the index degree of numerical differentiation and ignore rank discrimination.

In this regard, we hold the view that entropy-weighted methods provide only a limited viewpoint in the context of the multi-index decision-making problem under consideration.

As shown in Sect. 2.2, in the construction of the weighted MPI, the Shannon's entropy can be regarded as a diversity measure of DMUs. However, different indices could have been used to delineate such diversity, and different orderings of homogeneous DMUs could have been obtained depending on the diversity measures used.

To address this limitation, we propose a set of diversity indices based on a single continuous variable that graphically depict a diversity profile. Specifically, we refer to the diversity index of degree β , that is Δ_β , proposed by Patil and Taillie (1979, 1982) to quantify the diversity of an ecological population, composed by N units, partitioned into S species ($i = 1, 2, \dots, s$) and formally expressed as:

$$\Delta_\beta = \sum_{i=1}^s \frac{(1 - p_i^\beta)}{\beta} p_i, \quad \beta \geq -1. \quad (10)$$

In Eq. 10 the index captures the multidimensional aspect of diversity and can be considered a function of β with parameter p , where $p_i = \frac{N_i}{\sum_{i=1}^s N_i}$ and N_i represents the number of units belonging to the i -th species.

From a mathematical point of view, that index makes sense for any real number β , i.e. $-\infty < \beta < +\infty$; however, to guarantee that Δ_β has some desirable properties it is necessary to impose the restriction that $\Delta_\beta \geq -1$ (Patil and Taillie 1979, 1982). In addition, as stressed by the same authors, it may not be useful to calculate and plot Δ_β for $\beta \geq 1$ as these profiles tend to converge rapidly beyond this point (Patil and Taillie 1979, 1982). Therefore, we examine Δ_β as a function of β within the specified range of $[-1, 1]$. By plotting Δ_β versus β we graphically obtain a diversity profile. The profiles of the Δ_β family are decreasing and

convex curves. It is easy to verify that some of the most frequently used indices of diversity are special cases of the one-parameter family Δ_β of diversity indices. For each value of β we can obtain a diversity measure. To be specific, Shannon's entropy is a particular case of β diversity profile, resulting when $\beta \rightarrow 0$, whereas the Richness and Simpson indices are got posing $\beta = -1$ and $\beta = 1$, respectively.

Following Di Battista et al. (2017), the one parameter family Δ_β of diversity indices can be viewed as a function in a fixed domain rather than a sequence of observations. More in detail, the authors propose an alternative way to explain the diversity profile through the Functional Data Analysis (FDA). As pointed out in Ramsay and Silverman (2005), the underlying idea of FDA is to assume the existence of some functions, giving rise to the observed data. Analytic tools, such as the analysis of β profile derivatives graph, radius of curvature and length of a curve, which are essential components of FDA, can be exploited to further inspect the diversity profiles and rank the different units (Di Battista et al. 2017).

The framework described above was used to create a dynamic index, namely the FWMPI, that addresses the shortcomings of entropy-weighted methods. The FWMPI can be obtained by rewriting some steps of Fallahnejad's original procedure. To be more specific, we propose measuring the degree of differentiation using the one-parameter family Δ_β of diversity indices rather than the entropy value h_t , with the ultimate goal of accounting for all possible diversity measures instead of just one represented by Shannon's index. Accordingly, Steps 3–6 in the multi-index decision-making problem will be reformulated as follows. The degree of diversification is calculated as:

$$d_t = 1 - \Delta_\beta, \tag{11}$$

while the degree of importance of MPI at time t is obtained by setting

$$w_t(\beta) = \frac{d_t(\beta)}{\sum_{s=1}^k d_s(\beta)}, \quad \forall \beta. \tag{12}$$

Finally, the functional weighted MPI is calculated according to Eq. 13:

$$FWMPI_j(\beta) = \sum_{t=1}^k w_t(\beta)MPI_{jt}, \quad \forall \beta. \tag{13}$$

It is critical to note that the functional tools can now be used to rank DMUs. In this regard, we pursue two alternatives.

First, the comparative analysis of the FWMPI curves is facilitated by considering the area under the curve (Di Battista et al. 2017) which provides a ranking mirroring both the level and the evolutionary dynamics of functions.

Given a set of n FWMPI functions, $FWMPI_1(\beta)$, $FWMPI_2(\beta)$... $FWMPI_n(\beta)$, the DMUs can be sorted in descending order according to the area under the curve, defined as:

Table 2 PANEL A-Simulation results under level of contamination $\epsilon=0$

	WPMI Shannon vs. Simpson	WMPI Shannon vs. Richness	WPMI Shannon vs. Simpson	WMPI Shannon vs. Rich- ness	
Size = 30			Size = 60		
ASR	1.47	1.67	ASR	2.07	2.14
SR max	5	6	SR max	8	8
SR > 5	0	3.33	SR > 5	5	6.67
SR > 10	0	0	SR > 10	0	0
Size=100			Size=150		
ASR	2.10	2.16	ASR	0.34	0.30
SR max	14	14	SR max	3	3
SR > 5	10	10	SR > 5	0	0
SR > 10	1	1	SR > 10	0	0
Size=200					
ASR	11.81	12.35			
SR max	58	58			
SR > 5	60	62			
SR > 10	39.50	42			

Table 3 PANEL A-Simulation results under level of contamination $\epsilon=0.025$

	WPMI Shannon vs. Simpson	WMPI Shannon vs. Richness	WPMI Shannon vs. Simpson	WMPI Shannon vs. Rich- ness	
Size = 30			Size = 60		
ASR	3.06	3.40	ASR	2.5	2.60
SR max	14	15	SR max	8	9
SR > 5	13.33	16.66	SR > 5	8.33	10
SR > 10	6.66	6.66	SR > 10	0	0
Size=100			Size=150		
ASR	2.44	2.58	ASR	4.29	4.41
SR max	15	15	SR max	21	24
SR > 5	9	10	SR > 5	31.33	32
SR > 10	2	2	SR > 10	6.66	7.33
Size=200					
ASR	9.01	9.57			
SR max	39	46			
SR > 5	56.50	59			
SR > 10	32.50	35			

Table 4 PANEL A-Simulation results under level of contamination $\epsilon = 0.05$

	WPMI Shannon vs. Simpson	WMPI Shannon vs. Richness	WPMI Shannon vs. Simpson	WMPI Shannon vs. Rich- ness
Size = 30			Size = 60	
ASR	2	2	ASR	0.16
SR max	8	8	SR max	2
SR > 5	3.33	3.33	SR > 5	0
SR > 10	0	0	SR > 10	0
Size=100			Size=150	
ASR	2	2.16	ASR	5.62
SR max	15	15	SR max	21
SR > 5	9	11	SR > 5	40
SR > 10	1	1	SR > 10	16.66
Size=200				
ASR	5.62	5.93		
SR max	26	27		
SR > 5	39.50	42.50		
SR > 10	16.50	18.50		

Table 5 PANEL B-Simulation results under level of correlation=0

	WPMI Shannon vs. Simpson	WMPI Shannon vs. Richness	WPMI Shannon vs. Simpson	WMPI Shannon vs. Rich- ness
Size = 30			Size = 60	
ASR	1.86	2.13	ASR	0.33
SR max	5	5	SR max	3
SR > 5	0	0	SR > 5	0
SR > 10	0	0	SR > 10	0
Size = 100			Size=150	
ASR	2.34	2.46	ASR	2.5
SR max	12	12	SR max	10
SR > 5	11	14	SR > 5	12
SR > 10	1	1	SR > 10	0
Size = 200				
ASR	2.08	2.18		
SR max	11	11		
SR > 5	9.50	10.50		
SR > 10	1	1.50		

Table 6 PANEL B-Simulation results under level of correlation=0.25

	WPMI Shannon vs. Simpson	WMPI Shannon vs. Richness	WPMI Shannon vs. Simpson	WMPI Shannon vs. Rich- ness
Size = 30			Size = 60	
ASR	2.26	2.73	ASR	4.16
SR _{max}	8	11	SR _{max}	11
SR > 5	6.66	13.33	SR > 5	33.33
SR > 10	0	0.33	SR > 10	3.33
Size = 100			Size=150	
ASR	2.62	2.80	ASR	3
SR _{max}	10	12	SR _{max}	12
SR > 5	14	17	SR > 5	17.33
SR > 10	0	1	SR > 10	2
Size = 200				
ASR	2.80	2.90		
SR _{max}	10	11		
SR > 5	13.50	16		
SR > 10	0	0.50		

Table 7 PANEL B-Simulation results under level of correlation=0.50

	WPMI Shannon vs. Simpson	WMPI Shannon vs. Richness	WPMI Shannon vs. Simpson	WMPI Shannon vs. Rich- ness
Size = 30			Size = 60	
ASR	2.66	2.86	ASR	2.06
SR _{max}	9	10	SR _{max}	8
SR > 5	16.66	20	SR > 5	6.66
SR > 10	0	0	SR > 10	0
Size = 100			Size=150	
ASR	2.34	2.70	ASR	2.56
SR _{max}	9	11	SR _{max}	12
SR > 5	12	13	SR > 5	10
SR > 10	0	1	SR > 10	0.66
Size = 200				
ASR	9.40	10.41		
SR _{max}	37	43		
SR > 5	61.5	66.5		
SR > 10	36.5	41		

Table 8 PANEL B-Simulation results under level of correlation=0.85

	WPMI Shannon vs. Simpson	WMPI Shannon vs. Richness	WPMI Shannon vs. Simpson	WMPI Shannon vs. Rich- ness	
Size = 30			Size = 60		
ASR	1.60	1.83	ASR	4	4.83
SR max	6	7	SR max	12	14
SR > 5	6.66	10	SR > 5	30	38.83
SR > 10	0	0	SR > 10	6.66	10
Size = 100			Size=150		
ASR	2.72	3.36	ASR	5.79	12.88
SR max	12	15	SR max	25	50
SR > 5	18	24	SR > 5	38	66.66
SR > 10	1	4	SR > 10	18	47.33
Size = 200					
ASR	2.14	2.94			
SR max	9	14			
SR > 5	9	16			
SR > 10	0	2			

$$A_i = \int_B FWMPI_i(\beta)d\beta, \quad i = 1, \dots, n. \tag{14}$$

Furthermore, by utilising the FDA approach, we can compute the functional depth rank. When dealing with a group of functions, the notion of depth for functional data enables us to establish the centrality of a function and generates an ordering of the sample curves from the center outwards. Functional depth is used to rank functional observations from most unusual to most common. The underlying concept is to figure out “how long” a curve stays in the middle of a group of them. Research on data depth has gained significant attention over the years, as evidenced by Cuevas et al. (2007), who offer a comprehensive review of data depth for high-dimensional or functional data.

In this paper, we employ the functional integrated depth introduced by Fraiman and Muniz (2001) to compute the integration of a univariate depth along the β axis. Let $F_{n,\beta}$ be the empirical distribution of sample $x_1(\beta), \dots, x_n(\beta)$, the functional integrated depth can be defined as follows:

$$I_i = \int_B D_i(\beta)d\beta, \quad \forall \beta \in B, \quad i = 1, \dots, n, \tag{15}$$

where $D_i(\beta) = 1 - |0.5 - F_{n,\beta}(x_i(\beta))|$. According to the values I_i it is possible to rank the sample curves in descending order, from the most central to the most outlying.

Table 9 PANEL C-Simulation results from a WEIBULL with shape=2 and scale=1

	WPMI Shannon vs. Simpson	WMPI Shannon vs. Richness	WPMI Shannon vs. Simpson	WMPI Shannon vs. Rich- ness
Size = 30			Size = 60	
ASR	0.06	0.13	ASR	3.60
SR max	1	1	SR max	14
SR > 5	0	0	SR > 5	20
SR > 10	0	0	SR > 10	8.33
Size = 100			Size=150	
ASR	3.02	3.98	ASR	2.70
SR max	11	14	SR max	10
SR > 5	20	27	SR > 5	9.33
SR > 10	2	10	SR > 10	0
Size = 200				
ASR	9.86	11.83		
SR max	39	49		
SR > 5	59.50	65.50		
SR > 10	40.50	46.50		

Table 10 PANEL C-Simulation results from a WEIBULL with shape = 3 and scale = 2

	WPMI Shannon vs. Simpson	WMPI Shannon vs. Richness	WPMI Shannon vs. Simpson	WMPI Shannon vs. Rich- ness
Size = 30			Size = 60	
ASR	0.06	0.06	ASR	4.33
SR max	1	1	SR max	20
SR > 5	0	0	SR > 5	30
SR > 10	0	0	SR > 10	8.33
Size = 100			Size=150	
ASR	3.16	3.36	ASR	2.74
SR max	11	12	SR max	12
SR > 5	19	21	SR > 5	11.33
SR > 10	2	4	SR > 10	0.66
Size = 200				
ASR	10.97	11.81		
SR max	43	45		
SR > 5	57.5	61		
SR > 10	42.50	44.50		

Table 11 PANEL C-Simulation results from a WEIBULL with shape = 4 and scale = 3

	WPMI Shannon vs. Simpson	WMPI Shannon vs. Richness	WPMI Shannon vs. Simpson	WMPI Shannon vs. Rich- ness
Size = 30			Size = 60	
ASR	0	0	ASR	4.80
SR max	0	0	SR max	18
SR > 5	0	0	SR > 5	36.66
SR > 10	0	0	SR > 10	10
Size = 100			Size=150	
ASR	3.04	3.14	ASR	2.48
SR max	10	10	SR max	8
SR > 5	20	21	SR > 5	9.33
SR > 10	0	0	SR > 10	0
Size = 200				
ASR	11.43	11.86		
SR max	53	55		
SR > 5	62.50	63		
SR > 10	41.50	43		

3 Simulation study

It is widely acknowledged that for CIs to be effective, they must be developed using robust methods that ensure both benchmarking and stability over time. In this section, we provide evidence to justify the credibility of our proposal by demonstrating how various orderings of homogeneous DMUs can arise due to the use of different diversity measures.

We set up a simulation experiment in the R environment (R Core Team 2023) to gather information on the ranking robustness when a dynamic weighted MPI is constructed using Shannon, Simpson, and Richness indices. To track the shift in ranking, we used the absolute value of the average shift in ranking (ARS) as a general measure of divergence. The ARS obeys the following formula:

$$ARS = \frac{1}{n} \sum_{i=1}^n |Rank_i(CIs_1) - Rank_i(CIs_2)|. \tag{16}$$

Here, n represents the number of DMUs, while $Rank_i(CIs_1)$ and $Rank_i(CIs_2)$ refer to the position of the i -th DMU, as determined by the CIs_1 obtained through a specific diversity index and the CIs_2 obtained through another diversity index, respectively. The greater the divergence measure, the greater the sensitivity of CI to changes in the way indicators are weighted and aggregated; otherwise, the greater the robustness.

Table 12 PANEL A ($\epsilon = 0$ vs. $\epsilon = 0.025$)

	FWPMI	WMPI Shannon	WPMI Richness	WMPI Simpson
Size = 30				
ASR	9	11.07	10.87	10.67
SR max	28	28	29	27
SR > 5	63.33	60	60	63.33
SR > 10	33.33	40	43.33	43.33
Size = 60				
ASR	19.27	22.37	19.77	20
SR max	48	55	55	56
SR > 5	78.33	83.33	83.33	76.67
SR > 10	70	76.67	68.33	66.67
Size = 100				
ASR	34.80	32.22	36.38	32.92
SR max	91	94	87	91
SR > 5	94	89	89	90
SR > 10	82	83	80	83
Size = 150				
ASR	46.89	49.01	52.23	52.97
SR max	138	139	141	141
SR > 5	94	94	94.67	92
SR > 10	88.67	89.33	89.33	88.67
Size = 200				
ASR	64.10	69.65	67.45	68.13
SR max	193	190	195	192
SR > 5	95	94	94	93
SR > 10	89.50	89.50	88.50	88.50

In addition, we have included some specific divergence measures in our study. This decision is based on the observation that metrics which indicate a particular outcome on average can sometimes be contradicted by an analysis of the individual DMUs average shift in ranking. Therefore, our simulation study also incorporates the maximum absolute shift in ranking (SR max), the proportion of absolute shifts in ranking that exceed five positions (SSR > 5), and the proportion of absolute shifts in ranking that exceed ten positions (SSR > 10).

In order to assess the ability of the dynamic CIs constructed using diversity indices to handle potential challenges, we created various scenarios. Specifically, we aimed to determine the degree to which the ranking of units is affected by the presence of outliers, collinearity and skewness between indicators, and variations in sample size. We perform a dataset where four single indicators (I_1, I_2, I_3, I_4) are analysed over three different time periods.

The initial scenario was created to examine how extreme values affect the ranking of CIs constructed using different diversity measures based indices. In

Table 13 PANEL A ($\epsilon = 0$ vs. $\epsilon = 0.05$)

	FWPMI	WMPI Shannon	WPMI Richness	WMPI Simpson
Size = 30				
ASR	9.40	8.87	11	11
SR max	28	21	28	28
SR > 5	63.33	70	73.33	73.33
SR > 10	33.33	36.67	43.33	43.33
Size = 60				
ASR	20.24	19.40	19.47	24.87
SR max	56	55	52	56
SR > 5	86.67	88.33	706	96.67
SR > 10	75	68.33	65	85
Size = 100				
ASR	31.04	37.14	35.66	31.78
SR max	78	93	95	85
SR > 5	88	93	92	88
SR > 10	75	82	85	79
Size = 150				
ASR	51.85	47.63	48.33	51.39
SR max	141	129	139	131
SR > 5	89.30	94	92	96.67
SR > 10	85.30	91.33	85.33	91.33
Size = 200				
ASR	64.88	70.51	68.87	65.99
SR max	173	185	196	192
SR > 5	95.50	96.50	95.50	92.50
SR > 10	90.50	93	92	87

this scenario, we generated indicators from a Uniform distribution within the range of [5,15], without any collinearity present between them. Next, we introduced extreme values into the simulated dataset for each indicator, drawing from a Uniform distribution in accordance with the following scheme:

$$\begin{aligned}
 I_1c &\sim U(k_1max(I_1); k_2max(I_1)), \\
 I_2c &\sim U(k_1max(I_2); k_2max(I_2)), \\
 I_3c &\sim U(l_1min(I_3); b) \\
 I_4c &\sim U(l_1min(I_4); l_2min(I_4)).
 \end{aligned}$$

It is worth noting that the parameters k_1 and k_2 contribute to the outlying nature of extreme values on the right tail of the distribution, while l_1 and l_2 are responsible for the left tail. For this study, we have set $k_1 = 2$, $k_2 = 3$, $l_1 = 0.2$, and $l_2 = 0.3$. Additionally, we have increased the outlyingness of extreme values over time by raising k_1 and k_2 to 4 and 6 for the second time period, and to 6 and 10 for the third time period, respectively. To randomly select which observations would be contaminated, we have defined a contamination level of ϵ . Specifically,

Table 14 PANEL B (correlation = 0 vs. correlation = 0.25)

	FWPMI	WMPI Shannon	WPMI Richness	WMPI Simpson
Size = 30				
ASR	11.13	11.67	11.20	11.27
SR max	27	22	21	21
SR > 5	70	86.67	80	80
SR > 10	63.33	56.67	63.33	63.33
Size = 60				
ASR	18.83	20.17	20.40	20.57
SR max	56	58	58	44
SR > 5	73.33	81.67	80	86.67
SR > 10	58.33	61.67	68.33	78.33
Size = 100				
ASR	37.02	33.94	33	34.16
SR max	85	84	86	95
SR > 5	86	92	91	87
SR > 10	81	82	77	78
Size = 150				
ASR	49.44	51.99	51.51	47.81
SR max	128	148	149	129
SR > 5	93.33	95.33	93.33	89.33
SR > 10	86.66	89.33	84.67	87.33
Size = 200				
ASR	68.28	67.67	67.83	68.03
SR max	192	186	178	185
SR > 5	96	92.50	96.50	95
SR > 10	88.50	90	93	90.50

we have analyzed three different situations: no contamination ($\epsilon = 0$), and contamination levels of 2.5% ($\epsilon = 0.025$) and 5% ($\epsilon = 0.05$), respectively.

In the second scenario, the interest was in generating correlated normal distributed indicators. To denote the strength of the linear relationship between two different indicators, we varied the collinearity between them, setting four values, namely 0, 0.25, 0.5 and 0.85.

The third scenario involved examining the changes in ranking position when dealing with skewed indicators. To achieve this, we generated values from a Weibull distribution, varying the shape and scale parameters.

Additionally, we compared the outcomes across different sample sizes for each data-generating mechanism, producing data with sample sizes of $n = 30$, $n = 60$, $n = 100$, $n = 150$, and $n = 200$.

Tables 2, 3, and 4 display the simulation outcomes for the scenario involving extreme values generated by the Uniform distribution as described earlier. These results are presented in Panel A.

Table 15 PANEL B (correlation = 0 vs. correlation = 0.50)

	FWPMI	WPMI Shannon	WPMI Richness	WPMI Simpson
Size = 30				
ASR	9.2	9.93	10.40	10.40
SR max	25	22	20	20
SR > 5	60	70	76.67	76.67
SR > 10	33.33	46.67	46.67	46.67
Size = 60				
ASR	20.16	20.37	20.3	21.47
SR max	56	52	50	55
SR > 5	88.33	76.67	80	88.33
SR > 10	76.66	66.67	66.67	80
Size = 100				
ASR	34.42	33.44	33.54	33.46
SR max	91	86	88	93
SR > 5	90	87	88	88
SR > 10	78	80	81	80
Size = 150				
ASR	53	46.51	45.81	51
SR max	138	131	132	134
SR > 5	92.66	88.67	94	92.67
SR > 10	85.33	86	84	86.67
Size = 200				
ASR	69.16	67.95	69.07	65.23
SR max	185	187	187	192
SR > 5	99	94.50	96.50	94
SR > 10	95	91	93.50	87.50

According to the average shift in ranking measures, there are differences in the rankings observed for both comparisons: Shannon vs. Simpson and Shannon vs. Richness. The largest values in the metrics utilized are observed when the sample size is increased and when we introduce a contamination level of 5% of extreme values into the simulated data.

Upon analyzing the sensitivity of the outcomes under the correlation scenario (displayed in Panel B of Tables 5, 6, 7, and 8), we observe that the ranking divergence is primarily identified through specific SR metrics that indicate a notable degree of volatility as the sample size and correlation between indicators increase.

In addition, we examine the level of stability of the rankings in cases where we are dealing with skewed indicators generated from a Weibull distribution. The outcomes, which are presented in Tables 9, 10, and 11, indicate that there is a considerable amount of variability in the ranking positions of each DMU, particularly when the indicators are drawn from a Weibull distribution with a slope parameter of 4 and

Table 16 PANEL B (correlation = 0 vs. correlation = 0.85)

	FWPMI	WMPI Shannon	WPMI Richness	WMPI Simpson
Size = 30				
ASR	10.67	11.47	11.13	11.13
SR max	25	23	22	22
SR > 5	73.33	80	70	70
SR > 10	46.67	53.33	56.67	56.67
Size = 60				
ASR	18.87	19.90	20.57	20.67
SR max	51	58	58	44
SR > 5	80	80	80	93.33
SR > 10	61.67	66.67	68.33	73.33
Size = 100				
ASR	34.08	33.70	33.50	33.58
SR max	94	85	85	94
SR > 5	90	91	95	87
SR > 10	79	82	84	80
Size = 150				
ASR	50.15	52.61	52.53	47.33
SR max	140	148	149	134
SR > 5	90.67	96.67	96	89.33
SR > 10	83.33	91.33	88.67	82.67
Size = 200				
ASR	66.81	67.41	66.87	68.25
SR max	184	184	178	183
SR > 5	94.50	94	95	96
SR > 10	89	90.50	90.50	88.50

a scale parameter of 3. This variability has a significant impact on the divergence measures, particularly when the sample size is increased to $n=200$.

Besides, in order to assess the effectiveness of the proposed methodology, we conduct a comparison between the performance of FWPMI and other methods that rely on a single diversity measure. This comparison is carried out across the various scenarios described earlier.

In particular, for the initial scenario, we compare the metrics used on rankings generated from uncontaminated artificial data with the metrics obtained when we introduced outliers into the data (Panel A: Tables 12, 13).

In the second scenario, for each index, we contrast the rankings resulting from uncorrelated data with those obtained when different levels of correlation are taken into consideration (Panel B: Tables 14, 15, 16).

Finally, we replicated the comparison of metrics for artificial data that exhibit different levels of indicator skewness (Panel C: Tables 17, 18).

Table 17 PANEL C (WEIBULL with shape = 2 and scale = 1 vs. WEIBULL with shape = 3 and scale = 2)

	FWPMI	WMPI Shannon	WPMI Richness	WMPI Simpson
Size = 30				
ASR	12	12.67	12.60	12.60
SR max	26	26	26	26
SR > 5	80	83.33	76.67	76.67
SR > 10	53.33	60	66.67	66.67
Size = 60				
ASR	21.23	20.57	23.03	23.03
SR max	54	58	49	49
SR > 5	80	81.67	91.67	91.67
SR > 10	71.67	66.67	76.67	76.67
Size = 100				
ASR	35.60	33.66	32.34	32.34
SR max	93	90	85	85
SR > 5	91	84	84	36.84
SR > 10	85	76	79	79
Size = 150				
ASR	47.88	50.72	47.85	47.84
SR max	146	141	138	138
SR > 5	93.33	94	90	90
SR > 10	88.67	86	82.67	82.67
Size = 200				
ASR	66.61	65.17	67.64	67.64
SR max	169	187	197	197
SR > 5	96.50	96	95.50	95.50
SR > 10	94	90	88.50	88.50

In the initial scenario, the results suggest that when outliers are present, the FWPMI methodology generally outperforms the alternative methods most of the time. Upon examining the outcomes of the second scenario, it becomes evident that the performance of the novel approach decreases as the sample size and level of correlation increase.

The findings from the last scenario imply that the proposed methodology demonstrates superior performance when dealing with smaller sample sizes and minimal skewness among the indicators.

4 Empirical evidence: application to BES data

In this section, we present the findings of our FWPMI proposal on the well-being dataset. In recent years, several scholars have expressed concerns about the limitations of GDP as a measure of well-being and as a benchmark for evaluating and

Table 18 PANEL C (WEIBULL with shape = 2 and scale = 1 vs. WEIBULL with shape = 4 and scale = 3)

	FWPMI	WMPI Shannon	WPMI Richness	WMPI Simpson
Size = 30				
ASR	11.53	12.80	12.80	12.80
SR max	26	26	26	26
SR > 5	73.33	83.33	80	80
SR > 10	46.67	60	63.33	63.33
Size = 60				
ASR	22.57	20.33	23	23
SR max	54	58	50	50
SR > 5	86.67	80	91.67	91.67
SR > 10	68.33	66.67	75	75
Size = 100				
ASR	34.18	33.70	32.30	32.30
SR max	87	91	86	86
SR > 5	88	82	87	87
SR > 10	81	77	78	78
Size = 150				
ASR	49.05	50.69	47.63	47.63
SR max	143	141	138	138
SR > 5	94	94	92	92
SR > 10	84.67	88.67	83.33	83.33
Size = 200				
ASR	67.60	64.98	67.83	67.83
SR max	186	187	198	197
SR > 5	95.50	95.50	94	94
SR > 10	89.50	91	90	90

comparing the development of regions and countries (see, among others, Larraz and Pavia (2010); Costanza et al. (2009); Fleurbaey (2009)).

The concept of well-being and its measurement have been significant topics of discussion in European research debates. Following the EU “beyond GDP” initiative in 2009, many projects have been suggested and implemented to combine indicators, datasets, domains, and dimensions. The Stiglitz Commission, established in France, made the most notable contribution to this effort (Stiglitz et al. 2009).

In response to the need for measurements of individual and societal well-being that surpass conventional measures, such as GDP, Italy developed the BES framework (short for Benessere Equo e Sostenibile, which means equitable and sustainable well-being). According to Riccardini and De Rosa (2016), the BES is “a measurement tool for progress in Italy” and is based on the theoretical model published by the OECD (Hall et al. 2010). The BES conceptualises well-being as a phenomenon



Fig. 1 Graphical representation of BESdT framework

comprised of two fundamental elements: equity within and between generations, and sustainability from an environmental, economic, and social perspective.

It is worth mentioning that the objective of the BES project is to create statistical indicators that are deemed important for a country progress, using a formative approach (Diamantopoulos et al. 2008). This type of approach assumes that the indicators define the underlying trait that represents the phenomenon, as opposed to the reflexive approach, which assumes that the indicators reflect the phenomenon itself (Diamantopoulos and Winklhofer 2001; Diamantopoulos and Siguaw 2006; Maggino 2017). As per Diamantopoulos' approach, the internal consistency of the formative indicators is of minimal relevance since two non-correlated indicators can both be significant for the same construct.

The BES indicators were chosen through a participatory process in which all sectors of society, including academics, institutions, associations, and citizens, expressed their preferences. The elementary indicators are clustered in 12 domains, namely: Health, Education and training, Work and work-life balance, Economic prosperity, Social relationships, Politics and institutions, Security, Subjective well-being, Landscape and Cultural heritage, Environment, Innovation, Research and Creativity, Quality of services (see BES (2013)).

Subsequently, the Italian National Institute of Statistics (ISTAT) has expanded the BES project by introducing other initiatives that focus more on the local level. The "Benessere Equo e Sostenibile dei Territori (BESdT)" is one such initiative, which applies the BES framework to a provincial (NUTS 3) scale.

Measuring well-being at the local level is highly significant in formulating policies aimed at achieving fair and sustainable development while taking into account

Table 19 Descriptive statistics of Indicators-year 2017

Indicator	Min	Max	Mean	Median	Q.5	Q.25	Q.75	Q.95	IQ.range	Std.dev
PeopleTerzEduc	12.0	41.2	24.9	24.5	17.0	21.0	28.6	34.6	7.7	5.6
NEET	11.3	44.9	22.9	20.5	12.4	16.5	27.8	39.2	11.4	8.5
UpperSecondEdu	43.4	75.7	60.4	61.3	46.4	55.6	65.8	70.8	10.2	7.4
PartLifeLongLearn	3.9	14.4	7.8	7.5	4.3	6.5	9.2	11.4	2.7	2.1
LandfillWaste	0.0	574.0	31.4	11.4	0.0	0.0	33.7	113.5	33.7	68.0
SepCollecWaste	11.3	87.8	55.2	58.9	23.0	44.2	68.5	77.8	24.3	17.4
LifeExp	80.7	84.0	82.7	82.7	81.5	82.2	83.3	83.9	1.1	0.8
InfMortRate	0.0	6.4	2.5	2.2	0.8	1.7	3.4	4.9	1.7	1.2
RoadAccidMortRate	0.0	2.4	0.7	0.7	0.0	0.4	1.0	1.8	0.6	0.5
CancerMortRate	6.6	11.3	8.5	8.4	6.9	7.8	9.0	10.3	1.2	1.0
MortRateDementia	22.5	60.5	34.0	33.4	24.7	29.2	37.9	44.4	8.8	6.4
WomenMunCouc	21.7	39.9	32.0	32.1	24.9	28.9	35.6	38.8	6.7	4.4
MunCoucUnder40Ys	19.0	44.9	30.8	31.0	22.4	27.7	33.9	40.0	6.3	5.2
PrisonDensity	0.0	187.9	116.8	116.1	62.5	102.8	133.2	177.6	30.4	33.9
EarlyChilhoodServ	0.5	36.1	13.6	11.9	2.7	7.3	18.0	28.3	10.7	7.9
ExtraRegHospMigr	1.5	25.3	8.2	6.6	2.2	4.1	10.7	19.5	6.6	5.5
IrregElectriPower	0.5	5.7	2.1	1.7	0.9	1.2	2.6	4.5	1.4	1.2
SeatKmPubTransp	0.5	5.7	2.1	1.7	0.9	1.2	2.6	4.5	1.4	1.2
HomicideRate	0.0	4.3	0.6	0.5	0.0	0.2	0.7	1.3	0.5	0.6
ViolentCrime	8.5	31.5	15.0	14.3	9.5	12.5	17.0	22.2	4.5	4.1
RoadMort	1.0	15.5	4.9	4.5	2.4	3.3	6.0	9.1	2.7	2.4
EmployRate	40.7	78.4	62.4	67.6	43.0	53.5	71.3	74.1	17.8	10.9
NonPartRate	4.3	44.3	20.6	15.5	9.5	12.0	28.9	41.1	17.0	11.0
YouthEmployRate	12.0	48.9	31.4	33.4	17.6	24.5	38.1	43.0	13.6	8.5
YouthNonPartRate	8.0	76.0	37.6	32.8	17.7	24.8	51.7	65.4	26.9	16.1

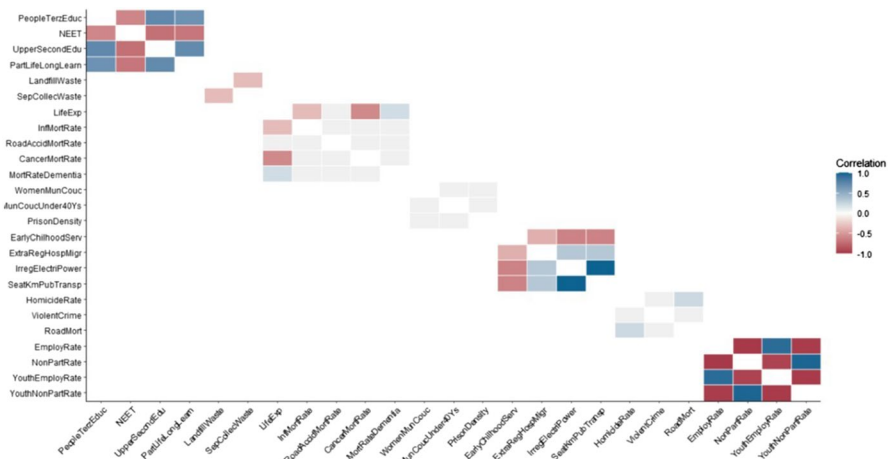


Fig. 2 Heatmap of correlation matrices

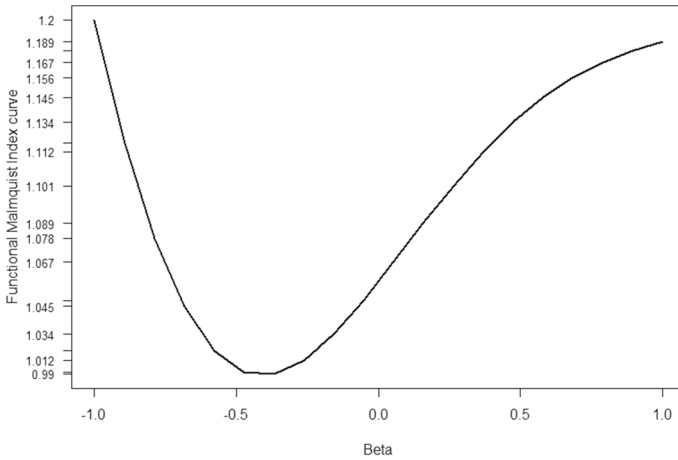


Fig. 3 FWMPI of Avellino city from 2004 to 2017 across the β domain

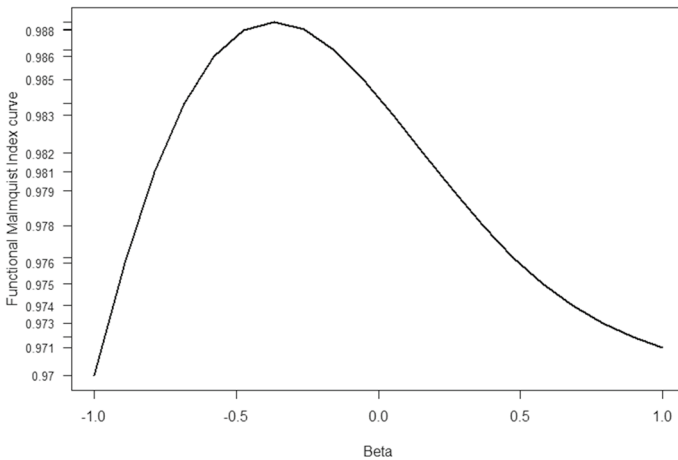


Fig. 4 FWMPI of Piacenza city from 2004 to 2017 across the β domain

the distinctive characteristics of each region (Calcagnini and Perugini 2019; Cracolici et al. 2018; Scott and Bell 2013; Nissi and Sarra 2018; Sarra and Nissi 2020). In Italy, the BES and BESdT frameworks are the only institutional statistical resources that systematically work towards this goal.

To apply our functional dynamic index, we utilize the BESdT framework, which evaluates well-being through a range of variables. Due to a lack of data and missing values, our analysis is confined to 103 provincial capital cities and only seven of the twelve domains from the original Ur-Bes dataset. In particular,

Table 20 Ranking of Provinces according to the area under the curve

Rank	Province/District	Rank	Province/District
1	Avellino	53	Palermo
2	Livorno	54	Lodi
3	Milano	55	Napoli
4	Lecco	56	Massa-Carrara
5	Bergamo	57	Prato
6	Campobasso	58	Belluno
7	Trieste	59	Pisa
8	Vercelli	60	Ragusa
9	Vicenza	61	Biella
10	Perugia	62	Torino
11	Modena	63	Treviso
12	Trapani	64	Salerno
13	Venezia	65	Lucca
14	Caltanissetta	66	Isernia
15	Frosinone	67	Siracusa
16	Bolzano/Bozen	68	Sondrio
17	Savona	69	Siena
18	La Spezia	70	Verbano-Cusio-Ossola
19	Grosseto	71	Pordenone
20	Trento	72	Padova
21	Rieti	73	Aosta
22	Firenze	74	Pesaro e Urbino
23	Ferrara	75	Viterbo
24	Lecce	76	Potenza
25	Novara	77	Chieti
26	Mantova	78	Catania
27	Cosenza	79	Pavia
28	Udine	80	Matera
29	Ravenna	81	Terni
30	Macerata	82	Taranto
31	Gorizia	83	Reggio nell'Emilia
32	Pescara	84	Rovigo
33	Cremona	85	Cuneo
34	Parma	86	Agrigento
35	Como	87	Catanzaro
36	Caserta	88	Nuoro
37	L'Aquila	89	Latina
38	Rimini	90	Benevento
39	Ascoli Piceno	91	Cagliari
40	Foggia	92	Oristano
41	Teramo	93	Arezzo
42	Asti	94	Forlì-Cesena
43	Genova	95	Reggio Calabria

Table 20 (continued)

Rank	Province/District	Rank	Province/District
44	Bologna	96	Sassari
45	Brindisi	97	Crotone
46	Brescia	98	Messina
47	Pistoia	99	Bari
48	Alessandria	100	Ancona
49	Roma	101	Enna
50	Vibo Valentia	102	Varese
51	Imperia	103	Piacenza
52	Verona		

we focus on the following pillars: “Health”, “Education and Training”, “Work and Life Balance”, “Environment”, “Safety”, “Politics”, “Quality of Services” (see Fig. 1).

Our goal is to employ the FWMPI to assess and illustrate the advancement of Italian provincial capital cities in promoting equitable and sustainable well-being between 2004 and 2017.

Table 19 presents the basic indicators that are used in constructing the dynamic composite index, and it also displays the summary statistics for the year 2017. Additionally, Fig. 2 depict the correlation matrices heatmap, categorised according to the BES pillars.

By following our approach, it is possible to visualise the variations of the FWMPI within the range of $[-1,1]$. This graphical representation facilitates a thorough comparison of the index values across different provinces. As an illustration, we have included in Figs 3 and 4 the FWMPI values for a couple of provinces as β changes. This example demonstrates how the graphical representation can capture the subtleties of the index values and provides a framework for conducting similar analyses on other provinces.

Specifically, we notice a contrasting pattern in the dynamic index within the beta domain (ranging from -1 to $+1$) for the two provinces being studied. Through the use of FDA tools, we can reconstruct the FWMPI profile within the β domain and derive a singular ranking of Italian provinces by calculating the area under the curve, as described in Sect. 2. This ranking is highly beneficial as it retains the valuable information contained within the entire domain of the diversity profile. In our context, it enables a precise and comprehensive assessment of the advancements made by Italian provincial capital cities in promoting equitable and sustainable well-being. Based on the area under the curve metric, as defined in Eq. 14, the cities of Avellino, Livorno, Milano, Lucca, and Bergamo demonstrated the most significant strides in promoting equitable and sustainable well-being during the studied time frame. Conversely, the cities with the lowest area under the curve were Bari, Ancona, Enna, Varese and Piacenza, experiencing a decline throughout the entire period. The comprehensive ranking obtained by calculating the area under the curve is presented in Table 20.

Table 21 Ranking of Provinces according to the depth measure

Rank	Province/District	Rank	Province/District
1	Piacenza	53	Foggia
2	Bolzano/Bozen	54	Messina
3	Parma	55	Brescia
4	Aosta	56	Salerno
5	Viterbo	57	Reggio Calabria
6	Milano	58	Ferrara
7	Novara	59	Livorno
8	Cremona	60	Siracusa
9	Cuneo	61	Genova
10	Firenze	62	Catania
11	Alessandria	63	Trieste
12	Caserta	64	Nuoro
13	Potenza	65	Vercelli
14	Bergamo	66	Enna
15	Treviso	67	L'Aquila
16	Chieti	68	Trapani
17	Roma	69	Rieti
18	Lecco	70	Catanzaro
19	Verbano-Cusio-Ossola	71	Rovigo
20	Pesaro e Urbino	72	Ascoli Piceno
21	Padova	73	Napoli
22	Mantova	74	Como
23	Savona	75	Biella
24	Verona	76	Frosinone
25	Venezia	77	Ragusa
26	Rimini	78	Ancona
27	Pordenone	79	Terni
28	Reggio nell'Emilia	80	Lucca
29	Ravenna	81	Caltanissetta
30	Torino	82	Lecce
31	Teramo	83	Benevento
32	Pavia	84	Avellino
33	La Spezia	85	Pisa
34	Arezzo	86	Sassari
35	Grosseto	87	Oristano
36	Bologna	88	Cosenza
37	Massa-Carrara	89	Isernia
38	Crotone	90	Cagliari
39	Belluno	91	Vibo Valentia
40	Modena	92	Palermo
41	Prato	93	Bari
42	Udine	94	Campobasso
43	Macerata	95	Siena

Table 21 (continued)

Rank	Province/District	Rank	Province/District
44	Forli-Cesena	96	Agrigento
45	Imperia	97	Lodi
46	Perugia	98	Varese
47	Sondrio	99	Pescara
48	Gorizia	100	Matera
49	Taranto	101	Brindisi
50	Vicenza	102	Pistoia
51	Trento	103	Asti
52	Latina		

The functional integrated depth, defined in Eq. 15, can be used to obtain an additional overall ranking. Table 21 displays the depth-ranking of the Italian province capital cities according to the FWMPI.

Piacenza, Bolzano, Parma, Aosta, Viterbo, and Milano are the most central cities, with Pescara, Matera, Brindisi, Pistoia, and Asti ranking last.

As a result, in the last positions there are the most outlying province capital cities. It should be noted that the depth-based rank does not allow for the inference of the direction of departure from the central observations.

5 Concluding remarks

The existing literature offers many different methods for constructing CIs. In accordance with ideas discussed in the DEA-framework, we addressed the issue of change of performance of units over time and proposed a novel method for capturing the dynamic of composite indices.

We accomplished this goal by developing a FWMPI. Our proposal is built on an extension of Shannon's entropy, which is widely used in multi-index decision making problems. To assess the degree of differentiation of DMUs, we used a family of diversity indices based on a single continuous variable. Unlike entropy-based methods, which only provide a partial view of the multi-index decision making problem, our new dynamic CI captures the multidimensional aspect of diversity and can graphically depict the composite index trend. As a result, we have the distinct advantage of supplementing the analysis with FDA tools.

We documented some pitfalls in using entropy-weighted based methods to further support the utility of our novel approach.

A simulation experiment has shown how different rankings of homogeneous DMUs can occur when different diversity measures are used. We ascertained the loss in robustness of CIs rankings under various scenarios that take into account outlier contamination, correlation and skewness among indicators, and sample size variation. The applicability and versatility of our proposal are validated by applying FWMPI to BESdT data. FDA tools like the area under the curve and the depth

measure give us useful information about the order and dynamics of the FWMPI trajectories. The depth measure describes the centrality of functional data, whereas the area under the FWMPI curves allows for a clear, unambiguous ranking of Italian province capital cities across the entire domain.

An intriguing future research direction would be to apply the novel approach to other CIs to judge the evolutionary dynamic in various application areas.

Funding Open access funding provided by Università degli Studi G. D'Annunzio Chieti Pescara within the CRUI-CARE Agreement.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Badea AC, Sanseverino CMR, Tarantola S, Bolado R (2011) Composite indicators for security of energy supply using ordered weighted averaging. *Reliab Eng Syst Saf* 96(6):651–662
- Becker W, Saisana M, Paruolo P, Vandecasteele I (2017) Weights and importance in composite indicators: closing the gap. *Ecol Indic* 80:12–22
- BES 2013 (2013) *Il Benessere Equo e Sostenibile in Italia*. ISTAT and CNEL. Rome, Italy
- Calcagnini G, Perugini F (2019) A well-being indicator for the Italian provinces. *Soc Indic Res* 142:149–177
- Caves DW, Christensen LR, Diewert WE (1982) The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica* 50(6):1393–1414
- Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision-making units. *Eur J Oper Res* 2:429–444
- Cherchye L, Moesen W, Rogge N, Van Puyenbroeck T (2007) An introduction to 'benefit of the doubt' composite indicator. *Soc Indic Res* 82:111–145
- Cherchye L, Moesen W, Rogge N, Van Puyenbroeck T, Saisana M et al (2008) Creating composite indicators with DEA and robustness analysis: the case of the technology achievement index. *J Oper Res Soc* 59(2):239–251
- Coelli TJ, Prasada Rao DS (2005) Total factor productivity growth in agriculture: a Malmquist index analysis of 93 countries, 1980–2000. *Agric Econ* 32(1):115–134
- Cook WD, Seiford LM (2009) Data envelopment analysis (DEA)-thirty years on. *Eur J Oper Res* 192(1):1–17
- Costa DS (2015) Reflective, causal, and composite indicators of quality of life: a conceptual or an empirical distinction? *Qual Life Res* 24(9):2057–2065
- Costanza R, Hart M, Posner S et al (2009) *Beyond GDP: The need for new measures of progress*. Boston University Creative Services, Boston
- Cracolici MF, Cuffaro M, Lacagnina V (2018) Assessment of Sustainable Well-being in the Italian Regions: an activity analysis model. *Ecol Econ* 143:105–110
- Cuevas A, Febrero M, Fraiman R (2007) Robust estimation and classification for functional data via projection-based depth notions. *Comput Stat* 22:481–496

- Diamantopoulos A, Riefler P, Roth KP (2008) Advancing formative measurement models. *J Bus Res* 61(12):1203–1218
- Diamantopoulos A, Siguaw JA (2006) Formative versus reflective indicators in organizational measure development: a comparison and empirical illustration. *Br J Manag* 17(4):263–282
- Diamantopoulos A, Winklhofer HM (2001) Index construction with formative indicators: an alternative to scale development. *J Mark Res* 38(2):269–277
- Di Battista T, Fortuna F, Maturo F (2017) BioFTF: an R package for biodiversity assessment with the functional data analysis approach. *Ecol Indic* 73:726–732
- Fallahnejad R (2017) Entropy based malmquist productivity index in data envelopment analysis. *Int J Data Envel Anal* 4:1425–1434
- Färe R, Grosskopf S, Lindgren B, Roos P (1994) Productivity Developments in Swedish Hospitals: A Malmquist Output Index Approach. In: Charnes A, Cooper WW, Lewin AY, and Seiford LM (eds) *Data Envelopment Analysis: Theory, Methodology, and Applications*, pp 253–272, Berlin, Germany. Springer Netherlands
- Färe R, Karagiannis G, Hasannasab M, Margaritis D (2019) A benefit-of-the-doubt model with reverse indicators. *Eur J Oper Res* 278(2):394–400
- Färe R, Grosskopf S (2004) *New Directions: Efficiency and Productivity*. Kluwer Academic Publishers, Massachusetts
- Farrell MJ (1957) The measurement of productivity efficiency. *J R Stat Soc Ser A* 120(3):253–281
- Filippetti A, Peyrache A (2011) The patterns of technological capabilities of countries: a dual approach using composite indicators and data envelopment analysis. *World Dev* 39(7):1108–1121
- Fleurbaey M (2009) Beyond GDP: the quest for a measure of social welfare. *J Econ Lit* 47(4):1029–1075
- Fraiman R, Muniz G (2001) Trimmed means for functional data. *TEST* 10:419–440
- Fortuna F, Naccarato A, Terzi S (2022) Country rankings according to well-being evolution: composite indicators from a functional data analysis perspective. *Ann Oper Res* 419
- Greco S, Ishizaka A, Tasiou M, Torrisi G (2019) On the methodological framework of composite indices: a review of the issues of weighting, aggregation, and robustness. *Soc Indic Res* 141:61–94
- Hall J, Giovannini E, Morrone A, Ranuzzi G (2010) *A framework to measure the progress of societies*. OECD statistics working papers, 2010/5. OECD Publishing
- Larraz IB, Pavia JM (2010) Classifying regions for European development funding. *Eur Urban Reg Stud* 17(1):99–106
- Keogh S, O'Neill S, Walsh K (2021) Composite measures for assessing multidimensional social exclusion in later life: conceptual and methodological challenges. *Soc Indic Res* 155:389–410
- Koopmans TC (1951) An analysis of production as an efficient combination of activities. In: Koopmans TC (ed) *Activity Analysis of Production and Allocation*, page No 13, Wiley. Cowles Commission for Research in Economics, New York
- Maggino F (2017) Developing indicators and Managing the Complexity. In Maggino F (eds), *Complexity in Society: From Indicators Construction to their Synthesis*, pp 87–114, Berlin, Germany. Springer
- Mahlberg B, Obersteiner M (2001) Remeasuring the hdi by data envelopment analysis. IIASA, (Interim Report IR-01-06)
- Malmquist S (1953) Index numbers and indifference surfaces. *Trabajos de Estadística* 4:209–242
- Melyn W, Moesen W (1991) Towards a synthetic indicator of macroeconomic performance: unequal weighting when limited information is available. Public economics research paper, CES 17, KU Leuven
- Nissi E, Sarra A (2018) A measure of well-being across the Italian urban areas: an integrated DEA-entropy approach. *Soc Indic Res* 136:1183–1209
- OECD (2008) *Handbook on constructing composite indicators*. OECD Publishing, Paris
- Patil GP, Taillie C (1979) An overview of diversity. In: Grassle JF, Patil GP, Smith W, Taillie C (eds) *Ecological Diversity in Theory and Practice*. Fairland. International Co-operative Publishing House, pp 23–48
- Patil GP, Taillie C (1982) Diversity as a concept and its measurement. *J Am Stat Assoc* 77:548–567
- Peykani P, Seyed Esmaeili FS, Mirmozaafari M, Jabbarzadeh A, Khamechian M (2022) Input/Output variables selection in data envelopment analysis: a Shannon entropy approach. *Mach Learn Knowl Extr* 4:688–699
- Ramsay JO, Silverman BW (2005) *Functional Data Analysis*, 2nd edn. Springer-Verlag, New York

- R Core Team: R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2023). www.R-project.org/
- Riccardini F, De Rosa D (2016) How the nexus of water/food/energy can be seen with the perspective of people well being and the Italian BES framework. *Agric Agric Sci Procedia* 8:732–740
- Sahoo BK, Singh R, Mishra B, Sankaran K (2017) Research productivity in management schools of India during 1968–2015: a directional benefit-of-doubt model analysis. *Omega* 66(Part A):118–139
- Saisana M, Saltelli A, Tarantola S (2005) Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *J R Stat Soc Ser A* 168(2):307–323
- Sarra A, Nissi E (2020) A spatial composite indicator for human and ecosystem well-being in the Italian urban areas. *Soc Indic Res* 148:353–377
- Scott K, Bell D (2013) Trying to measure local well-being: indicator development as a site of discursive struggles. *Environ Plan C Gov Policy* 31:522–539
- Smith P (2002) Developing composite indicators for assessing health system efficiency. In: Smith PC (ed) *Measuring up: Improving the performance of health systems in OECD countries*. Paris. OECD, pp 295–309
- Staessens M, Kerstens PJ, Bruneel J, Cherchye L (2019) Data envelopment analysis and social enterprises: analysing performance, strategic orientation and mission drift. *J Bus Ethics* 159:325–341
- Stiglitz JE, Sen A, Fitoussi JP (2009) Report by the Commission on the Measurement of Economic Performance and Social Progress
- Zhu Y, Tian D, Yan F (2020) Effectiveness of entropy weight method in decision-making. *Math Probl Eng* 7:1–5

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