Appendix II. Preliminary Sensitivity Analysis

Michaela Saisana & Andrea Saltelli
European Commission – Joint Research Centre – IPSC, ITALY

The main advantage and added value of the Environmental Performance Index (EPI) is that an aggregated index, with a set of environmental indicators measuring different aspects of sustainability, is more reliable than looking at each indicator separately. The Pilot Trend EPI, with information on the trends of nations' sustainability levels over the last eleven years (2000-2010), is a particularly valuable addition to the 2012 EPI. There are, however, practical challenges in the EPI related to the quality of available data and the aggregation of these into a single number.

Assessing the conceptual and statistical coherence of the EPI and estimating the impact of modelling assumptions on a nation's sustainability level serves a two-fold purpose: (a) it ensures the transparency and reliability of the EPI, and (b) it enables policymakers to derive more accurate and meaningful conclusions. Yale and Columbia Universities have invited the European Commission Joint Research Centre (JRC) in Ispra-Italy to assess each EPI report since its launch in 2006. The JRC researched extensively the quality of composite indicators and ranking systems that classify countries' performances along policy lines (OECD, 2008; Saisana *et al*., 2005; 2011; Saltelli *et al.* 2008, Paruolo *et al*., 2012).[i]

The statistical assessment of the 2012 EPI was done along three main avenues: an evaluation of conceptual/statistical coherence of its structure, an interpretation of the rankings based on significance tests, and an evaluation of the impact of key modelling assumptions (e.g., weighting and aggregation) on nations' EPI scores and ranks. This short note summarises the main findings from the first analysis on the conceptual/statistical coherence of the EPI structure. Detailed findings on all three types of analysis will be available online at www.epi.yale.edu by mid-March 2012.

Conceptual and statistical coherence in the EPI

As described in the main text of the EPI report, the EPI scores for nations worldwide are computed as the simple (or weighted) averages within and across ten policy categories and two objectives (*Ecosystem Vitality* and *Environmental Health*) for a total of 22 indicators. Each of those indicators offers a partial picture of a nation's sustainability level. The intention of the EPI is to provide a more reliable overall picture of sustainability levels around the world than any single indicator would provide taken independently.

The data delivered to the JRC at the time of writing represented normalized values (target-driven min-max method) of 22 treated variables (e.g., logarithmic transformation) together with the country scores on ten policy categories, two objectives and the overall EPI on an annual basis between 2000-2010. These normalized indicators are not affected by outliers or skewed distributions,[ii] except for *outdoor air pollution* (described by $PM_{2.5}$) and *CO2 emissions per kWh*. However, the skewed distributions of those variables do not bias significantly the results of the respective EPI objective (i.e. Environmental Health in the first case or Ecosystem Vitality in the second case). The 2000-2010 dataset is characterized by excellent data coverage (93 percent data availability in a matrix of 22 variables × 132 countries × 11 years). Data coverage per EPI objective, country, or year is also very good or excellent.

Researchers used principal component analysis (PCA) on the 2000-2010 dataset to assess the extent to which the conceptual framework is confirmed by statistical approaches and to identify eventual pitfalls. The analysis confirms, in part, the EPI structure: for Environmental Health, the first latent factor of the three policy categories captures 83 percent of the variance; for Ecosystem Vitality, the first latent factor of the seven policy categories describes only 31 percent of the total variance. These results suggest the use of arithmetic average across the policy categories is statistically justified for Environmental Health but questionable for Ecosystem Vitality.

Next, tests focused on identifying whether the EPI and the two EPI objectives are statistically well-balanced in the underlying components. Unlike past releases of the EPI where the two objectives received equal weights, the 2012 weights them at 3/10 and 7/10, respectively. The EPI team was aiming for scores that were not dominated by one of the two objectives, but the weights also reflect the number of policy categories included in each objective. The same goal guided the choice of the weights at the policy category level. Hence, in the present context, our analysis answers the question: 'is the EPI country classification dominated by just one of the two EPI objectives (or just one or two policy categories)?' We have used a non-linear 'importance measure' (henceforth S_i), known as correlation ratio or first order sensitivity measure (Saltelli *et al.*, 2008). The S_i describes 'the expected reduction in the variance of EPI scores that would be obtained if a given objective (or policy category) could be fixed'. As discussed in Paruolo *et al.*, 2012, we can take this as a measure of importance; thus, if the two EPI objectives or the ten EPI policy categories are all expected to contribute significantly to determining the EPI country classification, their S_i values should not differ too much. A more detailed discussion of this non-linear analysis will be available by March 2012 on the EPI website.

Results are reassuring for the overall EPI. The EPI objectives are both important in classifying countries on an annual basis in the overall EPI, (S_i values between 0.2 and 0.5 over the years; see Table 1 for results in the latest year, second column), although Ecosystem Vitality appears to have a greater impact. For simplicity, one may look at the linear approximation to S_i (i.e. the squared Pearson product moment correlation coefficients; see Table 1, third and sixth column) with the caveat that these are more suitable for linear relations.

When looking at the impact of the ten policy categories on the overall EPI, there is no dominance issue, though Child Mortality and Water (ecosystem) appear to be slightly more important, while Air Pollution (ecosystem) and Forestry have the least impact on the variance of the EPI scores.

Environmental Health is balanced with respect to Water & Sanitation and Child Mortality (S_i values close to 0.9; see fifth column), but Air Pollution has less impact than expected (S_i~0.4).

Ecosystem Vitality appears to be less balanced. Although there is no particular dominance issue in the four policy categories – Air, Water, Biodiversity and Climate Change – all have the same impact on the Ecosystem Vitality score, but the remaining three policy categories – Forestry, Marine & Fisheries and Agriculture – have practically “no saying” on the *Ecosystem Vitality* classification.

Table 1. Importance measures for the EPI 2012 components

EPI component	**Importance measures for EPI**		**Weights within EPI**	**Importance measures for the two EPI Objectives**		**Weights within objectives**
	S_i non linear[1]	**S_i linear[2]**		**S_i non linear[1]**	**S_i linear[2]**	
Environmental Health	0.231 (0.057)	0.329	30%			
Ecosystem Vitality	0.489 (0.076)	0.415	70%			
				Environmental Health		
Air Pollution (health)	0.165 (0.092)	0.267	8%	0.455 (0.100)	0.661	25%
Water & Sanitation (health)	0.279 (0.122)	0.289	8%	0.925 (0.045)	0.886	25%
Child Mortality	0.415 (0.078)	0.300	15%	0.938 (0.022)	0.918	50%
			9%	**Ecosystem Vitality 2012**		
Air pollution (ecosystem)	0.108 (0.051)	0.135	9%	0.410 (0.081)	0.363	13%
Water (ecosystem)	0.074 (0.059)	0.166	18%	0.342 (0.066)	0.388	13%
Biodiversity & Habitat	0.438 (0.080)	0.448	6%	0.484 (0.091)	0.444	25%
Forestry	0.121 (0.063)	0.000	6%	0.076 (0.038)	0.081	8%
Marine & Fisheries	0.041 (0.032)	0.015	6%	0.021 (0.026)	0.001	8%

Agriculture	0.166 (0.067)	0.005	18%	0.051 (0.055)	0.022	8%
Climate change	0.116 (0.042)	0.008	8%	0.461 (0.081)	0.446	25%

Source: European Commission Joint Research Centre

Notes: (1) Numbers represent the average kernel estimates of the Pearson correlation ratio ( ) calculated by bootstrap (1000 samples). (2) Numbers represent the Pearson correlation coefficient (squared). (3) Bootstrap standard deviations for the correlation ratio are given in parenthesis. (4) Results are based on the data reported for 2010.

he same type of analysis across the five variables underlying *Environmental Health* shows that outdoor air pollution ($PM_{2.5}$) seems to have a much lower impact with respect to the other four variables in the Environmental Health country classification. For Ecosystem Vitality, seven out of 17 variables are randomly associated to Ecosystem Vitality, which suggests that even if countries make an effort to improve in those variables, this will not necessarily be translated into an improvement in their Ecosystem Vitality classification.

The negative association between the two EPI objectives (ranging between -0.5 and -0.2 over the years 2000-2010) strongly suggests that Environmental Health and Ecosystem Vitality should not be aggregated linearly into a single number but rather presented separately or treated with a different, less compensatory aggregation strategy.

To understand this point, consider whether countries with similar overall EPI scores but very different performance on the two objectives should actually be placed on the same level of sustainability. Take, for example, Congo and Armenia:

Congo - Environmental Health =12, Ecosystem Vitality = 62, EPI=47.5
Armenia - Environmental Health =63, Ecosystem Vitality =40, EPI=47.5

Armenia is relatively more balanced in performance across the two EPI objectives, while one may argue that Congo should somehow be penalised for very low performance on Environmental Health.

This is the kind of consideration that led the authors of the Human Development Index to switch from linear to geometric aggregation between the 2009 and 2010 release of the index (see Paruolo et al., 2012 for a discussion). In the case discussed above, Congo would get 37.8 and Armenia 45.8 (using the 30-70 weights of the EPI). Again such important differences associated to the aggregation formula suggest caution – if the two objectives were positively correlated with one another this issue would be less critical, i.e. there would be less countries for which this would be an issue.

Linear aggregation is a simple and easy to communicate, but it is very demanding in terms of the type of data that can be confidently aggregated. Where the data are more complex and with unavoidable trade-offs, as is the case with the rich structure of the EPI data set, linear aggregation does not favour coherence, as discussed for the two countries above.

If that is the case, perhaps future releases of EPI should reflect the compensation issue and switch to a less compensatory aggregation than the linear weighted average. The weighted geometric average of the objectives and/or the policy categories presented above by way of illustration is just one of the possibilities. The consideration of logarithms for most variables in the 2012 EPI methodology is already a step in this direction, and the 2012 EPI, overall, appears a decisive improvement over the 2010 EPI (see the 2010 EPI validation report in Saisana and Saltelli, 2010).

Conclusion

The JRC analysis suggests that the 2012 EPI structure (tested on an eleven year period over 2000-2010) appears a decisive improvement over the 2010 EPI. The 2012 EPI is statistically coherent and balanced with respect to the two objectives on Environmental Health and Ecosystem Vitality and also within Environmental Health. Yet, some reflection is still needed on the construction of the Ecosystem Vitality objective where the use of arithmetic average in combining the information appears problematic due to the negative or random associations between the policy categories. These trade-offs within Ecosystem Vitality are a reminder of the danger of compensating between policy categories while also identifying the areas where more work is needed to achieve a coherent framework – particularly regarding the relative importance of the indicators that compose the EPI framework. Finally, the negative association between the two EPI objectives (ranging between -0.5 and -0.2 over 2000-2010) might be seen as a warning that Environmental Health and Ecosystem Vitality should not be aggregated into a single number but rather presented separately, e.g. with countries displayed on a simple plot (bi-dimensional radar plot) where the Euclidean distance from the origin illustrates the sustainability of the country.

References

OECD/EC JRC (Organisation for Economic Co-operation and Development / European Commission Joint Research Centre). 2008. *Handbook on Constructing Composite Indicators: Methodology and User Guide*. Paris: OECD.

Paruolo, P., Saisana, M., Saltelli, A. 2012. Ratings and rankings: Voodoo or Science? *Journal of the Royal Statistical Society* A (under second revision).

Saisana, M., B. D'Hombres, and A. Saltelli. 2011. 'Rickety Numbers: Volatility of University Rankings and Policy Implications'. *Research Policy* 40: 165–77.
Saisana, M., A. Saltelli. 2010. *Uncertainty and Sensitivity Analysis of the 2010 Environmental Performance Index*. EUR 24269, European Commission Joint Research Centre.
Saisana, M., A. Saltelli, and S. Tarantola. 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): 307–23.
Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. 2008. *Global Sensitivity Analysis: The Primer*. Chichester, England: John Wiley & Sons.

ENDNOTES

---

[i] JRC auditing studies of composite indicators are available at http://composite-indicators.jrc.ec.europa.eu/. JRC has co-authored, with the Organisation for Economic Co-operation and Development (OECD), a Handbook on Constructing Composite Indicators: Methodology and User Guide, whose methodology has largely been used for the present analysis.

[ii] Groeneveld and Meeden (1984) set the criteria for absolute skewness above 1 and kurtosis above 3.5. The skewness criterion was relaxed to 'above 2' to account for the small sample (132 countries).