METHODS FOR CONSTRUCTING COMPOSITE INDICES: ONE FOR ALL OR ALL FOR ONE?[1]

Matteo Mazziotta, Adriano Pareto

1. Introduction

In last years, the debate on the measurement of multidimensional phenomena has caused, within the worldwide scientific Community, a renewed interest thanks to the publication, in September 2009, of the Stiglitz report and, in March 2013, of the first report on "Equitable and Sustainable Well-being" (BES) by the Committee composed by Istat (Italian National Institute of Statistics) and CNEL (Italian Council for Economics and Labour). It is common awareness that a number of socio-economic phenomena cannot be measured by a single descriptive indicator and that, instead, they should be represented with multiple dimensions. Phenomena such as development, progress, poverty, social inequality, well-being, quality of life, provision of infrastructures, etc., require, to be measured, the 'combination' of different dimensions, to be considered together as the proxy of the phenomenon. This combination can be obtained by applying methodologies known as composite indices (Salzman, 2003; Mazziotta and Pareto, 2011).

This paper addresses the problem of summarizing a set of socio-economic indicators and aims to provide some general guidelines for the construction of a composite index. In particular, the attention is focused on the search of the most suitable method depending on the following factors: type of indicators (substitutable/non-substitutable), type of aggregation (simple/complex), type of comparisons to be made (relative/absolute), type of weights of the indicators (subjective /objective). As is known, in fact, building a composite index is a delicate task and full of pitfalls: from the obstacles regarding the availability of data and the choice of individual indicators, to their treatment in order to compare (normalization) and aggregate them (weighting and aggregation). Despite the problems mentioned, the composite indices are widely used by several international organizations for meas-

[1] The paper is the result of combined work of the authors: M. Mazziotta has written Sects. 1, 2 and 5; A. Pareto has written Sects. 3 and 4.

uring economic, environmental and social phenomena and, therefore, they provide an extremely relevant tool and in the course of evolution (OECD, 2008).

2. Well-being measures and composite indices

In recent years, there has been a lively debate about the use of the most famous indicator of well-being: the "Gross Domestic Product" (GDP). For decades, the economic measure for excellence is not able to represent the well-being or the progress of a society, much less to express the quality of life of a geographical area or a community. This debate has produced worldwide, a considerable literature which can be detected more than a hundred alternative indices, adopted by government organizations (and others), academia and business press (Bandura, 2008), but despite this, it seems that the popularity of GDP has not been minimally scratched.

In fact, the GDP is based on very solid theoretical bases, while many alternative indices guilty of clarity from the stage of definition of the phenomenon; in many circumstances (for example, the Human Wellbeing Index or HWI; Prescott-Allen, 2001), not having a shared theory behind , taking into account dozens of indicators so that all possible aspects are considered. This approach raises a number of important problems related to the allocation of weights and the aggregation of many variables. The only alternative that has been successful, globally, is the Human Development Index or HDI (UNDP, 2010); it is published annually by the United Nations and it considers 3 individual indicators: "Life expectancy at birth", "Education" and "GDP per capita". It is a composite index itself (expressed in absolute form) since the 3 indicators were aggregated, until 2009, through a simple arithmetic mean. This aggregation function has attracted much criticism since the arithmetic mean performs a compensation between indicators that are not-substitutable for each other. In the previous method of calculating the HDI, a high value of the "GDP per capita" could compensate a low value of the "Life expectancy at birth" or vice versa. Since 2010 United Nations Development Programme (UNDP) Report, the index has changed aggregation method: the arithmetic mean has been replaced with a geometric mean. In this way the problems of compensation are solved, but other problems are introduced, for example the case in which there are null values (or very close to zero). A further difficulty is linked to the same nature of the geometric mean, which assumes a multiplicative relationship of the variables rather than an additive, as stated in the assumptions of the HDI calculation.

As regards the measurement of poverty, rather than well-being or progress, in the Human Poverty Index or HPI (UNDP, 2007) a solid theoretical basis is largely respected. The index is based on the capabilities of Sen (1985); for developing countries, 3 non-substitutable indicators are chosen ("Deprivation in longevity", "Deprivation in knowledge" and "Deprivation of a decent standard of living") and the aggregation function adopted, the mean of order three, does not allow compensation among them.

Many scientists dispute the use of composite indices that lead to the determination of a single value for each geographic area, preferring the so-called dashboard (as in the case of monitoring the state of health of a vehicle: oil level, gasoline, water temperature, etc.). In the case of dashboard, it is possible to identify various dimensions of the phenomenon, all relevant, without that they are further aggregated. From the statistical point of view, it is an incontrovertible choice but from the standpoint of political and media is an heavy limitation. The easy-disclosure in the media and the immediate understanding by the user are certainly the strengths of a unique index.

One of the indices with greater media coverage in Italy is the measure of the Quality of Life (QoL) which, every year, the economic newspaper "Il Sole 24ore" publishes at the provincial level: in this case, 6 relevant dimensions ("Living standard", "Job and business", "Environment and health", "Public order", "Population", "Free time") are identified and, for each dimension, 6 representative indicators are considered for a total, therefore, of 36 individual indicators. After a phase of normalization of the indicators, the arithmetic mean is calculated within each dimension, and thereafter the arithmetic mean between the dimensions. In essence, the final index is calculated as the arithmetic mean of the 36 normalized indicators (Lun *et al.*, 2006).

Also in Italy, since 2003, the "Campaign Sbilanciamoci!" has published the Index of the Regional Quality of Development (QUARS) with the aim of providing a multidimensional measure of the development of Italian regions, based on 41 individual indicators from different sources. The considered dimensions are 7: "Environment", "Economy and labour", "Rights and citizenship", "Health", "Education and culture", "Equal opportunities", "Participation". The composite index is equal to the arithmetic mean of 7 macro-indicators, each of which corresponds to the mean of the standardized values of the indicators that compose it (Gnesi *et al.*, 2010).

The main weakness of the indices mentioned above is the use of a compensatory approach. Not assigning a weight to the indicators and dimensions, each variable has the same importance, so, for example, "Bank deposits", "GDP" and "Cinemas per 100,000 inhabitants" are considered the same way. This constitutes a limit in the moment in which, by calculating the arithmetic mean, it admit a compensation: a low value of "GDP" is compensated by a high value of "Cinemas per 100,000 inhabitants" or vice versa.

As part of partially compensatory or non-compensatory approach, different methodologies have been proposed in the literature, ranging from simple mathematical formulas, such as the Mean-Min function (Casadio Tarabusi and Guarini, 2012), to complex procedures, such as the Multicriteria Analysis (Munda and Nardo, 2009).

In a recent work (De Muro *et al.*, 2010), the authors proposed a non-compensatory composite index, called MPI (Mazziotta Pareto-Index) which consists of an arithmetic mean adjusted by a function of variability that penalizes the geographical areas with a unbalanced distribution of the indicators. In other words, if an Italian province has, as mentioned before, a low value of "GDP", and a high value of "Cinemas per 100,000 inhabitants", then the same province receives a penalty without compensation. The underlying principle is that, in order to obtain a high value of the index, all the individual indicators must assume high values, assuming that the variables themselves have equal importance.

3. Steps for constructing a composite index

Constructing a composite index is a complex task whose phases involve several alternatives and possibilities that affect the quality and reliability of the results. The main problems, in this approach, concern the choice of theoretical framework, the availability of the data, the selection of the more representative indicators and their treatment in order to compare and aggregate them.

It is possible, shortly, to individuate the following steps to tackle (Mazziotta and Pareto, 2012):

1) Defining the phenomenon to be measured. The definition of the concept should give a clear sense of what is being measured by the composite index. It should refer to a theoretical framework, linking various sub-groups and underlying indicators.
2) Selecting a group of individual indicators. Ideally, indicators should be selected according to their relevance, analytical soundness, timeliness, accessibility, etc. The selection step is the result of a trade-off between possible redundancies caused by overlapping information and the risk of losing information. A statistical approach to indicators choice involves calculating correlation between potential indicators and then including the ones that are less correlated in order to minimize the redundancy (Salzman, 2003).
3) Normalizing the individual indicators. This step aims to make the indicators comparable. Normalization is required prior to any data aggregation as the indicators in a data set often have different measurement units. Therefore, it is necessary to bring the indicators to the same standard, by transforming them into pure, dimensionless, numbers. Another motivation for the normalization is the fact that some indicators may be positively correlated with the phenomenon to

be measured (positive ‘polarity’), whereas others may be negatively correlated with it (negative ‘polarity’). We want normalize the indicators so that an increase in the normalized indicators corresponds to increase in composite index. There are various methods of normalization, such as ranking, re-scaling (or min-max transformation), standardization (or z-scores) and indicization (index number transformation or ‘distance’ to a reference).

4) Aggregating the normalized indicators. It is the combination of all the components to form one or more composite indices (mathematical functions). Different aggregation methods are possible. The most used are additive methods that range from summing up unit ranking in each indicator to aggregating weighted transformations of the original indicators. Multivariate techniques as Principal Component Analysis (Dunteman, 1989) are also often used.

It is important to emphasize that the theoretical part (definition of the phenomenon and selection of the indicators) is not separate from the statistical-methodological part: so, the choice of the individual indicators is not independent of the choice of the aggregation method.

No universal method exists for composite indices construction. In each case their construction is much determined by the particular application, including both formal and heurist elements, and incorporate some expert knowledge on the phenomenon. Nevertheless, the advantages of composite indices are clear, and they can be summarized in unidimensional measurement of the phenomenon, easy interpretation with respect to a battery of many individual indicators and simplification of the data analysis (e.g., ranking units and comparing their performance over time).

4. A guide for choosing the ‘best’ method

The main factors to take into account in the choice of the method to be adopted for summarizing individual indicators are as follows:

– type of indicators (substitutable/non-substitutable);
– type of aggregation (simple/complex);
– type of comparisons (absolute/relative);
– type of weights (objective/subjective).

There is not always a ‘well-established’ solution, and sometimes it may be necessary to renounce to some requirements, to satisfy others.

Type of indicators

It is one of the main factors that affect the choice of the aggregation method. The components of a composite index are called ‘substitutable’ if a deficit in one component may be compensated by a surplus in another (e.g., a low value of "People who have participated in religious or spiritual activities" can be offset by a high value of "People who have participated in meetings of cultural or recreational associations" and vice versa). Similarly, the components of a composite index are called 'non-substitutable' if a compensation among them is not allowed (e.g., a low value of "Hospital beds per 1,000 inhabitants" cannot be offset by a high value of "Medical doctors per 1000 inhabitants" and vice versa). So we can define an aggregation approach 'compensatory' or 'non-compensatory' depending on whether it permits compensability or not (Casadio Tarabusi and Guarini, 2012). A compensatory approach involves the use of additive methods, such as the arithmetic mean. For a partially compensatory or non-compensatory approach, non-linear methods are generally adopted, such as the geometric mean or the Multicriteria Analysis.

Type of aggregation

The choice of the 'best' aggregation method also depends on the aim of the work and on the type of 'users' (researchers or the general public). Generally, an aggregation method can be considered 'simple' or 'complex'. We say that an aggregation method is 'simple' when a easily understandable mathematical function is used (e.g., the HDI). On the contrary, an aggregation method is said to be 'complex' if a sophisticated model or multivariate method is used (e.g., Principal Components Analysis).

Type of comparisons

Data normalization firstly depends on the type of space-time comparisons requested: 'absolute' or 'relative'. Standardization or transformation in z-scores permits only to do 'relative' comparisons over time since it is based on the mean and the variance of the indicators at the time of reference (e.g., the QUARS index). Other methods, such as re-scaling and indicization, require that the definition of extreme values or of the base are independent from the data, in order to perform comparisons in 'absolute' terms (e.g., the HDI).

Type of weights

The question of the choice of a system of weights in order to weigh the individual indicators, according to their different importance in expressing the considered phenomenon, necessarily involves the introduction of an arbitrary component.

The easiest (but questionable) solution is to assign the same weight to all the components (equal weighting). In this case, the most suitable normalization method is the standardization that brings all the indicators to have the same variance. Alternatively, 'subjective' weights can be set by a group of specialists (e.g., policy makers) or social surveys about how important individual indicators are to the people. Finally, an 'objective' weighting can be used, choosing a methodology that assigns a weight proportional to the variability of the indicator (indicators with a low level of variability will have less weight and indicators with a high level of variability will have much more weight). Note that, although using a simple mean, it is possible to weigh implicitly the indicators through an appropriate normalization function.

Figure 1 shows the flow chart for the choice of the 'best' method in constructing a composite index, with the main possible solutions (normalization, weighting and aggregation) for each 'path' followed (assumptions and requirements).

If the phenomenon to be measured is decomposable into more dimensions, each of them is represented by a subset of individual indicators, it may be more convenient to build a composite index for each dimension (or 'pillar') and then obtain the overall index by means of the aggregation of the partial composite indices. In this case, it is possible to adopt a compensatory approach within each dimension and a non-compensatory or partially compensatory approach among the various dimensions.

The most used aggregation methods for substitutable indicators are the additive ones, such as the simple arithmetic mean or the Principal Component Analysis (PCA). For non-substitutable indicators, non-linear methods are instead used, such as multiplicative functions (partially compensatory approach) or the Multicriteria Analysis (non-compensatory approach).

Focusing on methods based on the use of mathematical functions, the type of normalization depends on the nature of the space-time comparisons to do and on the weight to be assigned to the individual indicators.

For relative comparisons with subjective weighting (equal or different weights), we recommend the rank, z-score or min-max transformation. For assigning objective weights proportional to the variability of the indicators is more suitable a index number transformation where it is assumed as a base the mean, the maximum value or another reference value of the distribution (endogenous base).

For absolute comparisons, it is not possible use ranking or standardization. In the case of subjective weighting, it is necessary to resort to a min-max transformation with minimum and maximum values independent of the distribution (exogenous benchmark), whereas in the case of objective weighting, a indicization with externally fixed base may be a good solution (exogenous base).

In Figures 2a-2d are shown, as an example, the 'paths' followed in the design of the following composite indices: Human Development Index, "Il Sole 24ore" Index, QUARS Index and Mazziotta-Pareto index.

It is noteworthy that each of the 4 composite indices follows a different 'path'.

Figure 1 – *Flow chart for the choice of the 'best' method*

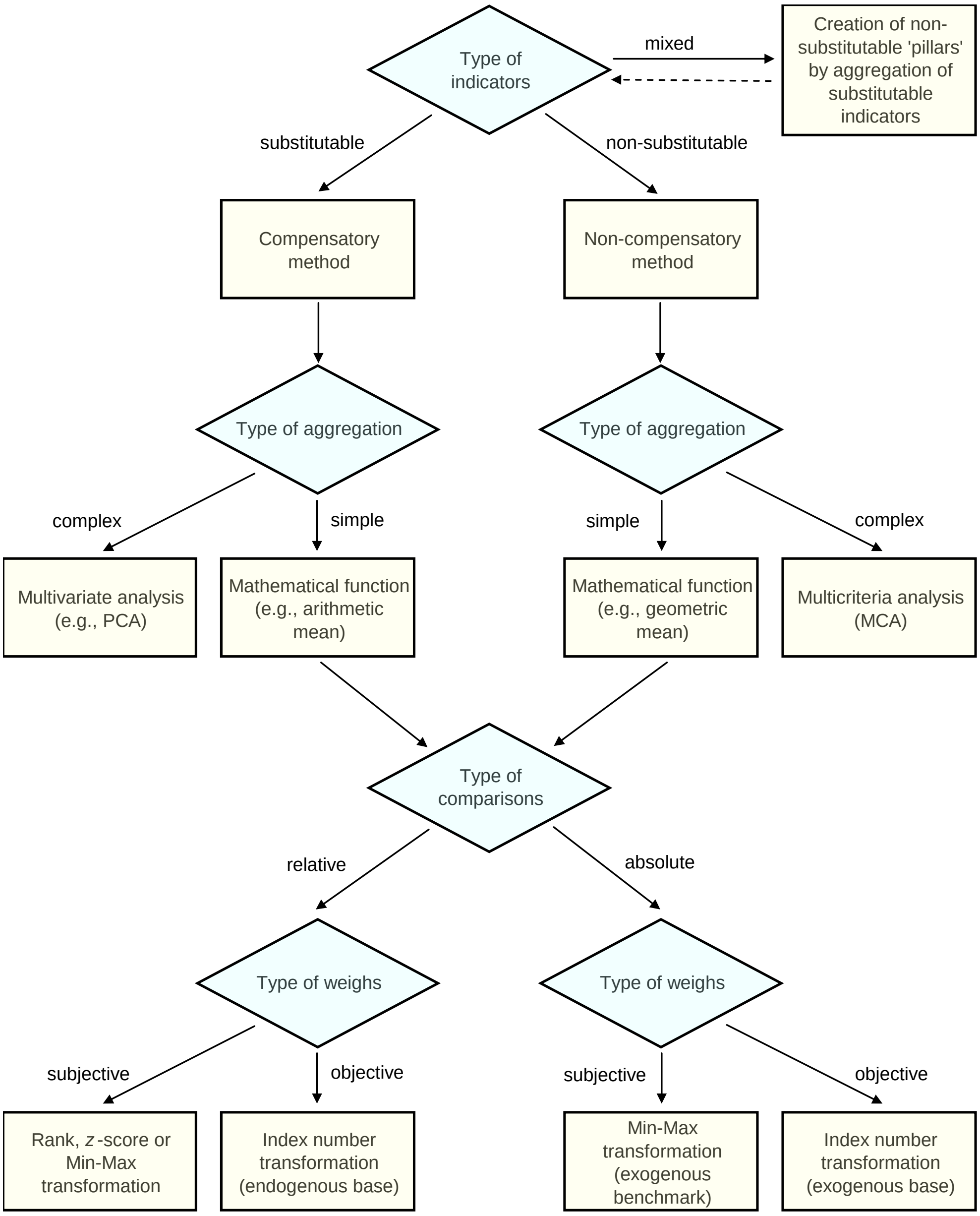


Figure 2a – *The 'path' of the Human Development Index (HDI)*

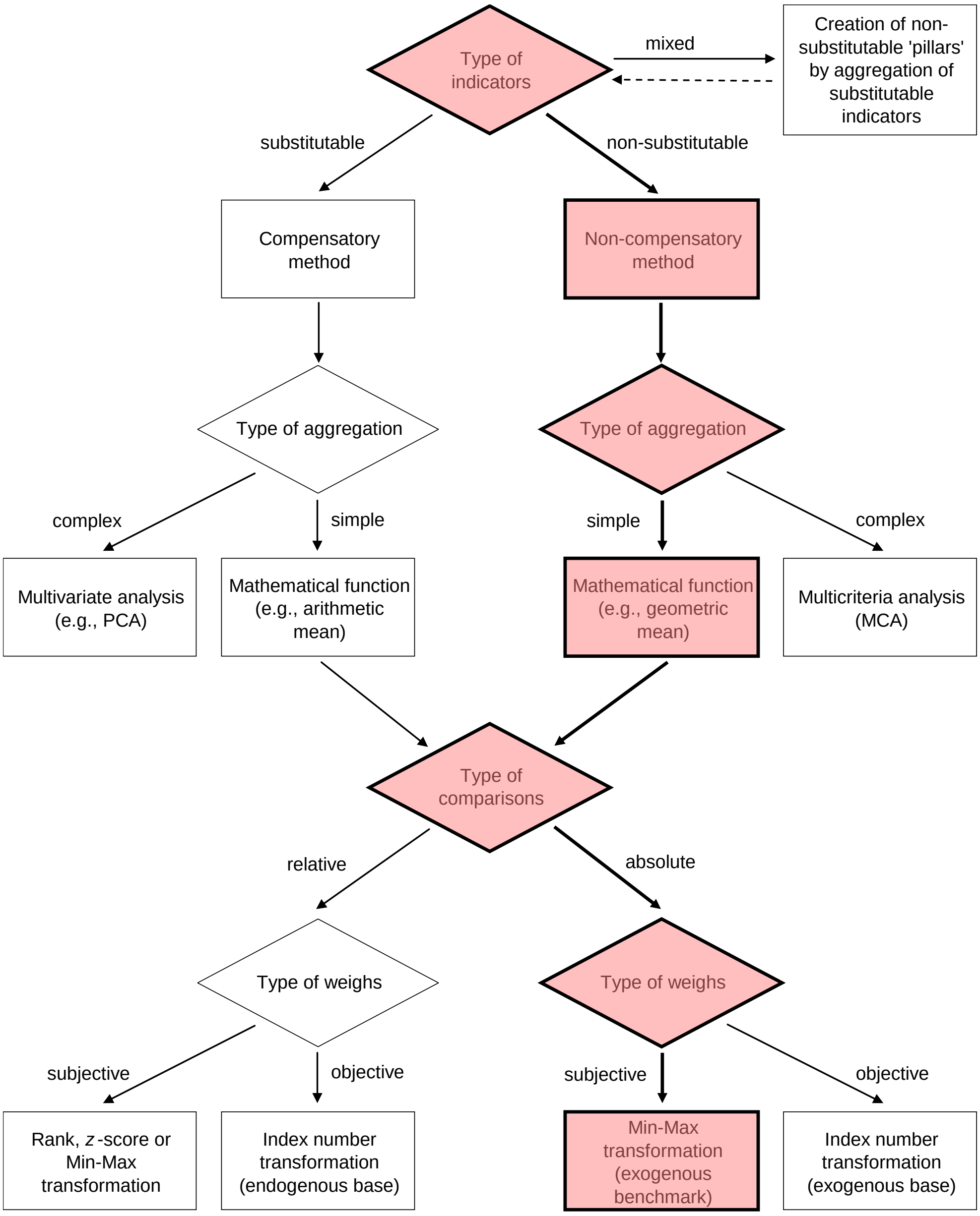


Figure 2b – *The 'path' of the "Il Sole 24ore" Index*

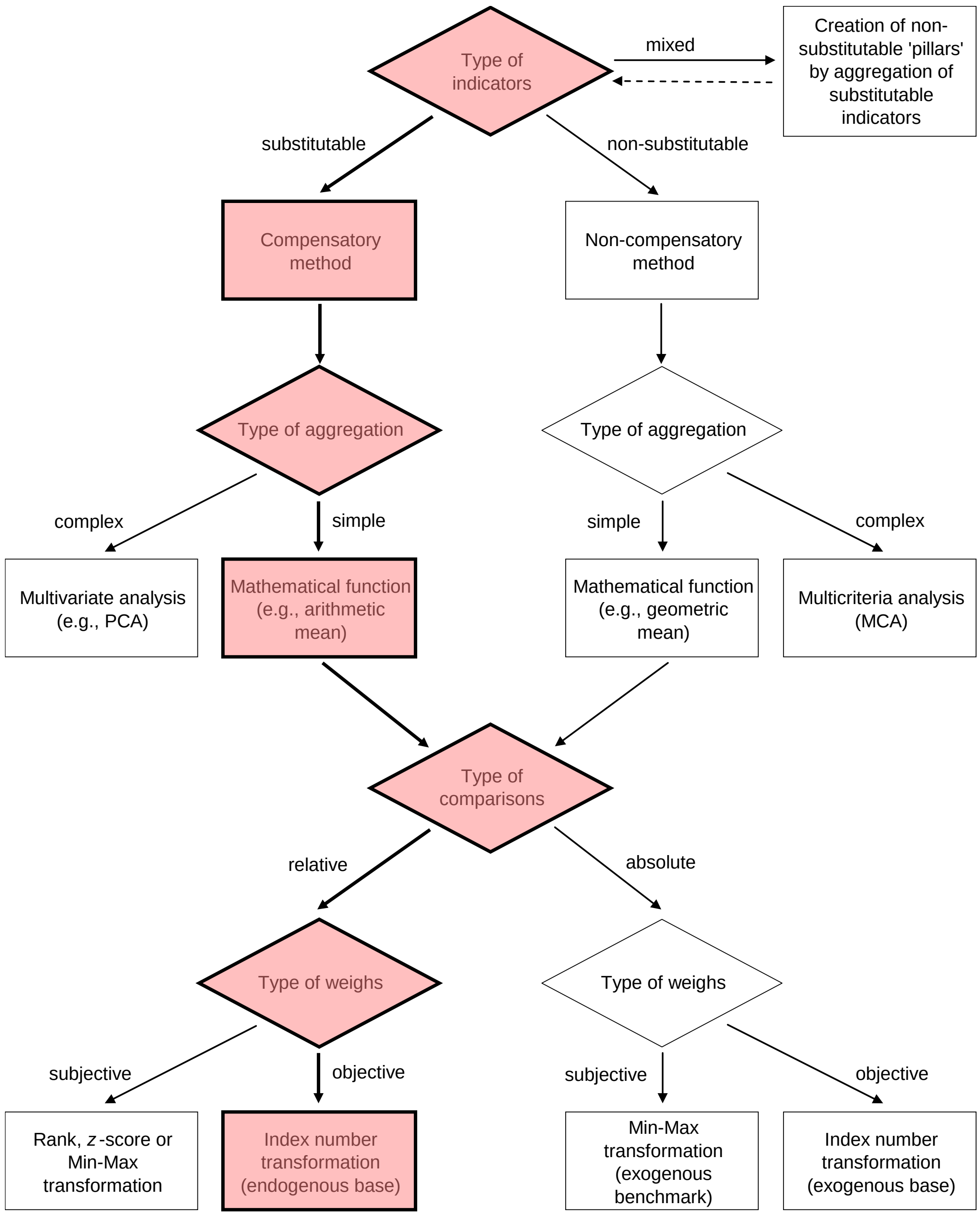


Figure 2c – *The 'path' of the Regional Quality of Development Index (QUARS)*

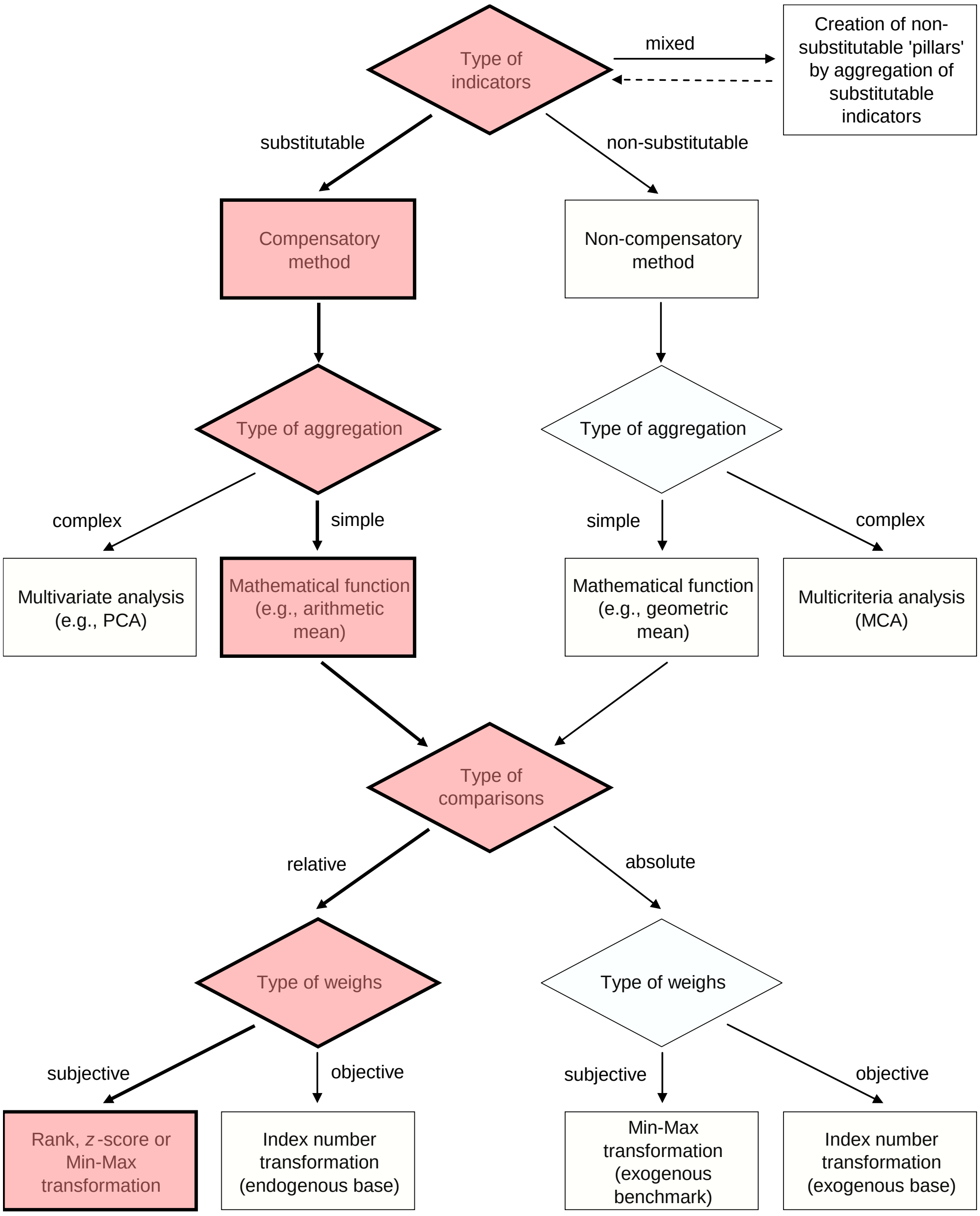
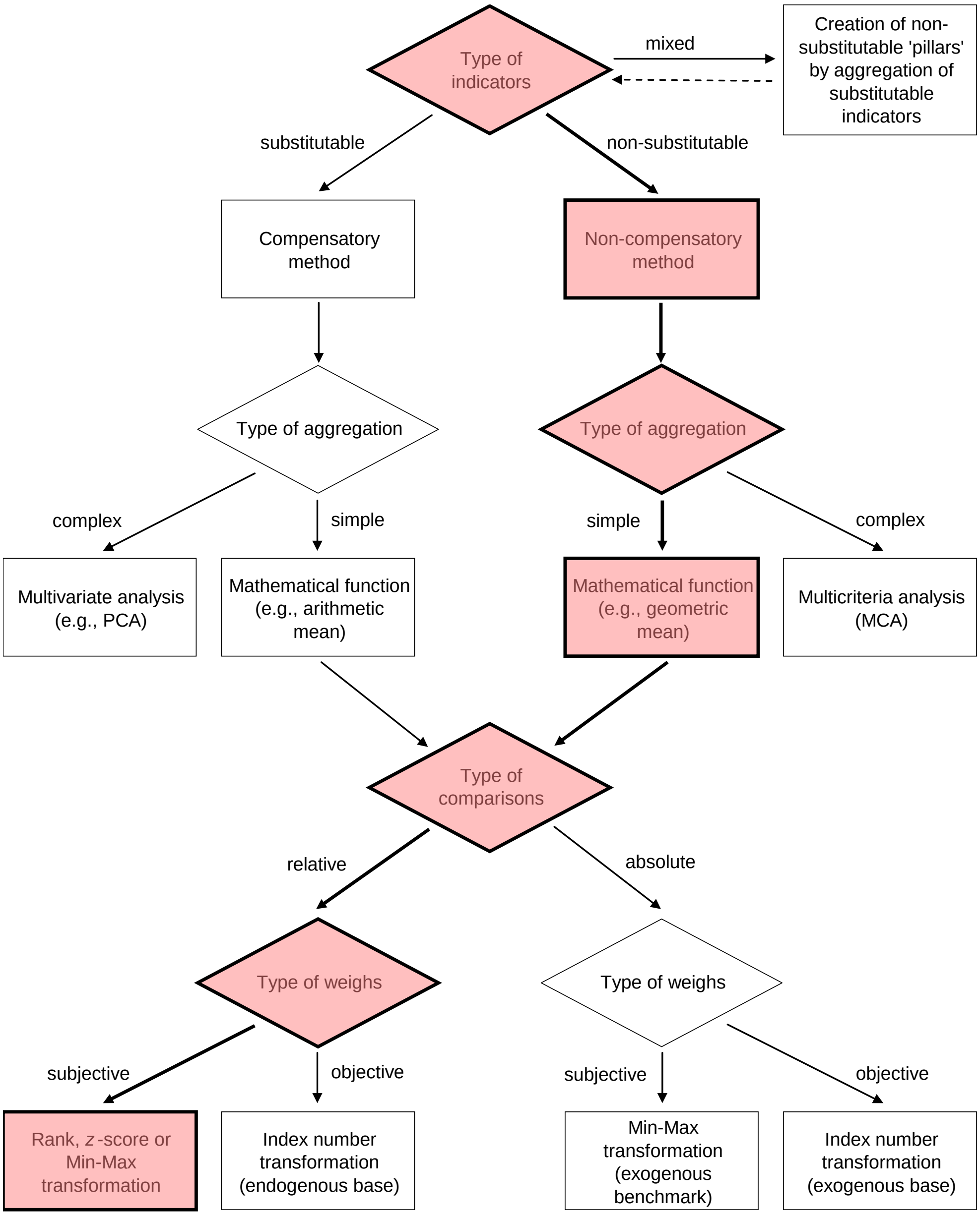


Figure 2d – *The 'path' of the Mazziotta-Pareto Index (MPI)*



5. Conclusions

As is known, the implementation of a composite index is a complex process that involves stages of work well defined, where the arbitrary choices of the researcher has a significant effect on the final results. The heated debate within the scientific Community, over the years, seems to converge towards the idea that there is not a composite index universally valid for all areas of application, and, therefore, its validity depends on the strategic objectives of the research.

In this paper we propose a scheme with some general guidelines to follow for summarizing a set of individual indicators. Beyond the procedure used, the composite indices provide an irreplaceable contribution to simplification; however, they are based on methods that flatten the basic information and they can lead to a myopic reading of reality, especially if not sustained, upstream, from an adequate step of selection and interpretation of the individual indicators.

Therefore, it is considered absolutely essential, in order to obtain valid and reliable results, to support the process of choosing the set of the individual indicators with a theoretical framework that defines the social reality in each of its dimensions (Delvecchio, 1995).

References

BANDURA R. 2008. *A Survey of Composite Indices Measuring Country Performance: 2008 Update*. New York: UNDP/ODS Working Papers.

CASADIO TARABUSI E., GUARINI G. 2012. An Unbalance Adjustment Method for Development Indicators. *Social Indicators Research*. doi: 10.1007/s11205-012-0070-4.

DELVECCHIO F. 1995. *Scale di misura e indicatori sociali*. Bari: Cacucci.

DE MURO P., MAZZIOTTA M., PARETO A. 2010. Composite Indices of Development and Poverty: An Application to MDGs. *Social Indicators Research*, Vol. 104, pp. 1-18.

DUNTEMAN G. H. 1989. *Principal Components Analysis*. Newbury Park: Sage Publications.

GNESI C., SEGRE E., VILLA A. 2010. *Come si vive in Italia? Rapporto QUARS 2010 – Indice di qualità regionale dello sviluppo*. www.sbilanciamoci.org.

LUN G., HOLZER D., TAPPEINER, G., TAPPEINER, U. 2006. The stability of rankings derived from composite indicators: analysis of the “Il Sole 24 ore” quality of life report. *Social Indicators Research*, Vol. 77, pp. 307-331.

MAZZIOTTA M., PARETO A. 2011. Nuove misure del benessere: dal quadro teorico alla sintesi degli indicatori, *SISmagazine, rivista on-line della Società Italiana di Statistica*, http://old.sis-statistica.org/magazine/spip.php?article194

MAZZIOTTA M., PARETO A. 2012. A Non-compensatory Approach for the Measurement of the Quality of Life. In: Maggino F. & Nuvolati G. (eds). *Quality of Life in Italy: Research and Reflections*. New York: Springer.

MUNDA G., NARDO M. 2009. Non-compensatory/non-linear composite indicators for ranking countries: A defensible setting. *Appl. Econ.*, Vol. 41, pp. 1513-1523.

PRESCOTT-ALLEN R. 2001. *The Wellbeing of Nations: A country-by-country index of quality of life and the environment*. Washington: Island Press.

SALZMAN J. 2003. *Methodological Choices Encountered in the Construction of Composite Indices of Economic and Social Well-Being*. Ottawa: Center for the Study of Living Standards.

SEN A. K. 1985. *Commodities and Capabilities*. Amsterdam: Elsevier.

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

UNDP 2007. *Human Development Report 2007/2008*. New York: Palgrave Macmillan.

UNDP 2010. *Human Development Report 2010*. New York: Palgrave Macmillan.

SUMMARY

The debate on the measurement of multidimensional socio-economic phenomena has had a strong acceleration in recent years thanks to the publication of the Stiglitz report and the first BES report (Equitable and Sustainable Well-being) by Istat and CNEL. The main objective is to find an alternative measure to GDP. Many attempts have been studied over the years but no one has really replaced the GDP. The reason is twofold: on the one hand, the socio-economic theories proposed do not seem to have a solid foundation; on the other hand, the statistical methods used to reduce multidimensionality are not always mathematically rigorous. This paper aims to provide some suggestions for constructing a composite index.

Matteo MAZZIOTTA, Italian National Institute of Statistics, mazziott@istat.it
Adriano PARETO, Italian National Institute of Statistics, pareto@istat.it