

Uncertainty and Sensitivity Analysis of the 2008 Environmental Performance Index

*by Michaela Saisana and Andrea Saltelli,
Econometrics and Applied Statistics Unit, Joint Research Centre of the
European Commission, Ispra, Italy*

The analysis presented in this Appendix aims at validating and critically assessing the methodological approach undertaken by the EPI team at Yale and Columbia University. Although this analysis was undertaken in the past versions of the Index, the new data and framework used in the 2008 EPI necessitates repeating this type analysis, so as to ensure that the methodology remains appropriate. At the same time, it aims at identifying those EPI countries with and without very robust ranks. For the first group, policy signals derived from the EPI can be taken with the confidence that changes in the EPI methodology would have a negligible effect on the country's measured performance, while for the latter a more cautious approach is advised vis-à-vis translating the EPI rank into policy actions.

A clear understanding of the EPI methodology is crucial to the success of the robustness assessment of the Index. In a first step, we thus considered if it is possible to reproduce the EPI results given the data and information provided to the public? The answer is "Yes". The EPI website provides enough information to the public that, with some statistical knowledge, anyone could replicate the entire EPI methodology and results.

Indisputably, the construction of the EPI demands a sensitive balance between simplifying an environmental system and still providing sufficient detail to detect characteristic differences (Diener and Suh, 1997). This leaves scientists and policymakers with a complex and synthetic measure that is almost impossible to verify against true conditions, particularly since environmental performance cannot be measured directly (Eyles and Furgal, 2002; von Schirnding 2002). It is therefore taken for granted that the EPI cannot be verified. Yet, in order to enable informed policymaking and be useful as an analytical assessment tool, the EPI needs to be reviewed in regard to its validity and potential biases. The first question to be answered was:

How is the EPI associated to its subcomponents and policy categories?

Following the replication process, correlation analysis is performed to examine the relationship between the EPI scores and the indicator scores, the policy scores, and finally the objectives scores. Correlation analysis is a basic but widely used tool for "confirming" the mathematical design of indices. Booyesen (2002) recommends that a weak correlation between an underlying indicator and an index should result in the exclusion of the respective indicator from the process. A major drawback of correlation analysis though is the fact that a strong correlation does not necessarily imply a strong influence or representation of the indicator in the overall index. In other words, any random variable could potentially show strong correlation with the index without actually being part of the index. A simple rank correlation analysis between the EPI scores and the category scores (Table 1) reveals that the EPI has very high correlation with the

Environmental Health category ($r_s = 0.90$) and the Water (Effects on Ecosystems) category ($r_s = 0.59$), and a fairly strong relationship with the Productive Natural Resources ($r_s = 0.34$) and the Climate Change ($r_s = 0.18$) categories. However, the relation of the EPI to the remaining two of the six policy categories, Air Pollution (Effects on Ecosystems) and Biodiversity & Habitat, appears to be random and non-significant at the 95% level. Relationships among the policy categories themselves vary, but they are in general low and in most cases random. Thus, it appears that the six policy categories represent totally different aspects of environmental performance – which is advantageous from an index development perspective. Although it is not desirable to have very high association between the main components of a composite indicator (since representing different dimensions is a key quality feature of a composite indicator), the negative association between several of the policy categories leads to a conclusion that there may be trade-offs between them. This creates an additional difficulty for an index that combines such different dimensions with the implicit assumption that strong performance on all policy categories is possible simultaneously. In this case it may be argued that there should be no single measure of environmental performance. Rather, the Index should focus on the six policy categories and identify linkages and trade-offs between them instead of attempting to aggregate them into a single score.

Table 1: Spearman rank correlation coefficients for the EPI, the two objectives and the six policy categories

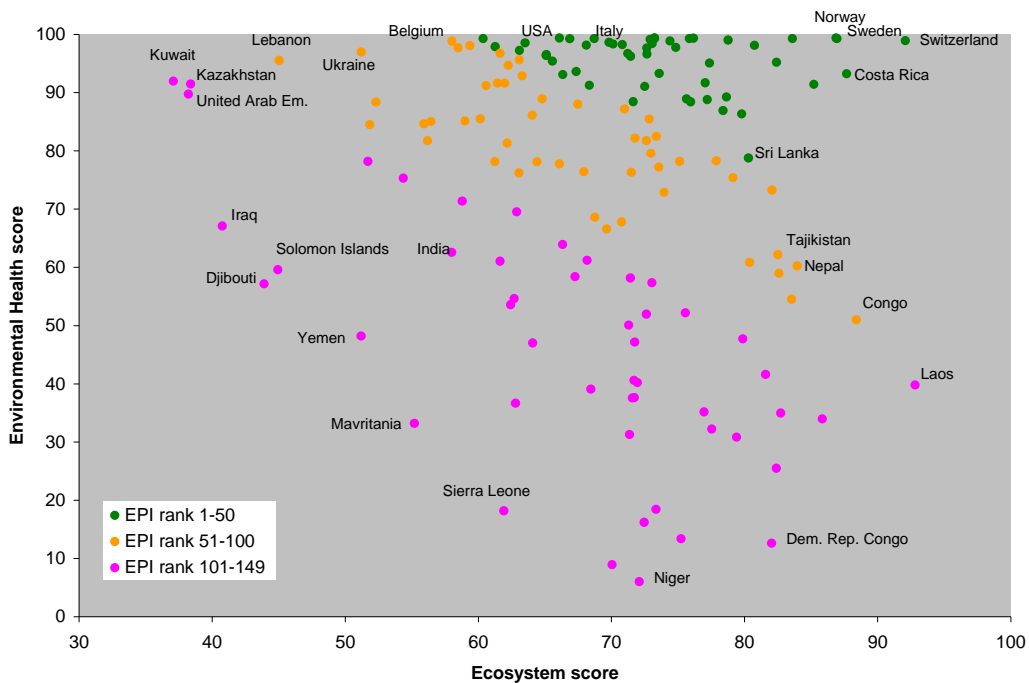
	<i>Policy categories</i>						<i>Objectives</i>	
	Environmental Health	Air Pollution (Effects on Ecosystems)	Water (Effects on Ecosystems)	Biodiversity & Habitat	Productive Natural Resources	Climate Change	ECOSYSTEM	ENVHEALTH
<i>EPI</i>	0.90	-0.09*	0.59	-0.04*	0.34	0.18	0.29	0.90
Environmental Health		-0.18	0.42	-0.22	0.29	-0.16	-0.08*	
Air Pollution (Effects on Ecosystems)			-0.06*	-0.12	0.05*	0.07*		
Water (Effects on Ecosystems)				-0.04*	0.18	0.26		
Biodiversity & Habitat					-0.01*	0.18		
Productive Natural Resources						-0.08*		

* coefficient not significant at the 95% level

The association between the EPI and the 25 underlying indicators reveals that just a few indicators strongly dominant the overall EPI scores. The primary drivers of the EPI rank are the Environmental Burden of Disease (DALY), the Adequate Sanitation (ACSAT), the Drinking Water (WATSUP) and the Indoor Air Pollution (INDOOR) indicators. Somewhat surprisingly, the three indicators related to climate change, although being weighted comparatively strongly, do not exert much influence on the EPI results. Parsimony principles would suggest excluding the non-influential indicators from the EPI framework (Gall, 2007). This, however, may not be advisable from a policy perspective, unless excluding certain indicators is supported by expert opinion on the relevance of the indicators to the issue.

A scatter plot between the two main objectives of the EPI, Environmental Health and Ecosystem Vitality, reveals an understandable - though problematic – trade-off between these two objectives (Figure 1). Countries may end up choosing one or the other path in pursuing environmental performance in a somewhat mutually exclusive pattern, perhaps descriptive of different scales and time horizons. The relationships between Environmental Health and Ecosystem Vitality scores indicate a major problem in translating sustainability-oriented performance into practice. At the same time, the high association between the EPI scores and the Environmental Health scores, and the random association between the EPI scores and the Ecosystem Vitality scores, leads to the conclusion that Ecosystem Vitality scores behaves as noise terms superimposed on Environmental Health scores.

Figure 1. Scatterplot of the Environmental Health scores versus the Ecosystem Vitality scores.



The conclusions from this preliminary analysis already point to the conclusion that the 2008 EPI has an architecture, highlighting the complexity of translating environmental stewardship into straightforward, clear-cut policy recipes. The trade-offs within the index dimensions are a reminder of the danger of compensability (one dimension compensating for another) while also identifying the areas where more work is needed to achieve a coherent framework. In particular, additional consideration should be given to determining the relative importance of the indicators that compose the framework.

Robustness of the EPI results to the methodological assumptions

There is ample evidence of the creativity in the community of composite indicators developers. This creativity arises as a response to the demands of the user/stakeholder community, but also reflects disagreements within the research community as to which indicators influence a particular phenomenon and how they should be weighted (Cutter et al., 2003). When building an

index to capture environmental performance, it is therefore necessary to take stock of existing methodologies to avoid skewing the assessment and decisionmaking.

Many methodological assumptions made when developing an index are intrinsic to policy research and value judgments. Varying these assumptions over a reasonable range of possibilities provides a measure of the importance of the chosen assumptions to the main results (Saisana *et al.*, 2005; Saisana and Tarantola, 2002; Saltelli *et al.*, 2000). The advantages offered by considering different scenarios to build the EPI could be: to gauge the robustness of the EPI results, to increase its transparency, to identify the countries whose performance improve or deteriorate under certain assumptions, and to help frame the debate around the use of the EPI for policymaking. The alternative scenarios to build the EPI should, however, bear certain quality features:

1. No strong dominance of a few indicators at the expense of others in the index.
2. No deliberate bias of the index results against a few countries.
3. Simplicity and easy reproduction of the index.

In the case of the 2008 EPI, the assumptions that needed to be tested, are: (1) the measurement error of the raw data, (2) the choice of capping the 25 indicators at the selected targets, (3) the choice to correct for skewed distributions in the indicator values, (4) the weights assigned to the indicators and/or to the subcomponents of the index, and finally (5) the aggregation function at the policy level. The analysis that we have undertaken maps the effects of these uncertainties and assumptions on the EPI country rankings. We also seek to use uncertainty and sensitivity analyses to assess whether useful conclusions can be drawn from the index given the construction methodology selected.

Sensitivity analysis is the study of how output variation in models such as the EPI can be apportioned, qualitatively or quantitatively, to different sources of variation in the assumptions. In addition, it measures the extent to which the composite index depends upon the information that composes it. Sensitivity analysis is closely related to uncertainty analysis, which aims to quantify the overall variation in the ranking resulting from uncertainties in the model input.

All of the five assumptions discussed above can heavily influence the output—and reliability—of the EPI. Using uncertainty and sensitivity analysis, we systematically evaluated the impact that the methodological and conceptual choices highlighted above have on the robustness of the EPI scoring and ranking. Our study aimed to answer four main questions.

1. How do the EPI ranks compare to the ranks under combinations of alternative scenarios derived from the 5 assumptions?
2. Which countries have the most volatile ranks and why?
3. What are the major sources of variability in the EPI rankings?
4. What associations are there between the EPI and its indicators and/or subcomponents?

Our approach

We tested the five central methodological assumptions by running 40,000 simulations to alter and combine the assumptions in different ways.

To be more specific, the measurement error was introduced by adding to each value in the dataset a random error with a mean equal to zero and standard deviation equal to the observed standard deviation of the corresponding indicator. Some thousands of alternative datasets that included error in some of the data values were generated. The two triggers on capping at target values and correcting for skewed data distributions were binary (yes/no). Regarding the weights to be attached to the indicators and/or the subcomponents, we identified four alternatives to the current one: Factor analysis-derived weights at the indicator level; equal weighting at the indicator level; equal weighting at the subcategory level (and relative weights within each subcategory as in the EPI); equal weighting at the policy level (and relative weights within each policy category and subcategory as in the EPI). Finally, a binary trigger determined the aggregation function to be an arithmetic or geometric mean. In the latter case, the use of a geometric aggregation would penalize countries that compensate very low performance in some policy categories with very high performance in other policy categories. Given that environmental excellence is understood to mean strong performance on the different EPI categories simultaneously, compensation at the policy level should be penalized. We undertook a saturated sampling of the space of input factors.

The combinations of the input factors were translated into a set of $N=40,000$ simulations in a Monte Carlo framework. The composite index was then evaluated N times, and the EPI scores and ranks obtained were associated with the corresponding draws of input factors to appraise their influence. When several layers of uncertainty were simultaneously activated, composite indicators turn out to be non-linear, possibly non-additive models, due to interactions between the input factors (Saisana et al. 2005). As a result, all EPI scores and ranks were non-linear functions of the input factors and the purpose of the uncertainty analysis was the estimation of their probability distribution functions (pdf's).

As argued by practitioners (Saltelli et al., 2000b; EPA, 2004), robust, “model-free” techniques for sensitivity analysis should be used for non-linear models. Variance-based techniques have been shown to yield useful results for sensitivity analysis. For more information the reader is referred elsewhere (e.g., Saltelli et al. 2008).

1. How do the EPI ranks compare to the ranks under all scenarios?

The uncertainty analysis results from the Monte Carlo simulations for the 149 countries are given in detail in Table 2. They reveal whether any deliberate bias against some countries was introduced by the methodological choices made in building the EPI and respond to arguments made by Andrews et al. (2004: 1323) that many indices “rarely have adequate scientific foundations to support precise rankings: [...] typical practice is to acknowledge uncertainty in the text of the report and then to present a table with unambiguous rankings”. The countries shown in Table 2 are ordered by their original EPI score. The numbers in Table 2 represent the probability of a country being among the top 10, top 10-20, and so on. For example, New Zealand has a 98% probability of being among the top 10 performing countries. Costa Rica and Finland follow, with an 81% probability of being ranked among the top 10. Interestingly, Switzerland, which scores top in the original EPI, is almost as likely to be among the top 10, top 10-20, or top 20-30 countries. These probabilities indicate the uncertainty about the country scores in the EPI. In fact, approximately half of the countries in the EPI are placed correctly in the environmental performance ladder, while the other half of the countries can fluctuate

significantly between various positions, and any conclusion on the performance of these countries should be drawn with great caution. The results presented in Table 2 depend on the theoretical framework and the indicators, but are independent of the methodology (methodology-free results), given that they represent a whole set of alternative scenarios. The observed deviations primarily arise from the choice of weights, combined with the choice of the aggregation function at the policy level. As Table 2 demonstrates, countries with high or low performance in the EPI do not have wide variations in their ranks under alternative scenarios. The exceptions to this rule are Austria, Canada, and Iceland. In our simulations Austria ranked between the top 10 to top 40-50. Another interesting example is Iceland (rank: 11) whose score can be anywhere within the top 10-20 to top 80-90. Canada, on the other hand (rank: 12) has a 58% probability to be ranked in the top10 and 33% to be ranked among the top10-20. This result suggests that in fact Canada outperforms Iceland on the environmental issues measured in the EPI given the current framework.

Table 2. Probabilities of country ranks in the Environmental Performance Index under all tested combinations of input factors

	Rank 1-10	Rank 11-20	Rank 21-30	Rank 31-40	Rank 41-50	Rank 51-60	Rank 61-70	Rank 71-80	Rank 81-90	Rank 91-100	Rank 101-110	Rank 111-120	Rank 121-130	Rank 131-140	Rank 141-149
Switzerland	31	30	20	11	6										
Sweden	63	25	10												
Norway	55	31	6												
Finland	81	16													
Costa Rica	81	16													
Austria	15	19	16	21	18										
New Zealand	98														
Latvia	25	39	26	6											
Colombia	74	18	5												
France	15	26	30	14	13										
Iceland	11	15	5	14	15	9	10	13							
Canada	58	33	9												
Germany	14	40	21	20											
United Kingdom	11	44	29	11											
Slovenia	9	18	25	23	8	10									
Lithuania	16	20	14	9	8	9	6	9							
Slovakia	15	21	14	25	5	6	8								
Portugal	20	46	16	11	5										
Estonia	56	34													
Croatia	16	19	23	10	6	9	5	5	5						
Japan	6	38	35	14	5										
Ecuador	63	26	5												
Hungary		6	6	13	16	20	6	10	9						
Italy		6	28	24	16	13	5								
Denmark		8	9	6	15	13	14	8	6	11					
Malaysia	31	48	15	5											
Albania		9	11	6	13	10	16	5	6	9	5				
Russia	9	33	43	9											
Chile	16	46	25	8											
Spain	5	30	18	19	14	11									
Luxembourg	9	15	16	20	26	5	5								
Panama	73	20													
Dominican R	18	54	21	6											
Ireland		5	16	13	15	13	13		9	5					
Brazil	5	20	29	24	11										
Uruguay			11	15	9	8	9	10	9		14				
Georgia		8	8	19	15	16	10	13	5						
Argentina		10	23	28	24	11									
United States	5	23	19	24	13	8									
Taiwan			20	13	19	16	10	13							
Cuba	5	24	29	19	13	5									
Poland	5	11	20	35	15										
Belarus			11	10	10	18	16	16	13						
Greece		8	18	14	19	15	5	10	6						
Venezuela	5	11	36	25	18	5									
Australia	30	30	14	10	9										
Mexico	11	15	34	28	6										
Bosnia & Her			5	10	11	24	9	6	8	14	6				
Israel		5	31	19	19	13	5	6							
Sri Lanka	19	36	16	16	10										
South Korea			6	14	14	19	9	8	13	8					
Cyprus		10	9	25	14	28	6								
Thailand	8	30	35	11	11										
Jamaica		8	15	24	11	11	9	10	5						
Netherlands			9	11	14	10	21	9	11	9					
Bulgaria		5	19	25	15	8	10	6							
Belgium			13	6	11	6	6	16	10	13	9				
Mauritius		6	9	19	18	8	16	15							
Tunisia		5	10	10	10	14	19	18	9						
Peru		15	30	18	30										
Philippines	6	13	26	21	16	9	5								
Armenia			6	13	19	8	16	18	8	6					
Paraguay			11	18	20	18	9	8	5	6					
Gabon	6	35	28	16	5	6									
El Salvador	5	6	13	16	9	10	9	8	9	8	5				
Algeria		5	5	15	26	24	11	6	5						
Iran		11	23	26	18	16									
Czech Rep.			9	8	15	11	13	19	15	10					
Guatemala	10	16	23	26	14	8									
Jordan			8	14	24	20	6	14	8						
Egypt		19	21	24	13	10	6								
Turkey				18	15	18	16	9	6	6					
Honduras	9	28	20	15	13	8	5								
Macedonia			5	5	15	10	18	21	13	6	5				
Ukraine				8	15	6	23	11	10	13	10				

	Rank 1-10	Rank 11-20	Rank 21-30	Rank 31-40	Rank 41-50	Rank 51-60	Rank 61-70	Rank 71-80	Rank 81-90	Rank 91-100	Rank 101-110	Rank 111-120	Rank 121-130	Rank 131-140	Rank 141-149
Viet Nam				5	10	18	29	20	8						
Nicaragua			8	21	28	14	10	14							
Saudi Arabia						11	13	23	16	10	11				
Tajikistan						9	6	14	19	18	19	11			
Azerbaijan						10	23	9	24	26	5				
Nepal					5	9	8	15	28	14	13	6			
Morocco							14	20	15	13	15	16			
Romania						14	10	33	16	11	5				
Belize			9	29	16	13	14	6							
Turkmenistan					5	6	14	8	19	25	11	9			
Ghana	11	14	16	10	10	13	9	8	5						
Moldova								8		10	13	30	31		
Namibia						16	16	25	18	16					
Trinidad & To			6	20	11	23	8	13	8	5					
Lebanon					5	15	13	13	5	5	6	20	8	11	
Oman					10	25	18	24	5	10					
Fiji					6	15	13	19	14	20	5				
Congo			9		23	18	13	13	13	9					
Kyrgyzstan					5	6	15	13	30	15	9				
Zimbabwe				21	11	15	10	15	13	6					
Kenya	18	15	11	11	11	11	6	8	6						
South Africa				14	16	20	11	16	11						
Botswana					5		13	18	18	19	16	6			
Syria								16	11	21	13	30	6		
Mongolia			5		13	9	25	18	15	10					
Laos			10	8	9	10	6	10	9	19	11				
Indonesia			9	10	11	23	19	14	5	6					
Côte d'Ivoire		10	13	15	20	8	11	9	6						
Myanmar					6	5	16	24	26	15	5				
China					9	13	9	25	19	13	5				
Uzbekistan								14	16	29	29	11			
Kazakhstan						5	15	16	36	24					
Guyana			6	10	19	15	20	20	5						
Papua New G			6	10	14	9	11	20	20	8					
Bolivia			8	10	23	13	20	11							
Kuwait						9	5	15	28	41					
United Arab E								10	9	36	19	11	11		
Tanzania			11	13	16	9	11	11	8	6	6				
Cameroon			6	10	6	13	23	23	15						
Senegal			5	6	9	16	30	24	6						
Togo			8	6	18	18	18	18	10						
Uganda			8	5	5	6	11	20	21	11	10				
Swaziland							6	16	31	24	15				
Haiti								10	21	23	30	10			
India								11	15	31	25	13			
Malawi			9	13	13	14		11	15	9	6	8			
Eritrea					6	13	16	16	25	18					
Ethiopia					6	8	9	8	9	25	26	5			
Pakistan					23	9	26	18	18						
Bangladesh						9	18	24	48						
Nigeria					6	5	13	15	24	23	6	5			
Benin			10	11	10	14	13	9	11	13					
Central Afr. R								13	14	16	38	13			
Sudan								10	34	46	6				
Zambia						10	10	14	9	21	21	11			
Rwanda			6		11	18	11	18	5	13	6	9			
Burundi					9	8	15	9	18	29	11				
Madagascar					8	13	16	20	21	15					
Mozambique			6												

2. Which countries have the most volatile ranks and why?

We use the term “volatility” as a measure of the difference between a country’s best and worst rank, calculated from the fifth and the 95th percentiles of the rank distribution simulations. For Finland, Costa Rica, New Zealand, Colombia and Panama, we can reasonably state that they have a top 10 performance (probability greater than 70%) and very low volatility in their scores. Interestingly, Panama is ranked 32nd in the EPI – a rank that occurs less than 5% of the times in our simulations. Table 3 presents the 20 countries that are affected most strongly by the methodological choices made during the construction of the EPI. These countries, with a difference in their best and worst rank (5th and 95th percentiles) of at least 80 positions, are ranked between 11th (Iceland) and 131st (Rwanda). A number of these countries, such as Lithuania, Hungary, Denmark, Albania, Ireland, Uruguay, and Bosnia & Herzegovina are ranked among the top 50 in the EPI. The volatility of these countries’ ranks can be attributed mainly to the choice of the weighting combined with the aggregation scheme at the policy level.

Table 3. Most volatile countries in the EPI

Country	EPI Rank	Range of Simulation Ranks	Country	EPI Rank	Range of Simulation Ranks
Iceland	11	[14,95]	El Salvador	65	[31,129]
Lithuania	16	[16,98]	Ghana	86	[12,93]
Hungary	23	[33,129]	Lebanon	89	[62,143]
Denmark	25	[25,131]	Kenya	96	[13,98]
Albania	27	[25,132]	Laos	101	[29,116]
Ireland	34	[24,114]	Côte d'Ivoire	103	[21,103]
Uruguay	36	[31,139]	Tanzania	113	[23,113]
Bosnia & Herzegovina	48	[48,141]	Uganda	117	[55,134]
South Korea	51	[42,125]	Malawi	121	[48,132]
Belgium	57	[42,137]	Benin	127	[51,130]
			Rwanda	131	[45,131]

3. What are the sources of major impact on the variability of the EPI ranking?

We now focus on assessing the impact of each of the five assumptions individually, which amounts to a total of eight different scenarios. We undertake the following comparisons:

Measurement error

- current case without measurement error in the data vs. measurement error in the data;

Winsorisation

- current winsorisation approach vs. no winsorisation;

Target values

- current target values v. no target values;

Weighting

- current weighting vs. FA-derived weights at the indicator level;
- current weighting vs. equal weighting at the indicator level;
- current weighting vs. equal weighting at the subcategory level;
- current weighting vs. equal weighting at the policy level;

Aggregation

- current arithmetic aggregation vs. geometric aggregation at the policy level.

Measurement error

It is reasonable to assume that the raw data are not flawless and that despite efforts to guarantee the most reliable sources for them, errors may still be present. To account for this, we have added a normally distributed random error term to the raw data with a mean zero and a standard deviation equal to the observed one for each indicator. Table 4 presents the countries that are mostly affected by this assumption. Most notably, Luxembourg (rank: 31) would deteriorate its rank by 53 positions. On the other extreme, the Philippines (rank: 61) would improve its rank and be placed in the 10th position. Overall, the introduction of measurement error in the raw data has a median impact of 9 ranks and a 90th percentile impact of 29 positions. In other words, this assumptions leaves 1 out of 2 countries almost unaffected (less than 9 positions change), but 1 out of 10 countries would shift more than 29 positions.

Table 4: Countries most affected by measurement error compared to the original EPI.

	EPI rank	Rank	Difference	Top five countries
Colombia	9	42	-33	Costa Rica
Iceland	11	47	-36	Dominican Rep.
Estonia	19	60	-41	Norway
Luxembourg	31	84	-53	Finland
Dominican Rep.	33	2	31	Canada
Cuba	41	74	-33	
Poland	42	83	-41	Bottom five countries
South Korea	51	18	33	Cambodia
Peru	60	27	33	Mauritania
Philippines	61	10	51	Angola
Iran	67	32	35	Burkina Faso
Honduras	73	38	35	Sierra Leone
Nepal	81	115	-34	
Fiji	94	54	40	Median: 9 ranks
South Africa	97	57	40	90 th percentile: 29 ranks

Winsorization

Winsorization is also expected to have an impact on the rankings, particularly for those countries that present a few extreme values. Table 5 presents the countries that are mostly affected by the choice of not winsorizing, as opposed to the current one. In the best case, South Africa (rank: 97) improves its position by 16, while in the worst case, Botswana (rank: 98) declines by 21 ranks. For 1 out of 2 countries, the impact of this assumption is only 5 positions, while 1 out of 10 countries shift by more than 11 positions, but not more than 21.

Table 5: Countries most affected by not winsorizing skewed distributions compared to the original EPI.

	EPI rank	Rank	Difference	Top five countries
--	-----------------	-------------	-------------------	---------------------------

Hungary	23	39	-16	Sweden
Luxembourg	31	48	-17	Norway
Georgia	37	50	-13	Switzerland
Belarus	43	56	-13	New Zealand
Bosnia & Herzegovina	48	61	-13	Costa Rica
Tajikistan	79	95	-16	
Azerbaijan	80	96	-16	Bottom five countries
Lebanon	89	75	14	Mali
Fiji	94	107	-13	Chad
South Africa	97	81	16	Sierra Leone
Botswana	98	119	-21	Niger
Indonesia	102	87	15	Angola
Côte d'Ivoire	103	91	12	
Uzbekistan	106	125	-19	Median:5
Tanzania	113	99	14	90 th percentile: 11

Targets

Allowing for “extra credit” when exceeding the indicator targets is also expected to have an impact on the results. Table 6 presents the countries that are mostly affected by this assumption. Luxembourg (rank: 31) and Laos (rank: 101) would see the greatest shift in their ranks (a decline of 12 and 15 positions respectively). In the best case, El Salvador (rank: 65) will improve by 9 positions. Overall, for 1 out of 2 countries, the impact of this assumption is only 3 positions, while 1 out of 10 countries shift by more than 7 positions, but not more than 15. The two assumptions on the use of target values and on the winsorization are thus by far the least influential methodological decision in the EPI, a result that we will confirm below.

Table 6: Countries most affected by not capping the indicators at the performance target compared to the original EPI.

	EPI rank	Rank	Difference	Top five countries
Slovakia	17	28	-11	Norway
Hungary	23	33	-10	Sweden
Luxembourg	31	43	-12	Switzerland
Bosnia & Herzegovina	48	57	-9	Costa Rica
Sri Lanka	50	40	10	New Zealand
Jamaica	53	61	-8	
Philippines	61	53	8	Bottom five countries
El Salvador	65	56	9	Mali
Saudi Arabia	78	86	-8	Burkina Faso
Azerbaijan	80	89	-9	Sierra Leone
Trinidad & Tobago	91	83	8	Angola
Lebanon	89	81	8	Niger
Laos	101	116	-15	

Cameroon	114	105	9	Median: 3
Central Afr. Rep.	128	136	-8	90 th percentile: 7

Alternative weighting schemes

Four alternative weighting schemes, all with their implications and advantages, are deemed as the most representative in the literature of composite indicators and worth being tested in our current analysis.

- current weighting vs. FA-derived weights at the indicator level;
- current weighting vs. equal weighting at the indicator level;
- current weighting vs. equal weighting at the subcategory level;
- current weighting vs. equal weighting at the policy level;

Using FA-derived weights at the indicator level significantly affects the country rankings. Half of the countries shift fewer than 16 positions but 15 countries shift more than 47 positions. The following Table 7 shows the countries that experience the biggest shift in their rank due to this assumption.

Table 7: Countries most affected by the FA weights compared to the original EPI.

	EPI rank	Rank	Difference	Top five countries
Lithuania	16	63	-47	Switzerland
Hungary	23	75	-52	Finland
Denmark	25	79	-54	New Zealand
Albania	27	93	-66	Estonia
Georgia	37	87	-50	Austria
Bosnia & Herzegovina	48	99	-51	
South Korea	51	105	-54	Bottom five countries
Egypt	71	23	48	Angola
Saudi Arabia	78	17	61	Yemen
Belize	84	21	63	Bangladesh
Moldova	87	134	-47	Solomon Islands
Trinidad & Tobago	91	40	51	Sierra Leone
Zimbabwe	95	48	47	
Kenya	96	45	51	Median:16
Mongolia	100	33	67	90 th percentile: 47

Equal weighting at the indicator level would increase the weight of the indicators in the Air Pollution (effects on nature) subcategory, the Water (effects on nature), the Biodiversity and Habitat category, and the Productive Natural Resources category. A total of seventeen indicators will increase their weight, as opposed to the current weighting scheme. The remaining eight indicators will reduce their weight, in particular, the DALY indicator and the three indicators related to Climate Change. The countries whose EPI ranks are most affected by this change are

shown in Table 8. The countries that improve their ranks the most are Laos, Kenya, Mongolia and Malawi (by more than 60 positions upwards). On the other hand, Denmark and South Korea decline more than 70 positions. Overall, for 1 out of 2 countries, the impact of this assumption is 15 positions, while 1 out of 10 countries shift by more than 48 positions (up to 72 positions).

Table 8: Countries most affected by using equal weights at the indicator level compared to the original EPI.

	EPI rank	Rank	Difference	Top five countries
Hungary	23	80	-57	Switzerland
Denmark	25	97	-72	Finland
South Korea	51	122	-71	New Zealand
Belgium	57	115	-58	Estonia
Tunisia	59	117	-58	Colombia
Ukraine	75	124	-49	
Belize	84	35	49	Bottom five countries
Moldova	87	139	-52	Yemen
Congo	92	39	53	Angola
Kenya	96	29	67	Iraq
Mongolia	100	33	67	Bangladesh
Laos	101	17	84	Solomon Islands
Côte d'Ivoire	103	49	54	
Malawi	121	55	66	Median: 15
Rwanda	131	77	54	90 th percentile: 48

We next tested the impact of an equal weighting at the subcategory level, while the relative weights for the indicators within each subcategory remain as in the original EPI. This is expected to have a less pronounced impact on the EPI ranks because this assumption assigns greater weight to the six of the ten subcategories and reduces the weight of the other four, particularly reducing the weight of the Climate Change and DALY indicators. As a consequence, the countries whose EPI ranks are most affected by this change are given in Table 9. The countries that improve their ranks the most are Trinidad & Tobago and Laos (improvements of more than 38 positions). On the other hand, Denmark and Taiwan decline more than 50 positions. Overall, for 1 out of 2 countries, the impact of this assumption is 9 positions, while 1 out of 10 countries shift by more than 26 positions (up to 51 positions).

Table 9: Countries most affected by equal weighting at the subcategory level compared to the original EPI.

	EPI rank	Rank	Difference	Top five countries
Denmark	25	76	-51	Switzerland
Argentina	38	65	-27	Finland
Taiwan	40	90	-50	New Zealand
Australia	46	18	28	Sweden
South Korea	51	100	-49	Colombia

Netherlands	54	86	-32	
Belgium	57	101	-44	Bottom five countries
Mauritius	58	29	29	Dem. Rep. Congo
Tunisia	59	92	-33	Niger
Gabon	64	37	27	Bangladesh
Belize	84	49	35	Angola
Trinidad & Tobago	91	50	41	Mauritania
Fiji	94	66	28	
Mongolia	100	72	28	Median: 9
Laos	101	63	38	90 th percentile: 26

We conclude our assessment of the impact of different weighting methods by evaluating the impact of equal weighting at the policy level. The relative weights within the policy categories and within the subcategories remain the same as in the EPI. A weight of 1/6 is thus assigned to each policy category, thus reducing significantly the previously assigned weight of .50 to the environmental health and the weight of 0.25 assigned original to climate change. All policy categories now have a weight of 1/6 = .167. The countries whose EPI ranks are most affected by this change are given in Table 10. The countries with the most notable improvement in their ranks are Laos and Kenya (improvement of more than 78 positions). On the other hand, Belgium and South Korea decline more than 75 positions. Overall, for 1 out of 2 countries, the impact of this assumption is 18 positions, while 1 out of 10 countries shift by more than 486 positions (up to 91 positions).

Table 10: Countries most affected by equal weighting at the policy category level compared to the original EPI.

	EPI rank	Rank	Difference	Top five countries
Denmark	25	77	-52	Switzerland
United States	39	87	-48	Finland
Taiwan	40	101	-61	Sweden
South Korea	51	126	-75	Norway
Netherlands	54	122	-68	New Zealand
Belgium	57	138	-81	
Tunisia	59	111	-52	Bottom five countries
Armenia	62	110	-48	Solomon Islands
Ukraine	75	123	-48	Djibouti
Belize	84	30	54	Yemen
Lebanon	89	137	-48	Iraq
Congo	92	23	69	Kuwait
Kenya	96	18	78	
Mongolia	100	35	65	Median: 18
Laos	101	10	91	90 th percentile: 48

Aggregation scheme at the policy level

Assuming that compensability is allowed among the indicators within each policy category but not desirable across all the policy categories is consistent with the current theories that environmental aspects should be non-compensatory. Table 11 presents those countries for which the most notable shift in the country rank occurs when a non-compensatory aggregation is performed at the policy level, i.e., a geometric mean function instead of an arithmetic mean function. Sri Lanka, Peru, and Egypt improve their ranks by 18 positions or more, while the greatest decline is observed for Uruguay (down more than 51 positions). Overall, for 1 out of 2 countries, the impact of this assumption is merely 5 positions, while 1 out of 10 countries shift by more than 18 positions (up to 51 positions).

Table 11: Countries most affected by geometric aggregation at the policy level compared to the original EPI.

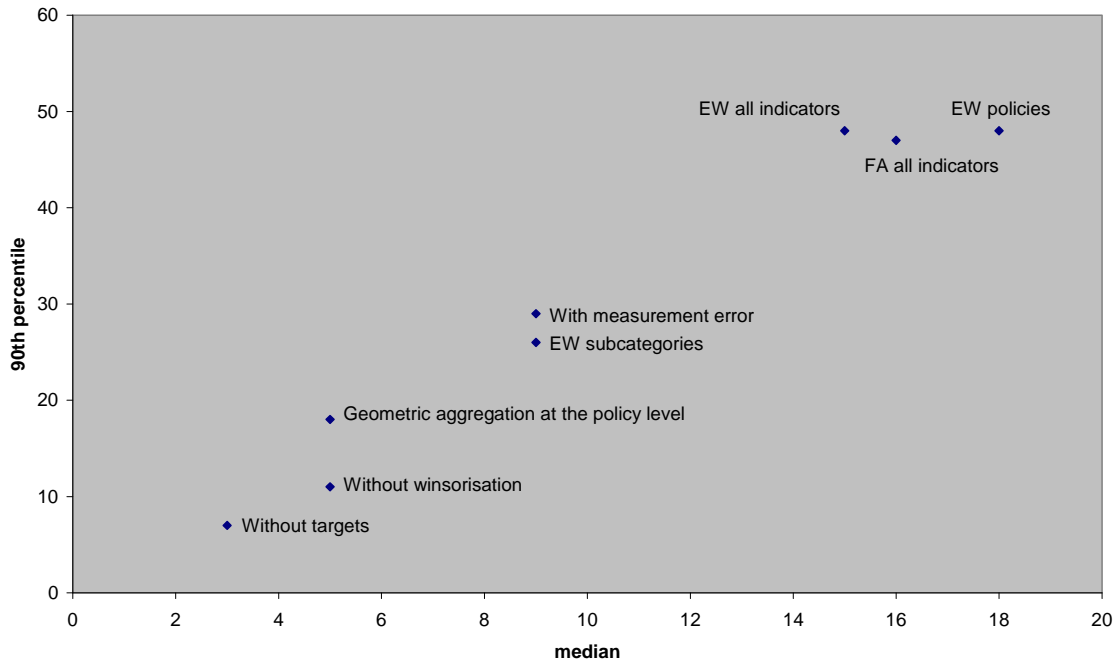
	EPI rank	Rank	Difference	Top five countries
Hungary	23	45	-22	Switzerland
Albania	27	62	-35	Norway
Ireland	34	58	-24	Sweden
Uruguay	36	87	-51	Finland
Greece	44	66	-22	Costa Rica
Bosnia & Herzegovina	48	94	-46	
Sri Lanka	50	31	19	Bottom five countries
Peru	60	42	18	Dem. Rep. Congo
El Salvador	65	83	-18	Mali
Egypt	71	51	20	Sierra Leone
Turkey	72	91	-19	Angola
Ukraine	75	96	-21	Niger
Moldova	87	113	-26	
Lebanon	89	119	-30	Median: 5
Kazakhstan	107	126	-19	90 th percentile: 18

As expected and confirmed in all cases discussed above, middle-of-the-road performers display higher variability than the top and bottom countries.

When only one input factor is changed at a time, the most significant impact to the EPI ranking is attributable to the weighting method. This is especially true when choosing equal weights at the policy level (and original weights within each policy) compared to the original EPI, equally weighting all indicators, or using factor analysis derived weights at the indicators level. In any of these three cases, 1 out of 2 countries shifts less than 15 positions with respect to the original EPI ranking, while 1 out of 10 countries shifts more than 50 positions. The addition of measurement error and the impact of an equal weighting at the subcategories also have significant impact on the EPI ranking (1 out of 2 countries shifts less than 9 positions, but 1 out of 10 countries shift close to 30 positions or more). The least influential input factor is the decision on whether to cap performance at the indicator targets and winsorisation. In fact, 1 out of 2 countries shift less than

five positions in the overall ranking and 1 out of 10 countries shift more than 10 positions, but not more than 21 positions.

Figure 2. Sensitivity analysis: impact of one-at-a-time changes in the five tested assumptions on the EPI ranking.



Note: median versus 90th percentile of the absolute differences in the rank score between a given scenario and the EPI. EW stands for equal weighting.

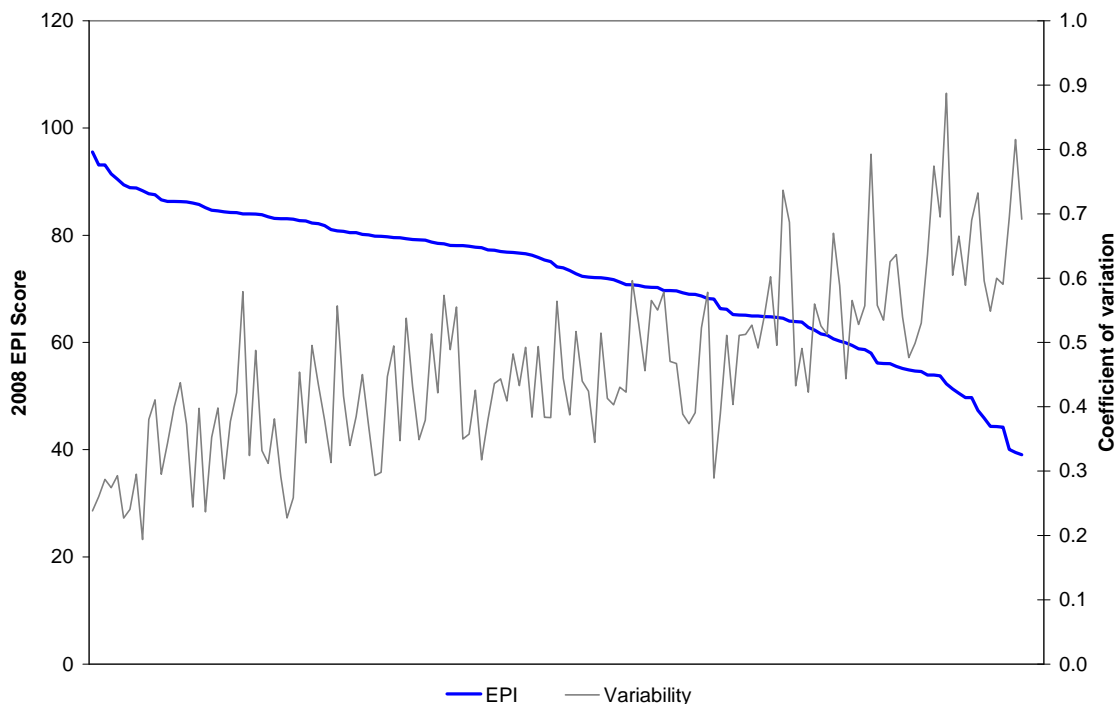
When all sources of uncertainty are allowed to vary simultaneously their combined effect becomes even more important. The use of geometric aggregation combined with equal weighting at the policy level, with/without targets, without winsorization, and without measurement error affects half of the countries by more than 39 positions, of which 1 out of 10 is affected by a median shift of 69 positions. The main graph which we propose as representative of the environmental performance of the countries world-wide, given the current framework, but free of methodological choices (since these choices have already been summarized by the different scenarios) shows the probabilities that a country is ranked in the 1-10 position, or 11-20, etc.

EPI and Variability

Countries that are situated in the top or mid-way in the EPI ranking tend to score uniformly high on the various indicators. In other words, these countries display a relatively low variability, which equals the coefficient of variation across the 25 indicator values for a given country. Figure 3 shows that the variability increases further down the EPI ranking. This scissors pattern is evident, and pronounced. The correlation coefficient between the EPI and the coefficient of variation series is equal to $r = -0.78$, indicating a fairly high degree of reverse association between the EPI scores and the variability in the underlying indicators. For comparison

purposes, in the case of the Trade and Development Index (UNCTAD, 2005) that is based on eleven components and developed for 110 countries, the correlation coefficient between the index scores and the coefficients of variation series was much higher and equal to $r = -0.93$.

Figure 3. The scissor diagram of EPI and variability



An implication of this finding is that while changes in the EPI scores over time could be regarded as a quantitative indication of trends in environmental performance, those with respect to the variability of the ranks could be seen as qualitative changes. Reducing even further the variability in the indicators should be among the objectives of environmental policies and strategies. To be successful, a country must simultaneously invest in multiple goals within a coherent environmental performance strategy while emphasizing reduction of the existing gaps in areas where performance is lagging. By demonstrating significant inter-country differences in the values of the coefficient of variation, the scissors diagram (Figure 3) points to the importance of country-specific approaches to environmental strategies. At the same time, though, it is unlikely that these variations will be reduced without coherent environmental policies and decisionmaking.

Concluding remarks

The methodological approach used to construct the 2008 EPI was studied in this section. The “statistical” filters of index quality show that, although the theoretical framework and the indicators were carefully chosen by experts, the issue of weighting is crucial to obtain a robust performance index. The current weighting scheme results in an EPI that is dominated by very few indicators while having an almost random association with several other underlying

indicators. With respect to the five input factors tested in the sensitivity and uncertainty analysis, the country rankings are relatively reliable for approximately half of the countries, while any conclusion on the ranking for the other half of the countries should be made with great caution. An equal weighting approach at the indicator level or at the policy level, as opposed to the current weighting scheme, greatly influences the ranks. Thus, the choice of the weights must be evaluated according to its analytical rationale, policy relevance, and implied value judgments. The real value of the EPI lies not in the overall ranking of the countries, but rather in the solid framework and construction of the indicators. It is from this perspective that further revisions of the Index should be considered if the goal is to arrive at a single number that provides meaningful input to policy-making.