WEIGHTING IN COMPOSITE INDICES CONSTRUCTION: THE CASE OF THE MAZZIOTTA-PARETO INDEX[1]

Matteo Mazziotta, Adriano Pareto

1. Introduction

Assigning weights to indicators is a very difficult operation and not without risks of a conceptual and methodological nature (Booysen, 2002; Salzman, 2003; OECD, 2008). Even no weighing still means assigning a weight, i.e., the same for all indicators (Greco et al., 2019).

The issue of choosing a weighting system for individual indicators that represents their different importance in expressing the phenomenon considered, necessarily involves the introduction of an arbitrary component. Subjective weighting can be adopted, implicitly, by assigning the same weight to all components (equal weights) or, explicitly, by means of a group of experts who establish the weight of each elementary indicator. Alternatively, an objective weighting can be adopted, implicitly, by choosing a normalization method that assigns a weight proportional to the variability of the elementary indicators or, explicitly, by calculating the weights using a multivariate analysis method, such as Principal Components Analysis (PCA).

The purpose of the explicit weighting is that each weight should represent the relative theoretical importance of the corresponding individual indicator. The explicit weights assigned to the individual indicators heavily influence the values of the composite index. Hence, the weights should be defined on the basis of a well-defined theoretical framework.

The techniques most used to explicitly weight the individual indicators are the following: a) no weighting or assignment of 'equal' weights, b) subjective or expert weighting, and c) weighting by PCA (Mazziotta and Pareto, 2017).

In the case a) if no explicit weight is defined, in addition to the implicit weighting induced by normalization, the individual indicators are weighted with equal weights. This implies that all the components of the composite index have the same importance, except for the implied weight, and this may not be correct.

[1] The paper is the result of the common work of the authors: in particular M. Mazziotta has written Sections 1, 4 and A. Pareto has written Sections 2, 3.

However, if there are no precise theoretical or empirical reasons for choosing different weights, this can be a valid solution in various contexts.

In the case b), subjective or expert weighting is an arbitrary weighting carried out by the researcher or by specialists in the phenomenon who define the weight of each individual indicator. The values obtained are then summarized using a specific function. Sometimes, the weights are defined by policy makers or through sample surveys in which the interviewed is asked to evaluate the importance of the various aspects that make up the phenomenon.

In the case c) PCA can be used to define the weights of the individual indicators by means of the coefficients obtained for the first main component. This empirical solution is relatively more objective than the others and has the advantage of considering the set of weights that ‘explain’ most of the variance of the indicators. However, the reliability of the weights obtained depends on the variance explained by the first component and on the structure of the correlations between the individual indicators which does not remain constant over time. And above all, it is absolutely not true that variability is a synonym of theoretical importance of the individual indicators.

In short, as mentioned, the issue of weighting is very complex, and each solution has advantages and disadvantages. Moreover, if it is true (and it is true) that the perfect composite index does not exist then it is equally true that it is not possible to have a system of weights without arbitrariness.

In this paper, the methodological problem of the assignment of weights is faced with respect to the two versions of the Mazziotta-Pareto Index (Mazziotta and Pareto, 2016).

2. Weighting the Mazziotta-Pareto Index

The Mazziotta-Pareto Index (MPI) - and its variant Adjusted MPI (AMPI) - is a composite index for summarizing a set of indicators that are assumed to be not fully substitutable. It is based on a non-linear function which, starting from the simple arithmetic mean of the normalized indicators, introduces a penalty for the units with unbalanced values of the indicators (Mazziotta and Pareto, 2016). This methodology is often applied to the calculation of both non-compensatory[2] composite indices of ‘positive’ phenomena, such as well-being and sustainable development, and ‘negative’ phenomena, such as poverty.

[2] In a non-compensatory approach, all the dimensions of the phenomenon must be balanced and an aggregation function that takes unbalance into account, in terms of penalization, is used (Casadio Tarabusi and Guarini, 2013).

In the MPI and AMPI, all components are assumed to have equal importance, which may not be the case (Boysen, 2002; Mazziotta and Pareto, 2020). In this Section a weighted version of the two indices (WMPI and WAMPI, where W stands for "Weighted") is proposed, when a set of weights is available.

2.1. The WMPI

Given the matrix $\mathbf{X}=\{x_{ij}\}$ with n rows (statistical units) and m columns (individual indicators), we calculate the standardized matrix $\mathbf{Z}=\{z_{ij}\}$ as follows:

$$z_{ij} = 100 \pm \frac{(x_{ij} - \mathrm{M}_{x_j})}{\mathrm{S}_{x_j}} 10 \qquad (1)$$

where M_{x_j} and S_{x_j} are, respectively, the mean and standard deviation of the indicator j and the sign ± is the polarity[3] of the indicator j.

If a set of weights w_j (j=1, ..., m) is available; where w_j is the weight of individual indicator j ($0 < w_j < 1$) and $\sum_{j=1}^{m} w_j = 1$, we can calculate the weighted mean and weighted standard deviation of the standardized values of unit i (i=1, ..., n) as follows:

$$\overline{\mathrm{M}}_{z_i} = \sum_{j=1}^{m} z_{ij} w_j \; ;$$

$$\bar{\mathrm{S}}_{z_i} = \sqrt{\sum_{j=1}^{m} (z_{ij} - \mathrm{M}_{z_i})^2 w_j}$$

Denoting with $\overline{\mathrm{cv}}_i = \bar{\mathrm{S}}_{z_i} / \overline{\mathrm{M}}_{z_i}$ the weighted coefficient of variation (Sheret, 1984) for unit i, the generalized form of the WMPI is given by:

$$\mathrm{WMPI}_i^{+/-} = \overline{\mathrm{M}}_{z_i} \pm \bar{\mathrm{S}}_{z_i} \overline{\mathrm{cv}}_i \qquad (2)$$

[3] The polarity of an individual indicator is the sign of the relation between the indicator and the phenomenon to be measured (+ if the individual indicator represents a dimension considered positive and - if it represents a dimension considered negative).

where $\overline{\mathrm{M}}_{z_i}$ is the 'mean level', $\bar{\mathrm{S}}_{z_i}\overline{\mathrm{cv}}_i$ is the 'penalty' (i.e., the 'horizontal variability')[4] and the sign ± depends on the kind of phenomenon to be measured. If it is 'positive', then the WMPI⁻ is used; else the WMPI⁺ is used.

If $w_j = \dfrac{1}{m}$ (j=1, ..., m), we have: $\mathrm{WMPI}_i^{+/-} = \mathrm{MPI}_i^{+/-}$ (i.e., the classical MPI).

2.2. The WAMPI

Given the matrix **X**={x_{ij}}, we calculate the normalized matrix **R**={r_{ij}} as follow:

$$r_{ij} = 100 \pm \frac{x_{ij} - \mathrm{Ref}_{x_j}}{\mathrm{Max}_{x_j} - \mathrm{Min}_{x_j}} 60 \tag{3}$$

where Min_{x_j} and Max_{x_j} are the 'goalposts' for indicator j. (e.g., the minimum and maximum of indicator j), Ref_{x_j} is a reference value[5] for indicator j (e.g., the mean of indicator j) and the sign ± depends on the polarity of indicator j.

Denoting with $\overline{\mathrm{M}}_{r_i}$ and $\bar{\mathrm{S}}_{r_i}$, respectively, the weighted mean and weighted standard deviation of the normalized values of unit i, the generalized form of WAMPI is given by:

$$\mathrm{WAMPI}_i^{+/-} = \overline{\mathrm{M}}_{r_i} \pm \bar{\mathrm{S}}_{r_i}\overline{\mathrm{cv}}_i \tag{4}$$

where $\overline{\mathrm{cv}}_i = \bar{\mathrm{S}}_{r_i} / \overline{\mathrm{M}}_{r_i}$ is the weighted coefficient of variation for unit i.

If $w_j = \dfrac{1}{m}$ (j=1, ..., m), we have: $\mathrm{WAMPI}_i^{+/-} = \mathrm{AMPI}_i^{+/-}$ (i.e., the classical AMPI).

---

[4] The penalty is a function of the indicators' variability in relation to the mean value and is used to penalize the units. The aim is to reward the units that, mean being equal, have a greater balance among the individual indicators.

[5] Note that the reference value is very important, since the set of reference values of all the indicators defines the 'balancing model' (Mazziotta and Pareto, 2021).

3. Some numerical examples

In this section, we present, through some numerical examples, the calculation of WMPI and WAMPI with negative penalty. Similar results are obtained for the two versions with positive penalty.

Table 1 – *Computing the WMPI⁻ con different weights.*

Unit	Original indicators			Normalized indicators			Mean	Std	CV	Penalty	WMPI⁻	
	X_1	X_2	X_3	Z_1	Z_2	Z_3					Value	Rank
			Weights	**0.33**	**0.33**	**0.33**						
1	110	1	0.4	114.1	87.8	100.0	100.6	10.8	0.107	1.16	99.5	3
2	90	3	0.2	107.1	108.2	90.9	102.0	7.9	0.077	0.61	101.4	2
3	70	3	0.8	100.0	108.2	118.3	108.8	7.5	0.069	0.51	108.3	1
4	50	3	0.2	92.9	108.2	90.9	97.3	7.7	0.079	0.61	96.7	4
5	30	1	0.4	85.9	87.8	100.0	91.2	6.3	0.069	0.43	90.8	5
r	0.545	0.599	0.587									
			Weights	**0.60**	**0.20**	**0.20**						
1	110	1	0.4	114.1	87.8	100.0	106.0	10.7	0.101	1.07	105.0	1
2	90	3	0.2	107.1	108.2	90.9	104.0	6.6	0.063	0.42	103.6	3
3	70	3	0.8	100.0	108.2	118.3	105.3	7.2	0.069	0.49	104.8	2
4	50	3	0.2	92.9	108.2	90.9	95.6	6.4	0.066	0.42	95.1	4
5	30	1	0.4	85.9	87.8	100.0	89.1	5.5	0.062	0.34	88.7	5
r	0.894	0.328	0.304									
			Weights	**0.20**	**0.60**	**0.20**						
1	110	1	0.4	114.1	87.8	100.0	95.5	10.5	0.110	1.15	94.3	4
2	90	3	0.2	107.1	108.2	90.9	104.5	6.8	0.065	0.45	104.0	2
3	70	3	0.8	100.0	108.2	118.3	108.6	5.8	0.053	0.31	108.2	1
4	50	3	0.2	92.9	108.2	90.9	101.7	8.0	0.079	0.63	101.0	3
5	30	1	0.4	85.9	87.8	100.0	89.8	5.1	0.057	0.29	89.5	5
r	0.266	0.912	0.310									
			Weights	**0.20**	**0.20**	**0.60**						
1	110	1	0.4	114.1	87.8	100.0	100.4	8.4	0.083	0.70	99.7	2
2	90	3	0.2	107.1	108.2	90.9	97.6	8.2	0.084	0.69	96.9	3
3	70	3	0.8	100.0	108.2	118.3	112.6	7.4	0.066	0.49	112.1	1
4	50	3	0.2	92.9	108.2	90.9	94.7	6.8	0.071	0.48	94.3	5
5	30	1	0.4	85.9	87.8	100.0	94.7	6.5	0.069	0.44	94.3	4
r	0.286	0.302	0.909									

In Table 1, the $WMPI^-$ is calculated for a matrix **X** of 5 statistical units and 3 incorrelated individual indicators, with four different sets of weights. In the first case, an equal weighting approach is followed (0.33 for each individual indicator), and we have $WMPI^- = MPI^-$. In the other three cases, we give the greatest weight (0.60), each time, to a different individual indicator and equal weights (0.20) to the other two. For each case, the table reports the individual indicators (X_1-X_3), the normalized indicators (Z_1-Z_3), the weighted mean, standard deviation and coefficient of variation and the WMPI$^-$ (value and rank). In the last row is also shown the correlation (*r*) between the WMPI$^-$ and the original indicators.

Table 2 – *Computing the WAMPI$^-$ con different weights.*

Unit	Original indicators			Normalized indicators			Mean	Std	CV	Penalty	WAMPI$^-$	
	X_1	X_2	X_3	Z_1	Z_2	Z_3					Value	Rank
			Weights	**0.33**	**0.33**	**0.33**						
1	110	1	0.4	130.0	64.0	100.0	98.0	27.0	0.275	7.43	90.6	4
2	90	3	0.2	115.0	124.0	80.0	106.3	19.0	0.178	3.39	102.9	2
3	70	3	0.8	100.0	124.0	140.0	121.3	16.4	0.135	2.23	119.1	1
4	50	3	0.2	85.0	124.0	80.0	96.3	19.7	0.204	4.02	92.3	3
5	30	1	0.4	70.0	64.0	100.0	78.0	15.7	0.202	3.18	74.8	5
r	0.407	0.739	0.535									
			Weights	**0.60**	**0.20**	**0.20**						
1	110	1	0.4	130.0	64.0	100.0	110.8	26.1	0.236	6.16	104.6	3
2	90	3	0.2	115.0	124.0	80.0	109.8	15.3	0.139	2.13	107.7	2
3	70	3	0.8	100.0	124.0	140.0	112.8	16.5	0.146	2.41	110.4	1
4	50	3	0.2	85.0	124.0	80.0	91.8	16.2	0.177	2.86	88.9	4
5	30	1	0.4	70.0	64.0	100.0	74.8	12.8	0.171	2.19	72.6	5
r	0.823	0.472	0.310									
			Weights	**0.20**	**0.60**	**0.20**						
1	110	1	0.4	130.0	64.0	100.0	84.4	26.7	0.317	8.46	75.9	4
2	90	3	0.2	115.0	124.0	80.0	113.4	17.1	0.150	2.57	110.8	2
3	70	3	0.8	100.0	124.0	140.0	122.4	12.8	0.105	1.34	121.1	1
4	50	3	0.2	85.0	124.0	80.0	107.4	20.4	0.190	3.87	103.5	3
5	30	1	0.4	70.0	64.0	100.0	72.4	14.0	0.193	2.70	69.7	5
r	0.140	0.955	0.253									
			Weights	**0.20**	**0.20**	**0.60**						
1	110	1	0.4	130.0	64.0	100.0	98.8	20.9	0.212	4.43	94.4	2
2	90	3	0.2	115.0	124.0	80.0	95.8	19.6	0.204	3.99	91.8	3
3	70	3	0.8	100.0	124.0	140.0	128.8	15.7	0.122	1.91	126.9	1
4	50	3	0.2	85.0	124.0	80.0	89.8	17.2	0.192	3.30	86.5	4
5	30	1	0.4	70.0	64.0	100.0	86.8	16.3	0.188	3.05	83.7	5
r	0.241	0.399	0.885									

As we can see, when the individual indicators have the same weight (0.33), the correlations with the WMPI$^-$ (that is the MPI$^-$) are very similar. In particular, we have r(WMPI$^-$, X_1)=0.545; r(WMPI$^-$, X_2)=0.599 and r(WMPI$^-$, X_3)=0.587.

On the contrary, when one individual indicator has a weight greater (i.e., is more important) than the others (0.66), the WMPI$^-$ is biased towards it. For instance, when w_1=0.60, w_2=0.20 w_3=0.20, we have r(WMPI$^-$, X_1)=0.894;

r(WMPI$^-$, X_2)=0.328 and *r*(WMPI$^-$, X_3)=0.304. Therefore, the ranking according to the WMPI$^-$ is much the same as that based on the most important individual indicator.

However, it is interesting to note that when individual indicators have different weights, the penalty changes. Thus, one unit that is more balanced with a certain set of weights (and then less penalized) may be more unbalanced with another set of weights (and then more penalized). It is the case of unit 1 that, with equal weighting, has a penalty of 10.8·0.107=1.16; whereas with weight w_3=0.60 has a penalty of 8.4·0.083=0.70[6].

Table 2 shows the calculation of the WAMPI$^-$, where a different normalization method is used (see formula 3). In this case, the 'goalposts' for each indicator are the minimum and maximum, and the reference value is the mean. Moreover, normalized indicators have the same range (and not the same variance as in the WMPI$^-$) and then the penalties are larger overall. Nevertheless, the result does not change and also the WAMPI$^-$ is most correlated with the individual indicators with the greatest weight. Indeed, if w_1=0.60 then *r*(WAMPI$^-$, X_1)=0.823, if w_2=0.60 then *r*(WAMPI$^-$, X_2)=0.955, and if w_3=0.60 then *r*(WAMPI$^-$, X_3)=0.885.

4. Concluding remarks

A composite index is a mathematical combination (or aggregation as it is defined) of a set of elementary indicators (or variables) that represent the different components of a multidimensional phenomenon to be measured (e.g., development, well-being, quality of life, corruption, etc.). Therefore, composite indices are used to measure concepts that cannot be captured by a single indicator.

Ideally, a composite index should be based on a theoretical framework that allows individual indicators to be selected, combined and weighted to reflect the size or structure of the phenomenon being measured. However, its construction is not simple and often requires a series of decisions / choices (methodological or not) to be made.

The decision to weigh in the same way both the individual indicators within any domains and the composite domain indices that will subsequently calculate the composite of the composites is justified by the arbitrariness that would be introduced if an objective approach were used (statistical techniques) or subjective (panel of experts) of weight assignment. However, weighing in the same way is a

[6] Note that with w_2=0.6 the penalty of unit 1 is similar to the penalty with equal weighting (1.15 vs 1.16), since the weighted standard deviation is less, but the weighted coefficient of variation is greater, and therefore the product does not change.

"non-neutral" choice, since it is decided to place both individual indicators and composite domain indices on the same level of importance.

The need to adapt these two composite indices with a system of weights arose during the pandemic when numerous international institutions had the need to measure the performance of the health system during the Covid emergency. In this context, the delicacy of the topic dealt with and the different nature of the individual indicators have focused attention on the different theoretical importance of the indicators themselves.

Introducing a weighting system in the MPI and AMPI allows to summarize a set of partially substitutable indicators that are assumed to have different importance. The experimentation of assigning weights to the MPI and the AMPI has brought comforting results, since the final composite indices actually undergo changes as a function of the intensity of the weight.

Certainly, the paper does not solve the problem of which system of weights is more appropriate, but it demonstrates how these two methodologies can be adaptable to a set of weights coming from objective and subjective approaches.

However, some aspects need to be further investigated. In particular:

- how the properties of the indices change with the introduction of weights;
- how outliers in indicators with high weights affect the values of the indices;
- how much the values of the indices differ from the values of other aggregation functions, such as the weighted geometric mean.

References

BOOYSEN F. 2002. An overview and evaluation of composite indices of development, *Social Indicators Research*, Vol. 59, pp. 115–151.

CASADIO TARABUSI E., GUARINI G. 2013. An unbalance adjustment method for development indicators, *Social Indicators Research*, Vol. 112, pp. 19–45.

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 Indicators Research*, Vol. 141, pp. 61–94.

MAZZIOTTA M., PARETO A. 2016. On a Generalized Non-compensatory Composite Index for Measuring Socio-economic Phenomena, *Social Indicators Research*, Vol. 127, pp. 983–1003.

MAZZIOTTA M., PARETO A. 2017. Synthesis of Indicators: The Composite Indicators Approach, in F. MAGGINO (eds.), *Complexity in Society: From Indicators Construction to their Synthesis*. Cham: Springer.

MAZZIOTTA M., PARETO A. (a cura di) 2020. *Gli indici sintetici*. Torino: Giappichelli.

MAZZIOTTA M., PARETO A. 2021. Everything you always wanted to know about normalization (but were afraid to ask). *Italian Review of Economics, Demography and Statistics*, Vol. LXXV, No.1, pp. 41–52.

OECD 2008. *Handbook on Constructing Composite Indicators. Methodology and user guide*. Paris: OECD Publications.

SALZMAN J. 2003. *Methodological choices encountered in the construction of composite indices of economic and social well-being*. Technical Report, Center for the Study of Living Standards, Ottawa.

SHERET M. 1984. The Coefficient of Variation: Weighting Considerations, *Social Indicators Research*, Vol. 15, pp. 289–295.

SUMMARY

This paper presents a weighted version of the Mazziotta-Pareto Index (MPI) and Adjusted MPI (AMPI). Since the MPI and AMPI are based on the calculation of the mean and standard deviation of normalized values, for each unit, we calculate the weighted mean and weighted standard deviation of normalized values. The weighted coefficient of variation is then obtained by simply dividing the weighted standard deviation by the weighted mean. Finally, the two standard formulas can be applied. Some numerical examples are also shown, in order to assess the effect of different weighting schemes on the results.

Matteo MAZZIOTTA, Istat, mazziott@istat.it
Adriano PARETO, Istat, pareto@istat.it