ORIGINAL PAPER

Principal component analysis for constructing socio-economic composite indicators: theoretical and empirical considerations

Matteo Mazziotta[1] · Adriano Pareto[1]

Abstract

Composite indicators for measuring multidimensional phenomena have become very popular in various social, economic and political fields. This increasing popularity has led to the frequent use of Principal Component Analysis (PCA) to aggregate a set of socio-economic indicators into a composite index. However, a PCA-based composite index must be supported by an appropriate measurement model to function adequately, and this aspect is almost always ignored, both by those teaching PCA and by manuals for constructing composite indicators. In this paper, we discuss the importance of the measurement model for the proper construction of a PCA-based composite index and show that PCA can lead to serious problems in the construction of composite indicators in formative models. A simple numerical example and an application to real data are also shown, where a formative model is required, and the PCA-based composite index produces incorrect results.

Keywords PCA · Data reduction · Composite index · Measurement model

Introduction

Composite indicators are constructed with the aim of obtaining a single 'synoptic' or 'comprehensive' number, that represents a large set of measurements (individual indicators) on the multiple aspects of a phenomenon (conceptual entity), such as human development, competitiveness, happiness, quality of life and well-being (Mishra 2007). These measures have been called 'pragmatic' because they meet a practical need to rate or rank individual units (e.g., countries, cities, universities or hospitals) for a specific purpose (Paruolo et al. 2013). Sociologists, economists, and

---

✉ Adriano Pareto
  pareto@istat.it

[1] Italian National Institute of Statistics, Via Cesare Balbo, 16, Rome 00184, Italy

policy makers use composite indicators for obtaining useful tools for social, economic, and political decision making (Somarriba and Pena 2009). For this reason, socio-economic indicators are often analyzed using multivariate statistical methods, such as Principal Components Analysis (PCA), in order to summarize the data and create composite indices (Ram 1982; Slottje et al. 1991; McGillivray 2005; Li et al. 2012; Ferrara and Nisticò 2014; Yadav and Velan 2021; da Vieira et al. 2022).

Nevertheless, the idea of summarizing complex phenomena into single numbers is not straightforward (Saltelli 2007), and this disciplinary field is still a 'black box' for some researchers (Dialga and Thi Hang Giang 2017). The construction of a composite index involves both theoretical and methodological assumptions that must be carefully assessed to avoid results of dubious analytical rigour (Nardo et al. 2005).

One of the best known guides for the construction and use of composite indicators, if not the best known, is the "Handbook on Constructing Composite Indicators. Methodology and User Guide" by the OECD (2008). It considers the multivariate analysis as a preliminary step that should be used to study the overall structure of the dataset, assess its suitability, and guide subsequent methodological decisions (e.g., weighting, aggregation). However, PCA is also cited as a method for weighting and aggregating individual indicators (OECD 2008; UNECE 2019). In this regard, a distinction must be made between dimensionality reduction and the construction of composite indicators.

Dimensionality reduction is a purely mathematical operation that consists of combining a set of individual indicators, in such a way that most of the information in the data is retained. PCA is the most common method developed for this purpose (Hotelling 1933). The basic idea is simple: to reduce the dimensionality of a dataset so that the original 'variability' is reproduced as well as possible. This means finding new variables that are linear functions of the original variables, that sequentially maximize variance and are uncorrelated with each other. Since the new variables are defined by the dataset at hand and not *a priori*, PCA can be considered a data-driven tool (Greco et al. 2019).

The construction of composite indicators (or composite indices) is a conceptual and mathematical process that consists of combining (or aggregating as it is termed) a set of individual indicators based on a well-defined measurement model that can be formative or reflective (Michalos 2014; Mazziotta and Pareto 2017). A composite indicator is thus formed when individual indicators are combined into a single index based on an underlying model of the multidimensional concept being measured (UNECE 2019).

Of course, a composite index can be obtained by dimensionality reduction (with an appropriate measurement model), but dimensionality reduction does not necessarily lead to a composite index. In this case, there may be potential difficulties in interpretations, inaccurate ranking, and conflicts with the theoretical framework. Recently, Boudt et al. (2022) proposed an adjustment in the construction of a PCA-based composite index to avoid the presence of positive and negative weights, but they do not deal with the definition of the correct measurement model. In fact, if the measurement model is formative, PCA should not be used.

In this paper, we discuss the use of PCA to construct a composite index and show that it can be used improperly if a proper measurement model has not been defined.

The paper is organized as follows. Section "The measurement model" explains the difference between formative and reflective measurement models, while section "PCA and composite indices" describes the use of PCA to construct composite indices and its relationship with the measurement models. In section "A numerical example", a simple numerical example is illustrated, while in section "An application to real data" an application to real data is provided in order to show that PCA can perform poorly when a formative approach is followed. Finally, some concluding remarks are given in section "Conclusions".

The measurement model

It is well known that a measurement model[1] can be conceptualized through two different approaches: reflective or formative (Edwards and Bagozzi 2000; Jarvis et al. 2003; Coltman et al. 2008; Diamantopoulos et al. 2008).

The most widely used approach is the reflective model, in which individual indicators represent the effects (or manifestations) of an underlying latent variable (manifest variables). Therefore, causality is from the concept to the indicators and a change in the phenomenon causes variations in all its measures (i.e., covariation among indicators reflects variation in the latent factor). In this model, the construct exists (in an absolute sense) independent of the researcher's perception or interpretation, even if it is not directly measurable. Specifically, the latent variable R represents the common cause shared by all indicators X_i that reflect the construct, with each indicator corresponding to a linear function of the underlying variable plus a measurement error:

$$X_i = \lambda_i R + \varepsilon_i \tag{1}$$

where X_i is indicator i, λ_i is a coefficient (loading) that captures the effect of R on X_i and ε_i is the measurement error for indicator i. It is assumed that the measurement errors are independent (i.e., $r(\varepsilon_i, \varepsilon_j)=0$, for $i \neq j$) and unrelated to the latent variable (i.e., $r(R, \varepsilon_i)=0$, for all i).

A fundamental characteristic of reflective models is that the change in the latent variable must precede the change in the indicators. Thus, the indicators share a common theme and are interchangeable (adding or removing an indicator does not change the essential nature of the underlying concept). It follows that all indicators must be highly correlated with each other, and the correlations are explained by the measurement model.

Another important issue concerns the polarity of the individual indicators. The 'polarity' of an individual indicator is the sign of the relationship between the indicator and the concept being measured. In a reflective model, indicators with the same polarity must be positively correlated, while indicators with opposite polarity must be negatively correlated. Otherwise, the model will produce inconsistent results (Mazziotta and Pareto 2019).

---

[1] Note that only linear models are considered here.

A typical example of a reflective model is the measurement of a person's intelligence. In this case, it is intelligence that influences responses and reaction times to a IQ (intelligence quotient) test, and not vice versa. Hence, if a person's intelligence increases, this is accompanied by an increase in correct answers to all questions and a decrease in response times. So, the "Percentage of correct answers" has positive polarity, while the "Average response time" has negative polarity. And these two indicators are negatively correlated.

Other appropriate applications of the reflective model include concepts such as attitudes and purchase intentions (Jarvis et al. 2003).

PCA and Factor Analysis[2] (FA) are often used to find sets of correlated indicators that are thought to reflect underlying latent constructs (Shwartz et al. 2015).

The second approach is the formative model, in which individual indicators are causes of an underlying latent variable, rather than its effects (causal variables). Therefore, causality is from the indicators to the concept and a change in the phenomenon does not necessarily imply variations in all its measures (i.e., the latent factor is not assumed to explain the variances of the indicators or their covariation). In this model, the construct is defined by, or is a function of, the observed variables. It depends on an operationalist or instrumentalist interpretation by the researcher (Borsboom et al. 2003). The specification of the formative model is:

$$R = \sum_i \lambda_i X_i + \zeta \tag{2}$$

where λ_i is a coefficient capturing the effect of X_i on R, and ζ is an error term[3]. ζ includes all remaining causes of the construct that are not represented in the indicators and are not correlated with them (i.e., $r(X_i, \zeta)=0$), while the X_i are considered error-free.

Since, in this case, the indicators define the construct, its meaning depends on the number and type of indicators and they are not interchangeable (adding or removing an indicator may change the underlying concept). It follows that indicators can have any intercorrelation pattern (high correlations between indicators are possible, but generally not expected) and the correlations are not explained by the measurement model. Therefore, in a formative model, polarities and correlations are independent and indicators can have positive, negative or zero correlations.

A typical example of a formative model is the measurement of the well-being of society. It depends on health, income, occupation, services, environment, etc., and not vice versa. Hence, if one of these factors improves, well-being increases (even if the other factors do not change). However, an increase in well-being is not necessarily accompanied by an improvement in all factors. So, for example, "GDP per capita" has a positive polarity, while "CO_2 emissions" has a negative polarity. But these two

---

[2] Factor Analysis is an explanatory model in which the observed variables (indicators) are assumed to be (linear) functions of a certain (fewer) number of unobserved variables (latent factors).

[3] Some authors exclude the error term so that (2) reduces to a weighted linear combination of the X_i (Diamantopoulos 2006).

indicators are not necessarily negatively correlated (in fact, they are generally positively correlated).

Most of socio-economic composite indicators, such the Human Development Index (UNDP 1990, 2010) are based on formative models (UNECE 2019).

Since a formative model is not based on the hypothesis that the indicators are correlated, the correlation structure of the data cannot be used to determine the latent variable. Rather, the latent variable can be estimated by taking a weighted[4] average of the indicators that encompass the concept (Shwartz et al. 2015).

It is important to note that (1) is a system of simple regression equations where each individual indicator is the dependent variable and the latent variable is the explanatory variable; while (2) is a multiple regression equation where the latent variable is the dependent variable and the indicators are the explanatory variables.

Table 1 summarizes the main differences between the two types of models.

The choice between reflective and formative models substantially affects the results, but has received little attention in the literature (Jarvis et al. 2003). Most researchers apply procedures without even questioning their appropriateness for the specific construct, and Diamantopoulos and Winklhofer (2001) speak of an "almost automatic acceptance of reflective indicators". Consequently, misspecification is often the adoption of a reflective model where a formative approach would be appropriate. The other case of misspecification, i.e., the erroneous adoption of a formative model where a reflective approach would be appropriate, is rather negligible. One explanation is that procedures for developing and evaluating measures for reflective latent factor models have a long tradition in the social sciences and have become established over the years (Diamantopoulos et al. 2008).

In the next Section, we show that misspecification of the measurement model in the construction of a PCA-based composite index, can lead to some potentially serious consequences. Therefore, researchers need to think carefully about the direction of causality between concept and individual indicators.

Table 1 Reflective model versus formative model

Characteristic	Reflective model	Formative model
Nature of construct	Existing, but not directly measurable	Defined by the researcher
Direction of causality	From construct to indicators	From indicators to construct
Indicators interchangeability	Yes	No
Indicators intercorrelation	Indicators should be correlated	Indicators can be incorrelated
Relationship between polarities and correlations	Indicators with equal polarities expected to be positively correlated, whereas indicators with opposite polarities expected to be negatively correlated	Polarities and correlations are independent
Measurement error	At the indicators level	At the construct level

---

[4] Weights could be determined *a priori*, according to the theoretical contribution of the indicators to the concept (Howell et al. 2007).

PCA and composite indices

The simplest and common formula for constructing a composite index (CI) is the sum of weighted and normalized[5] indicators (OECD 2008):

$$\text{CI} = \sum_i w_i Z_i \tag{3}$$

where Z_i is the normalized indicator i, and w_i is the weight of normalized indicator i, with:

$$\sum_i w_i = 1 \text{ and } 0 \leqslant w_i \leqslant 1 \text{ for all } i \tag{4}$$

Normalized indicators can be obtained through standardization (z-scores), re-scaling or other methods (OECD 2008; Mazziotta and Pareto 2017; Terzi et al. 2021). In any case, individual indicators must be transformed so that an increase in the normalized indicators corresponds to an increase in the composite index (Salzman 2003). Therefore, it is necessary to 'invert' the sign of indicators with negative polarity. For example, in the case of well-being, "CO_2 emissions" has negative polarity (the lower the CO_2 emissions, the greater the well-being), so it can be standardized (i.e., transformed in z-score) and multiplied by -1. Alternatively, indicators with negative polarity should have negative weights in (3), but condition (4) cannot be satisfied.

PCA is a multivariate statistical method that allows a large number of quantitative individual indicators to be transformed into a set of new uncorrelated variables (principal components or factors), ordered so that the first few explain most of the observed variance (Dunteman 1989). The first principal component is often used as the 'best' composite index[6] (Booysen 2002; Mishra 2007, 2008; Somarriba and Pena 2009). It is defined as:

$$\text{PC}_1 = \sum_i a_{i1} Z_i \tag{5}$$

where a_{i1} is the weight (loading) of indicator i for the first principal component, with:

$$\sum_i a_{i1}^2 = 1$$

[5] Normalization aims to make the indicators comparable and is required before any data aggregation when the indicators have different measurement units and ranges (OECD 2008).

[6] Note that if the first principal component accounts for a limited part of the variance in the data, other principal components, such as the second and the third, can also be used as composite indices. Otherwise, a single composite index can be obtained as a weighted average of all the principal components with weights given by the proportion of the variance explained by each of them (Krishnakumar and Nagar 2008).

and the weights are determined in such a way that the sum of the squared correlation coefficients between the index and the individual indicators (used to construct the index) is maximized:

$$\sum_i r^2(\text{PC}_1, Z_i) = \max \tag{6}$$

The solution is given by the eigenvector corresponding to the largest eigenvalue of the correlation (or covariance) matrix of the individual indicators (Jolliffe 2002).

The definition of principal component (5) as a weighted sum of individual indicators is similar to Eq. (2) and suggests that a PCA-based composite index can be used with a formative measurement model (Edwards and Bagozzi 2000; Zumbo 2007).

Nevertheless, some important issues need to be considered.

First, in a PCA-based composite index, the weights depend on the correlations between indicators, as they are given by the eigenvectors of the correlation (or covariance) matrix. However, correlations between individual indicators are not relevant in a formative model and cannot be explained by it, because indicators do not necessarily share the same theme and hence do not have a 'preconceived' pattern of intercorrelation (Coltman et al. 2008). In this regard, already 20 years ago, Fayers and Hand (2002) in an article on "Journal of the Royal Statistical Society" pointed out that the main focus of methods of scale construction and assessment of standard psychometric approaches "*has been the correlation structure of the data – which is inappropriate for causal variables*". And they added that "*Also, it is disturbing to note that anyone developing a scale by using traditional methods would remain blissfully unaware that they may be omitting important items or including inappropriate items*". The idea of PCA is to account for the largest possible variation in the individual indicators with the smallest possible number of factors. Therefore, in PCA, the weights are used only to correct for overlap between two or more correlated indicators and are not a measure of the theoretical importance of the associated indicators (OECD 2008). In fact, if no correlation is found between the individual indicators, the weights cannot be estimated using this method.

Second, by construction (see Eq. 6), a PCA index assigns larger weights to highly correlated indicators (because they help maximize the sum of squared correlation coefficients between the index and the individual indicators) and marginal weights to poorly correlated indicators. As a result, the index so constructed is inherently 'elitist' since it favors the highly correlated subset over the poorly correlated subset of variables, regardless of the (possible) contextual importance of the latter subset of variables (Mishra 2007, 2008). However, in a multiple regression, such as (2), the individual indicators should have little or no correlation with each other to avoid multicollinearity[7]. In contrast to reflective models, where each individual indicator is by design collinear with the others, multicollinearity in formatively measured constructs can potentially lead to unstable weights (Tabachnick and Fidell 2001). Moreover, an excessive multicollinearity makes it difficult to separate the distinct influence of the

---

[7] Note that under reflective measurement model, multicollinearity is not an issue, because only simple regressions are involved (Diamantopoulos and Winklhofer 2001).

individual indicators on the latent variable (Bollen 1989; Diamantopoulos and Winklhofer 2001). Thus, collinearity among individual indicators challenges the interpretation of formative composite indices (Cenfetelli and Bassellier 2009).

Third, under certain conditions, the principal components are equivalent to the factor scores obtained by FA and can then be considered as estimators of latent factors (Krishnakumar and Nagar 2008). For example, FA and PCA produce similar results in cases with a large number of variables (e.g., 30 or more) and/or high estimated communality (Gorsuch 1983; Hair et al. 2006). Moreover, both types of analysis have a common mathematical core, are based on the correlation (or covariance) matrix of the data and use methods of matrix decomposition to obtain the components or factors. If 'principal components' is used as a factor extraction method, the matrix of factor loadings obtained in FA is identical to matrix of correlation coefficients between original variables and principal components obtained in PCA. Hence, the factors, before any rotation, are identical to the first few principal components with the largest variances (Gniazdowski 2017, 2021). However, FA is based on a reflective measurement model, therefore PCA cannot be used correctly in a formative approach (Mazziotta and Pareto 2019).

Finally, it is also important to note that the signs of the weights in (4) are based on observed covariances, and not on user-defined polarities, as required in a formative model. Indeed, even if individual indicators with negative polarities are 'reversed' by normalization, a PCA-based composite index ignores the polarities of the individual indicators and yields consistent results only if:

$$\operatorname{sgn}(r(\mathrm{X}_i, \mathrm{X}_j)) = p_i p_j \quad \text{for } i \neq j \tag{7}$$

where p_i e p_j are the polarities of indicators i and j.

Equation (7) is the formalization of item 5 in Table 1 for reflective models. If a formative model is assumed where (7) is not satisfied, a PCA-based index will produce incorrect results.

Because of these features, a PCA-based composite index should only be used in reflective models, as is the case of FA. Indeed, PCA is frequently used to evaluate reflective measurement models (Götz et al. 2010) and is considered an appropriate method for examining the latent structure underlying a set of indicators (Bohrnstedt 1970; Vinzi et al. 2003). The use of PCA to test the 'content validity' or 'construct validity' of reflective models can be found in several studies in the literature, since the 1970s, especially in psychological and clinical fields (Harbison et al. 1974; Raskin and Terry 1988; Toledano and Pfaus 2006; Klingstedt et al. 2020; Ghazali et al. 2021).

A numerical example

In this Section, we consider a numerical example where a formative composite index is required. A simple arithmetic mean and the first principal component are compared as composite indices, but the PCA-based composite index fails because it is not consistent with the polarities of the individual indicators.

Suppose that we want to construct a composite index of well-being for 7 world regions, in 2018, based on the following individual indicators (Source: World Bank, World Development Indicators):

- X_1=GDP per capita, PPP (current international $);
- X_2=CO_2 emissions (metric tons per capita).

These two indicators are not manifestations of an underlying latent variable, but determine the latent variable that gets its meaning from them, i.e., we create an 'induced' latent variable that is an aggregation of observed variables (Heise 1972). Thus, causality is from the indicators to the concept and indicator X_1 has a positive polarity, whereas indicator X_2 has a negative polarity.

In Table 2 are reported the original data (mean and standard deviation are in bold). The table also provides the normalized indicators[8] Z_1 and Z_2, the ranks R_1 and R_2, the arithmetic mean of the normalized values M_1, and the first principal component scores PC_1.

Note that $r(X_1, X_2)$=0.95, i.e., X_1 and X_2 are positively correlated, so that the higher the GDP per capita, the greater the CO_2 emissions. On the other hand, we have $r(Z_1, Z_2)$=-0.95, because the polarity of X_2 has been inverted in order to construct the composite index.

In a formative approach, as Eq. (2), we can form a composite index by the arithmetic mean. However, the first principal component might be the best solution, as it accounts for 97.7% of the variance in the data.

The rankings by normalized indicators and by composite index are shown in Fig. 1. As we can see, North America has the highest GDP per capita (61,762) and CO_2 emissions (15,3), while Sub-Saharan Africa has the lowest GDP per capita (3,965.6) and CO_2 emissions (0.8) (Fig. 1a). This means that North America is ranked first by X_1, while Sub-Saharan Africa is ranked first by X_2. Nevertheless, North America is ranked first by PC_1, and Sub-Saharan is ranked seventh (Fig. 1b). Thus, the ranking by PC_1 accurately reflects the ranking by X_1, but neglects the ranking by X_2. This is due to the fact that PCA ignores the polarity of the individual indicators and normalized indicators (i.e., indicators that both have positive polarity) are not positively correlated (i.e., Eq. 7 is not satisfied). Therefore, using PC_1 to aggregate X_1 and X_2 results in an inconsistent composite index and an unrealistic ranking of units.

Another very important point is that a PCA-based composite index can be non-monotone[9]. In our application, the formulas used to calculate the two composite indices are as follows:

$$M_1 = (Z_1 + Z_2)/2 = 0.5Z_1 + 0.5Z_2$$
$$PC_1 = a_{11}Z_1 + a_{21}Z_2 = 0.707Z_1 - 0.707Z_2$$

---

[8] Individual indicators were normalized as *z*-scores. The signs were reversed if the polarity is negative.

[9] An aggregation function $f(x_1, x_2,\ldots, x_m)$ is called monotone (increasing or non-decreasing) if, for any $i\epsilon\{1, 2,\ldots, m\}$ such that $x_i \le y_i$, then $f(x_1, x_2,\ldots, x_m) \le f(y_1, y_2,\ldots, y_m)$.

Table 2 Comparing arithmetic mean and first component score as composite indices

Region	Original value		Normalized value		Rank		M_1		PC_1	
	X_1	X_2	Z_1	Z_2	R_1	R_2	Value	Rank	Value	Rank
East Asia & Pacific	17,369.1	6.3	-0.28	-0.17	3	5	-0.23	7	-0.05	3
Europe & Central Asia	35,761.0	6.7	0.71	-0.25	2	6	0.23	1	0.49	2
Latin America & Caribbean	16,398.4	2.6	-0.33	0.64	5	3	0.16	2	-0.49	5
Middle East & North Africa	16,503.3	5.6	-0.33	-0.02	4	4	-0.17	6	-0.16	4
North America	61,762.0	15.3	2.11	-2.14	1	7	-0.02	5	2.15	1
South Asia	6,218.6	1.5	-0.88	0.89	6	2	0.00	4	-0.89	6
Sub-Saharan Africa	3,965.6	0.8	-1.00	1.06	7	1	0.03	3	-1.04	7
Mean	**22,568.3**	**5.6**	**0.00**	**0.00**						
Standard dev.	**18,614.3**	**4.5**	**1.00**	**1.00**						

where a_{i1} is the weight of indicator i, as used in the creation of the first principal component[10]. In both composite indices, the weights of the two variables are equal, but in PC_1 the weight of Z_2 is negative. Thus, if Z_2 increases, and Z_1 remains constant, PC_1 decreases. This means that if CO_2 emissions decrease (i.e., Z_2 increases) and GDP per capita does not change, M_1 correctly increases, while PC_1 incorrectly decreases.

An application to real data

A typical case in which the use of PCA can have potentially critical consequences and lead to misleading results is the construction of a health index.

Consider the set of 15 individual indicators of the "Health" domain of the 10th Italian report[11] on Equitable and Sustainable Well-being (BES) for the year 2022 (Istat 2023). The BES project was launched in 2010 to evaluate the progress of society not only from an economic point of view, but also from a social and environmental one. To this end, traditional economic indicators, such as GDP per capita, were integrated with measures of people's quality of life and the environment. The report provides an integrated picture of the main economic, social and environmental phenomena that characterize Italy, through the analysis of a large number of indicators for the Italian regions, divided into 12 domains.

The indicators selected for the "Health" domain describe essential elements of the population's health profile in the main dimensions: objective, functional and subjective health. They are divided into three groups: global outcome indicators, specific indicators for life cycle stages, and indicators of lifestyles-related risk or health protection factors.

The indicators have different units of measurement and ranges; some have positive polarity (e.g., "Life expectancy at birth"), while others have negative polarity (e.g., "Infant mortality rate"). Therefore, they were normalized into z-scores and the signs of the indicators with negative polarity were reversed.

A PCA was performed on both the set of original indicators and the set of normalized indicators, and the first principal component was used as the health index. The results are presented in Table 3, which shows the polarities and weights of the original and normalized indicators[12].

As we know, in a correct composite index, indicators with positive polarity should have positive weight, while indicators with negative polarity should have negative weight. Otherwise, if all indicators are normalized to have positive polarity, each of them must have a positive weight, as in Eq. (3). However, this is not the case because the weights of "Age-standardized mortality rate for dementia and nervous system diseases" and "Alcohol consumption" do not have the correct sign (indicators and values in bold).

---

[10] The weights a_{j1} are constructed so that the variance of PC_1 is maximized.

[11] https://www.istat.it/en/well-being-and-sustainability/the-measurement-of-well-being/bes-report.

[12] Note that the results of the two analyses are identical, except for the signs of the weights of the indicators with negative polarity.

a) Normalized indicators
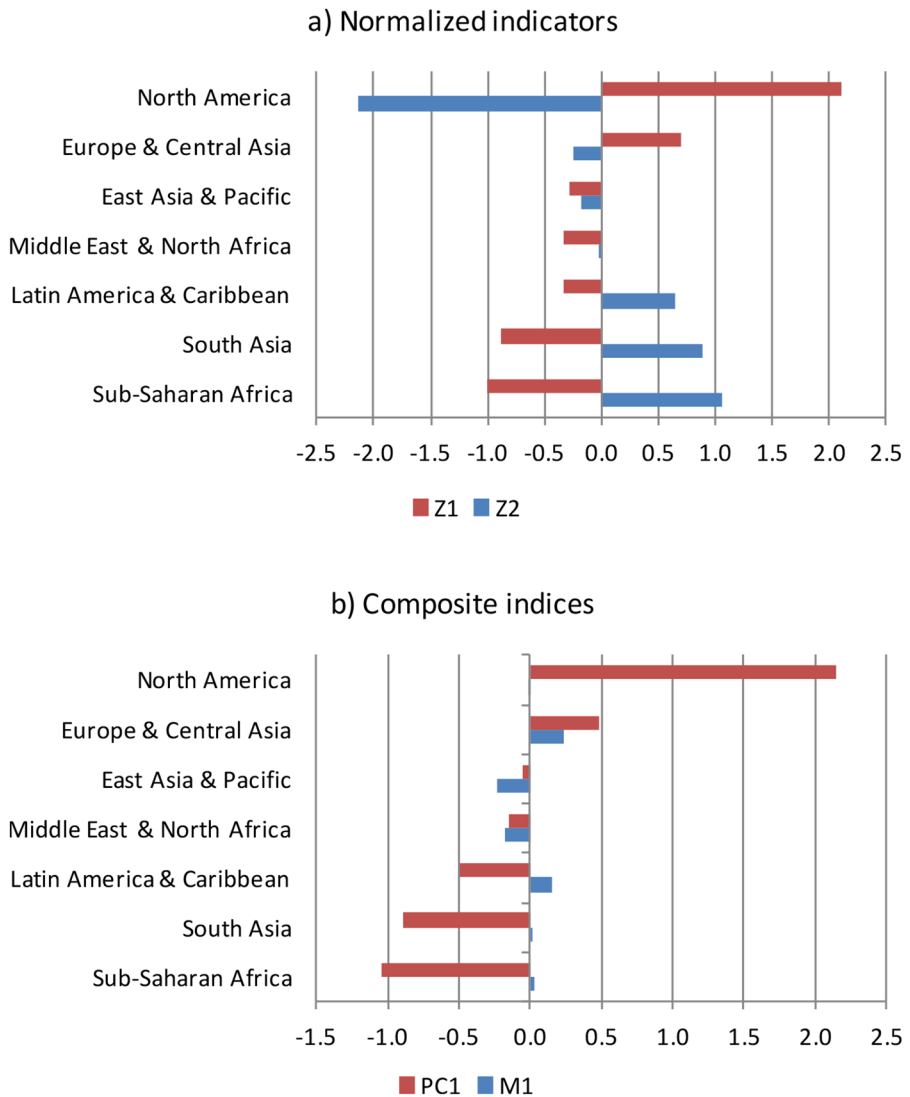

b) Composite indices

Fig. 1 Normalized indicators and composite indices

This occurs because Eq. 7 is not satisfied for each pair of individual indicators. For example, "Life expectancy at birth" and "Alcohol consumption" are positively correlated ($r=0.67$), but the first indicator has positive polarity and the second has negative polarity. In other words, a positive correlation indicates that an increase in alcohol consumption is associated with an increase in life expectancy, which is contrary to what the theoretical framework assumes. The same is true for "Life expectancy at birth" and "Age-standardized mortality rate for dementia and nervous system diseases" ($r=0.60$). But the most interesting aspect is that reversing the sign of indicators with negative polarity, through normalization, does not solve the problem.

Table 3 Polarities and weights of wealth indicators in a PCA-based composite index

Individual indicator	Original indicators		Normalized indicators	
	Polarity	Weight	Polarity	Weight
Life expectancy at birth	Positive	0.316	Positive	0.316
Healthy life expectancy at birth	Positive	0.248	Positive	0.248
Mental health index	Positive	0.144	Positive	0.144
Avoidable mortality	Negative	-0.288	Positive	0.288
Infant mortality rate	Negative	-0.220	Positive	0.220
Road accidents mortality rate (15–34 years old)	Negative	-0.136	Positive	0.136
Age-standardized cancer mortality rate	Negative	-0.266	Positive	0.266
Age-standardized mortality rate for dementia and nervous system diseases	**Negative**	**0.255**	**Positive**	**-0.255**
Multichronicity and severe limitations (75 years and over)	Negative	-0.306	Positive	0.306
Life expectancy without activity limitations at 65 years of age	Positive	0.316	Positive	0.316
Overweight or obesity	Negative	-0.292	Positive	0.292
Smoking	Negative	-0.063	Positive	0.063
Alcohol consumption	**Negative**	**0.269**	**Positive**	**-0.269**
Sedentariness	Negative	-0.325	Positive	0.325
Adequate nutrition	Positive	0.269	Positive	0.269

This result shows that data-driven methods such as PCA, which rely solely on the observed correlation (or covariance) matrix, are inappropriate for constructing a formative composite index in which the nature of the concept and the polarity of each individual indicator are defined by the researcher.

The experiment can be replicated with any other set of indicators, and it is always possible that some correlations do not match polarities. Of course, the researcher could remove the indicators with undesirable weights and still use a PCA-based composite index, as many do[13], but in a formative model, the individual indicators are not interchangeable and "*omitting an indicator is omitting a part of the construct*" (Bollen and Lennox 1991).

Conclusions

The construction of composite indices to measure multidimensional concepts is a common problem in data analysis. Researcher cannot easily solve this issue using PCA or related methods, such as FA, because they are used for a reflective approach; while composite indices are generally based on a formative approach.

PCA is essentially a data reduction technique for summarizing a set of correlated individual indicators, improving interpretability and minimizing information loss. So, it can be very useful for eliminating redundant indicators. However, in a formative measurement model, the individual indicators define the concept being measured and they are not redundant.

---

[13] For example, a constrained PCA can be used that converts negative weights to zero and ensures that all weights are non-negative (Boudt et al. 2022).

Individual indicators may have high, low or no intercorrelation, so it does not makes sense to aggregate them with a PCA-based composite index. It does not take into account all non-redundant information, as it only explains most of the variance, and can therefore remove useful information. For example, if some individual indicators are poorly correlated with others, PCA can assign them very small weights, irrespective of their importance.

Moreover, although PCA allows the construction of subsequent indices (orthogonal to the first principal component), it may not be possible to use them for any comprehensive analysis, as there is no reliable and well-established procedure to construct a single composite index by merging several principal component indices derived from the data (Mishra 2007).

Finally, in a formative approach, the polarities of the individual indicators are defined by the researcher, whereas the signs of the weights obtained in a PCA-based composite index depend on the observed correlations, regardless of the polarities.

Therefore, a PCA-based composite index can provide very misleading information about the latent variable of interest, as it is based solely on the covariance structure between individual indicators.

In light of the above, although PCA and related methods are often used to construct socio-economic composite indicators, they should be used only for reflective purposes, and not to build formative composite indices.

Declarations

References

Booysen F (2002) An overview and evaluation of composite indices of development. Social Indic Res 59:115–151

Bohrnstedt GW (1970) Reliability and validity assessment in attitude measurement. In: Summers GF (ed) Attitude measurement. Rand McNally, London, pp 80–99

Bollen K (1989) Structural equations with latent variables, 1st edn. Wiley-Interscience, Malden

Bollen K, Lennox R (1991) Conventional wisdom on measurement: a structural equation perspective. Psychol Bull 110:305–314

Borsboom D, Mellenbergh GJ, Heerden JV (2003) The theoretical status of latent variables. Psychol Rev 110:203–219

Boudt K, d'Errico M, Luu HA, Pietrelli R (2022) Interpretability of Composite indicators based on principal components. J Probab Stat, Hindawi

Cenfetelli RT, Bassellier G (2009) Interpretation of formative measurement in Information Systems Research. MIS Q: Manag Inf Syst 33:689–707

Coltman T, Devinney TM, Midgley DF, Venaik S (2008) Formative versus reflective measurement models: two applications of formative measurement. J Bus Res 61:1250–1262

da Vieira C, Ferreira Neto W, Roque JA, da Rocha MPB BD (2022) Using principal component analysis to build socioeconomic status indices. Int J Stat Appl 12:77–82

Dialga I, Thi Hang Giang L (2017) Highlighting Methodological limitations in the steps of Composite indicators Construction. Social Indic Res 131:441–465

Diamantopoulos A (2006) The error term in formative measurement models: interpretation and modeling implications. J Model Manag 1:7–17

Diamantopoulos A, Winklhofer HM (2001) Index Construction with formative indicators: an alternative to Scale Development. J Mark Res 38:269–277

Diamantopoulos A, Riefler P, Roth KP (2008) Advancing formative measurement models. J Bus Res 61:1203–1218

Dunteman GH (1989) Principal components analysis. Sage, Newbury Park

Edwards JR, Bagozzi RP (2000) On the Nature and Direction of relationships between constructs and measures. Psychol Methods 5:155–174

Fayers PM, Hand DJ (2002) Causal variables, indicator variables and measurement scales: an example from quality of life. J R Stat Soc Ser A 165:233–261

Ferrara AR, Nisticò R (2014) Measuring well-being. A multidimensional perspective: a Multivariate Statistical Application to Italian regions. Dipartimento di Economia, Statistica e Finanza, Università della Calabria, Arcavacata

Ghazali FB, Ramlee SNS, Alwi N, Hizan H (2021) Content validity and test–retest reliability with principal component analysis of the translated malay four-item version of Paffenbarger physical activity questionnaire. J Health Res 35:493–505

Gniazdowski Z (2017) New interpretation of principal components analysis. Zeszyty Naukowe WWSI 11:43–65

Gniazdowski Z (2021) Principal Component Analysis versus factor analysis. Zeszyty Naukowe WWSI 15:35–88

Gorsuch RL (1983) Factor analysis, 2nd edn. Lawrence Erlbaum Associates, Hillsdale

Götz O, Liehr-Gobbers K, Krafft M (2010) Evaluation of structural equation models using the partial least squares (PLS) approach. In: Esposito Vinzi V, Chin WW, Henseler J, Wang H (eds) Handbook of partial least squares: concepts, methods, and applications. Springer-, Berlin, pp 691–711

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. Social Indic Res 141:61–94

Hair JF, Black WC, Babin BJ, Anderson RE, Tatham RL (2006) Multivariate Data Analysis, 6th edn. Pearson Prentice Hall, Upper Saddle River

Harbison JJM, Graham PJ, Quinn JT, McAllister H, Woodward RA (1974) A questionnaire measure of sexual interest. Arch Sex Behav 3:357–366

Heise DR (1972) Employing nominal variables, induced variables, and block variables in path analysis. Sociol Methods Res 1:147–173

Hotelling H (1933) Analysis of a complex of statistical variables into principal components. J Educ Psychol 24:417–441

Howell RD, Breivik E, Wilcox JB (2007) Reconsidering formative measurement. Psychol Methods 12:205–218. https://doi.org/10.1155/2022/4155384

Istat (2023) BES 2022: Equitable and sustainable well-being in Italy. Italian National Institute of Statistics, Rome

Jarvis CB, Mackenzie SB, Podsakoff PM (2003) A critical review of Construct indicators and Measurement Model Misspecification in Marketing and Consumer Research. J Consum Res 30:199–218

Jolliffe IT (2002) Principal components analysis, 2nd edn. Springer, New York

Klingstedt ML, Wångby-Lundh M, Olsson T, Ferrer-Wreder L (2020) Reliability and construct validity of five life domains in the adolescent drug abuse diagnosis instrument in a sample of Swedish adolescent girls in special residential care. NAD Nord Stud Alcohol Drugs 37:411–426

Krishnakumar J, Nagar AL (2008) On exact statistical properties of Multidimensional Indices based on Principal Components, factor analysis, MIMIC and structural equation models. Social Indic Res 86:481–496

Li T, Zhang H, Yuan C, Liu Z, Fan C (2012) A PCA-based method for construction of composite sustainability indicators. Int J LCA 17:593–603

Mazziotta M, Pareto A (2017) Synthesis of indicators: the Composite indicators Approach. In: Maggino F (ed) Complexity in Society: from indicators construction to their synthesis. Springer, Cham, pp 159–191

Mazziotta M, Pareto A (2019) Use and Misuse of PCA for Measuring Well-Being. Social Indic Res 142:451–476

McGillivray M (2005) Measuring non-economic well-being achievement. Rev Income Wealth 51:337–364

Michalos AC (ed) (2014) Encyclopedia of Quality of Life and Well-Being Research. Springer, New York

Mishra SK (2007) A Comparative Study of Various Inclusive Indices and the Index Constructed by the Principal Components Analysis. MPRA Paper 3377. Available via DIALOG. http://mpra.ub.uni-muenchen.de/3377

Mishra SK (2008) On Construction of Robust Composite Indices by Linear Aggregation. SSRN Paper. Available via DIALOG. http://ssrn.com/abstract=1147964

Nardo M, Saisana M, Saltelli A, Tarantola S (2005) Tools for Composite indicators Building. European Commission-JRC, Ispra

OECD (2008) Handbook on constructing Composite indicators. Methodology and user guide. OECD, Paris

Paruolo P, Saisana M, Saltelli A (2013) Ratings and rankings: voodoo or science? J R Stat Soc Ser A 176:609–634

Ram R (1982) Composite indices of physical quality of life, basic needs fulfilment, and income: a principal component representation. J Dev Econ 11:227–247

Raskin R, Terry H (1988) A principal-components analysis of the narcissistic personality inventory and further evidence of its Construct Validity. J Pers Soc Psychol 54:890–902

Saltelli A (2007) Composite indicators between analysis and advocacy. Social Indic Res 81:65–77

Salzman J (2003) Methodological choices encountered in the construction of Composite indices of Economic and Social Well-Being. Center for the Study of Living Standards, Ottawa

Shwartz M, Restuccia JD, Rosen AK (2015) Composite measures of Health Care Provider performance: a description of approaches. Milbank Q 93:788–825

Slottje DJ (1991) Measuring the quality of life across countries. Rev Econ Stat 73:684–693

Somarriba N, Pena B (2009) Synthetic indicators of quality of life in Europe. Social Indic Res 94:115–133

Tabachnick BG, Fidell LS (2001) Using Multivariate statistics, 4th edn. HarperCollins, New York

Terzi S, Otoiu A, Grimaccia E, Mazziotta M, Pareto A (2021) Open issues in composite indicators. A starting point and a reference on some state-of-the-art issues. Roma TrE-, Rome

Toledano R, Pfaus J (2006) Original research-outcomes assessment: the sexual arousal and desire inventory (SADI): a multidimensional scale to assess subjective sexual arousal and desire. J Sex Med 3:853–877

UNDP (1990) Human development report 1990: Concept and measurement of human development. Oxford University Press, Oxford

UNDP (2010) Human development report 2010. Palgrave MacMillan, New York

UNECE (2019) Guidelines on producing leading, composite and sentiment indicators. United Nations, Geneva

Vinzi VE, Lauro C, Tenenhaus M (2003) PLS path modeling. DMS - University of Naples, HEC - School of Management, Jouy-en-Josas

World Bank, World Development Indicators Available via DIALOG https://databank.worldbank.org/source/world-development-indicators

Yadav V, Velan N (2021) Computing a composite financial inclusion index for the Indian states: a principal component analysis. PalArch's J Archaeol Egypt/Egyptol 18:779–799

Zumbo BD (2007) Validity: foundational issues and statistical methodology. In: Rao CR, Sinharay S (eds) Handbook of statistics. Elsevier, Boston, pp 45–79