Weights and importance in linear aggregation

Saltelli, Andrea
European Commission, Joint Research Centre,
Via E. Fermi 2749
Ispra 21027 (VA), Italy
E-mail: andrea.saltelli@jrc.ec.europa.eu (corresponding author)

Saisana, Michaela
European Commission, Joint Research Centre,
Via E. Fermi 2749
Ispra 21027 (VA), Italy
E-mail: michaela.saisana@jrc.ec.europa.eu

Paruolo, Paolo
University of Insubria, Department of Economics
Via Monte Generoso 71
Varese 21100, Italy
E-mail: paolo.paruolo@uninsubria.it

1. Introduction

Composite indicators are built by combining variables with different possible aggregation strategies (Nardo et al., 2008). Each variable is customarily attached a weight purportedly meant to appreciate the 'importance' of that variable. By far the most common strategy to aggregate variables seen in existing composite indicators (CI) is by means of linear combination of weighted variables, whereby the composite index y is derived from a set of k variables x_i via:

$$y_j = \sum_{i=1}^{k} w_i x_{ji}, \tag{1}$$

where j is one of the individuals being measured by the composite indicator, y_j its score, and $x_{j1}, x_{j2}, \cdots, x_{jk}$ are its normalized scores on the k variables. The weights w_i are attached by the developers on the basis of different strategies, be those statistical such as factor analysis or based on expert evaluation. Weights represent a form of judgement of the relative importance of the different variables, including the case of equal weights where all variables are (in theory, though rarely in practice as well shall see below) equally important.

According to most practitioners (see e.g. Munda (2008)) weights as used in (1) are not measures of importance, but rather measures of trade off among variables. More precisely the ratio $\frac{w_i}{w_l}$ measures how much of x_i must be given up to offset or balance a unit increase in x_l. If $\frac{w_i}{w_l} = 2$ one has to take away 0.5 from x_i to balance for a unit increase in x_l.

While this explains clearly what weights are in (1), it is perhaps still not so clear what they are not, and in particular why they are not measures of importance. Perhaps the problem here is what is meant by importance. To grasp this it is useful now to consider an alternative strategy of aggregation, a so-called 'non compensatory' one: the outranking matrix of Condorcet (see Munda, 2008 for a review). We give here a nutshell description of the Condorcet matrix for the sake of the discussion.

Condorcet's outranking matrix is built starting by pairwise comparisons of individuals, say individual A against individual B. The comparison is performed as follows: all the weights w_i for which the relative indicator/variable x_{Ai} is bigger than x_{Bi} are added up in favor of A, while all weights for which x_{Bi} is better than x_{Ai} are added up in favor of B, irrespective of the distance between x_{Ai} and x_{Bi}. Note that depending on the meaning of the indicators, the expression 'x_{Bi} better than x_{Ai}' can either correspond to $x_{Bi} > x_{Ai}$ or $x_{Ai} > x_{Bi}$. In case of tie, namely $x_{Bi} = x_{Ai}$, the weight w_i is equally split between the two individuals. Note that the same result is obtained using the un-normalized score x'_{ji} in place of x_{ji}.

These sums of weights populate the $n \times n$ outranking matrix, whose element ab is the sum of all weights of those indicators which see A better than B. By definition, if the weights add up to one, the entries ab and ba of the outrank matrix sum up to one, and so on for the other couples of individuals. We are not concerned here with how the outranking matrix is used to order the individuals - a hint is given in Table 4.

Thus in building the outranking matrix – which is used as a basis to rank the individuals – the weights are truly used as importance measures. The entire weight is assigned to either an individual or to another depending of which of the two individuals has the highest score in the relative variable.

Is this not the case when using Eq. (1)? Clearly not, as in (1) the 'importance' of the variables depends much from how they are treated as well as from the covariance structure of the sample, i.e. how and how much the variables depends upon one another. For example, imagine that the variables have been normalized via the min-max approach, whereby each variable is subtracted its smallest element in the sample and divided by the difference between the maximum and the minimum element in the sample.

$$(2) \qquad x_{ji} = \frac{x'_{ji} - x'_{min,i}}{x'_{\max,i} - x'_{\min,i}}$$

Where $x'_{\max,i}$and $x'_{\min,i}$ are the maximum and minimum value respectively for the raw (untreated) variables x'_i. In this case all x_i vary in $[0, 1]$ but their variances differ. Thus a given variable x'_i could be totally non-influent with respect to y simply because its variance is small compared to that of the other variables, and this largely irrespective of its weight, unless the trivial case where the variable has all the weight to itself, i.e. $w_i \approx 1$. Note that the discussion developed thus far applies to pillars as well as to variables, meaning by a pillar a subset of variables which is identified by either experts opinion or by e.g. factors analysis as to represent a salient – possibly latent – characteristic of the composite. One might have a pillar whose collective weights amount to – say – 50% of the total, but which contributes little to the index because collectively the variance of the pillar is comparatively small.

This also explains why equal weights is not a sufficient condition for all variables being equally important. Also the trade off is influenced by the scale of variation of the variables. We just said that if $\frac{w_i}{w_i} = 2$ one has to take away 0.5 from x_i to balance for a unit increase in x_i. Imagine now being an individual (or a country) j trying to balance its score y_j by offsetting a 0.5 loss in x_i by a unit increase in x_i. This may be easy or impossible depending on the variance of x_i. If the variance of x_i over all individuals is e.g. 0.1 it will be highly unlikely that an individual can increase its score by one.

A popular alternative to the min-max normalization (Equation 2) which is less probe to the shortcomings just illustrated is standardization, whereby each variable's value is subtracted its mean and the result is divided by its standard deviation:

$$(3) \qquad x_{ji} = \frac{x'_{ji} - \mu_i}{\sigma_i},$$

where μ_i is the mean of x'_i and σ_i its standard deviation. It is evident from the discussion above on the Condorcet approach that the variables/indicators do not need to be normalized when using this method.

Let us now go back to the question of determining the importance of variables. How do we demonstrate that a variable or pillar is influent or non influent? If we were able to answer confidently to this question, then maybe we could qualify our statement that 'weights are not measures of importance in linear of aggregation' with a practical statement referred to a given linearly-built composite indicator y: 'When aggregating linearly the variables in composite indicate y the developers make an error *thus* quantifiable'. The *thus* is the subject of the present paper.

2. Weights versus importance

The developer of a composite indicator produces and assigns weights as measures of the importance of the various variables or pillars, whether or not a linear aggregation such as (1) is being used. Thus one would expect that a measure of correlation (Pearson or Spearman) between y and x_i will give a value which, while not identical to W_i, would at least not contradict it openly. We would for example call a contradiction a variable/pillar weighting 0.5, i.e. corresponding to 50% of the total of all weights, and correlating (either Pearson or Spearman) below the .01 level with y. Can we do more that? Especially for linear aggregation, can we develop or adopt a measure informing the developer of the percentage error made in assuming weights equal to importance? Normally $\sum_{i=1}^{k} w_i = 1$, thus if we could dispose of a likewise normalized importance measure S_i such that $\sum_{i=1}^{k} S_i = 1$ it would be natural to measure the error as some function of the distance between w_i and S_i as a measure of the error made in assigning that weight, and take an average of the above over the k weights as a measure of quality of the indicator y itself.

We propose to use as S_i the measure:

$$S_i = \frac{V_{x_i}\left(E_{\mathbf{x}_{\sim i}}\left(y \mid x_i\right)\right)}{V(y)} \tag{4}$$

Where $V_{x_i}\left(E_{\mathbf{x}_{\sim i}}\left(y \mid x_i\right)\right)$ is the reduction in variance which one can expect by fixing a factor, (Saltelli and Tarantola, 2002). This measure which is used in sensitivity analysis is also known as Karl Pearson's 'correlation ratio' η^2 (Pearson, 1905). The term $V(y)$ in (4) is simply the unconditional variance.

S_i is a powerful measure of importance, in that it offers a precise definition of importance – we can define a factor's (or pillar's) importance in a CI 'the expected reduction in variance which would be obtained if the factor / pillar could be fixed'. A shortcoming of the S_i is that they do not add up to one for the case of correlated variables, though we can re-normalize them for the purpose of the analysis, as shown in the next section. More details on the rational for this choice of S_i – and how it is computed from the indicator and its input variables, can be found in (Paruolo et al., 2011).

3. An application to the THES and SJTU ranking

The variance based importance measures of the first order (Eq. (4)) have been computed for two well known composite indicators of university performance: the Academic Ranking of World Universities by Shanghai's Jiao Tong University (SJTU) and the one associated to the UK's Times Higher Education Supplement (THES). Variables and weights used in the 2008 edition of the ranking are given in Tables 1 and 2. The following remarks can be made:

- There is much more correlation among the SJTU variables than in the THES ones.
- Although SJTU developers partition their set in more importance variables (weight 20%) and less important ones (weight 10%) the variables end up being more evenly distributed in importance. Normalized importance range between 15% and 19%, see Table 3.
- The problems are more serious for THES. Recruiter Review should weight only a 10% and seems to have instead an importance of more than double. Conversely the variable Full-time equivalent

faculty per student ratio which should weight for a 20% weights instead less than one half of that. This means that overall THES relies on peer review (Academic plus recruiter) more than implied by the developers weights.

4. Conclusions

In (Paruolo et al., 2011), of which this contribution is a summary, we have tackled the issue of testing the veracity of a composite indicator by comparing the importance of its dimensions – be these variables or pillars – against a plausible statistical measure. The issue of testing the quality of CI, also in relation to their normative implications, is a relevant one (Stilgitz et al., 2009, p.65), and variables' importance is clearly a normative element. Our approach to test the quality of a composite indicator is a non invasive one, e.g. it does not require assumptions about modeling the error in the construction of the index, as in standard sensitivity analysis (Saisana et al. 2005, Saltelli et al., 2008).

Tables

Indicator	Weight
Alumni of an institution winning Nobel Prizes and Fields Medals	10%
Staff of an institution winning Nobel Prizes and Fields Medals	20%
Highly cited researchers in 21 broad subject categories	20%
Articles published in Nature and Science	20%
Articles in Science Citation Index Expanded, Social Sciences Citation Index	20%
Academic performance with respect to the size of an institution	10%

Table 1: Variables in SJTU

Indicator	Weight
Academic Opinion: Peer review, 6354 academics	40%
Citations per Faculty: Total citation/ Full Time Equivalent faculty	20%
Recruiter Review: Employers' opinion, 2339 recruiters	10%
International Faculty: Percentage of full-time international staff	5%
International Students: Percentage of full-time international students	5%
Student Faculty: Full-time equivalent faculty/student ratio	20%

Table 2: Variables in THES

REFERENCES (RÉFERENCES)

Condorcet, M. de, 1785. Essai sur l'application de l'analyse la probabilit des dcisions rendues la probabilit des voix, De l' Imprimerie Royale, Paris.

Munda, G., 2008. Social Multi-criteria Evaluation for a Sustainable Economy, Springer, Berlin.

Paruolo, P., Saltelli, A., Saisana, A., 2011, Ratings and rankings: Voodoo or Science? Submitted to the Journal Royal Statistical Society A, March 2011.

Pearson, K. 1905, On the General Theory of Skew Correlation and Non-linear Regression, volume XIV of Mathematical Contributions to the Theory of Evolution, Drapers' Company Research Memoirs. Dulau & Co., London, Reprinted in: Early Statistical Papers, Cambridge University Press, Cambridge, UK, 1948.

OECD/EC-JRC, 2008, Nardo, M. M. Saisana, A. Saltelli and S. Tarantola (EC/JRC), A. Hoffman and E. Giovannini (OECD) eds, Handbook on Constructing Composite Indicators: Methodology and User Guide, 2008, JRC-OECD publication ISBN: 9789264043466.

Saisana, M., Saltelli, A., Tarantola, S., 2005. Uncertainty and sensitivity analysis techniques as tools for the analysis and validation of composite indicators. Journal of the Royal Statistical Society A, 168(2), 307-323.

SJTU (503 universities)	Original weights	S_i (rescaled)	S_i (Eq. (4))
Medals alumni	10%	15%	0.70
Medals staff	20%	16%	0.72
Highly cited	20%	19%	0.87
Nature&Science	20%	19%	0.89
Articles	20%	15%	0.69
Size adjusted	10%	16%	0.74
	100%	100%	4.61
THES (400 universities)			
Peer review	40%	36%	0.82
Citation per faculty	20%	17%	0.38
Recruiter review	10%	24%	0.54
International faculty	5%	5%	0.12
International student	5%	7%	0.16
Faculty per student	20%	9%	0.21
	100%	100%	2.22

Table 3: Main effects versus nominal weights.

Saltelli, A., 2007. Composite indicators between analysis and advocacy. Social Indicators Research. 81 (1), 65-77.

Saltelli, A. , M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, 2008 Global Sensitivity Analysis. The Primer. John Wiley and Sons.

Saltelli, A. and S. Tarantola, 2002, On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal. Journal of American Statistical Association, 97:702–709.

Stiglitz, J.E., Sen, A., Fitoussi JP, 2009, Report by the Commission on the Measurement of Economic Performance and Social Progress, www.stiglitz-sen-fitoussi.fr.

RÉSUMÉ (ABSTRACT)

Composite indicators are built by combining input variables within a mathematical model. The mathematical model can be thought of as made up of all treatments applied to the data as well as of their combination in the index. Each variable in the index is customarily attached a weight purportedly meant to appreciate the 'importance' of that variable. By far the most common strategy to aggregate variables seen in existing composite indicators is by a weighted arithmetic average of normalized variables, whereby the composite index y is derived from a set of k normalized variables. Weights are attached by the developers on the basis of different strategies, be those statistical, such as e.g. factor analysis, or based on expert evaluation, such as e.g. analytic hierarchy process. Weights represents a form of judgement of the relative importance of the different variables, including the case of equal weights where all variables are (in theory) equally important. Using methods derived from global sensitivity analysis we show that important discrepancies exist in most composite indicators between declared weights and effective importance of variables or pillars.

Step 1 - Input table to the multicriteria analysis example. Three countries (A,B,C) need to be ranked according to five variables (GDP, Unemployment rate, Solid waste, Income Disparity, Crime Rate) and a set of weights. Example from [?].

Country	GDP (+)	Unemployment (-)	Solid waste (-)	Disparity (-)	Crime (-)
A	25,000	0.15	0.4	9.2	40
B	45,000	0.10	0.7	13.2	52
C	20,000	0.08	0.35	5.3	80
weights	.166	.166	0.333	.166	.166

Step 2 - Countries are compared pairwise. For each comparison, e.g. AB, all the weights corresponding to indicators that favour A versus B are added up as evidence in favour of 'A better than B'. In this case AB gets the weights of Waste, Disparity and Crime. BA gets the weights of the remaining indicators, GDP and unemployment

Couple	Evidence
AB	0.333+0.166+0.166=0.666
BA	0.166+0.166=0.333
AC	0.166+0.166=0.333
CA	0.166+0.333+0.166=0.666
BC	0.166+0.166=0.333
CB	0.166+0.333+0.166=0.666

Step 3 - The resulting outranking matrix $\mathbf{O}$. Note that entries $\mathbf{O}_{ij}$ and $\mathbf{O}_{ji}$ add up to one.

	A	B	C
A	0	0.666	0.333
B	0.333	0	0.333
C	0.666	0.666	0

Step 4 - The matrix $\mathbf{O}$ is used to compare the orderings. For example the ordering BCA gets as support the sum of the entries $\mathbf{O}_{BC}, \mathbf{O}_{BA}, \mathbf{O}_{BA}$. The preferred countries' ordering is hence CAB (support=2).

Order	Support
ABC	0.666 + 0.333 + 0.333 = 1.333
ACB	0.333 + 0.666 + 0.666 = 1.666
BAC	0.333 + 0.333 + 0.333 = 1
BCA	0.333 + 0.666 + 0.333 = 1.333
CAB	0.666 + 0.666 + 0.666 = 2
CBA	0.666 + 0.333 + 0.666 = 1.666

Table 4: A non compensatory multi criteria ordering. When using this approach, due to Condorcet and known as Condorcet-Kemeny-Young-Levenglick (CKYL) method (Munda, 2008), the input variables do not need to be normalized. In the comparison of country A against country B, the weight of each variable is unambiguously assigned to country A or B depending on whether the variable is higher for A or B, irrespective of the entity of the difference. In this sense one can say that in this approach weights are truly measures of importance.