Constructing Consistent Composite Indicators: the Issue of Weights

Giuseppe Munda and Michela Nardo

Institute for the Protection and Security of the Citizen

2005

EUR 21834 EN

European Commission
Directorate-General Joint Research Centre
Institute for the Protection and Security of the Citizen

EUR 21834 EN
Luxembourg: Office for Official Publications of the European Communities

Printed in Italy

Constructing Consistent Composite Indicators: the Issue of Weights

Giuseppe Munda[1,2] **and Michela Nardo**[2]

1)Universitat Autonoma de Barcelona, Dept. of Economics and Economic History, building B,08193 Bellaterra (Barcelona), Spain, e_mail: giuseppe.munda@jrc.it

2) European Commission , Joint Research Centre, TP 361, 21020 Ispra (VA), Italy, e_mail: michela.nardo@jrc.it

Abstract: Composite indicators are very common in economic and business statistics for benchmarking the mutual and relative progress of countries in a variety of policy domains such as industrial competitiveness, sustainable development, globalisation and innovation. The proliferation of the production of composite indicators by all the major international organizations is a clear symptom of their political importance and operational relevance in policy-making. As a consequence, improvements in the way these indicators are constructed and used seem to be a very important research issue from both the theoretical and operational points of view . This paper shows that a theoretical inconsistency exists between the real theoretical meaning of weights and the meaning that is generally attributed to them by the standard practice in constructing composite indicators; thus, a recursive important mistake is present in most of the empirical applications. Guidelines to solve this drawback are given.

Key Words: *Composite indicators, Measurement Theory, Multi-Criteria Analysis, Social Choice*

JEL Classification Numbers: *C43, C82*

1. Introduction

Composite indicators (or indexes) are used whenever a plurality of variables is needed for the evaluation of a macroeconomic dimension. Composite indicators are very common in fields such as economic and business statistics (e.g., the OECD Composite of Leading Indicators) and are used in a variety of policy domains such as industrial competitiveness, sustainable development, quality of life assessment, globalisation, innovation or academic performance (see Cox and others 1992, Cribari-Neto et al 1999, Färe et al. 1994, Griliches 1990, Forni et al. 2001, Huggins 2003, Grupp and Mogee 2004, Lovell et al. 1995, Munda 2005, Nardo et al. 2005, Saisana and Tarantola 2002, and Wilson and Jones 2002, among others). The proliferation of these indicators is a clear symptom of their importance in policy-making, and operational relevance in macroeconomics in general (see e.g. Granger, 2001). All the major international organizations such as OECD, the EU, the World Economic Forum or the IMF are producing composite indicators in a wide variety of fields (Nardo et al., 2005). A general objective of most of these indicators is the ranking of countries and their benchmarking according to some aggregated dimensions (see e.g. Cherchye, 2001, Kleinknecht 2002 and OECD, 2003). As a consequence, the improvement of the way these indicators are constructed and used seems to be a very important research issue from both theoretical and operational points of view. Our main objective here is to contribute to the improvement of the overall quality of composite indicators by looking at one of their technical weaknesses, that is, the consistency between the mathematical aggregation rule used for their construction and the meaning of weights. Along the paper, concepts coming from measurement theory, multi-criteria decision analysis and social choice are used.

2. Linear Aggregation Rules and Meaning of Weights

Although various functional forms for the underlying aggregation rules of a composite indicator have been developed in the literature (e.g. Diewert, 1976, Journal of Economic and Social Measurement, 2002), in the standard practice, a composite indicator, *I*, can be considered a *weighted linear aggregation rule* applied to a set of variables (OECD, 2003, p. 5):

$$I = \sum_{i=1}^{N} w_i x_i \tag{1}$$

where x_i is a scale adjusted variable (e.g. GDP per capita) normalised between zero and one, and w_i a weight attached to x_i, usually with $\sum_{i=1}^{N} w_i = 1$ and $0 \le w_i \le 1$, $i = 1,2,..., N$. In this framework, a crucial role is played by the concept of weight.

The common practice in attaching weights is well synthesised by a recent OECD document: *"Greater weight should be given to components which are considered to be* <u>*more*</u> *significant in the context of the particular composite indicator"* (OECD, 2003, p. 10). In the decision theory literature, this concept of weights is usually referred to as *symmetrical importance*, that is *"... if we have two non-equal numbers to construct a vector in* R^2*, then it is preferable to place the greatest number in the position corresponding to the most important criterion."* (Podinovskii, 1994, p. 241). Let's try to put some light on this issue, by proving formally that the concept of symmetrical importance is incompatible with a linear aggregation rule, given that in a linear aggregation rule, weights can only have the meaning of a trade-off ratio (see also Vincke, 1992, pp. 36-37).

Suppose that country *a* is evaluated according to some variables $(x_1(a),...,x_n(a))$, then the *substitution rate at a*, of the variable *j* with respect to the variable *r* (taken as a reference variable) is the amount $S_{jr}(a)$ such that, country *b* whose evaluations are: $x_l(a) = x_l(b), \forall l \neq j, r$; $x_j(b) = x_j(a) - 1$; and $x_r(b) = x_r(a) + S_{jr}(a)$ is indifferent to country *a*. Therefore, $S_{jr}(a)$ is the amount which must be added to the variable *r* in order to compensate the loss of one unit on variable *j* for country *a*. Consider now a composite indicator $I(x_1, x_2, ..., x_n)$ and suppose that the score of this indicator is the same for the two countries. Let $z(a) = (x_1(a), x_2(a), ..., x_n(a))$ and $z(b) = (x_1(b), x_2(b), ..., x_n(b))$, then as a first approximation one has:

$$0 = I(z_b) - I(z_a) = \sum_{i=1}^{n} \left(\frac{\partial I}{\partial x_i}\right)_{z_a} \left(x_i(b) - x_i(a)\right) = -\left(\frac{\partial I}{\partial x_j}\right)_{z_a} + S_{jr}(a)\left(\frac{\partial I}{\partial x_r}\right)_{z_a}$$

and manipulating

$$S_{jr}(a) = \frac{\left(\partial I / \partial x_j\right)(z_a)}{\left(\partial I / \partial x_r\right)(z_a)} \qquad (2)$$

When the function I is a weighted sum of all the normalised variables, i.e.

$$I(x_1, x_2, ..., x_n) = \sum_{i=1}^{n} w_i x_i \qquad (3)$$

from expression (2) one obtains:

$$S_{jr}(a) = \frac{w_j}{w_r} = \text{constant.} \qquad (4)$$

This means that in the weighted linear aggregation, the substitution rates equal the weights of the variables up to a multiplicative coefficient. As a consequence, ***the estimation of weights is equivalent to that of substitution rates***, implying a compensatory logic. *Compensability* refers to the existence of trade-offs, i.e. the possibility of offsetting a disadvantage on some variables by a sufficiently large advantage on another variable. Therefore, the ***use of weights in combination with intensity of preference (given that variables are always supposed to be measured on an interval or ratio scale) within a linear aggregation rule originates compensatory aggregation conventions and gives the meaning of trade-offs to the weights.***

In other words, in a linear aggregation framework, the weights always depend on the value of the trade-off. Such a trade-off holds a constant value, since in this context, the *local trade-off* (i.e. the marginal rate of substitution) is also the *global one*, i.e. it does not depend on the values that variable scores may have in a given point. However, one has to note that the trade-off always depends on *the* *measurement scale* used for measuring the variable scores and on the *range* that the measurements of variable scores may present.

To clarify the issue consider the hypothetical example presented in Table 1.

	GDP (Millions of Euro)	**Populations (Number of Inhabitants)**	**Percentage of Protected Species**
A	32,000	1,000,000	60%
B	80,000	3,000,000	70%
C	100,000	5,000,000	40%

Table 1. Illustrative Example with Three Countries and Three Variables

Consider first the measurement scale. Suppose that in the construction of a sustainability composite indicator, the trade-off between protected species and GDP is set such that a decrease of 1 point in the percentage of protected species can be compensated by an increase of 100,000,000 Euro of GDP. This trade-off can be expressed as $\frac{w_{species}}{w_{GDP}} = 100{,}000{,}000$. If instead the measurement scales of GDP is changed and this variable is measured per capita, the same trade-off indicated above now would be modified e.g. in "1% of protected species less can be compensated by 100 Euro of GDP per capita more". Thus in this case one has $\frac{w_{species}}{w_{GDP}} = 100$. Since the measurement scale of the variable *protected species* has not changed, the only weight that must change value is the one attached to GDP, that in the second case has to increase considerably (since the numerator remain constant and the value of the ratio decrease).

One obvious observation might be that in a composite indicator variables are normalized and thus effects due to measurement scales should disappear. This, however is not true. Consider for example the normalization technique *distance from the group leader,* which assigns 100 to the leading alternative and other alternatives are ranked as percentage points away from the leader (Saisana and Tarantola, 2002), that is $100\left(\frac{actualvalue}{maximum\ value}\right)$. By applying this normalization technique while keeping the original trade-off " 1% decrease in species versus

100 million Euro GDP" one has to standardize the value 100 Mill. Euros according to the new scale. This is equivalent to dividing this value by the score of the country with the highest GDP: $\frac{100}{100{,}000} = 10^{-3}$. When income is expressed as GDP per capita, then the trade off would now be "1% decrease in species versus $\frac{100}{32{,}000} = 3.125 \bullet 10^{-3}$ increase in GDP per capita"[1]. Again trade-offs and corresponding weights must change according to the range of variation of the measurement scale considered. One may easily check that this kind of consequences apply independently to the normalization technique chosen. The conclusion is that in the case of a linear aggregation rule, trade-offs depend on the scales of measurement, and since weights are connected to the values of trade-offs they also depend on the scales of measurement.

Clearly trade-offs can be evaluated only if one knows the quantitative scores of the variables involved without any uncertainty. On the contrary, the concept of importance is connected to the variable itself and NOT with its quantification. If protected species are considered *more, equal or less important* than GDP, this is a quality of the variables which is independent from any measurement scale one may use. As clearly shown by Anderson and Zalinski (1988), when weights depend on the range of variable scores, such as in the context of a linear aggregation rule, the interpretation of weights as a measurement of the psychological concept of importance *is always completely inappropriate.*

More formally, to use the compensatory approach in practice, such as the linear aggregation rule, one has to determine for each individual indicator, a mapping $f_i : x_i \to R$ which provides at least an interval scale of measurement and to assess scaling constants (i.e. weights) in order to specify how the compensability should be accomplished, given the scales f_i between the different individual indicators (Roberts, 1979). Note that the scaling constants which appear in the compensatory approach depend on the scales f_i, thus they do not characterise the intrinsic relative importance of individual indicators. The

[1] A is the country with the highest GDP per capita with 32,000 Euro, followed by B with 26,667 and C with 20,000 Euro.

implication is ***the existence of a theoretical inconsistency in the way weights are actually used and their real theoretical meaning.***

An overview of methods to attach weights in a multi-attribute value function framework (the general framework to which the linear aggregation rule belongs) can be found in Beinat (1997) and Keeney and Raiffa (1976). There is unanimous agreement in the literature that the only method where weights are computed as scaling constants and there is no ambiguous interpretation is the so-called trade-off method starting with revealed preferences. No weight importance judgment is required in this method. The trade-off method can be briefly described as follows. Let's consider two countries *A* and *B*, differing only for the scores of variables x_k and x_t. The problem is then to adjust the score of say x_k for *B*, in such a way that *A* and *B* become indifferent. Formally, it is:

$$I(A) = I(B) \Leftrightarrow I(x_1,\dots,x_k^{'},\dots,x_t^{'},\dots,x_n) = I(x_1,\dots,x_k^{''},\dots,x_t^{''},\dots,x_n) \Rightarrow \tag{5}$$

$$\Rightarrow \sum_{\substack{i=1 \\ i\neq k,t}}^{N} w_i x_i + w_k x_k^{'} + w_t x_t^{'} = \sum_{\substack{i=1 \\ i\neq k,t}}^{N} w_i x_i + w_k x_k^{''} + w_t x_t^{''} \Rightarrow \tag{6}$$

$$\Rightarrow w_k x_k^{'} + w_t x_t^{'} = w_k x_k^{''} + w_t x_t^{''} \tag{7}$$

Equation (7) is an equation in the unknown w_k and w_t. To compute the *N* weights as trade-offs, it is necessary to assess *N-1* equivalence relations which together with the usual normalisation constraint $w_i + \dots + w_n = 1$ determine a linear system of *N* equations in the *N* unknown weights. Of course if some uncertainty on the variable scores exists, this method cannot be applied.

As one can easily understand to assess weights as trade-offs, as it should be always done when using a linear aggregation rule, it is a much harder job than to use weights as importance coefficients. This is probably the main reason why the standard practice tends to use weights as importance coefficients, ***but unfortunately this practice is not defensible on theoretical grounds.***

Vansnick (1990) showed that the two main approaches in multi-criteria aggregation procedures i.e., the compensatory and non-compensatory ones can be directly derived from the seminal work of Borda and Condorcet. If one wants the weights to be interpreted as "importance coefficients" (or equivalently *symmetrical importance* of variables) non-compensatory aggregation procedures must be used (Bouyssou, 1986; Bouyssou and Vansnick, 1986). From a social choice point of view, these non-compensatory rules are always Condorcet consistent rules; their use in the framework of composite indicators, can be corroborated by referring to a clear result of social choice literature. The majority rule is theoretically the most desirable aggregation rule, but practically often produces undesirable intransitivities, thus "*more limited ambitions are compulsory. The next highest ambition for an aggregation algorithm is to be Condorcet*" (Arrow and Raynaud, 1986, p. 77).

Thus we can conclude that the use of non-compensatory aggregation rules to construct composite indicators is **compulsory** for reasons of theoretical consistency when weights with the meaning of importance coefficients are used. Moreover the use of Condorcet consistent rules is also **desirable** in general as advised by social choice literature. Unfortunately these considerations are completely neglected by the standard practice on composite indicators[2].

3. Conclusion

The following main conclusions can be drawn:

1. Weights in linear aggregation rules have always the meaning of trade-off ratio. Therefore, given that in the standard practice of constructions of a composite indicator, weights are used as importance coefficients in combination with linear aggregation rules, a theoretical inconsistency exists. This inconsistency applies to most of the empirical applications.
2. When a linear aggregation rule is used, the only method able to derive theoretically consistent weights is the so-called trade-off method. Operationally this method is very complex. Moreover the assumption that the variable scores are measured on an interval or ratio scale of measurement and no uncertainty exists must always apply. Rarely this happens in the practice of composite indicators, where for instance, sometimes quantitative scores are arbitrarily given to variable scores originally measured on an ordinal measurement scale (see e.g. Nicoletti et al., 2000).

---

[2]The only exception probably being the 2005 Environmental Sustainability Index, where at least some methodological weaknesses are acknowledged. http://www.yale.edu/esi/a_methodology.pdf

3. In standard composite indicators based on the linear aggregation rule, compensability among the different individual indicators must always be assumed; this implies complete substitutability among the various components considered. For example, in a hypothetical sustainability index, economic growth can always substitute any environmental destruction or inside e.g., the environmental dimension, clean air can compensate for a loss of potable water. From a descriptive point of view, such a complete compensability is often not desirable.
4. Whenever weights are used with the meaning of importance coefficients, the aggregation algorithm must be a Condorcet consistent rule. The use of these rules is desirable on more general grounds too. In particular, it should be noted that by using Condorcet aggregation rules no limitation on the measurement scale of the variable scores exists. The cost to pay is that information on the intensity of preference may be lost.

ACKNOWLEDGEMENTS: This research has been partially developed in the framework of the European Commission financially supported project "*Knowledge Economy Indicators: Development of Innovative and Reliable Indicator Systems (KEI)*" (KEI SCS8 2004-502529).

References

Anderson N.H., Zalinski J. (1988) – Functional measurement approach to self-estimation in multiattribute evaluation, *Journal of Behavioural Decision Making* (1), pp. 191-221.

Arrow K.J., Raynaud H. (1986) - *Social choice and multicriterion decision making*, M.I.T. Press, Cambridge.

Beinat E. (1997) – *Value functions for environmental management*, Kluwer, Dordrecht.

Bouyssou D. (1986) – Some remarks on the notion of compensation in MCDM, *European Journal of Operational Research*, 26 pp. 150-160.

Bouyssou D., Vansnick J.C. (1986) – Noncompensatory and generalized noncompensatory preference structures, *Theory and Decision*, 21, pp. 251-266.

Cherchye L. (2001) – Using data envelopment analysis to assess macroeconomic policy performance, *Applied Economics*, 33, pp. 407-416.

Cox D., Fitzpatrick R., Fletcher A., Gore S., Spiegelhalter D. And Jones D. (1992) – Quality-of-life assessment: can we keep it simple?, *Journal of the Royal Statistical Society*, 155(3), pp. 353-393.

Cribari-Neto F., Jensen M.J., Novo A. (1999) - Research in econometric theory: quantitative and qualitative productivity rankings, *Econometric Theory*, Volume 15, Issue 05, pp 719-752.

Diewert W.E. (1976) - Exact and superlative index numbers, *Journal of Econometrics*, Volume 4, Issue 2 , Pages 115-145.

Färe R., Grosskopf S., Norris M., Zhang Z. (1994) – Productivity growth, technical progress and efficiency change in industrialised countries, *American Economic Review*, 84(1), pp. 66-83.

Forni M., Hallin M., Lippi M., Reichlin L. (2001) Coincident and Leading Indicators for the Euro Area, *The Economic Journal*, Volume 111, Issue 471, Page 62.

Granger W.J. (2001) - Macroeconometrics – Past and future, *Journal of Econometrics*, Volume 100, Issue 1 , Pages 17-19.

Griliches Z. (1990) – Patent statistics as economic indicators, *Journal of Economic Literature*, 28, pp. 1661-1707.

Grupp H., Mogee M.E. (2004) - Indicators for national science and technology policy: how robust are composite indicators?, *Research Policy*, vol. 33(9), pages 1373-1384.

Huggins R. (2003) – Creating a UK competitive index: regional and local benchmarking, *Regional Studies*, 37, pp.89-96.

Journal of Economic and Social Measurement (2002) **-** *Special Issue on Index Number Aggregation*, Volume 28, Numbers 1-2.

Keeney R., Raiffa H. (1976) - *Decision with multiple objectives: preferences and value trade-offs*, Wiley, New York.

Kleinknecht A., Van Montfort K., Brouwer E. (2002) – The non trivial choice between innovation indicators, *Economic Innovation and New Technologies*, 11(2), pp. 109-121.

Lovell C.A.K., Pastor J.T., Turner J.A. (1995) – Measuring macroeconomic performance in the OECD: a comparison of European and non-European countries, *European Journal of Operational Research*, 87, pp. 507-518.

Munda G. (2005) – Multi-Criteria Decision Analysis and Sustainable Development, in J. Figueira, S. Greco and M. Ehrgott (eds.) –*Multiple-criteria decision analysis. State of the art surveys*, Springer International Series in Operations Research and Management Science, New York, pp. 953-986.

Nardo M., Saisana M., Saltelli A., Tarantola S., Hoffman A., Giovannini E. (2005) **–** *Handbook on constructing composite indicators: methodology and user guide*, OECD Statistics Working Paper, Paris.

Nicoletti G., Scarpetta S. and Boylaud O. (2000) – *Summary indicators of product market regulation with an extension to employment protection legislation*, OECD Economics Department Working papers No. 226.

OECD (2003) – *Composite indicators of country performance: a critical assessment*, DST/IND(2003)5, Paris.

Podinovskii V.V. (1994) -Criteria importance theory, *Mathematical Social Sciences*, 27, pp. 237 - 252.

Roberts F. S. (1979) **-** *Measurement theory with applications to decision making, utility and the social sciences*, Addison-Wesley, London.

Saisana M. and Tarantola S. (2002) – *State-of-the-art report on current methodologies and practices for composite indicator development,* EUR 20408 EN Report, European Commission, JRC, Ispra, Italy.

Vansnick J. C. (1990) **-** Measurement theory and decision aid - in Bana e Costa C.A. (ed.)- *Readings in multiple criteria decision aid,* Springer-Verlag, Berlin, pp. 81-100.

Vincke Ph. (1992) - *Multicriteria decision aid,* Wiley, New York.

Wilson J.W., **and Jones C.P.** (2002) – An analysis of the S&P 500 index and Cowles's extensions: price indexes and stock returns, 1870-1999, *Journal of Business*, 75, pp. 505-533.

The mission of the Joint Research Centre is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of European Union policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.