WEIGHT OPTIMIZATION FOR COMPOSITE INDICATORS BASED ON VARIABLE IMPORTANCE: AN APPLICATION TO MEASURING WELL-BEING IN EUROPEAN REGIONS

Viet Duong Nguyen, Chiara Gigliarano

Abstract. Composite indicators are widely recognized as effective tools for representing complex assessments in the form of a one-dimensional measure. The proliferation of related theoretical frameworks and methodologies has been accompanied by a growing debate around the determination of optimal weights in developing composite indicators. This paper introduces two weighting procedures aimed at assisting developers in attaining the most plausible solution, which closes the disparity between the importance of input features and their corresponding weights. The first technique involves utilizing variance-based sensitivity analysis and calibrating the weights in accordance with the contribution of each input to the output uncertainty. Alternatively, the second approach employs a combination of cluster analysis and predictive modeling to evaluate the relative capability of individual features in differentiating observations within the multidimensional context, thereby informing a proper weight assignment. To demonstrate the practical application of these weighting procedures, a composite indicator has been developed to assess the level of well-being in large European regions during the ten-year period from 2010 to 2019. Despite differences in the weighting schemes used to calculate the final index values, the empirical results indicate a general consensus regarding the allocation of welfare across the territories.

1. Introduction

Composite indicators are basically models used to measure the performance of objects in complex concepts which are not able to judge based on a single aspect. The role of composite indicators is to provide a proper aggregation that combines the conduct of objects in different dimensions into only one scalar. On the one hand, composite indicators are useful to support decision makers in capturing multidimensional realities and comparing object performance straightforwardly. On the other hand, they might provide incorrect benchmarks and misleading policy messages if they are poorly constructed, induced by inefficient input selection or misinterpreted model configuration (Nardo *et al*., 2005b).

Whereas the selection of inputs is primarily contingent upon the definition of the interested phenomenon, the configuration of weights and aggregation functions largely falls within the purview of modelers. A multidimensional problem entails many possible measurement approaches, leading to a certain degree of subjectivity when imposing judgments on its constituent components. Consequently, weights can be attained from any consideration such as statistical models, participatory methods, or expert opinions. Since weights highly impact the result of composite scores and the ranking of units in benchmarking exercises, it is imperative that the assumptions and implications of the employed weighting scheme are transparent and rigorously tested for robustness (Nardo *et al.*, 2005a).

In this study, we introduce two data-driven weighting methods for constructing composite indicators. The first approach involves an optimization procedure based on variance-based sensitivity measures. The application of variance-based sensitivity analysis to composite indicators has been pioneered in a number of studies that primarily focuses on assessing uncertainty in model output (Grupp and Mogee, 2004; Saisana *et al.*, 2005). Nevertheless, this technique has not gained widespread adoption in weight elicitation due to the necessity of having a predefined set of weights prior to conducting the analysis. Becker *et al*. (2017) has devised a strategy for calibrating weights to ascertain the empirical significance of each input so that it is aligned with the value recommended by expert opinions. Despite the merits, this method tends to generate gaps between the estimated importance of variables and their corresponding weights, diminishing the transparency when interpreting the composite index. Our proposed method closes the gaps by seeking the most effective configuration for the multidimensional context, wherein a set of weights is tuned to achieve no difference between the weights and the normalized sensitivity score of inputs.

The second approach, adopting an alternative perspective, leverages valuable information garnered from unsupervised learning techniques to derive appropriate weights for a composite indicator. Among these techniques, principal component analysis (PCA) and factor analysis are two prominent candidates thanks to their ability of dimensionality reduction. Some noteworthy instances of PCA weighting can be found in the works of Klasen (2000) and Nicoletti *et al.* (2000). However, PCA and factor analysis exhibit discernible limitations, such as inapplicability to low-correlated data, susceptibility to outliers, and the potential to produce negative weights (Nardo *et al.*, 2005b). As a viable substitute, we advocate employing a combination of cluster analysis and predictive modeling to gauge the importance of input variables in distinguishing objects in the multidimensional space. Subsequently, this information serves as the basis for defining the weights based on the rationale that variables of higher significance in classification ought to be more amplified within the composite measure.

For a practical application, we present a composite indicator designed to measure well-being in 198 large European regions over the decade from 2010 to 2019. The equal weighting, the PCA weighting, and the two proposed methods were applied to provide a comprehensive view of welfare allocation across the European territories at both regional and national levels.

2. Measuring Variable Importance

2.1. Sensitivity Analysis Approach

Sensitivity analysis involves the study of how uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model input (Saltelli, 2002). This approach allows for identifying which variables have the greatest influence on the composite indicator score, thereby providing insights into the model's validity and reliability. In this paper, we focus on the technique of measuring sensibility using conditional variances. The foundation of this approach was introduced by Sobol' (2001), who devised a measure that bears his name. The Sobol' method is based on an essential assumption that all the input features are mutually independent, which might be unrealistic in practice. Mara *et al.* (2015) developed a methodology to overcome this issue by proposing a strategy of estimating importance indices that account for the dependency of input factors. Let Y denote a composite indicator obtained from a square integrable function $f(X)$ where the input $X = (X_1, X_2, \dots, X_n)$ is a random vector, the authors provided an improvement of the original Sobol' indices:

$$
\begin{aligned}
S_i^{full} &= \frac{\mathrm{Var}(\mathrm{E}_{X_{\sim i}}(Y|X_i))}{\mathrm{Var}(Y)},\\
ST_i^{full} &= \frac{\mathrm{E}(\mathrm{Var}_{X_i}(Y|(X_{\sim i}|X_i)))}{\mathrm{Var}(Y)},\\
S_i^{ind} &= \frac{\mathrm{Var}(\mathrm{E}_{X_{\sim i}}(Y|(X_i|X_{\sim i})))}{\mathrm{Var}(Y)},\\
ST_i^{ind} &= \frac{\mathrm{E}(\mathrm{Var}_{X_i}(Y|X_{\sim i}))}{\mathrm{Var}(Y)}.
\end{aligned}
\tag{1}
$$

The measures S_i^{full} and ST_i^{full}, called the full Sobol' indices, reflect the main and the total contribution of X_i to the output variance, taking into account its dependency with the other inputs. On the other hand, S_i^{ind} and ST_i^{ind}, called the independent Sobol' indices, respectively measure the main and total contributions of X_i that does not account for its mutual dependence on all the other inputs. With respect to the variable X_i, denote w_i as the weight and I_i as the variable importance measured by one of the four indices. The importance measures for all the variables are normalized by $\widetilde{I_i} = I_i / \sum_{k=1}^{n} I_k$ to make them comparable to the value of weights. Denote a loss function

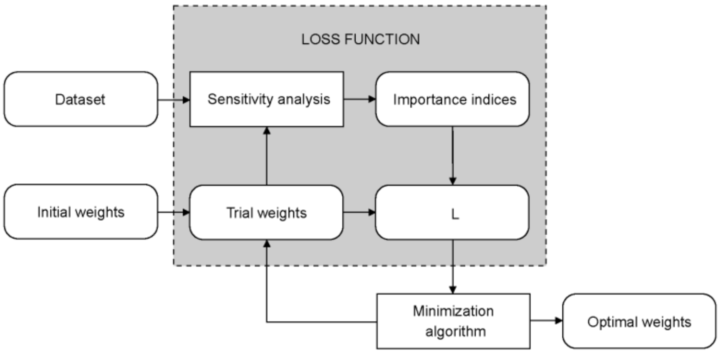
$$L = d^2(w, \tilde{I}) = \sum_{i=1}^{n} \left(w_i - \widetilde{I_i}\right)^2, \quad (2)$$

which is the squared Euclidean distance between two vectors $w = (w_1, \dots, w_n)$ and $\tilde{I} = \left(\widetilde{I_1}, \dots, \widetilde{I_n}\right)$. The optimal set of weights is defined by

$$w^* = \underset{w_1,\dots,w_n}{\operatorname{argmin}} L \quad \text{s.t.} \quad w_i \in (0,1), \ \sum_{i=1}^{n} w_i = 1. \quad (3)$$

At $L_{\min}$, the distance between the two vectors is minimal and hence we attain the set w^* as close as possible to $\tilde{I}$. In case $L_{\min} = 0$ that is equivalent to $w^* \equiv \tilde{I}$, the weights obtained are exactly proportional to the measures of importance. Figure 1 gives an illustration of the optimization procedure. Please note that the rounded boxes indicate the inputs/outputs while the rectangle boxes demonstrate the functions. At the beginning, a sample of X and an initial set of weights are fed into the loss function to estimate the distance L. The trial weights are then updated using a minimization algorithm based on the estimated values of L until the loss function achieves its minimum, which indicates the best course of action.

Figure 1 - *Weight optimization procedure based on sensitivity analysis.*



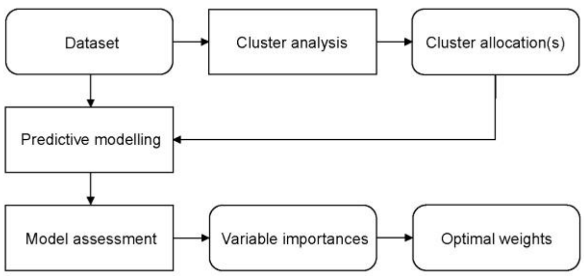
2.2. Unsupervised Learning Approach

The techniques such as PCA or factor analysis encompass extracting a small number of principal components (factors) that capture sufficiently large variation in the data and evaluating the relative correlation of each variable with the identified components. In this context, the importance of a variable reflects how well it might help in explaining the major variance in the multidimensional data space. A standard procedure when employing PCA/factor analysis to build composite indicators is using the loadings on the first component as the weights assigned to the variables (Klasen, 2000; Greyling and Tregenna, 2017).

Despite of popularity, the use of PCA and factor analysis is limited in the case of low correlation among the feature dimensions. To handle this problem, we suggest using cluster analysis as an alternative solution in measuring variable importance. The notion of variable importance in clustering refers to the extent to which individual features contribute to the formation of distinct clusters. Given a cluster structure, it is possible to fit a model to predict the cluster labels from the input features, and variable importance can be calculated as the mean decrease in prediction accuracy when a particular feature is permuted (Breiman, 2001). This value provides information regarding the variable's efficacy in discriminating data points based on common patterns. The weights derived from the cluster-based technique thus reflect the ability of input features to distinguish observations in a multidimensional context.

The illustration of the weighting procedure based on cluster analysis is depicted in Figure 2, with the inputs/outputs presented in the rounded boxes and the used functions given in the rectangle boxes. The key idea behind this approach is to train a mapping function employing the variables in the dataset to predict the cluster membership, which is derived from the clustering process. By evaluating the mean decrease in prediction accuracy by each variable, the importance measures can be defined and then normalized to obtain the weights for the composite indicator.

Figure 2 - *Weight optimization procedure based on cluster analysis.*



3. Data

Table 1 provides a concise overview of the dimensions utilized to create the composite indicator for measuring well-being in European regions. According to the OECD well-being framework (OECD, 2020), the dimensions of income and jobs pertain to material conditions that shape people's economic sustainability. Whereas the aspects of health, education, environment, and safety refer to the fundamental measures of life quality. Civic engagement reflects the degree of public trust in government and of voters' participation in the political process. The society dimension measures the severity of social exclusion, which is represented by the share of young people not in employment, education, or training (NEET). Lastly, digital accessibility is an essential topic to be considered, as it promotes social inclusion, access to information resources, economic opportunities, and personal empowerment for individuals with disabilities.

Table 1 – *Dimension structure for measuring regional well-being.*

Dimension	Measuring indicator	
Income	*income:*	household disposable income per capita (real USD PPP)
Jobs	*emp_rate:* *unemp_rate:*	employment rate (%) unemployment rate (%)
Health	*life_exp:* *mort_rate:*	life expectancy at birth (years) age adjusted mortality rate (per 1000 population)
Education	*sec_edu*:	share of population from 25-64 years old with at least secondary education (%)
Environment	*air_pol:*	air pollution in $PM_{2.5}$ (average level in µg/m^3 experienced by the population)
Safety	*hom_rate:*	intentional homicide rate (per 100 000 population)
Civic engagement	*vote:*	voter turnout to general elections (%)
Society	*soc_exc:*	share of population from 18-24 years old not in employment and not in any education and training (%)
Digital accessibility	*bb_acc:*	share of households with broadband access (%)

Owing to the fact that the eleven indicators are gauged in different units, it is required to make them comparable by converting all the features into the same scale $[0,1]$ using the formulas:

$$\bar{x} = \frac{x - \min(x)}{\max(x) - \min(x)}, \text{ or} \qquad (4)$$

$$\bar{x} = \frac{max(x) - x}{\max(x) - \min(x)} \qquad (5)$$

The min-max normalization (4) is applied to the transformation of features that are positively correlated with well-being, including income, employment rate, life expectancy, education attainment, voter turnout, and broadband access. On the contrary, the max-min normalization (5) is implemented for the remaining features, which are considered to exert a negative effect on well-being. If a dimension is constituted by a pair of indicators, such as jobs and health, the dimension score is computed by averaging the normalized values of both components, then applying the min-max normalization again to ensure conformity to the identical scale.

The data for the well-being features is collected from the OECD Regional Statistics database (OECD, 2023), which encompasses yearly time-series for the variables of demography, economy, labor market, social and innovation themes in the OECD member countries. The original dataset comprises 2250 observations, measuring the eleven well-being factors for 225 large (TL2) regions in 29 European countries over a ten-year period from 2010 to 2019. We performed few data replacement for the Netherlands and Greece, where the features of homicide rate and broadband access are not available at the regional level in some time ranges, by utilizing the figures at their national level. We decided to remove from the original dataset the regions that are subjected to the following two conditions: the number of missing values is greater than 15% of the total data cells; and the information in at least one variable is completely unavailable. With these criteria, 27 regions from 11 countries were disposed of, leading to the absence of five countries including Bulgaria, Ireland, Iceland, Lithuania, and Malta.

To thoroughly address the problem of missing data, we applied the k-nearest neighbors (k-NN) imputation technique with the Euclidean distance metric. An encoding technique was also employed to take into account the regional and temporal effects. The factor variables of time and location were first converted using dummy encoding, subsequently fed into the k-NN algorithm along with all the other indicators to define the proximate data points of observations with missing information. Following this, the unknown cells were filled by a distance-weighted average of the values from their closest neighbors. Finally, the complete features in the imputed dataset were used to compute the nine well-being dimensions.

4. Well-being Scores by Composite Indicators

To establish the composite indicator for regional well-being, we used the weighted arithmetic mean function

$$C^{(r,t)} = \sum_{i=1}^{n} w_i X_i^{(r,t)}, \tag{6}$$

where C is the composite score, X_i is the normalized score in dimension i, w_i is the weight assigned to dimension i, and the term (r, t) denotes the region and time allocated to the observation. Table 2 presents the sets of weights for the well-being composite indicator derived from four different weighting methods. The first column simply contains the equal weights, which can be used as a baseline for comparison. The second column shows the weights calculated from the standard weighting approach using PCA. The values in the third column are the result from the sensitivity-based weighting procedure using the full main Sobol' index (S_i^{full}). The last column displays the clustering-based weights utilizing permutation importance estimated by a random forest classifier for the three-cluster allocation.

Table 2 – *Weights by different weighting methods.*

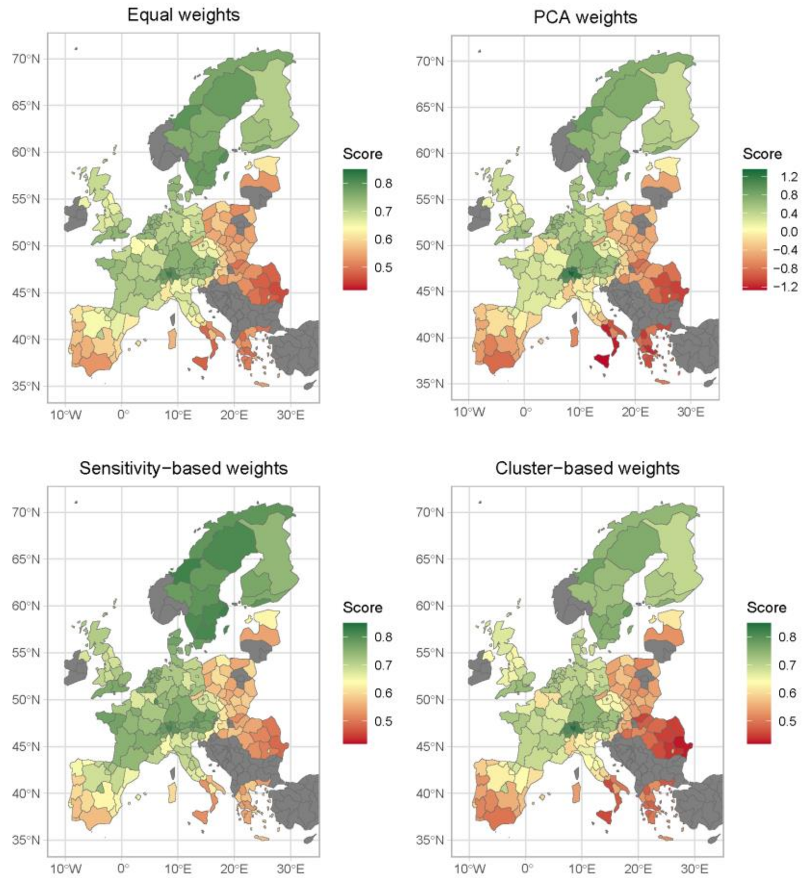
	Equal	PCA[1]	Sensitivity-based[2]	Cluster-based[3]
Income	0.111	0.128	0.067	0.128
Jobs	0.111	0.161	0.091	0.118
Health	0.111	0.079	0.133	0.176
Education	0.111	0.103	0.138	0.165
Environment	0.111	0.084	0.130	0.101
Safety	0.111	0.047	0.155	0.048
Civic engagement	0.111	0.081	0.117	0.067
Society	0.111	0.161	0.090	0.093
Digital accessibility	0.111	0.156	0.081	0.104
Sum	1.000	1.000	1.000	1.000

There are differences in the level of importance each method assigns to the variables, as manifested through the corresponding weights. This inconsistency is the result of the distinct mechanism employed by each method. However, a common pattern in the allocation of well-being composite scores is found (see Figure 1) due to the compensation between the features in the aggregation process. Northern Europe is the area that displays the highest average scores in well-being, followed by Western European regions and the British Isles. In Eastern Europe, most territories show a welfare level below the regional average, and places in the southeast notably exhibit the lowest scores throughout the entire observed regions. In Southern Europe, a north-south gradient in well-being is visible, where the northern part produces mostly above-moderate scores while the southern part

[1] Weights are the normalized factor loadings of the first principal component (37% of total variance).
[2] Weights are the normalized S_i^{full} estimated by penalized cubic splines via generalized cross-validation. Note that $S_i^{full} = ST_i^{full}$ in purely addictive models.
[3] Weights are the normalized permutation importances calculated by a random forest ensemble that grows 500 trees and randomly samples three features as candidates at each split. The allocation in the three-cluster solution by k-means clustering is chosen as the response based on the elbow method.

predominantly records subpar status. At the TL2 regional details, the regions in Switzerland, Sweden, and Norway consistently show a superior level of well-being while the most deprived ones are in Eastern Romania, Southern Italy, and Greece.

Figure 1 – *Average well-being composite scores for European regions in the period 2010-2019.*



Note: the grey areas denote territories with missing in- formation or territories not in European regions.

With respect to national well-being, Figure 2 shows the alterations in rankings for 24 European countries in the observed period. A country's score is computed by taking the population-weighted average of the scores from all its constituent regions. Switzerland, Sweden, Norway, and Luxembourg consistently rank at the highest positions in the charts. These nations tend to display robust performance on the composite indicator regardless of the weighting schemes used, proving their status as the most-welfare countries in Europe. On the other hand, the countries of Romania, Latvia, Hungary, and Greek frequently appear at the lower end of the rankings. This evidence implies that these nations face challenges in various well-being dimensions, which diminishes their overall prosperity compared to the other European members.

Figure 2 – *Annual rankings by national average scores in well-being.*

5. Conclusion

This paper introduces two innovative weighting procedures for composite indicators, focusing on the quantification of variable importance across diverse conceptual frameworks. The first approach, sensitivity-based weighting, enables researchers to derive a solution wherein the magnitude of weights corresponds to the contribution of input features to the variance in composite scores. This method is designed to work compatibly with any single-valued function, independent of its complexity and parameter configuration, making it applicable to all models that return a scalar output. The second approach, cluster-based weighting, addresses multidimensional challenges by investigating the underlying cluster structure in data and estimating the impact of each dimension on predicting cluster membership. The optimal number of clusters can be determined through clustering validation indices or by examining the association between various clustering schemes and the performance of prediction models. The cluster-based weights obtained from this process can serve as a measure of each variable's ability to differentiate observations in the multidimensional space representing the phenomena of interest.

We have developed a composite indicator for measuring the well-being of inhabitants in large European regions. An imputed dataset containing information on nine well-being dimensions for 198 regions during the 2010-2019 period was used for computing the composite scores. Four weighting methods were employed, including equal weighting, PCA weighting, and the two novel techniques proposed in our study. Despite the variations in weighting schemes and score outcomes, all four methods collectively reveal similar patterns of welfare allocation throughout the regions under evaluation. Regions in Northern Europe exhibit the highest average well-being scores, followed by their Western European counterparts and those within the British Isles. Southern Europe holds the third position with a clear north-south differentiation while Eastern European regions experiences the lowest levels of well-being. At the level of national well-being, Switzerland, Sweden, Norway, and Luxembourg maintain their prominence by consistently securing top positions in the annual ranking charts, whereas Romania, Latvia, Hungary, and Greek frequently appear as the most deprived nations in these figures.

References

BECKER W., SAISANA M., PARUOLO P., VANDECASTEELE I. 2017. Weights and Importance in Composite Indicators: Closing the Gap, *Ecological Indicators*, Vol. 80, pp. 12-22.

BREIMAN L. 2001. Random Forests, *Machine Learning*, Vol. 45, pp. 5-32.

GREYLING T., TREGENNA F. 2017. Construction and analysis of a composite quality of life index for a region of South Africa, *Social Indicators Research*, Vol. 131, pp. 887-930.

GRUPP H., MOGEE M. E. 2004. Indicators for National Science and Technology Policy: How Robust are Composite Indicators? *Research Policy*, Vol. 33, No. 9, pp. 1373-1384.

KLASEN S. 2000. Measuring Poverty and Deprivation in South Africa, *Review of Income and Wealth*, Vol. 46, No. 1, pp. 33-58.

MARA T. A., TARANTOLA S., ANNONI P. 2015. Variance-Based Sensitivity Indices for Models with Dependent Inputs, *Environmental Modelling & Software*, Vol. 72, pp. 173-183.

NARDO M., SAISANA M., SALTELLI A., TARANTOLA S. 2005a. Tools for Composite Indicators Building. *Technical Report EUR 21682 EN*, JRC-EC, Ispra, Italy.

NARDO M., SAISANA M., SALTELLI A., TARANTOLA S., HOFFMAN A., GIOVANNINI E. 2005b. Handbook on Constructing Composite Indicators: Methodology and User Guide. *OECD Statistics Working Paper No. 2005/03*, OECD Publishing.

NICOLETTI G., SCARPETTA S., BOYLAUD O. 2000. Summary Indicators of Product Market Reg- Ulation with an Extension to Employment Protection Legislation. *ECO Working Papers No. 226*, OECD.

OECD. 2020. *How's Life? 2020: Measuring Well-being*. OECD Publishing.

OECD. 2023. OCED Regional Statistics (database). https://doi.org/10.1787/region-data-en, accessed on 01 march 2023.

SAISANA M., SALTELLI A., TARANTOLA S. 2005. Uncertainty And Sensitivity Analysis Techniques as Tools For The Quality Assessment Of Composite Indicators, *Journal of the Royal Statistical Society: Series A*, Vol. 168, No. 2, pp. 307-323.

SALTELLI A. 2002. Sensitivity Analysis for Importance Assessment, *Risk Analysis*, Vol. 22, No. 3, pp.579-590.

SOBOL' I. M. 2001. Global Sensitivity Indices for Nonlinear Mathematical Models and their Monte Carlo Estimates, *Mathematics and Computers in Simulation*, Vol. 55, No. 1-3, pp. 271-280.

Viet Duong NGUYEN, University of Insubria, vdnguyen@uninsubria.it
Chiara GIGLIARANO, University Carlo Cattaneo, cgigliarano@liuc.it