Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators

A. Saltelli
European Commission,
Joint Research Centre of Ispra,
Italy
andrea.saltelli@jrc.it

Prepared with Michaela Saisana and Stefano Tarantola - based on a paper revised for JRSS-A

Outline

• CI controversy
• Composite Indicators as models
• Wackernagel’s critique of ESDI ...
• Methods
• Put the critique into practice: the TAI example
• Conclusions
Pros

- Composite indicators can be used to summarize complex or multi-dimensional issues, in view of supporting decision-makers.
- Composite indicators provide the big picture. They can be easier to interpret than trying to find a trend in many separate indicators. They facilitate the task of ranking countries on complex issues.
- Composite indicators can help attracting public interest by providing a summary figure with which to compare the performance across countries and their progress over time.
- Composite indicators could help to reduce the size of a list of indicators or to include more information within the existing size limit.
Cons

• Composite indicators **may send misleading, non-robust policy messages** if they are poorly constructed or misinterpreted. Sensitivity analysis can be used to test composite indicators for robustness.

• The simple “big picture” results which composite indicators show may invite politicians to draw **simplistic policy conclusions**. Composite indicators should be used in combination with the sub-indicators to draw sophisticated policy conclusions.

• The construction of composite indicators involves stages where **judgement** has to be made: the selection of sub-indicators, choice of model, weighting indicators and treatment of missing values etc. These judgements should be transparent and based on sound statistical principles.

• There could be **more scope for Member States** about composite indicators than on individual indicators. The selection of sub-indicators and weights could be the target of political challenge.

• The composite indicators increase the **quantity of data** needed because data are required for all the sub-indicators and for a statistically significant analysis.
Pros & Cons
(JRSS paper)

“[...] it is hard to imagine that debate on the use of composite indicators will ever be settled [...] official statisticians may tend to resent composite indicators, whereby a lot of work in data collection and editing is “wasted” or “hidden” behind a single number of dubious significance. On the other hand, the temptation of stakeholders and practitioners to summarise complex and sometime elusive processes (e.g. sustainability, single market policy, etc.) into a single figure to benchmark country performance for policy consumption seems likewise irresistible.”
Indicators as models … and the critique of models

The nature of models, after Rosen
The critique of models

After Rosen, 1991, “World” (the natural system) and “Model” (the formal system) are internally entailed - driven by a causal structure. [Efficient, material, final for 'world' - formal for 'model']

Nothing entails with one another “World” and “Model”; the association is hence the result of a craftsmanship.
The critique of indicators

Environmental sustainability Index, figure from The Economist, Green and growing, The Economist, Jan 25th 2001,

Produced on behalf of the World Economic Forum (WEF), and presented to the annual Davos summit this year.
Mathis Wackernagel, mental father of the “Ecological Footprint” and thus an authoritative source in the Sustainable Development expert community, concludes an argumented critique of the study done presented at Davos by noting:
"Overall, the report would gain from a more extensive peer review and a sensitivity analysis. The lacking sensitivity analysis undermines the confidence in the results since small changes in the index architecture or the weighting could dramatically alter the ranking of the nations."
How to tackle it: the TAI example

We have tried to address the issue of robustness taking as example the UN Technology Achievement index (2001), Human Development Report series
The TAI example

Technology achievement index: 4 dimensions, 8 indicators

I) Creation of technology

1-Number of patents granted per capita (to reflect the current level of invention activities),

2-Receipts of royalty and license fees from abroad per capita (to reflect the stock of successful innovations of the past that are still useful and hence have market value).
The TAI example

Technology achievement index: 4 dimensions:

II) Diffusion of recent innovations

3-Diffusion of the Internet (indispensable to participation),

4-Exports of high-and medium-technology products as a share of all exports.
The TAI example

Technology achievement index: 4 dimensions:

III) Diffusion of old innovations

5-Telephones

6-Electricity

(Epecially important because they are needed to use newer technologies and are also pervasive inputs to a multitude of human activities. Both indicators are expressed as logarithms, as they are important at the earlier stages of technological advance but not at the most advanced stages)
The TAI example

Technology achievement index: 4 dimensions:

IV) Human skills

7-Mean years of schooling

8-Gross enrolment ratio of tertiary students enrolled in science, mathematics and engineering.
The TAI example: data covariance

Technology Achievement index.

• Correlation analysis reveals that the eight sub-indicators have an average bivariate correlation of 0.55 and that 6 pairs of indicators have a correlation coefficient higher than 0.70.
The TAI example: digression on data covariance

- Depending on the school of thought, one may see a high correlation among sub-indicators as something to correct for, e.g. by making the weights inversely proportional to the strength of the overall correlation for a given sub-indicator, see e.g. Index of Relative Intensity of Regional Problems in the EU (Commission of the European Communities, 1984).
The TAI example: digression on data covariance

- Practitioners of multicriteria decision analysis would instead tend to consider the existence of correlations as a feature of the problem, not to be corrected for, as correlated sub-indicators may indeed reflect non-compensable different features of the problem.
How to tackle it: the TA example

Tools:

Uncertainty Analysis (UA) and Sensitivity Analysis (SA). UA focuses on how uncertainty in the input factors propagates through the structure of the composite indicator and affects the composite indicator values. SA studies how much each individual source of uncertainty contributes to the output variance.
Edward E. Leamer, 1990

"I have proposed a form of organised sensitivity analysis that I call "global sensitivity analysis" in which a neighborhood of alternative assumptions is selected and the corresponding interval of inferences is identified. Conclusions are judged to be sturdy only if the neighborhood of assumptions is wide enough to be credible and the corresponding interval of inferences is narrow enough to be useful."
Uncertainty analysis = Mapping assumptions onto inferences
Sensitivity analysis = The reverse process

Simplified diagram - fixed model
Sources for SA:

- book (2000),
- primer (2004)
- free software
How to tackle it: the TA example

Two normalisation methods (rescaling and standardising) and two participatory approaches for assigning weights to the sub-indicators (budget allocation and analytic hierarchy process) have been applied.
Answers being sought:

(1) Does the use of one normalisation method and one set of weights in the development of the composite indicator (e.g. original TAI) provide a biased picture of the countries’ performance?

(2) To what extent do the uncertain input factors (normalisation methods, weighting schemes and weights) affect the countries’ ranks with respect to the original TAI?
Set up of the analysis - Weighting schemes

Two participatory methods:

Budget allocation (experts allocate a finite number of points among the set of indicators)

or

Analytic hierarchy process (expert compare the indicators pairwise, and express numerically the relative importance on a 1 [indifference]-9 [much more important] scale. Possibility of inconsistencies.)
Set up of the analysis - Normalisation

Rescaling or standardising

\[ Y_c = \sum_{q=1}^{Q} I_{q,c} \cdot w_q, \text{ where} \]

\[ I_{q,c} = \begin{cases} 
\frac{x_{q,c} - \min(x_q)}{\text{range}(x_q)} \\
\frac{x_{q,c} - \text{mean}(x_q)}{\text{std}(x_q)} 
\end{cases} \]
1. selection of sub-indicators
2. data selection
3. data editing
4. data normalisation
5. weighting scheme
6. weights’ values
7. composite indicator formula

Limits of the analysis: we considered alternative weighting approaches
(1) alternative normalisation methods for the sub-indicators values,
(2) uncertainty in the weights of the sub-indicators (both scheme and value).
Steps in building a composite indicator

1. selection of sub-indicators
2. data selection
3. data editing
4. data normalisation
5. weighting scheme
6. weights’ values
7. composite indicator formula

\[ Y_c = \sum_{q=1}^{8} I_{q,c} \cdot w_q \]

\( x_i \) Inputs

\( w_i \) Weights

\( Y_c \) Model

\( I_{q,c} \) indicators

\( w_q \) weights

Values of Technology Achievement Index for different countries
## Parameterisation of uncertain inputs

### 10 uncertain input factors for the analysis

<table>
<thead>
<tr>
<th>Input factor</th>
<th>Definition</th>
<th>pdf</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Re/St</td>
<td>Normalisation method</td>
<td>Uniform</td>
<td>[0, 1], where [0, 0.5] ≡ Re-scaling</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.5, 1] ≡ Standardising</td>
</tr>
<tr>
<td>BA/AHP</td>
<td>Weighting scheme</td>
<td>Uniform</td>
<td>[0, 1], where [0, 0.5] ≡ BA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.5, 1] ≡ AHP</td>
</tr>
<tr>
<td>W-patents</td>
<td>Weights’ list for Patents</td>
<td>Discrete uniform</td>
<td>[1, 2, …, 20]</td>
</tr>
<tr>
<td>W-royalties</td>
<td>Weights’ list for Royalties</td>
<td>Discrete uniform</td>
<td>[1, 2, …, 20]</td>
</tr>
<tr>
<td>W-internet</td>
<td>Weights’ list for Internet</td>
<td>Discrete uniform</td>
<td>[1, 2, …, 20]</td>
</tr>
<tr>
<td>W-exports</td>
<td>Weights’ list for Exports</td>
<td>Discrete uniform</td>
<td>[1, 2, …, 20]</td>
</tr>
<tr>
<td>W-telephone</td>
<td>Weights’ list for Telephone</td>
<td>Discrete uniform</td>
<td>[1, 2, …, 20]</td>
</tr>
<tr>
<td>W-electricity</td>
<td>Weights’ list for Electricity</td>
<td>Discrete uniform</td>
<td>[1, 2, …, 20]</td>
</tr>
<tr>
<td>W-schooling</td>
<td>Weights’ list for Schooling</td>
<td>Discrete uniform</td>
<td>[1, 2, …, 20]</td>
</tr>
<tr>
<td>W-enrolment</td>
<td>Weights’ list for Enrolment</td>
<td>Discrete uniform</td>
<td>[1, 2, …, 20]</td>
</tr>
</tbody>
</table>
### Zooming in the input factors

<table>
<thead>
<tr>
<th>Factor</th>
<th>Definition</th>
<th>pdf</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Re/St</td>
<td>Normalisation method</td>
<td>Uniform</td>
<td>[0, 1], where [0, 0.5] ≡ Re-scaled values (0.5, 1] ≡ Standardised values</td>
</tr>
<tr>
<td>BA/AHP</td>
<td>Weighting scheme</td>
<td>Uniform</td>
<td>[0, 1], where [0, 0.5] ≡ Budget Allocation (0.5, 1] ≡ Analytic Hierarchy Process</td>
</tr>
<tr>
<td>W-patents</td>
<td>Weights’ list for Patents</td>
<td>Discrete uniform</td>
<td>[1, 2, …, 20]</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>
Steps in the analysis

(a) Assign a pdf to each input factor $X_i$.

The first input factor $X_1$ is the trigger to select the type of normalisation method (0-1);

The second input factor $X_2$ is the trigger to select the weighting scheme (0-1).

Factors $X_3, \ldots X_{10}$ are random numbers used to select the uncertain weights (1,2,...20 because there are 20 experts);

2 triggers + 8 weights = 10 uncertain factors
(b) Generate randomly N combinations of independent input factors $\mathbf{X}^L$, with $L = 1, \ldots, N$

$$\mathbf{X}^L \equiv (X_1^L, X_2^L, \ldots X_{10}^L)$$

This is a sample; a trial set for the evaluation of the output.
Steps in the analysis

(c) For each sample $L$, select normalisation method

and weighing scheme based on sampled values of $X_1$, $X_2$

respectively;

(d) For each sample $L$ use factors $X_3, \ldots X_{10}$ to select

the weights;
Steps in the analysis

(e) Evaluate the model computing the output value $Y^L$

(f) Close the loop over $L$, and analyse the resulting output vector $Y^L$, with $L = 1, ..., N$. 
Uncertainty analysis - Results for

The values of the composite indicator are displayed in the form of confidence bounds.

Blue - original TAI
Red - median of Monte Carlo TAI
Uncertainty analysis - Results for

Significant difference TAI – median of MonteCarlo TAI

A few countries show significant overlap and therefore the ranking is unclear

Blue - original TAI
Red - Monte Carlo TAI
Uncertainty analysis - Results for $Y_c$

Answer to question 1

For most countries, the original TAI value is very close to the MC-TAI median value. This implies that the original TAI (one normalisation method and one set of equal weights), provides a picture of the countries' technological achievements that is not generally biased.

(Exception: Netherlands - Singapore)
Uncertainty Analysis for $D_{AB} = \sum_{q=1}^{Q} (I_{q,A} - I_{q,B}) \cdot w_q$

In cases where partial overlapping between two countries occurs, the difference in the TAI values for that pair of countries can be further analyzed in via sensitivity analysis.
Moving from uncertainty to sensitivity analysis

Uncertainty Analysis

Sensitivity analysis

Each slice is the fractional contribution of a factor [or group of factors] to variance of the output
Our recommended practice is based on two fractional variance indices—one is a first order effect— one factor influence by itself

\[
S_i \equiv \frac{V_{X_i} \left( E_{X_{-i}} \left( Y \mid X_i \right) \right)}{V_Y}
\]

The other is a factors total influence inclusive of all interaction with other factors

\[
S_{Ti} \equiv \frac{E_{X_{-i}} \left( V_{X_i} \left( Y \mid X_{-i} \right) \right)}{V_Y}
\]
Sensitivity analysis - which uncertain input affects the difference Singapore - Netherlands?

Fixing non influential factors

<table>
<thead>
<tr>
<th>Input factors</th>
<th>First order ($S_i$)</th>
<th>Total effect ($S_{Ti}$)</th>
<th>$S_{Ti} - S_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Re/St</td>
<td>0.00</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>BA/AHP</td>
<td>0.14</td>
<td>0.56</td>
<td>0.42</td>
</tr>
<tr>
<td>W-patents</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>W-royalties</td>
<td>0.01</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>W-internet</td>
<td>0.02</td>
<td>0.05</td>
<td>0.03</td>
</tr>
<tr>
<td>W-exports</td>
<td>0.02</td>
<td>0.17</td>
<td>0.15</td>
</tr>
<tr>
<td>W-telephones</td>
<td>0.02</td>
<td>0.07</td>
<td>0.06</td>
</tr>
<tr>
<td>W-electricity</td>
<td>0.17</td>
<td>0.37</td>
<td>0.20</td>
</tr>
<tr>
<td>W-schooling</td>
<td>0.02</td>
<td>0.10</td>
<td>0.08</td>
</tr>
<tr>
<td>W-enrolment</td>
<td>0.12</td>
<td>0.29</td>
<td>0.16</td>
</tr>
<tr>
<td><strong>Sum</strong></td>
<td><strong>0.52</strong></td>
<td><strong>1.69</strong></td>
<td><strong>1.17</strong></td>
</tr>
</tbody>
</table>
Which uncertain input affects the difference Singapore – Netherlands?

Further analysis (e.g. scatter-plots) reveals that MC-TAI favours Singapore when high weights are assigned to Enrolment sub-indicator, for which Singapore is much better than the Netherlands, and/or to Electricity sub-indicator, for which Singapore is marginally better than the Netherlands.
Trigger_weighting

Weight of Enrolment

Netherlands better  ▲ Similar  └ Singapore better

Budget Allocation  Analytic Hierarchy Process
Inference from sensitivity analysis for the difference Singapore - Netherlands? (Question 2)

- The weights for Electricity and Enrolment and the type of weighting approach (BA or AHP) are important.
- The selection of the normalisation method does not affect the output variance.
- All the input factors, taken singularly, explain 52% of the output variance. The remaining 48% is explained by interactions among the factors.
- The trigger for the weighting scheme has a strong interaction with other factors, mainly with the weights for Electricity, Enrolment and Exports.
- The high value of $S_{Ti}$ for the weight of Export tells us that it cannot be fixed in spite of its low $S_{i}$.
Worthiness of the MC version of the TAI

- With MC, now many countries overlap – does the MC-TAI hence take away meaning from TAI?

- Not necessarily – there are considerable correlations among TA’s, so overlapping bands do not imply overlapping performance (→ scatter-plots)

- Does the MC TAI impair using the index in a name-and-shame fashion?

- Not necessarily – the comparison is apparently weaker, but it is in fact more robust (defensible). You can still tell the leaders from the laggards (see ranks).
Conclusions

The analysis was useful in showing that:

• The original TAI was not generally biased.
• Exceptions (Netherlands and Singapore) can be identified and explained = identifying regions in the space of the weights that favour one country with respect to another → input for iterative CI building
• The weighting approach matters, normalisation does not → input for iterative CI building
• An iterative use of uncertainty and sensitivity analysis in the phase of indicators building can help structuring the composite indicators, providing information on whether the countries’ ranking measures anything meaningful and reducing the possibility that composite indicators may send misleading or non-robust policy messages.
Conclusions

The analysis was nevertheless partial in that:

• No uncertainty in the sub-indicators
• Variability in the weights to capture all the debate plurality ...
• What if the very concept of composite indicator is rejected?
• What if weighting is rejected?
Conclusions

What if weighting is rejected?
• Munda and Nardo, 2003 have argued that even if weights are customarily assigned as measure of relative importance when using linear aggregation, they have in fact a meaning as substitution rate, whereby e.g. an equal weight for two indicators would mean that we are willing to trade one unit down in one indicator for one unit up in another.
• Ebert and Welsch 2003 claim that if you have ratio scale non comparable variables, linear aggregation (even if with normalisation) is still scale sensitive ...
References


References


References


