Ecological Indicators 117 (2020) 106603

Contents lists available at ScienceDirect

Ecological Indicators

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

The evolution of composite indices of well-being: An application to Italy☆

Fabio Bacchini[a,*], Barbara Baldazzi[b], Lorenzo Di Biagio[a]

[a] Italian National Institute of Statistics — Viale Liegi 13, 00198 Rome, Italy
[b] Italian National Institute of Statistics — Piazza G. Marconi, 54, 00146 Rome, Italy

ARTICLE INFO

Keywords:
Composite indices
Multidimensional well-being
Normalization-aggregation-unbalance interplay
Unbalance adjustment
Beyond GDP
Time series

JEL classification codes:
C43
I31

ABSTRACT

Prompted by the work of the Stiglitz's Commission, the growing attention to the beyond-GDP measures has led to the inclusion of well-being indicators in the policy agenda. This innovation asks for an improvement of the existing methodology to produce composite indices, in order to correctly address spatial and temporal comparisons as well as tackling for unbalances. Following a short review of the main international experiences, this paper will investigate these issues considering the methodology currently adopted to normalize and aggregate the selected individual indicators included in the Italian well-being. We study the properties of this methodology looking at different normalization and aggregation approaches and underlining some drawbacks, mostly due to the way in which time dimension, normalization, aggregation and unbalance adjustment interplay with each other. We illustrate our findings by means of examples related also to the ecological side. We argue that new efforts should be done to overcome these drawbacks extending the research agenda toward new non-compensatory approaches. Testing for time series methods, such as dynamic factor models could represent another important step forward. Meanwhile the introduction of a more traditional framework for the composite indicators for Italian well-being could be considered.

1. Introduction

Nowadays, driven by the work of the Stiglitz's Commission (Stiglitz et al., 2009), it is widely accepted to consider well-being as a multidimensional phenomenon. It means that different dimensions are measured on a micro or macro population (i.e. households, regions, countries) using a dashboard of indicators, often across time.

The growing attention to the beyond-GDP measures has led to progressively include well-being indicators in the policy agenda. In Italy, for example, the Ministry of Treasury has started to use well-being indicators in the evaluation of fiscal policies (see Ministero dell'Economia e delle Finanze, 2018) while the European Commission has funded a new project MAKSWELL (MAKing Sustainable development and WELL-being frameworks work for policy, see Bacchini et al., 2018) that aims to improve data and methodologies to relate policy analysis and well-being.

In literature there is a wide debate and some researchers support the idea of deriving, from a multidimensional framework, a single metric that makes it easy to compute the progress/decline in well-being over time. But the identification of a metric, similar to the integrated system currently adopted to produce GDP measures, is a hard task. Meanwhile, a number of composite indices have been introduced both by international organizations (see for example UNDP, 2016) and by national Institutes of Statistics (Quality of Life Spain (INE – Spain, 2019), Bes Italy (Istat, 2015) and WBI Portugal (INE - Portugal, 2017)). The introduction of composite indices and their use to measure the effects of policy programs, require a framework that makes it possible to clearly assess their evolution between two different periods. This is the traditional approach for which GDP measure is useful, allowing comparisons both over time and across countries.

Although there is not a common standard to measure well-being across countries, the various experiences share some common char-

☆ Even if the paper is the result of the combined effort of all the three authors, Sects. 1, 3, 4, 4.1, 5, 6 can be mostly attributed to F. Bacchini, Sect. 2 can be mostly attributed to B. Baldazzi, Sect. 5.1 can be mostly attributed to B. Baldazzi and L. Di Biagio and Sects. 3.1, 4.2, 4.3, 4.4, 5.1.1, 5.2 and App. A can be mostly attributed to L. Di Biagio.

* Corresponding author.
E-mail addresses: bacchini@istat.it (F. Bacchini), baldazzi@istat.it (B. Baldazzi), lorenzo.dibiagio@istat.it (L. Di Biagio).

https://doi.org/10.1016/j.ecolind.2020.106603
Received 11 September 2019; Received in revised form 28 May 2020; Accepted 2 June 2020
1470-160X/

acteristics such as the definition of domains and of the individual indicators. In this paper we consider as done the aforementioned steps but rather we focus on the methodology to compute composite indices, mainly on the normalization, aggregation and unbalance adjustment techniques.

To investigate these issues we consider the example of Italy that is based on a consolidated framework for the measure of well-being both at national and territorial level (see Calcagnini and Perugini, 2019; Casadio Tarabusi and Guarini, 2013; Istat, 2018). Moreover, according to the international review, Italy presents a high standard legislation that explicitly includes well-being indicators in the policy evaluation cycle (Bacchini et al., 2018; Bacchini et al., 2020; Ministero dell'Economia e delle Finanze, 2018). Concerning the dissemination of the Italian well-being, the annual reports (*Rapporto Bes* — Istat, 2015; Istat, 2016; Istat, 2017; Istat, 2018; Istat, 2019) present both a dashboard representation as well as composite indices computed for each domain.

In order to normalize and aggregate the individual indicators into a composite index, the methodology adopted in the Italian annual reports is the one proposed by Mazziotta and Pareto (see Mazziotta and Pareto, 2016 and Istat, 2015, p. 49), the so-called AMPI (*Adjusted Mazziotta–-Pareto Index*). Althought AMPI does not represent a consolidated standard in international literature on well-being, it has been adopted in several studies related to Italy (see for example Ciommi et al., 2017) and it is quoted in recent surveys on aggregation methods (Casadio Tarabusi and Guarini, 2013; Greco et al., 2019).

This work pinpoints some drawbacks in the use of this approach both looking at the desirable properties of the composite index and at the relationship amid the unbalance adjustment and the time evolution. We illustrate our findings by means of examples drawn from the different dimensions of well-being,with attention to the ecological side. The results presented in the paper suggest that the current state of the art in composite indices claims for an agenda where the interplay between normalization, aggregation and time dimension is correctly addressed. Without these improvements it will be very difficult to respond to the policy evaluation cycle related to well-being as in the spirit proposed by Stiglitz et al. (2018).

This paper is organized as follows. Section 2 will briefly review the international experiences, Section 3 will introduce the general characteristics of composite indices, extending the notation to the time domain while Section 4 presents the main characteristics of Italian well-being and AMPI. Section 5 will analyze, respectively, the implications of normalization and aggregation, together with their interplay with unbalance, focusing on the AMPI method while Section 6 will draw conclusions.

2. International experiences on well-being measures

The ongoing international initiatives on the development of measurement systems on well-being and sustainability are heterogeneous across countries and international institutions. Different projects have provided a mapping of European experiences (see for example Andreoni and Galmarini, 2015). Recently, the MAKSWELL project, funded by the European Commission (2017–2019), has released a report with an up-to-date mapping of well-being frameworks at European level (Tinto et al., 2018). Some of these experiences explicitly introduce composite indices to summarize the different indicators chosen to measure well-being.

Concerning international experiences, some of them are listed in Table 1 such as those elaborated by international organizations (*HDI* (UNDP, 2016), *Better life index* (OECD, 2017), *Quality of life* (Eurostat, 2018), *SGD Index* (Bertelsmann Stiftung and Sustainable Development

Table 1
Selected international composite indices on well-being: main characteristics.

Name	Domains	Indicators	Composite Index	Normalization	Method of synthesis	Time series	Developer
HDI	3	4	yes	Min–Max with goalposts	geometric mean	1990–2018	UNDP
Better life index	11	24	graphic representation	Min–Max	weighted arithmetic mean (with subjective weights)	last year available	OECD
Quality of life	9	21	no			last year available	Eurostat
SDG index	17	114	yes	Min–Max with goalposts	arithmetic mean	2019	Bertelsmann Stiftung and the Sustainable Development Solutions Network
Quality of life Spain	9	57	yes	Min–Max with reference	arithmetic mean with penalty	2008–2018	Instituto Nacional de Estadística - Spain
Bes	12	130	yes - only for each domain	Min–Max with reference	arithmetic mean with penalty	2010–2018	Istat - Italy
WBI Portugal	10	79	yes	fixed base index 2004 = 100	weighted arithmetic mean	2004–2016	Instituto Nacional de Estadística - Portugal
CIW	8	64	yes	fixed base index 1994 = 100	arithmetic mean	1994–2014	Univ. Waterloo

Solutions Network, 2019)), three composite indices provided by national Institutes of Statistics (*Quality of Life Spain* (INE – Spain, 2019), *Bes Italy* (Istat, 2019) and *Well Being Index (WBI) Portugal* (INE - Portugal, 2017)) and one from the University of Waterloo (*The Canadian Index of Well-being, CIW* (University of Waterloo, 2016))[1]

From Table 1 the heterogeneity of the approaches clearly emerges for all the characteristics reported: number of domains and indicators, systems of normalization and aggregation. Even the use of composite indices does not appear a common strategy (Better Life, Quality of life do not provide a standard composite index).

Started in 1990, HDI represents the most known framework. It is a summary measure of average achievement in three key dimensions of human development: a long and healthy life, being knowledgeable and having a decent standard of living. The HDI is an index that assumes values between 0 (minimum development) and 1 (maximum development) obtained as the geometric mean of normalized indicators that are selected to represent the three key dimensions of human development. The indicators are standardized with the Min–Max method with minimum and maximum values (goalposts) set as "natural zeroes" and "aspirational targets".

Considering HDI as a benchmark experience, respect to the normalization procedure a Min–Max approach has been adopted also by the Better Life Index (but only in a graphical representation), the SDG index, while, for example, Quality of Life Spain in 2016 was based on z-scores. A different direction has been taken by WBI and CIW: both frameworks use a normalization based on index numbers and an aggregation based on the arithmetic mean.

Among the examples provided, in the Italian Bes the aggregation is performed with an arithmetic mean with penalty (the Mazziotta–Pareto Index), while the normalization has been specifically designed in order to consider time series. Quality of Life Spain is now adopting the same strategy, even if the normalization range is smaller.

The heterogeneity raising from these experiences emphasizes how, taking for granted the selection of indicators and domains, the interplay of normalization, aggregation, time dimension and unbalance adjustment represents a huge, but not well consolidated challenge. Moving from general considerations to the empirical facts, we argue below how these different characteristics have been faced by the Italian Well-Being.

3. From individual indicators to composite indices

Composite indices could be very useful for policy analysis: they allow measuring multidimensional concepts, also over time, in a way that is usually easier to interpret than finding common trends amid different dimensions. In this way they facilitate communication with the general public and promote accountability (see OECD and JRC, 2008).

In this work we focus on the characteristics of the composite indices introduced for *Rapporto Bes*, the well-being report elaborated each year by Istat (see Istat, 2015; Istat, 2016; Istat, 2017; Istat, 2018; Istat, 2019). According to the examples proposed in Table 1, this experience (embraced also by Quality of Life Spain, INE – Spain (2019)) seems unique in the use of a penalty approach to assess the evolution of well-being over time.

From a general point of view, composite indices could "differ in the dimensions and indicators selected, the transformations applied to the indicators, the assumed substitutability between indicators and the relative weights given to them" (Decancq and Lugo, 2013).

Introducing some notation, given a real-valued matrix $\mathbf{X} = \{x_{ij}^t\}$ with n rows (statistical units i, in our case Italian regions), m columns (individual indicators j) and k time periods t, in our case years, we define as 'synthesis' or 'composite index' a real function I defined on data matrices and that, for each unit and each time, computes a synthesis of all the individual indicators, i.e.:

$$I\colon \mathbf{R}^{n\times m\times k} \to \mathbf{R}^{n\times k}.$$

Function I is required to be theoretically consistent and able to generate values that are unambiguous and easy to communicate together with the methodological approach adopted. Following Mauro et al. (2018), I has to retain the full advantage of a continuous cardinal measure and providing, in a consistent way, both spatial and temporal comparisons (Mazziotta and Pareto, 2016, p. 989, Section 3.3).

Cardinal measures are more appropriate than counting measures for the measurement of well-being (Mauro et al., 2018) but at the expense of having to deal with potentially difficult and problematic issues, such as standardization of variables, implicit weighting, management of substitutability rates. More generally the function I can be decomposed into two functions, the normalization function $N\colon \mathbf{R}^{n\times m\times k} \to \mathbf{R}^{n\times m\times k}$ and the aggregation function $A\colon \mathbf{R}^{n\times m\times k} \to \mathbf{R}^{n\times k}$, so that I is just the latter applied to the former.

To emphasize the role of weights and time we can use the notation proposed by Decancq and Lugo (2013):

$$I(x) = \begin{cases} [w_1 I_1(x_1)^{\beta} + \ldots + w_m I_m(x_m)^{\beta}]^{\frac{1}{\beta}}, & \text{for } \beta \neq 0. \\ I_1(x_1)^{w_1} \cdot \ldots \cdot I_m(x_m)^{w_m} & \text{for } \beta = 0. \end{cases} \tag{1}$$

where $I_j(x_j)$ is the transform of the indicator x_j measured on the ith statistical unit. However if we are interested in the change of the composite index of well-being over time we require the possibility of a consistent comparison between $I^{t'}(x)$ and $I^t(x)$ with $t' > t$.

Comparison between $I^{t'}(x)$ and $I^t(x)$ should take at least two aspects into account: the stability of weights over time and the possibility to isolate the contribution of a single indicator $I_j(x_j)$ to the evolution of the composite index between t and t'. This issue is addressed, for example, both in the CIW and in the WBI using a fixed base and a weighted arithmetic mean.

Concerning the weights it will be important to define whether each w_m value will not change in time or if they follow a chain-linked approach, in which the w_m's are updated every year (see for example Eurostat, 2006). For a general discussion on weights and composite indices please refer to Decancq and Lugo (2013) and also to Becker et al. (2017).

3.1. Degree of substitutability and unbalance

As underlined by Decancq and Lugo (2013) the choice for the aggregation method is strictly related to our hypotheses on the degree of substitutability (compensability) between individual indicators, i.e., the possibility of compensating deficits and surpluses. In Eq. (1), the parameter β could be "related to the elasticity of substitution between the transformed achievements, σ, where $\sigma = \frac{1}{1-\beta}$" (Decancq and Lugo, 2013). For $\beta = 1$ we obtain the (weighted) arithmetic mean. In this case the elasticity of substitution is infinite, implying the possibility of offsetting any deficit in one indicator with a suitable surplus in another, where the trade off is constant regardless of the levels of the indicators (perfect substitutes). For $\beta = 0$ we obtain the (weighted) geometric mean that implies a unit elasticity of substitution (a one percent decrease in one of the indicators can be compensated by a one percent increase in another), while $\beta \to -\infty$ is related to an elasticity of sub-

[1] To have a more comprehensive look at the studies on well-being we suggest (Decancq and Lugo, 2013). Some updating could be find in the recent working group on "Guidelines on producing leading, composite and sentiment indicators" coordinated by UNECE.

stitution equal to zero "which means that there is no possible substitution between dimensions. In this case the well-being index becomes the minimum" among the normalized indicators.

Substitutability is closely related to the concept of unbalance, i.e., a disequilibrium among the variables that are used to build a given composite index. (see Casadio Tarabusi and Guarini, 2013). Since in a composite index each indicator is introduced to represent a relevant aspect of a phenomenon, then a perfect substitutability among factors might not be desirable, and the analyst might decide to introduce a degree of non-substitutability between individual indicators, especially if they pertain to different dimensions or if they represent goals that are considered equally legitimate and important.

Other than the geometric mean, several methods have been introduced in order to consider unbalances in the aggregation (see for example Casadio Tarabusi and Guarini, 2013, p.28, Sect. 5). Although it is not very easy to describe a specific function for measuring unbalances, it is possible to identify desirable properties for the aggregation.

Mauro et al. (2018, p. 80, Section 3), proposes three key technical properties that a composite index based on a cardinal measure must satisfy: *continuity*, so that jumps in the values of the composite index depend only on the jumps of the data and not on the aggregation function A; (strict) *monotonicity*, so that any improvement (worsening) in any indicator results in an increase (decrease) of the synthetic score and *heterogeneity penalization*, i.e., a penalization of unbalance between indicators, so that there is no perfect substitutability between individual indicators (dimensions) and the elasticity of substitution is not infinite.

One example of such an index is the Human Development Index (HDI, see Section 2) elaborated by the United Nations. Since it is computed by a simple geometric mean, it is continuous, monotone and penalizes heterogeneity. This index is also a good example in the shift of attention between fully compensatory and not fully compensatory composite indices. In fact the first release of the index was based on the arithmetic mean of the three different indicators that make it up.[2]

4. Main characteristics of Italian Well-being and its methodology for composite indices

Starting from its first dissemination in 2013, the annual report on the measurement of "equitable and sustainable well-being" (*Rapporto Bes* – Rapporto sul benessere equo e sostenibile, for example see Istat, 2019) has been characterized by a strong attention both to the revision of the indicators included in the domains and to the dissemination policy. Particularly, the dissemination stage aims at providing both a dashboard representation and (starting from 2015) an aggregation framework, using composite indices for each domain. After a short description of the main characteristics of the *Bes* framework, this paragraph illustrates the method that has been adopted until now to build up the composite indices, the so-called AMPI *Adjusted Mazziotta–Pareto index*.

4.1. Italian equitable and sustainable well-being

The Italian National Institute of Statistics (Istat), together with the National Council for Economics and Labor (CNEL), launched in December 2010 an inter-institutional initiative aimed at developing a multi-dimensional approach for the measurement of "equitable and sustainable well-being" (*Bes* — benessere equo e sostenibile), in line with the recommendations issued by the OECD and the Stiglitz Commission (see Stiglitz et al., 2009).[3]

Since the preliminary steps, Bes has had the ambition to measure not only the level of well-being, through the analysis of all relevant aspects of quality of life of the population, but also its equity amid social groups and geographic areas of the Country, and sustainability for future generations. This approach increases the complexity of the measurement but allows a more accurate analysis of the evolution of well-being in Italy.

In the context of recent international initiatives, the approach adopted with Bes has been characterized by a participative process, involving civil society and national experts in the definition of the framework and in the selection of indicators. This shared work together with the evidence coming from international experiences resulted in the identification of a total of 12 domains.

The 12 selected domains are divided into 2 typologies, 9 of them are defined as outcome domains and are those related to dimensions which have a direct impact on human and environmental well-being (Alkire, 2002); the remaining 3 domains are defined as drivers of well-being, measuring functional elements to improve the well-being of the community and the surrounding environment. The domains are:

- Outcome: health; education and training; work and life balance; economic well-being; social relationship; security; landscape and cultural heritage; environment; subjective well-being;
- Driver: politics and institutions; innovation, research and creativity; quality of services.

Overall, 134 indicators were originally identified to represent the 12 domains of well-being. However, the framework is considered as an open lab, and the set of indicators is reviewed annually to consider emerging information needs and methodologies.

In 2018 the importance attributed by citizens to each of the 12 domains of Bes in the individual perception of well-being was tested by a qualitative survey, which is an ideal update of that carried out in 2011 in the definition stage of the Bes domains.

Concerning the dissemination stage, the annual report on Bes proposes both a dashboard representation and a representation based on composite indices for each domain (Istat, 2018). Dashboard representation is driven by specific web graphs (seehttp://www.istat.it/it/benessere-e-sostenibilità/la-misurazione-del-benessere-(bes)/gli-indicatori-del-bes) and by the comparisons of indicators over time and across regions.

According to the data available at Dec. 2018 (Istat, 2018), compared to the previous year, nearly 40% of the indicators improved (43 on 110) while 31.8% deteriorated and 29.1% remained stable (see Fig. 1). Results show an even more positive evolution if values are compared with 2010: 53.4% of indicators improved (62 on 116).

As mentioned above, starting from the 2015 edition (Istat, 2015) the annual report on Bes has also introduced an aggregation measure for each domain based on composite indices. This has simplified the interpretation of the evolution of many different indicators (see Fig. 2). The aim of this paper is to present the methodology used to compute these composite indices together with its drawbacks.

[2] Since 2010 the arithmetic mean has been replaced by the geometric mean so that now poor performance in any dimension is directly reflected in the HDI, and there is no longer perfect substitutability across dimensions (cf. UNDP, 2010, p. 15).

[3] For an exhaustive description of the Bes project we refer to Bacchini et al. (2020) while a short description is contained in Sabbadini and Maggino (2018).

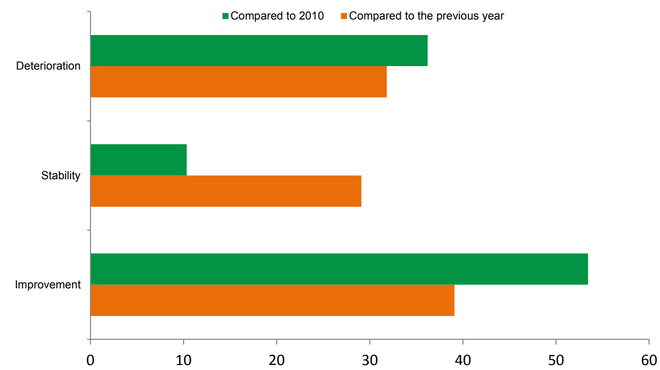

Fig. 1. Comparison of the values of all indicators in the Bes framework. Italy. Year 2010 and last year available. Source: Istat (2018).

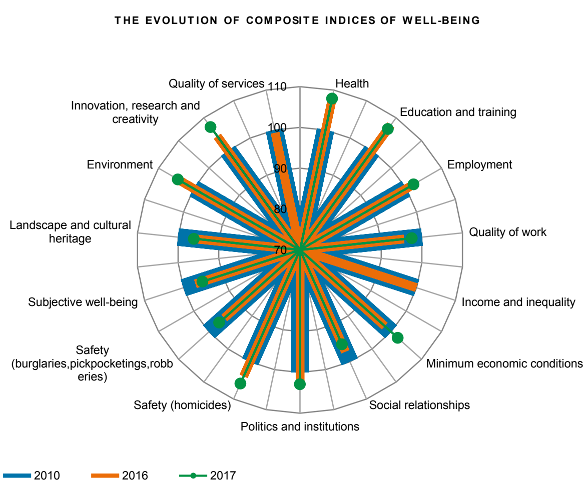

Fig. 2. Comparison of the values of all composite indices in the Bes framework. Italy. Year 2010, 2016 and 2017. Source: Istat (2018).

4.2. *Composite indices for the Italian well-being*

As we have seen (Section 2), most of the international experiences monitoring well-being over time use the arithmetic mean (WBI Portugal, CIW) or the geometric mean (HDI). However, as underlined by Casadio Tarabusi and Guarini (2013), when we talk about aggregation we need also to address unbalances between the indicators because "the balance among factors of a multidimensional socio-economic phenomenon reinforces the importance of the conglomerative perspective" (see also Greco et al., 2019).

To take unbalances into account, since the very first computation of composite indices in 2015, *Rapporto BES* has adopted (Istat, 2015) and is currently adopting (Istat, 2019) the AMPI (*Adjusted Mazziotta–Pareto Index*) aggregation method (the "minus" version), as explained below (see also Mazziotta and Pareto, 2016, p. 989, Section 3.3; Istat, 2015, p. 53). This method includes also a particular normalization procedure (AMPN — *Adjusted Mazziotta–Pareto Normalization*) that will be dealt with in the next sub-section.

Adopting the same notation as in Mazziotta and Pareto (2016, p. 990, Eq. (4)):

$$\mathrm{AMPI}^{-} = \mathrm{M}_{r_i} - \mathrm{S}_{r_i}\mathrm{cv}_{r_i} \tag{2}$$

where M_{r_i}, S_{r_i} and cv_{r_i} are, respectively, the mean, the standard deviation and the coefficient of variation of the normalized values of the individual indicators for the unit i and time t (the aggregation is independently computed for each unit i and time t). The "minus" exponent means that the composite index is "increasing" or "positive" i.e., increasing values of the index correspond to positive variations of the phenomenon. Since this is the only case in Bes (Istat, 2015, p. 53) then from now on we will refer to AMPI⁻ simply as AMPI.

AMPI decomposes the score of each statistical unit for each time period into two parts: the arithmetic mean (M) and the penalty (S·cv). The penalty is a function of the indicators' variability (with respect to the average value).

According to Greco et al. (2019) AMPI could be classified as a mixed strategy method based on a compensatory side (arithmetic mean) together with a non compensatory aggregation (the penalty). In details, since AMPI is an unbalanced-adjusted function (in terms of the penalty) then it does not allow perfect substitutability among factors: AMPI is a composite index "for summarizing a set of indicators that are assumed to be non-substitutable, i.e., all components must be balanced" (Mazziotta and Pareto, 2016, p. 986, Section 3.1). "The aim is to reward the units that, mean being equal, have a greater balance among the indicators values" (Mazziotta and Pareto, 2016, p. 988). "[AMPI] is characterized by the use of a function that penalizes units with unbalanced values of normalized indicators" (Istat, 2015, p. 53). Then, unlike other aggregation methods, AMPI aims to penalize those statistical units for which individual indicators, observed at time t, are unbalanced, i.e., their values are far apart. This might seem a simple concept, but since the aggregation is applied to normalized indicators then the definition of balance, or equilibrium, strictly depends on the normalization method adopted. Therefore we must understand what balance/unbalance among indicators actually means in the context of AMPN, in order to properly appreciate the effects of the penalty. In this regard see Section 5.1.

4.3. *The normalization stage*

Concerning the normalization stage, AMPN could be considered as a specific case of the Min–Max approach.[4] Actually AMPN[5] is a variant of the Min–Max procedure with two details: a, b are set to 70 and 130 (respectively) and in order to facilitate the interpretation of results, for each indicator a reference value $r_j = r_j^{t_b}$, related to a specific entity and time t_b (assumed as a benchmark), is introduced. At the end of the procedure this reference value will be transformed into $\frac{a+b}{2} = 100$.

AMPN works in two steps. In the former the normalized values both for each x_{ij}^t and for the reference value r_j are derived using the formula

$$z_{ij}^t = \frac{x_{ij}^t - m_j^*}{M_j^* - m_j^*}\left(b - a\right) + a. \tag{3}$$

[4] The *Min–Max* procedure is a well-known linear transformation to bring the values of the indicator back to a fixed range $[a, b]$, with $a, b \in \mathbf{R}$, $a < b$, usually $[0, 1]$ (see for example OECD and JRC, 2008, p. 85, subSection 5.3). Let x_{ij} be the value of the indicator j for the statistical unit i. The normalized value z_{ij} of x_{ij} is defined as $z_{ij} = \frac{x_{ij} - m_j}{M_j - m_j}\left(b - a\right) + a$ where m_j and M_j, with $m_j < M_j$, are the minimum and maximum (respectively) of all the values of the indicator j. As in Decancq and Lugo (2013) this procedure works in the simple case of cross-section observations while, when time is introduced, some modifications are required. Commonly we could define the minimum and maximum also over time, following these rules (see Tarantola, 2008): $m_j^* = \min_{\forall_i, \forall_t} x_{ij}^t$, $M_j^* = \max_{\forall_i, \forall_t} x_{ij}^t$.

Using these definitions the final normalization is:

[5] In terms of the terminology proposed by the Decancq and Lugo (2013), Table 2, we are talking about a method C ("linear transformation").

(3). We call these two values $x_{ij}^{t}{}'$ and r_j'. Then, in the latter, the final value z_{ij}^t is obtained as:

$$z_{ij}^t = x_{ij}^{t}{}' + \frac{a+b}{2} - r_j' = x_{ij}^{t}{}' + 100 - r_j'. \tag{4}$$

AMPN turns out to act like a Min–Max procedure in which instead of using the actual minimum m_j^* and maximum M_j^* two other goalposts are defined so that also the reference value r_j is considered.[6] Istat, for its *Rapporto Bes*, is currently adopting the AMPN methodology without necessarily updating the minimum and maximum every year, and setting the reference value r_j ad the value of the indicator j for Italy in 2010.

Eq. (4) can be re-written as $z_{ij}^t = sx_{ij}^t + p$, where $s = \frac{b-a}{M_j^* - m_j^*} = \frac{60}{M_j^* - m_j^*}$ is a scale factor that modifies the variability of the raw indicator j and $p = \frac{a+b}{2} - \frac{b-a}{M_j^* - m_j^*} r_j = 100 - \frac{60}{M_j^* - m_j^*} r_j$ is a position term that alters the ratios of the values of j.[7] Even if the AMPN-normalized values should in principle approximately fall in the interval [70, 130], they might well fall outside of this range. For two reasons. First of all (as in the Min–Max procedure) the minimum and maximum may not be computed over the whole time period (this is the case when new yearly data become available and an update does not take place). Secondly because the translation that gets r_j to 100 may move the interval, theoretically down to [40, 100] or up to [100, 160] and hence the actual values are less informative if compared with the alleged minimum (70) and maximum (130).[8] In particular the indicators are not always brought to a common interval, not even in the same domain: e.g., for the domain *Environment*, in Istat (2017) the indicator *AMB8-Urban green* is normalized between 97.5 and 157.5 while the indicator *AMB5-6-Quality of urban air* is normalized between 63.7 and 123.7.

The AMPN method, as the Min–Max method, partially control for the indicators' variability: in fact the scale factor s is proportional to the inverse of the range of the raw indicator, therefore the less variable indicators acquire variability, while the more variable ones lose it. At the same time both the methods alter the temporal variations between two periods (usually $p \neq 0$), making time series smoother or more erratic, but without inverting the trends. Also relative power relations among regions are not preserved. The intensities of these transformations depend on the raw data, and they can differ from one indicator to another.

Some concrete examples taken from *Rapporto Bes 2017* (Istat, 2017) can well illustrate the difference between the growth rate of the raw indicators and the normalized ones. According to suggestions from the Scientific Committee, in *Rapporto Bes* there are three individual indicators that are normalized but not aggregated with any other: *LAV1-Employment rate*, *BSO1-Life satisfaction* and *SIC1-Homicide rate*.

In the case of the *LAV1-Employment rate*, the AMPN normalization does not alter the annual growth rate (see Table 2). Employment rate in Italy grows from 2015 to 2016 by 1.8% either considering the raw indicator or the normalized one. This similarity depends on the relationships between the overall minimum (42.1), maximum (78.2) and the reference value (61) of the indicator. In other words, specifically for this case, the normalization is mostly a scale dilation ($s = 1.66$) since the translation ($p = -1.39$) is not significant if compared to the values of the indicator. Anyway normalization modifies the unit of measure and makes the scale of values less readable: if it is clear what we mean by an employment rate of 61.6%, it is less immediate to interpret the corresponding value of 101.0.

Table 2
Employment rate: raw and normalized indicator. Italy. Years 2008–2016. Data from Istat (2017).

	2008	2009	2010	2011	2012	2013	2014	2015	2016
Raw ind.	62.9	61.6	61.0	61.0	60.9	59.7	59.9	60.5	61.6
Norm. ind.	103.2	101.0	100.0	100.0	99.8	97.8	98.2	99.2	101.0

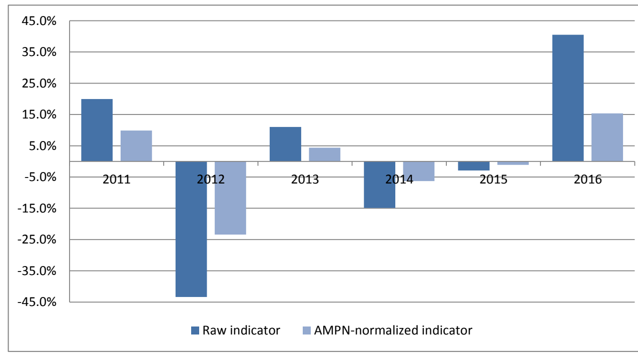

Fig. 3. Life satisfaction: raw and AMPN-normalized indicator. Campania. Years 2011–2016. Growth rates. Data from Istat (2017).

Now consider the indicator *BSO1-Life satisfaction*. For the Campania region we see that the raw indicator goes down by 43% between 2011 and 2012, while the normalized indicator just by 23%. Between 2015 and 2016 the raw indicator increases by 41% while the normalized one just by 15% (Fig. 3).[9]

Fig. 3 makes plain how the AMPN normalization accounts for differences in the variance of the indicators but it does not preserve the growth rate between the years.

An important element to point out is that the z-scores (if multiplied by 10 and adding 100, as suggested in Booysen (2002), in order to get more visually manageable scores and to avoid negative scores) and AMPN under certain hypotheses might lead to very similar results. In fact in AMPN the reference value r_j and the range $M_j^* - m_j^*$ are used, while in the computation of z-scores these are replaced by the mean $\mu_j^{t_b}$ and the standard deviation $\sigma_j^{t_b}$ (both computed across statistical units in the base year). If we suppose that $r_j \approx \mu_j^{t_b}$ and since the range can be usually estimated as six times the standard deviation (Hozo et al., 2005), then comparing the formulas for z-scores and AMPN we see that they are essentially the same. The main difference is that for AMPN the range is computed across all statistical units and all time periods, while for z-scores the standard deviation (and the mean) is computed only at the reference time.

Finally, one of main drawback of the design of AMPN, compared to other methods, is that it aims at dealing with both the spatial and temporal dimension (minima and maxima are computed on all time periods and for all Italian regions) but with a reference value (a spatial average) chosen for a given year. This two ways of considering the time

[6] If m_j^* and M_j^* are the overall minimum and maximum of the indicator j across all units and all time periods, define the goalpost $\widetilde{m}_j^*$ as $r_j - \frac{M_j^* - m_j^*}{2}$ and the goalpost $\widetilde{M}_j^*$ as $r_j + \frac{M_j^* - m_j^*}{2}$. Then in Eq. (3) use $\widetilde{m}_j^*$ and $\widetilde{M}_j^*$ instead of m_j^* and M_j^* to get the AMPN.

[7] If the indicator has negative polarity (i.e., the indicator is negatively correlated with the phenomenon to be measured) then the normalization is done in the same way except that the complement to $a + b = 200$ is considered.

[8] For example: in Istat (2017) the indicator *SIC4-Robbery rate* is normalized between 48.8 and 108.8; the indicator *AMB8-Urban green* is normalized between 97.5 and 157.5. The two numbers 97.5 and 108.8 are both around 100, but the first is a minimum and the second is a maximum. Actually in certain cases 70 or 130 won't ever be reached: as an example consider the negatively polarized raw indicator *AMB5-6-Quality of urban air*, which can attain values from 0 to 100. Since the reference value (UNDP, 2010) is 60 then the minimum of the normalized indicator is at least 76, and 70 will never occur.

[9] The situation is similar also with the indicator *SIC1-Homicide rate*. Between 2005 and 2006 the raw indicator improves by 20%, while the normalized one just by 3.2%. Between 2015 and 2016 the raw indicator improves by 14.3%, while the normalized one just by 1.5%.

Table 3
Three individual indicators: PAE3, PAE10, AMB14. Overall minimum and maximum (for the time span 2010–2018 and all the Italian regions), reference value (UNDP, 2010), raw and normalized values (standard AMPN) for Campania in 2017. Data from Istat (2019).

Indicator	Overall Min	Overall Max	Ref. value UNDP (2010))	Campania, 2017 Raw	Normalized
PAE3	2	68.4	12.2	67.8	49.76
PAE10	4.9	36.1	18.3	34.7	68.46
AMB14	12.2	37.5	21.6	35.3	132.49

dimension could conflict with each other, leading to unintended effects (see Section 5.1) while they do not address the issue of preservation of growth rates.

4.4. Monotonicity

According to Mauro et al. (2018, p. 80, Section 3), an aggregation function should satisfy three desirable properties: (strict) monotonicity, continuity and heterogeneity penalization. As it is easy to show, AMPI satisfies the proprieties of continuity and heterogeneity penalization: AMPI is continuos because is simply made up of (algebraic) sums and products of continuous functions and AMPI penalizes heterogeneity because, fixing the mean, the penalty increases with the variance. However monotonicity is satisfied only under specific conditions. In this regard, as a novelty in the literature on AMPI, we provide also the theoretical conditions under which monotonicity is satisfied.

We prove the following (for the proof see App. A):

Theorem 1. *If a set of (normalized) indicators have values in the range* $[a, b]$, *with* $a > 0$ *and* $b > a$ *then AMPI is certainly monotone if and only if* $b \leqslant \frac{3}{2}a$.

This condition could be very useful in particular for studies that aims at extending the AMPI approach in order to get new measures to capture inequalities across territories (Ciommi et al., 2017).

The theorem means that in principle, and when aggregating a suitable number of indicators, AMPI might not be monotone even in the range [70, 130].[10] But since AMPN-values can very well fall outside of this interval (see Section 4.3), especially if the goalposts are not updated every year, then the monotonicity of AMPI could be more easily undermined because it strongly depends on the range of the normalized indicators – the larger the worse. Therefore we can conclude that in certain cases the penalty inherent in AMPI is too excessive.

It is important to point out that no such monotonicity breaches ever occurred in the annual *Rapporto Bes* (Istat, 2015,…, Istat, 2019). But this does not mean that it cannot happen in the future.

Anyway we can provide an example based on actual data from Istat (2019) mixing together three ecological indicators taken from two Bes domains: *Landscape and cultural heritage* and *Environment*. Consider the individual indicators

1. PAE3 – Illegal building rate (per 100 building permits issued)
2. PAE10 – People that are not satisfied with the quality of landscape of the place where they live (percentage values)
3. AMB14 – Protected natural areas (percentage values)

for the region Campania in 2017. See Table 3. The first two indicators, taken from the domain *Landscape and cultural heritage* are negatively polarized (the lower, the better) while the third indicator, taken from *Environment*, is positively polarized (the higher, the better).

The AMPI aggregation for the three indicators amounts to 68.55, but any improvement in AMB14 makes AMPI go down. For example if the situation for Campania in 2017 were slightly better (e.g., with PAE3 and PAE10 the same as in Table 3 but AMB14 equals to 37.5, like Abruzzo in 2015) then AMPI would equal to 68.53, showing a marginal decrease.[11] With respect to Thm. 1, in this example the ratio b/a is approximately 2.77, hence above the 3/2 of the statement. Fig. A.2.

[10] We refer to Appendix A for an exhaustive discussion on monotonicity.

5. The interplay between normalization, aggregation and time

When we consider the evolution of composite indices drawn up from more than one individual indicator observed over several years, the effects due to the normalization process interact with the aggregation procedure. In this respect time represents a new dimension for the analysis with expected impacts also on the unbalance adjustment.

Time dimension is also a key characteristics when we use composite indicators for policy evaluation (see for example Ministero dell'Economia e delle Finanze, 2018), when the impact of the single policy over time needs to be strongly related to the movement of the indicator/composite index more than on its construction properties.

An example could help explaining this point. Suppose that the eight individual indicators for *Environment* are observed for the statistical unit i at time t. Recall that the $x's$ are the raw values, while the $z's$ are the normalized ones. If the aggregation scheme is the arithmetic mean the composite index will be:

$$Env_{i,t} = \frac{z_{i1}^{t} + z_{i2}^{t} + \cdots + z_{i8}^{t}}{8} \tag{5}$$

Now suppose to call *shock* the impulse, a real number expressed in percentage points, impressed on the raw indicator x_{i8} from t to t':

$$x_{i8}^{t'} = x_{i8}^{t}\cdot(1 + shock) \tag{6}$$

Considering the normalization based on index numbers[12] we obtain

$$z_{i8}^{t'} - z_{i8}^{t} = 100\frac{x_{i8}^{t'}}{r_8} - 100\frac{x_{i8}^{t}}{r_8} = shock\cdot z_{i8}^{t} \tag{7}$$

where r_8 represents the reference value of the indicator in the base year. Then, supposing the other indicators assume the same values between t and t' the effect on the composite index Env is:

$$Env_{i,t'} - Env_{i,t} = \frac{z_{i8}^{t'} - z_{i8}^{t}}{8} = \frac{shock\cdot z_{i8}^{t}}{8} = \frac{shock}{8}\cdot z_{i8}^{t} \tag{8}$$

Along this scheme it is feasible to relate a change in time on a single indicator to the change on the composite one interpreting, for example, the shock as the effect obtained by the policy.

An analogous interpretation holds even if the aggregation is performed using the geometric mean. But it does not hold for other normalization techniques (z-scores, Min-Max, AMPN) or aggregation methods (AMPI).

5.1. The effects of normalization on the measurement of unbalances

As we have seen before, AMPN requires that the minimum and the maximum are computed using the data available for each indicator in the selected time span, considering at the same time all the regions and

[11] For a thorough discussion on these "problematic" values refer to App. A and Fig. A.2.

[12] The *index numbers* method, or distance-to-a-reference method, is a well-known linear transformation that preserves ratios both in the spatial and temporal dimension (and hence also growth rates), see OECD and JRC, 2008, p. 85, subSection 5.4. Given a reference value r_j, the one that we want to normalize to 100, the normalized value z_{ij}^{t} of x_{ij}^{t} is given by $z_{ij}^{t} = 100\frac{x_{ij}^{t}}{r_j}$, where $r_j = r_j^{t_b}$ represents the reference value at the base year t_b. If the indicator has negative polarity then $\frac{r_j}{x_{ij}^{t}}$ is considered.

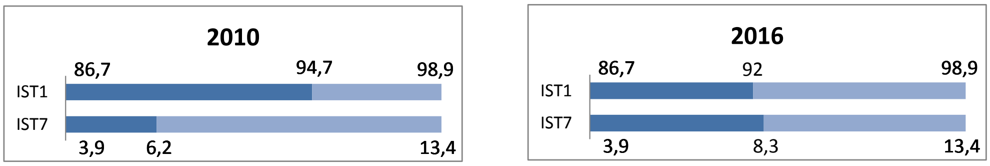

Fig. 4. Raw indicators IST1 and IST7. Relative positions of the indicators with respect to their own minima and maxima. Italy. 2010 and 2016. Data from Istat (2017).

Table 4
Rankings of Italian regions and geographical breakdowns according to *Politics and institutions*. Normalization: AMPN, reference: Italy 2010 and Italy 2015. Aggregation: AMPI. Year 2016. Data from Istat (2017).

Reference 2010	Reference 2015	Differences between rankings
Bolzano	Bolzano	=
Trentino	Trentino	=
Campania	Trento	−3
Trento	Emilia-Romagna	1
Emilia-Romagna	Piemonte	1
Piemonte	Campania	1
Liguria	Toscana	−2
Toscana	Valle d'Aosta	1
Lazio	Liguria	−1
Valle d'Aosta	Lazio	2
Center	*Center*	=
Friuli	*North*	−1
Italy	Friuli	−2
North	Marche	2
Sicilia	*Italy*	−1
Marche	Sicilia	2
Lombardia	Lombardia	=
South	Umbria	−1
Umbria	*South*	1
Calabria	Abruzzo	−1
Abruzzo	Calabria	1
Puglia	Puglia	=
Molise	Molise	=
Sardegna	Veneto	−1
Veneto	Sardegna	1
Basilicata	Basilicata	=

Italy. Then the value for Italy in 2010 is set equal to 100. Because of this latter rule, as a byproduct, the indicators for Italy in 2010 are assumed to be balanced, regardless of their actual values. In other words, this framework introduces an artificial notion of equilibrium based on the alleged harmony of all the Italian indicators in the reference year, and the aggregation would penalize the distance from that artificial equilibrium and not a disequilibrium measured against minima and maxima or against average values (like *Quality of life Spain 2016*) or defined a priori in some ways (like HDI, with its "natural zeroes" and "aspirational targets"). Therefore it is hard to justify the application of AMPI, whose main purpose is to penalize disequilibrium among indicators: if we aggregate two indicators for Italy, one already at its best in 2010, and stable over time, and another one that steadily improves from 2010 onwards, then AMPI would impose more and more burdensome penalties at the composite index as time goes by.

As an actual example consider the raw individual indicators *IST1-Participation in the school system of children aged 4–5* and *IST7-Participation in life-long learning* for Italy from the domain *Education and training* (Istat, 2017). See Fig. 4. IST1, among all years and all regions, spans between the minimum of 86.7 and the maximum of 98.9, while IST7 varies between the minimum of 3.9 and the maximum of 13.4. In 2010 IST1 attains 94.7 for Italy, a value nearer to the maximum than to the minimum. In 2016 IST1 goes down to 92, approaching the mid-range 92.8. In 2010 IST7 is 6.2 for Italy, a value nearer to the minimum than to the maximum, while in 2016 it goes up to 8.3, approaching the mid-range also in this case.

If we use the normalization procedure AMPN, both the indicators for Italy are set to 100 in 2010. The normalized IST1 goes from 100 in

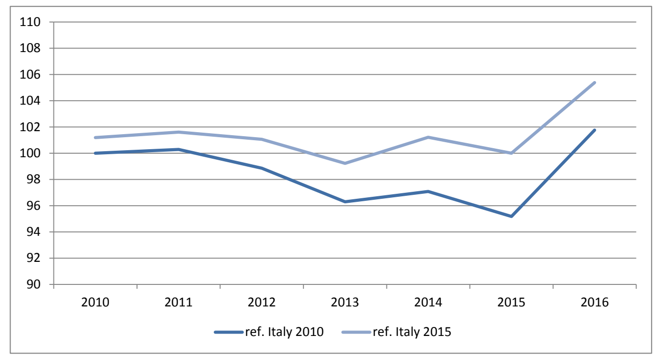

Fig. 5. Composite index *Politics and institutions*. Normalization: AMPN, reference: Italy 2010 and Italy 2015. Aggregation: AMPI. Italy. 2010–2016. Data from Istat (2017).

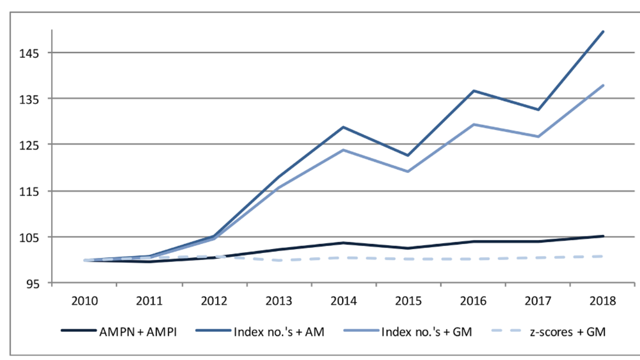

Fig. 6. Composite indices *Environment* computed with different normalization and aggregation techniques. Italy. Years 2010–2018. AMPN and AMPI; index numbers and arithmetic mean (AM); index numbers and geometric mean (GM); z-scores and geometric mean (GM). Data from Istat (2019).

2010 to 86.7 in 2016, while the normalized IST7 goes from 100 in 2010 to 113.3 in 2016.

Aggregating the two normalized indicators with the arithmetic mean we have 100 for both years. Applying geometric mean or AMPI we obtain 100 in the base year (2010) while in 2016 the aggregation of IST1 and IST7 goes down to 99.1 with the geometric mean and 98.2 with AMPI. AMPI in particular imposes a penalty of 1.8 in 2016.

However if we consider the penalty as a price to pay for unbalance (measured against minima and maxima), IST1 and IST7 are much more balanced in 2016 than in 2010, and the penalty looks unjustified.

5.1.1. *Impact on rankings*

The way in which the penalty is considered could have an impact also on spatial rankings (see for example Greco et al., 2019) or on temporal trends. This is true even if we make the simple hypothesis of a change of the base year.

As an example consider the composite index *Politics and institutions*. This index is the synthesis of seven individual indicators (see Istat, 2017, ch. 6 and App. A). In Table 4 Italian regions and geographical breakdowns (North, Center, South and Italy) are ranked, from the best to the worst, according to the values of the composite index for 2016. The index has been computed through AMPN normalization and AMPI

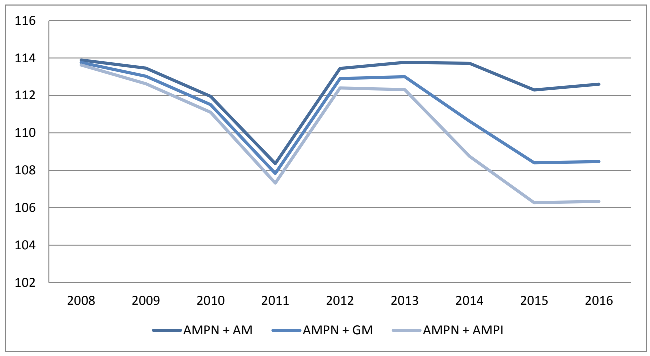

Fig. 7. Composite index *Education and training*. Normalization: usual AMPN between 70 and 130. Aggregation: arithmetic mean AM, geometric mean GM, AMPI. Lazio. 2008–2016. Data from Istat (2017).

aggregation. In the first column the reference value for AMPN is set to Italy 2010, in the second column it is set to Italy 2015. The third column shows the differences between rankings, i.e., the shift of positions if we change the reference year from 2010 to 2015. As we can see there are many discrepancies between the two lists, in particular Campania goes down three positions. Setting 2010 as the base year, from 2010 to 2016 Campania gains fifteen positions in the rankings, while it gains only six positions if we use 2015 as the base year.

This suggests that a change of reference might also affect the time profile of the composite index. Consider again the composite index *Politics and institutions* for Italy, from 2010 to 2016. If we choose 2010 as the reference year then the index drops by 1.2% from 2013 to 2015, while if the reference year is set to 2015 the index grows by 0.8% in the same period (see Fig. 5).[13]

5.2. Aggregation and compensation

To illustrate the performance of AMPI along time we consider again the Italian well-being, particularly the composite index *Environment* for the period 2010–2018, that is the combination of eight individual indicators: *AMB3-Water losses in urban supply system*, *AMB4-Landfill of waste*, *AMB5-6-Quality of urban air*, *AMB8-Urban green*, *AMB9-Satisfaction for the environment*, *AMB14-Protected natural areas*, *AMB16-Electricity from renewable sources*, *AMB17-Separate collection of municipal waste* (see Istat, 2019, ch. 10). The first three indicators are negatively polarized.

We computed composite indices using different normalization and aggregation techniques: AMPN and AMPI, index numbers and arithmetic mean, index numbers and geometric mean, z-scores and geometric mean (Fig. 6).

The arithmetic and geometric mean do not return very different pictures when they are applied to the same normalization framework, i.e. index numbers, while differences in the normalization method provide quite different results even if the aggregation scheme is the same (geometric mean in the example).

In particular between 2010 and 2018 *AMPN + AMPI* improves only by 5.2%, while index numbers + arithmetic mean improves by 49.6%. In fact AMPN normalization, as z-scores, modifies the intensity of the growth rates. So while AMB16 and AMB17 improve by more than 50% and AMB4 and AMB5-6 improve by more than 100%, after the AMPN normalization they improve no more than 25% and AMB16 increases just by 2.3%. AMB16 has an overall high coefficient of variation, but AMPN greatly reduces its variability and hence its weight in the aggregation, slowing down the AMPI composite index. Using index numbers and geometric (or arithmetic) mean the sharp increases in most of the indicators are reflected in the composite index.[14]

This example clearly shows the trade-off amid the control of variability and the preservation of the annual growth rate.

Moreover, AMPI does not provide a useful framework to evaluate the difference in the value of a composite index between two points in time, partly due to the chosen normalization (see Section 4.3) and partly due to the not-so-clear evolution of the penalty over time, that, as already discussed, can become too harsh (see Section 4.4) or not justified (see Section 5.1).

For example, consider the composite index *Education and training* for the region Lazio and for the period 2008–2016. (Istat, 2017, ch. 2). We compare three different composite indices, the traditional AMPI with two pseudo-indices obtained as a composition of the AMPN and arithmetic and geometric mean (see Fig. 7).

AMPI penalty reaches 6.3 points in 2016 (approx. 5.5% of the arithmetic mean) and this results in a very different picture from the one obtained with arithmetic or geometric mean. In particular between 2011 and 2016 four out of the five individual indicators grow, and both arithmetic and geometric mean increase, while AMPI gets worse.[15]

In general, considering all the composite indices computed for *Rapporto Bes* (for all domains, years and regions) it can be seen that the AMPI penalty is always approximately twice the implicit penalization inherent in the geometric mean (i.e., the difference between arithmetic and geometric mean). More precisely, in *Rapporto Bes 2019* (Istat, 2019), the ratio between the AMPI penalty and the geometric mean penalization, considered for all composite indices, has a mean of 1.98 and a cv of 0.05; the 95% of the values are between 1.75 and 2.15; the minimum is 1.38, and the maximum is 2.33.

6. Conclusions

In the last years there has been a growing number of frameworks to measure well-being at national and international level. At the same time well-being indicators are now related, with different national intensities, to the policy agenda. These elements have reinforced the attention on the research programs for the elaboration of composite indices.

Due to this pressure, choices for normalization, aggregation, unbalance adjustments and their interplay need clarity both in the

[13] It is important to remember that a normalization based on index numbers combined with an aggregation based on the geometric mean would not be affected by this kind of problems.

[14] Several examples could be provided to argue about the impact of differences in the normalization procedure. One of them is the composite on *Health* where, with AMPN the *SAL1-Life expectancy at birth* acquires more variability over time compared to the normalization based on index numbers.

[15] Similar considerations can be made also for other domains of well-being: for example in 2013 the AMPI composite index *Safety* for Puglia (Istat, 2017) has a penalty greater than the 7% of the arithmetic mean.

definition as well as in the communication phase, that is crucial when we are interested in the impact of a specific policy on well-being (see Ministero dell'Economia e delle Finanze, 2018).

This paper has aimed to shed some lights on the complexity of these topics especially when the time dimension is introduced. Together with an overview of the main issues that need to be addressed, looking more specifically at the Italian well-being experience we have pointed out some drawbacks of the method currently adopted, the so-called AMPI.

First of all we have contributed to the literature deriving the theoretical conditions under which AMPI is actually monotone. This condition could be very useful particularly for studies that want to extend the AMPI approach, including measures to capture inequalities across territories (Ciommi et al., 2017). Then we have illustrated the critical characteristics of the interaction between normalization and aggregation using a compensatory approach such as AMPI.

In particular, in the normalization stage, the search for the minimum and maximum of each indicator, to derive the goalposts, is performed along all the regions and all time periods but then a constraint on the so-called base year for Italy is introduced. These two ways of considering the time dimension have a significant impact on the unbalance adjustment scheme selected.

Although AMPI has the relevant characteristic of explicitly tackling the unbalance problem, we come to the conclusion that there could be the risk of being in a situation "when the unbalance problem is tackled, (but) the normative aspects deriving from the different methods used are usually not explained satisfactorily. In general this lack of clearness regards weighting procedures, but correspondingly it often concerns the entire aggregation methods linked to the balance problem" (Casadio Tarabusi and Guarini, 2013, p. 21). Facing these challenges inside the AMPI paradigm could be an important step forward considering also the agenda proposed by Greco et al. (2019) for the spectrum of non-compensatory approaches.

Finally we would like to point out how the time comparison will be an important task for the research agenda on well-being. This issue, associated to the availability of longer time series, will require a fruitful contamination with econometric methods. Cointegration, panel data, dynamic factors models could represent emerging methodological issues for well-being. To our knowledge some papers along these lines are already available (see for example Iglesias et al., 2017 and Bacchini et al., 2019) but more work remains to be done.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We are grateful to Rita De Carli, Marco Fattore, Roberto Monducci, Miria Savioli, Maria Pia Sorvillo, Alessandra Tinto and the anonymous reviewer for helpful comments and suggestions. We are grateful for the support from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 822781 GROWINPRO.

Appendix A. Monotonicity

The aggregation function AMPI has many mathematical important and desirable properties (see Bullen, 2013, ch. II, Section 1): it is continuous, homogeneous, reflexive and symmetric. In this section we study monotonicity. We recall that an aggregation function A is called *monotone* if

$$\forall\ i \in \{1,\ \dots, m\}\ z_i \leqslant s_i \Rightarrow A(z_1 \dots z_m) \leqslant A(s_1,\ \dots, s_m). \tag{9}$$

The property (9) is essential for the ease of interpretation of the aggregated values, in fact if $z_1, \dots, z_m$ and $s_1, \dots, s_m$ are the normalized values of m indicators for two different statistical units (regions) or two different time periods (years), the fact that $A(z_1 \dots z_m) > A(s_1, \dots, s_m)$ suggests that, at least for one of the individual indicators, the s-region does worse than the z-region or that the z-region has deteriorated from one year to the other. The (9) property is listed in Sen (1976, p. 219), Beliakov et al. (2007, p. 4, def. 1.5), OECD and JRC (2008, p. 105), Casadio Tarabusi and Guarini (2013, p. 25), Prop. (i), Mauro et al. (2018), p. 81, subSection 3.1 as one of the elementary requirements for a good (composite) index.[16]

In order to study under which circumstances AMPI is monotone[17] first of all we consider an equivalent formulation of Eq. (2). Let $z_1, \dots, z_m$ be the normalized values of m indicators for a given unit i and a given time t, then,

$$\mathrm{AMPI}\left(z_1,\ \dots, z_m\right) = \mathrm{AM}\left(z_1,\ \dots, z_m\right) - \frac{\mathrm{VAR}(z_1, \dots, z_m)}{\mathrm{AM}(z_1, \dots, z_m)}, \tag{A.1}$$

where $\mathrm{AM} = \frac{1}{m}\sum_{k=1}^{m} z_k$ is the arithmetic mean, $\mathrm{VAR} = \frac{1}{m}\sum_{k=1}^{m}(z_k - \mathrm{AM}(z_1, \dots, z_m))^2$ is the variance and the ratio VAR/AM is the penalty.

Secondly, we will use an inductive formula for its computation and then take its derivative. We use an incremental calculation of mean ad variance: see Finch (2009), Eqs. (6), (24), (25). Let $z_1, \dots, z_{m-1}$ and x be in $[a, b]$, with $a, b \in \mathbf{R}$, $a < b$. Let $\mu = \mathrm{AM}(z_1, \dots, z_{m-1})$ and $\sigma^2 = \mathrm{VAR}(z_1, \dots, z_{m-1})$. Suppose that $m \geqslant 2$ and $x \neq \mu(1 - m)$. Then

$$\mathrm{AMPI}\left(z_1,\ \dots, z_{m-1}, x\right) = \mu + \frac{1}{m}\left(x - \mu\right) - \frac{1}{m}\frac{\left(m-1\right)\sigma^2 + \left(x-\mu\right)\left(x - \mu - \frac{1}{m}(x-\mu)\right)}{\mu + \frac{1}{m}(x-\mu)}$$

The former equation can be re-written as

$$\mathrm{AMPI}\left(z_1,\ \dots, z_{m-1}, x\right) = \mu + \frac{1}{m}\left(x - \mu\right) - \frac{m-1}{m}\cdot\frac{\sigma^2 + \frac{1}{m}(x-\mu)^2}{\mu + \frac{1}{m}(x-\mu)} \tag{A.2}$$

[16] "[…] the motivation of our search for a new measure can be understood by noticing the violation of [monotonicity] by the poverty measures currently in wide use" (Sen, 1976)

[17] As everywhere in the rest of the paper, we analyze the AMPI aggregation function used in the *Rapporto Bes* (Istat, 2015), i.e., the "minus" version (see Eq. (2) and the related discussion). As for the "plus" version (Mazziotta and Pareto, 2016, p. 990, Eq. 4) Casadio Tarabusi and Guarini (2013) showed that is not monotone even if aggregating just two indicators "[the formula] has a specific disadvantage: the function does not fulfill positive monotonicity in a large portion of its domain."

Taking the first derivative in x we get

$$\frac{1}{m} - \frac{m-1}{m^2} \cdot \frac{2\left(x-\mu\right)\left(\mu + \frac{1}{m}(x-\mu)\right) - \sigma^2 - \frac{1}{m}(x-\mu)^2}{\left(\mu + \frac{1}{m}(x-\mu)\right)^2} \tag{A.3}$$

The derivative does not vanish if $m = 2$ and hence AMPI is monotone when it aggregates just two indicators (see also Mazziotta and Pareto, 2016, p. 992, subSection 3.4.2).

Let us assume that $m \geqslant 3$. If we further suppose that $a > 0$ (and hence $\mu > 0$, $x > 0$) we can take the non-negative root r of (A.3):

$$r = \mu\left(1-m\right) + \frac{\sqrt{m}\sqrt{m-1}}{\sqrt{m-2}} \cdot \sqrt{m\mu^2 + \sigma^2} \tag{A.4}$$

that can be conveniently re-written as

$$r = \mu\left(1 - m + \frac{\sqrt{m}\sqrt{m-1}\sqrt{m+c_v^2}}{\sqrt{m-2}}\right), \tag{A.5}$$

where c_v^2 is the square of the coefficient of variation, i.e., $c_v^2 = \frac{\sigma^2}{\mu^2}$. AMPI, as a function of x, is strictly increasing if $x \leqslant r$, while it is strictly decreasing if $x \geqslant r$.

Lemma A.1. *Let $m \geqslant 3$ and $a > 0$. Let r be as in (A.5). Then*

$$r \geqslant \frac{3}{2}\mu.$$

Proof.

$$\begin{aligned} r &\geqslant \mu\left(1 - m + \tfrac{\sqrt{m}\sqrt{m-1}\sqrt{m}}{\sqrt{m-2}}\right) \\ &= \mu\left(1 + \tfrac{m}{\sqrt{m-2}\,(\sqrt{m-1} + \sqrt{m-2})}\right) \\ &\geqslant \mu\left(1 + \tfrac{m}{2m}\right) = \tfrac{3}{2}\mu \end{aligned}$$

Theorem A.2. *Let $m \geqslant 3$ and let $z_1, \ldots, z_m \in [a, b]$, $s_1, \ldots, s_m \in [a, b]$, with $a, b \in \mathbf{R}$, $a > 0$, $b > a$ and $b \leqslant \frac{3}{2}a$. If $z_i \leqslant s_i$ for every i then*

$$\mathrm{AMPI}(z_1, \ldots, z_m) \leqslant \mathrm{AMPI}(s_1, \ldots, s_m).$$

Proof. As before let $\mu = \mathrm{AM}(z_1, \ldots, z_{m-1})$ and $\sigma^2 = \mathrm{VAR}(z_1, \ldots, z_{m-1})$.

$\mathrm{AMPI}(z_1, \ldots, z_{m-1}, x)$ is an increasing function in x when $x \in (0, r]$ with r defined as in (A.5). Since, by Lemma A.1, $r \geqslant \frac{3}{2}\mu \geqslant \frac{3}{2}a \geqslant b$ then

$$\mathrm{AMPI}(z_1, \ldots, z_{m-1}, z_m) \leqslant \mathrm{AMPI}(z_1, \ldots, z_{m-1}, s_m).$$

Since AMPI is a permutation-invariant function with respect to its arguments, the thesis easily follows. □

Theorem A.3. *Let $a > 0$ and $b > \frac{3}{2}a$. There exist $m \geqslant 3$, $z_1, \ldots, z_m \in [a, b]$, $s_1, \ldots, s_m \in [a, b]$ with $z_i \leqslant s_i$ for every i and such that*

$$\mathrm{AMPI}(z_1, \ldots, z_m) > \mathrm{AMPI}(s_1, \ldots, s_m).$$

Proof. Let $z_1, \ldots, z_{m-1} = a$, and so $\mu = \mathrm{AM}(z_1, \ldots, z_{m-1}) = a$, while $\sigma^2 = \mathrm{VAR}(z_1, \ldots, z_{m-1}) = 0$. Then by (A.5) $r = r(m) = a\left(1 - m + \frac{m\sqrt{m-1}}{\sqrt{m-2}}\right)$ and the function $\mathrm{AMPI}(z_1, \ldots, z_{m-1}, x)$ is strictly decreasing in x when $x \geqslant r$. Since

$$\begin{aligned} \frac{m\sqrt{m-1}}{\sqrt{m-2}} - m &= \frac{m(\sqrt{m-1} - \sqrt{m-2})}{\sqrt{m-2}} \\ &= \frac{m}{\sqrt{m-2}\,(\sqrt{m-1} + \sqrt{m-2})} \\ &= \frac{1}{\sqrt{1-2/m}\,(\sqrt{1-1/m} + \sqrt{1-2/m})}, \end{aligned}$$

then $\lim_{m\to\infty} \frac{m\sqrt{m-1}}{\sqrt{m-2}} - m = \frac{1}{2}$ and so $\lim_{m\to\infty} r(m) = \frac{3}{2}a < b$. This means that for m sufficiently large we can suppose that $r(m) < b$. Let $s_i = z_i$ for every $i \in \{1, \ldots, m-1\}$. Let $z_m = r(m)$ and $s_m = b$. The conditions on the z's and the s's, as stated in the theorem, are satisfied, and moreover by construction

$$\mathrm{AMPI}(z_1, \ldots, z_m) > \mathrm{AMPI}(s_1, \ldots, s_m). \quad \square$$

Theorem A.2 and Theorem A.3 together make

Theorem 1. *If a set of (normalized) indicators have values in the range $[a, b]$, with $a > 0$ and $b > a$ then AMPI is certainly monotone if and only if $b \leqslant \frac{3}{2}a$.*

Table A.1 shows some examples in which AMPI is not monotone. As we can see the intensity of the deviations from the property of monotonicity depends on the variability range of the indicators (namely the ratio b/a), i.e., on the normalization method adopted before aggregating individual indicators. This means that the aggregation method AMPI should be accompanied only by few, selected choices of normalization ranges, otherwise the aggregation could lead to far more incisive counterexamples to monotonicity and might even lead to a composite index that does not enjoy the

Table A.1
Normalized individual indicators and composite index AMPI. Theoretical examples of the violation of the property of monotonicity for different ranges $[a, b]$. Absolute values.

Range of indicators	m = No. of indicators	b/a	Values of indicators ($z_i \leqslant s_i$)		AMPI ($z > s$)	
			z_i	s_i	z	s
$a = 40$, $b = 130$	3	3.3	40 40 90	40 40 130	46.86	44.29
$a = 40$, $b = 160$	3	4.0	40 40 90	40 40 160	46.86	40.00
$a = 70$, $b = 130$	5	1.9	70 ··· 70 124	70 ··· 70 130	75.03	74.98
$a = 70$, $b = 120$	10	1.7	70 ··· 70 112	70 ··· 70 120	71.84	71.77
$a = 70$, $b = 130$	10	1.9	70 ··· 70 112	70 ··· 70 130	71.84	71.51
$a = 0.1$, $b = 0.9$	3	9.0	0.1 0.1 0.2	0.11 0.12 0.9	0.12	0.01
$a = 0.1$, $b = 0.9$	3	9.0	0.1 0.1 0.2	0.1 0.1 0.9	0.12	−0.02
$a = 1$, $b = 10$	3	10.0	1 1 1	1 1 8	1.00	0.07
$a = 1$, $b = 10$	3	10.0	1 1 1	1 2 10	1.00	0.59
$a = 1$, $b = 10$	3	10.0	1 1 8	1 1 10	0.07	−0.50

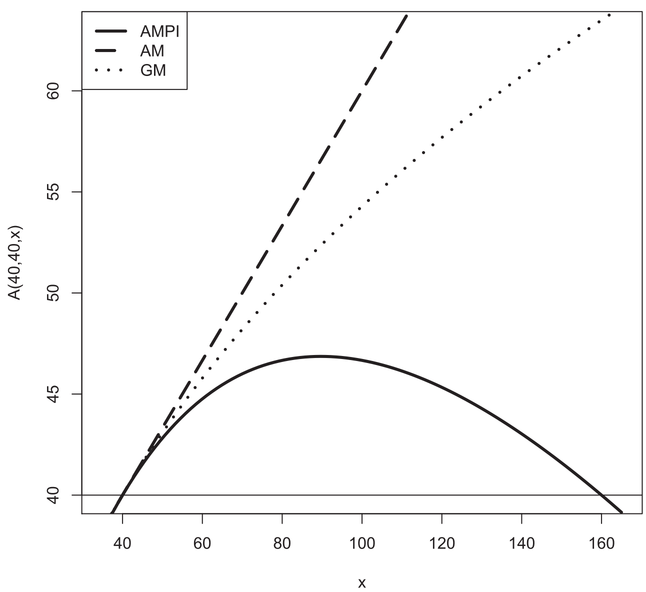

Fig. A.1. Value of the composite indices $A(40, 40, x)$, for x between 35 and 165 and A = AMPI, arithmetic mean AM or geometric mean GM.

usual property of internality (cf. Bullen, 2013, ch. II, Section 1), i.e., it may fall out of the range of the normalized indicators and it can even become negative.

Anyway, even if AMPI is accompanied by the usual AMPN-normalization, monotonicity might not be satisfied. First of all because contrary to what is stated in Mazziotta and Pareto (2016, p. 993) we are still in trouble (if $m \geqslant 5$) even when we restrict our attention to indicators between 70 and 130 (Table A.1, lines 3–5). But this is even more true if we consider that AMPN can in principle return values between 40 and 160 or also in a larger range if the minimum and maximum are not computed for the whole time span (see Section 4.3).

In Fig. A.1 the value of the aggregation of 40, 40, x is shown, with x that goes from 35 to 165 and the aggregation made by AMPI, arithmetic mean AM or geometric mean GM. The curve in bold is a hyperbola with a local maximum at 89.71.

To stress again the fact that the AMPI behavior depends on the chosen range for AMPN, and that this range must be rather rigid, we give an example of the violation of monotonicity based on Istat (2017).

Consider the Italian composite index *Politics and institutions*. This index is the aggregation of seven individual indicators (see Istat, 2017, ch. 6):

1. POL2 – Trust in the parliament
2. POL3 – Trust in the judicial system
3. POL4 – Trust in political parties
4. POL5 – Trust in police and fire brigade
5. POL7 – Women and political representation at regional level
6. POL11 – Length of civil proceedings
7. POL12 – Prison density

The indicators POL11 and POL12 are negatively polarized, all the others are positively polarized. Every indicator is available from 2011 or 2012, except POL12 that is available from 2010. All the missing data for 2010 have been interpolated, and in Istat (2017) the composite index has been released for the period 2010–2016. All the indicators are normalized by the procedure described in Section 4.3. However if we just changed the range

Table A.2
Individual indicators (raw and normalized values) and composite index AMPI for *Politics and institutions*. AMPN-normalization between $a = 0$ and $b = 100$, with reference value set to 50. Sardinia. 2010 and 2011. Absolute values and percentage variations. Data from Istat (2017).

	2010		2011		Variation %	
Indicator	**raw**	**normalized**	**raw**	**normalized**	**raw**	**normalized**
POL2	3.1	31.3	3.1	31.3	0.0	0.0
POL3	4.9	68.8	4.9	68.8	0.0	0.0
POL4	2.2	26.5	2.2	26.5	0.0	0.0
POL5	7.2	41.7	7.2	41.7	0.0	0.0
POL7	10.0	41.9	10.0	41.9	0.0	0.0
POL11	473.0	48.6	473.0	48.6	0.0	0.0
POL12	112.5	78.6	106.0	83.4	6.1	6.1
Composite AMPI		41.7		41.6		−0.3

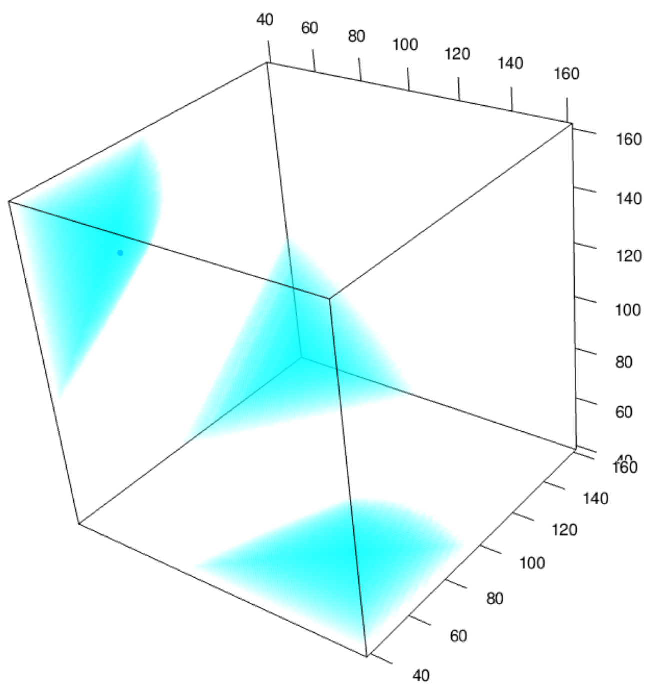

Fig. A.2. Problematic points (cyan) for the AMPI aggregation of $m = 3$ normalized indicators, each one in the range [40, 160]. The blue point is (49.76, 68.46, 132.49) of Table 3.

of the normalization (from 70–130 to the more common 0–100, and reference value set to 50) then for the Sardinia region we would have a decrease of the composite index by 0.3% between 2010 and 2011, while every individual indicator has remained constant, except POL12 (positively polarized) that has increased by 6.1%. See Table A.2.

Since the AMPI aggregation function is everywhere differentiable, property (9) is equivalent to asking for a non-negative gradient of AMPI everywhere, i.e.,

$$\frac{\partial \text{AMPI}}{\partial z_1} \geqslant 0, \ \ldots, \frac{\partial \text{AMPI}}{\partial z_m} \geqslant 0$$

(see Casadio Tarabusi and Guarini, 2013, p. 25). In other words, points $(z_1, \ldots, z_m)$ at which at least one of the partial derivatives is negative could be labelled as “problematic”, because the penalty is deemed too harsh.

Considering the actual AMPN range [40, 160], and $m = 3, 5, 7, 10, 15, 30, 50$ the problematic points represents approximately the 7.5%, 17.4%, 24.2%, 31.2%, 38.7%, 51.0%, 59.8%[18] (respectively) of the total volume of the hypercube $[40, 160]^m$. See Fig. A.2.

We see that the proportion of the problematic points increase with the number of aggregated indicators (or also by expanding the normalization range).

In the annual *Rapporto Bes* (Istat, 2015,…, Istat, 2019) problematic points have never occurred (yet). But we can give examples based on actual data from Istat (2019) either by taking a real Bes domain but mixing data from different regions/years or by aggregating individual indicators from different Bes domains, like we saw in Sect. 4.4.

[18] ± 0.25% with a confidence level of 95%. These proportions have been computed with a classical Monte Carlo method.

References

Alkire, S., 2002. Dimensions of human development. World Develop. 30 (2), 181–205.
Andreoni, V., Galmarini, S., 2015. Mapping the distribution of Well-Being in Europe beyond national borders. Science and Policy Report by the Joint Research Centre.
Bacchini, F., Baldazzi, B., De Carli, R., Di Biagio, L., Savioli, M., Sorvillo, M.P., Tinto, A., 2020. The evolution of the Italian framework to measure well-being. Under revision — Preliminary version: GrowInPro Working Paper, 21.http://www.growinpro.eu/the-italian-framework-to-measure-well-being-towards-the-2-0-version/.
Bacchini, F., Benedetti, I., Laureti, T., 2019. Makswell: An EU project on MAKing Spatial and temporal multivariate analysis of WELL-being: an application to the italian equitable and sustainable WELL-being. Technical report, Paper presented at ISI World Statistics Congresses.
Bacchini, F., Calza, M.G., Gandolfo, M., Sorvillo, M.P., Tinto, A., 2018. Makswell: An EU project on MAKing Sustainable development and WELL-being frameworks work for policy. Technical report, Paper presented at ISQOLS 2018.
Becker, W., Saisana, M., Paruolo, P., Vandecasteele, I., 2017. Weights and importance in composite indicators: Closing the gap. Ecolog. Indicators 80, 12–22.
Beliakov, G., Pradera, A., Calvo, T., 2007. Aggregation functions: a guide for practitioners, vol. 221 Springer.
Bertelsmann Stiftung and Sustainable Development Solutions Network, 2019. SDG Index and Dashboards Report 2019. Transformation to achieve the Sustainable Development Goals.
Booysen, F., 2002. An overview and evaluation of composite indices of development. Social Indicators Research 59 (2), 115–151.
Bullen, P.S., 2013. Handbook of means and their inequalities, vol. 560 Springer Science & Business Media.
Calcagnini, G., Perugini, F., 2019. A well-being indicator for the italian provinces. Soc. Indic. Res. 142, 149–177.
Casadio Tarabusi, E., Guarini, G., 2013. An unbalance adjustment method for development indicators. Soc. Ind. Res. 112, 19–45.
Ciommi, M., Gigliarano, C., Emili, A., Taralli, S., Chelli, F.M., 2017. A new class of composite indicators for measuring well-being at the local level: an application to the Equitable and Sustainable well-being (BES) of the italian provinces. Ecolog. Indicators 76, 281–296.
Decancq, K., Lugo, M.A., 2013. Weights in multidimensional indices of wellbeing: an overview. Econ. Rev. 32 (1), 7–34.
Eurostat, 2006. Methodology of short-term business statistics. Eurostat.http://ec.europa.eu/eurostat/documents/1916593/1917176/KS-BG-06-001-EN.pdf/973bf1c8-e598-4228-9d18-2f7704a95e45.
Eurostat, 2018. Quality of life indicators. Eurostat.http://ec.europa.eu/eurostat/statistics-explained/index.php/Quality_of_life_indicators.
Finch, T., 2009. Incremental calculation of weighted mean and variance. University of Cambridge 4 (11–5), 41–42 https://fanf2.user.srcf.net/hermes/doc/antiforgery/stats.pdf.
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.
Hozo, S.P., Djulbegovic, B., Hozo, I., 2005. Estimating the mean and variance from the median, range, and the size of a sample. BMC Med. Res. Methodol. 5 (1), 13.
Iglesias, K., Gazareth, P., Suter, C., 2017. Explaining the decline in subjective well-being over time in panel data. In: Metrics of Subjective Well-Being: Limits and Improvements. Springer, pp. 85–105.
INE – Portugal, 2017. Índice de bem estar. INE.https://www.ine.pt/ngt_server/attachfileu.jsp?look_parentBoui=313012429&att_display=n&att_download=y.
INE – Spain, 2019. Quality of life indicators. INE.https://www.ine.es/ss/Satellite?L=1&c=INEPublicacion_C&cid=1259937499084&p=1254735110672&pagename=ProductosYServicios%2FPYSLayout&param1=PYSDetalleGratuitas.
Istat, 2015. Rapporto Bes: Il benessere equo e sostenibile in Italia. Istat.http://www.istat.it/it/archivio/175169.
Istat, 2016. Rapporto Bes: Il benessere equo e sostenibile in Italia. Istat.http://www.istat.it/it/archivio/1940299.
Istat, 2017. Rapporto Bes: Il benessere equo e sostenibile in Italia. Istat.http://www.istat.it/it/archivio/207259.
Istat, 2018. Rapporto Bes: Il benessere equo e sostenibile in Italia. Istat.http://www.istat.it/it/archivio/207259.
Istat, 2019. Rapporto Bes: Il benessere equo e sostenibile in Italia. Istat.https://www.istat.it/it/archivio/236714.
Mauro, V., Biggeri, M., Maggino, F., 2018. Measuring and monitoring poverty and well-being: a new approach for the synthesis of multidimensionality. Soc. Indic. Res. 135 (1), 75–89.
Mazziotta, M., Pareto, A., 2016. On a generalized non-compensatory composite index for measuring socio-economic phenomena. Soc. Indic. Res. 127 (3), 983–1003. URL: http://link.springer.com/article/10.1007/s11205-015-0998-2.
Ministero dell'Economia e delle Finanze, 2018. Documento di Economia e Finanza 2018 – Allegato: Indicatori di benessere equo e sostenibile. Technical report, MEF.http://www.dt.mef.gov.it/modules/documenti_it/analisi_progammazione/documenti_programmatici/def_2018/Allegato_6_-_Indicatori_di_benessere_equo_e_sostenibile.pdf.
OECD, 2017. Better Life Index. OECD.http://www.oecdbetterlifeindex.org/.
OECD, JRC, 2008. Handbook on constructing composite indicators: methodology and user guide. OECD.http://www.oecd.org/sdd/42495745.pdf.
Sabbadini, L.L., Maggino, F., 2018. Quality of life in Italian official surveys. Soc. Indic. Res. 135 (3), 1043–1055.
Sen, A., 1976. Poverty: an ordinal approach to measurement. Econometrica: J. Econ. Soc. 219–231.
Stiglitz, J., Fitoussi, J.-P., Durand, M., 2018. Beyond GDP: Measuring What Counts for Economic and Social Performance. OECD Publishing.
Stiglitz, J., Sen, A., Fitoussi, J.-P., et al., 2009. The measurement of economic performance and social progress revisited. Reflections and overview. Commission on the Measurement of Economic Performance and Social Progress, Paris.
Tarantola, S., 2008. European Innovation Scoreboard: strategies to measure country progress over time. Joint Research Centre, mimeo.
Tinto, A., Bacchini, F., Baldazzi, B., Ferruzza, A., Van den Brakel, J., Willems, R., Rosenski, N., Zimmermann, T., Andrási, Z., Farkas, M., Fábián, Z., 2018. Report on international and national experiences and main insight for policy use of well-being and sustainability framework. Technical report, MAKSWELL—deliverable 1.1. https://www.makswell.eu/attached_documents/output_deliverables/deliverable_1.1.pdf.
UNDP, 2010. The Human Development Report 2010 'The Real Wealth of Nations: Pathways to Human Development'. UNDP.http://hdr.undp.org/sites/default/files/reports/270/hdr_2010_en_complete_reprint.pdf.
UNDP, 2016. The Human Development Report 2016 'Human Development for Everyone'. UNDP.http://hdr.undp.org/en/2016-report/download.
University of Waterloo, 2016. Canadian Index of Wellbeing National Report. University of Waterloo.https://uwaterloo.ca/canadian-index-wellbeing/reports/canadian-index-wellbeing-national-report.