Joint Research Centre Statistical Audit of the 2014 Global Innovation Index

Michaela Saisana and **Andrea Saltelli**, European Commission Joint Research Centre (Ispra, Italy)

Modelling the concepts underlying innovation at the national scale around the globe, as attempted in the Global Innovation Index (GII), raises both conceptual and practical challenges. The conceptual challenges are discussed in the main text of Chapter 1 of the GII 2014 report. In this annex, the focus is on the practical challenges related to the data quality and the methodological choices made by grouping these data into 21 sub-pillars, 7 pillars, 2 sub-indices, and an overall index.

We consider statistical soundness to be a necessary but not a sufficient condition for a sound GII. Given that the statistical analysis of an index is based primarily, but not solely, on correlations, correspondence of the GII with real-world phenomena needs to be critically addressed, whereas 'correlations need not necessarily represent the real influence of the individual indicators on the phenomenon being measured'.[1] The point we are making here is that the validity of the GII relies on the interplay between statistical and conceptual soundness. To this end, the development of the GII has followed an iterative process that went back and forth between a theoretical understanding of innovation on the one hand and empirical observations of the data underlying the variables on the other.

The Econometrics and Applied Statistics Unit at the European Commission Joint Research Centre (JRC) in Ispra (Italy) was invited for a fourth consecutive year to audit the GII following some adjustments that were made to the list of indicators included in the GII framework (see Chapter 1 for more details).

The JRC assessment of the 2014 GII focused on two main issues: the statistical soundness of its multi-level structure and the impact of key modelling assumptions on its scores and ranks.[2] These are necessary steps to ensure the transparency and reliability of the GII, to enable the public to derive more accurate and meaningful conclusions, and to support policy makers with choices on priority setting and policy formulation.

As in past GII reports, the JRC analysis complements the country rankings with confidence intervals for the GII, the Innovation Input Sub-Index, and the Innovation Output Sub-Index in order to better appreciate the robustness of these ranks to the computation methodology. In addition, the JRC analysis includes an assessment of potential redundancy of information in the GII and a measure of distance to the efficient frontier of innovation by using data envelopment analysis.

Conceptual and statistical coherence in the GII framework

An earlier version of the GII model was assessed by the JRC in April 2014. Fine-tuning suggestions were taken into account in the final computation of the rankings in an iterative process with the JRC aimed at establishing the foundation for a balanced index. The entire process followed four steps (see Figure 1):

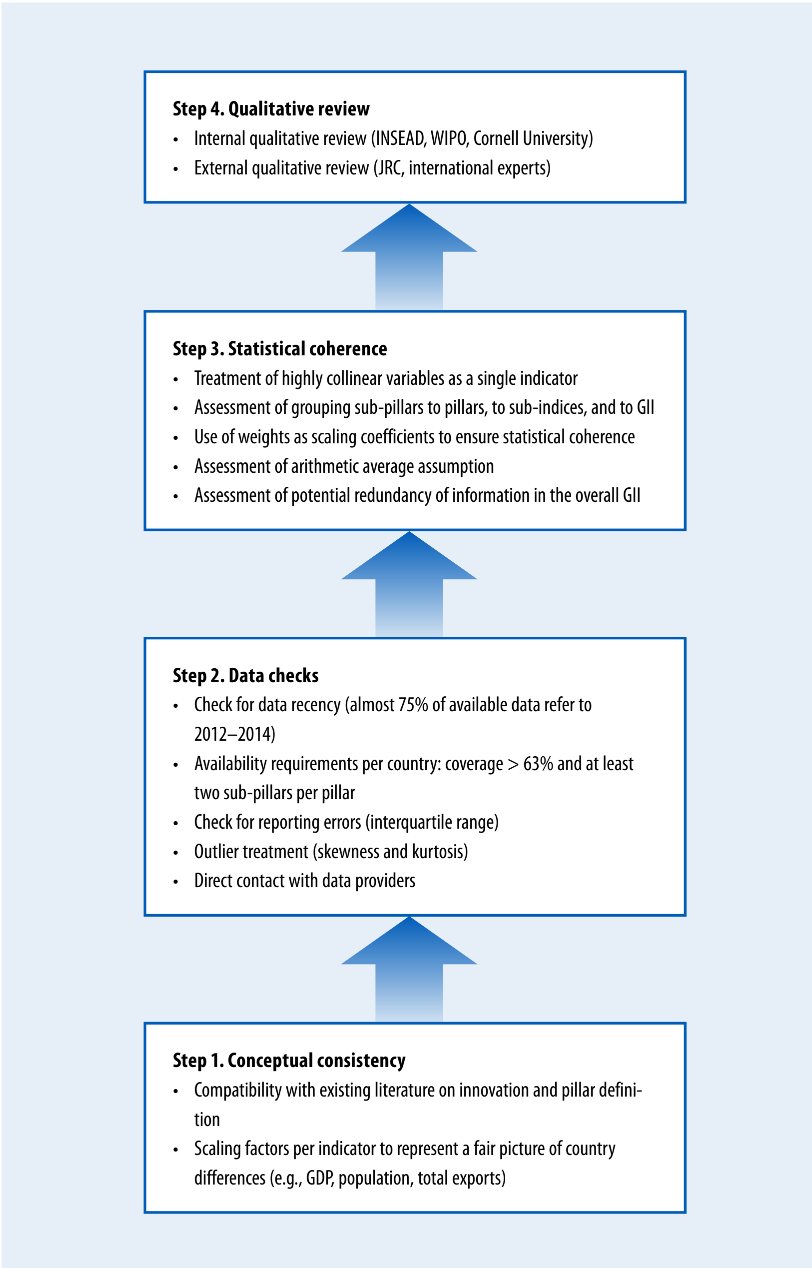
Step 1: Conceptual consistency

Eighty-one indicators were selected for their relevance to a specific innovation pillar on the basis of the literature review, expert opinion, country coverage, and timeliness. To represent a fair picture of country differences, indicators were scaled either at the source or by the GII team as appropriate and where needed.

Step 2: Data checks

The most recently released data were used for each country with a cut-off year of 2004. Almost 75% of the available data refer to 2012 or a more recent year. Countries were included if data availability was at least 63% (i.e., 51 out of 81 variables) and at least two of the three sub-pillars in each pillar could be computed. Potentially problematic indicators that could bias the overall results were identified as those having absolute skewness greater than 2 and kurtosis greater than 3.5;[3] these were treated either by winsorization or by taking the natural logarithm (in cases with more than five outliers). These criteria were decided jointly with the JRC in 2011 (see Appendix IV Technical Notes for details).

Figure 1: Conceptual and statistical coherence in the GII 2014 framework



Source: Saisana and Saltelli, European Commission Joint Research Centre, 2014.

Step 3: Statistical coherence

Weights as scaling coefficients

Weights of 0.5 or 1.0 were decided jointly with the JRC and the GII team in 2012 to be scaling coefficients and not importance coefficients, with the aim of arriving at sub-pillar and pillar scores that were balanced in their underlying components (i.e., that indicators and sub-pillars can explain a similar amount of variance in their respective sub-pillars/pillars). Paruolo et al. (2013) show that, in weighted arithmetic averages, the ratio of two nominal weights gives the rate of substitutability between the two indicators, and hence can be used to reveal the relative importance of individual indicators. This importance can then be compared with ex-post measures of variables' importance, such as the non-linear Pearson correlation ratio. As a result of this analysis, 36 out of 81 indicators and two sub-pillars—7.2 Creative goods and services and 7.3 Online creativity—were assigned half weights, while all other indicators and sub-pillars were assigned a weight 1.0.

Principal components analysis and reliability item analysis

Principal component analysis (PCA) was used to assess to what extent the conceptual framework is confirmed by statistical approaches. PCA results confirm the presence of a single latent dimension in each of the seven pillars (one component with an eigenvalue greater than 1.0) that captures between 57% (pillar 4: Market sophistication) up to 82% (pillar 1: Institutions) of the total variance in the three underlying sub-pillars. These results reveal that the adjustments made to the 2014 GII framework did not affect the solid statistical coherence properties of the previous version. Furthermore, results confirm the expectation that the sub-pillars are more correlated to their own pillar than to any other (see Table 1). It is interesting to note that sub-pillar 7.3 Online creativity has the same degree of correlation (0.86) with its own pillar Creative outputs (pillar 7) that it has with Human capital and research (pillar 2) and Infrastructure (pillar 3), which evidences an association between human capital and infrastructure on one hand and online content, such as Wikipedia monthly edits and video uploads on YouTube, on the other.

Table 1: Statistical coherence in the GII: Correlations between sub-pillars and pillars

	Sub-pillar	Institutions	Human capital and research	Infrastructure	Market sophistication	Business sophistication	Knowledge and technology outputs	Creative outputs
INPUT	Political environment	0.91	0.71	0.76	0.61	0.74	0.63	0.77
	Regulatory environment	0.93	0.65	0.69	0.61	0.67	0.56	0.65
	Business environment	0.88	0.75	0.78	0.71	0.63	0.60	0.66
	Education	0.62	0.77	0.63	0.43	0.50	0.59	0.54
	Tertiary education	0.57	0.81	0.68	0.49	0.56	0.47	0.54
	Research and development (R&D)	0.72	0.89	0.82	0.69	0.69	0.82	0.71
	Information and communication technologies (ICTs)	0.78	0.88	0.93	0.65	0.71	0.72	0.77
	General infrastructure	0.46	0.50	0.68	0.39	0.44	0.38	0.46
	Ecological sustainability	0.72	0.69	0.82	0.53	0.58	0.61	0.71
	Credit	0.68	0.68	0.64	0.86	0.56	0.62	0.60
	Investment	0.41	0.40	0.40	0.81	0.43	0.38	0.28
	Trade and competition	0.51	0.42	0.45	0.56	0.42	0.40	0.45
	Knowledge workers	0.74	0.79	0.75	0.62	0.87	0.72	0.70
	Innovation linkages	0.51	0.37	0.42	0.38	0.72	0.33	0.51
	Knowledge absorption	0.45	0.41	0.43	0.39	0.72	0.43	0.44
OUTPUT	Knowledge creation	0.61	0.78	0.67	0.60	0.61	0.85	0.62
	Knowledge impact	0.41	0.52	0.51	0.39	0.34	0.75	0.45
	Knowledge diffusion	0.49	0.46	0.45	0.44	0.63	0.71	0.51
	Intangible assets	0.44	0.34	0.42	0.29	0.49	0.32	0.75
	Creative goods and services	0.64	0.62	0.69	0.49	0.60	0.60	0.79
	Online creativity	0.81	0.86	0.86	0.63	0.73	0.78	0.86

Source: Saisana and Saltelli, European Commission Joint Research Centre, 2014.

The five input pillars share a single statistical dimension that summarizes 78% of the total variance, and the five loadings (correlation coefficients) of these pillars are all very similar. This similarity suggests that the five pillars make roughly equal contributions to the variation of the Innovation Input Sub-Index scores, as envisaged by the developing team. The reliability of the Input Sub-Index, measured by the Cronbach alpha value, is very high, at 0.93, which is well above the 0.70 threshold for a reliable aggregate.[4]

The two output pillars—Knowledge and technology outputs and Creative outputs—are sufficiently correlated with each other (0.67); they are also both strongly correlated with the Innovation Output Sub-Index (0.91). This result suggests that the Output Sub-Index is also well balanced in its two pillars.

Finally, building the GII as the simple average of the Input and Output Sub-Indices is also statistically justifiable because the Pearson correlation coefficient of either sub-index with the overall GII is 0.97; the two sub-indices have a correlation of 0.87. Thus far, results show that the grouping of sub-pillars into pillars, sub-indices, and the overall GII 2014 is statistically coherent, and that the GII has a balanced structure at each aggregation level.

Assessing potential redundancy of information in the GII

As already discussed, the Input and Output Sub-Indices correlate strongly with each other and with the overall GII. Furthermore, the five pillars in the Input Sub-Index have a very high statistical reliability. These results—the strong correlation between Input and Output Sub-Indices and the statistical reliability of the five Input pillars—may be interpreted by some as a sign of redundancy of information in the GII. Yet this is not the case here. In fact, for more than 51.7% (up to 74.1%) of the 143 economies included in the 2014 GII, the GII ranking and any of the seven pillar rankings differ by 10 positions or more (see Table 2). This is a desired outcome because it demonstrates the added value of the GII ranking, which helps to highlight other components of innovation that do

Table 2: Distribution of differences between pillar and GII rankings

	Innovation Input Sub-Index					Innovation Output Sub-Index	
Rank differences (positions)	Institutions (%)	Human capital and research (%)	Infrastructure (%)	Market sophistication (%)	Business sophistication (%)	Knowledge and technology outputs (%)	Creative outputs (%)
More than 29	18.2	14.0	12.6	32.9	23.8	22.4	8.4
20–29	10.5	18.2	11.9	18.2	15.4	10.5	12.6
10–19	24.5	25.2	30.1	23.1	22.4	21.7	30.8
10 or more*	**53.1**	**57.3**	**54.5**	**74.1**	**61.5**	**54.5**	**51.7**
5–9	21.0	18.2	21.0	16.1	19.6	23.8	24.5
Less than 5	22.4	22.4	21.7	9.1	16.1	17.5	23.1
Same rank	0.0	0.0	2.1	2.1	2.8	2.8	2.8
Total†	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**

Source: Saisana and Saltelli, European Commission Joint Research Centre, 2014.
* This column is the sum of the prior three rows.
† This column is the sum of all white rows.

not emerge directly by looking into the seven pillars separately.

Step 4: Qualitative review

Finally, the GII results—including overall country classifications and relative performances in terms of the Innovation Input or Output Sub-Indices—were evaluated to verify that the overall results were, to a great extent, consistent with current evidence, existing research, and prevailing theory. Notwithstanding these statistical tests and the positive outcomes on the statistical coherence of the GII structure, it is important to note that the GII model is and has to remain open for future improvements as better data, more comprehensive surveys and assessments, and new relevant research studies become available.

Impact of modelling assumptions on the GII results

Every economy score on the GII and its two sub-indices depends on modelling choices: the seven-pillar structure, the indicators selected, the imputation or not of missing data, the normalization, the weights, and the aggregation method, among other elements. These choices are based on expert opinion (e.g., selection of indicators), or common practice (e.g., min-max normalization in the [0, 100] range), driven by statistical analysis (e.g., treatment of outliers) or simplicity (e.g., no imputation of missing data). The robustness analysis performed by the JRC aimed at assessing the simultaneous and joint impact of these modelling choices on the rankings. It thus complements the GII 2014 ranks with error estimates stemming from the unavoidable uncertainty in the choices made.

The robustness assessment of the GII was based on the combination of a Monte Carlo experiment and a multi-modelling approach, following good practices suggested in the composite indicators literature.[5] We focused on three key issues: pillar weights, missing data, and the aggregation formula. The data are assumed to be error-free because potential outliers and eventual errors and typos were corrected during the computation phase (see Step 2 in Figure 1).

The Monte Carlo simulation related to the issue of weighting and comprised 1,000 runs, each corresponding to a different set of weights for each of the seven pillars, randomly sampled from uniform continuous distributions centred in the reference values. The choice of the range for the weights' variation was driven by two different needs: to ensure a wide enough interval to have meaningful robustness checks and to respect the rationale of the GII that places the Input Sub-Index and the Output Sub-Index on equal footings. Given these considerations, limit values of uncertainty intervals for the pillar weights are: 10%–30% for the five Input pillars and 40%–60% for the two Output pillars (see Table 3).

The GII developing team, for transparency and replicability, has always opted not to estimate missing data. The 'no imputation' choice, which is common in similar contexts, might encourage economies not to report low data values.[6] To overcome this limitation, the JRC estimated missing data using the Expectation Maximization (EM) algorithm.[7]

Regarding the aggregation formula, decision-theory practitioners have challenged the use of simple arithmetic averages because of their

Table 3: Uncertainty parameters: Missing values, aggregation, and weights

		Reference	Alternative
I. Uncertainty in the treatment of missing values		**No estimation of missing data**	**Expectation Maximization (EM)**
II. Uncertainty in the aggregation formula at the pillar level		**Arithmetic average**	**Geometric average**
III. Uncertainty intervals for the GII weights			
GII Sub-Index	Pillar	Reference value for the weight	Distribution assigned for robustness analysis
Innovation Input	Institutions	0.2	U[0.1, 0.3]
	Human capital and research	0.2	U[0.1, 0.3]
	Infrastructure	0.2	U[0.1, 0.3]
	Market sophistication	0.2	U[0.1, 0.3]
	Business sophistication	0.2	U[0.1, 0.3]
Innovation Output	Knowledge and technology outputs	0.5	U[0.4, 0.6]
	Creative outputs	0.5	U[0.4, 0.6]

Source: Saisana and Saltelli, European Commission Joint Research Centre, 2014.

fully compensatory nature, in which a comparative high advantage on a few indicators can compensate for a comparative disadvantage on many indicators.[8] We relaxed this strong perfect substitutability assumption inherent in the arithmetic average and we considered instead the geometric average, which is a partially compensatory approach that rewards economies with balanced profiles and motivates economies with unbalanced profiles to improve in the GII pillars in which they perform poorly, and not just in *any* GII pillar.[9]

Four models were tested based on the combination of no imputation versus EM imputation, and arithmetic versus geometric average, combined with 1,000 simulations per model (random weights versus fixed weights), for a total of 4,000 simulations for the GII and each of the two sub-indices (see Table 3 for a summary of the uncertainties considered in the GII 2014).

Uncertainty analysis results

The main results of the robustness analysis are shown in Figure 2 with median ranks and 90% confidence intervals computed across the 4,000 Monte Carlo simulations for the GII and the two sub-indices. Countries are ordered from best to worst according to their reference rank (black line), the dot being the median rank.

All published GII 2014 ranks lay within the simulated 90% confidence intervals, and for most economies these intervals are narrow enough for meaningful inferences to be drawn: there are fewer than 10 positions for 81 of the 143 economies. However, it is also true that some economy ranks vary significantly with changes in weights and aggregation function and, where applicable, they also vary because of the estimation of missing data. Indeed, 21 economies have 90% confidence interval widths between 20 and 29. Confidence interval widths for 6 of them lie between 30 and 39 (Bangladesh, Fiji, the Islamic Republic of Iran, Togo, Uganda, and the Bolivarian Republic of Venezuela), and for 2 countries the widths are over 40 (Bhutan, Tajikistan). For these countries, the GII ranks should be interpreted cautiously. Some caution is also warranted in the Input Sub-Index for 32 economies that have 90% confidence interval widths over 20 (up to 37 for Dominican Republic). The Output Sub-Index is more sensitive to the methodological choices: 40 economies have 90% confidence interval widths over 20 (up to 67 for Bhutan). This sensitivity is mostly the consequence of the estimation of missing data and the fact that there are only two pillars (with 0.68 correlation); hence changes to the imputation method, weights, or aggregation formula have a more notable impact on the country ranks.

Although some economy ranks, either in the GII 2014 or its two sub-indices, appear to be sensitive to the methodological choices, the published rankings for the vast majority can be considered representative of the plurality of scenarios we have simulated herein. Taking the median rank as our yardstick for an economy's average rank in the realm of the GII's unavoidable methodological uncertainties, we find that 75% of the economies shift fewer than five positions with respect to the median rank in the GII (four and seven positions in the Input and Output Sub-Index, respectively).

For full transparency and information, Table 4 reports the GII 2014

Table 4: GII 2014 and Input/Output Sub-Indices: Ranks and 90% confidence intervals

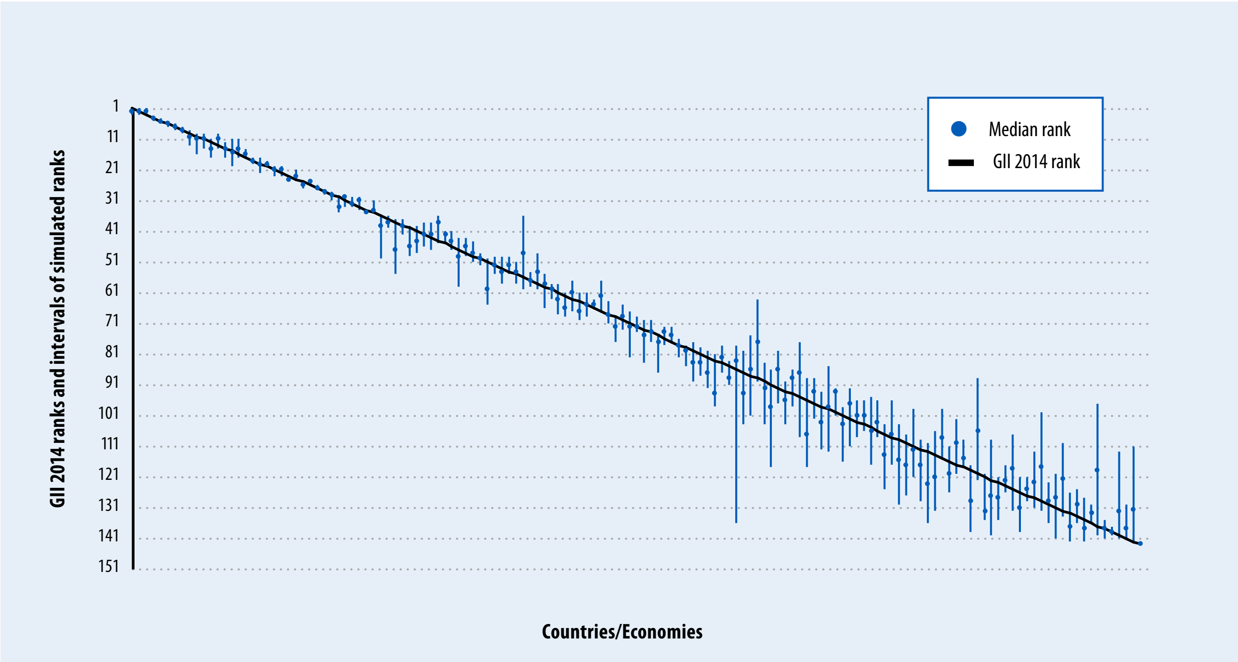
	GII 2014		Input Sub-Index		Output Sub-Index	
Country/Economy	Rank	Interval	Rank	Interval	Rank	Interval
Switzerland	1	[1, 3]	7	[6, 9]	1	[1, 3]
United Kingdom	2	[1, 3]	3	[2, 5]	4	[1, 4]
Sweden	3	[1, 3]	6	[3, 6]	3	[1, 3]
Finland	4	[4, 5]	5	[2, 8]	6	[5, 6]
Netherlands	5	[4, 6]	11	[10, 14]	2	[2, 4]
United States of America	6	[5, 7]	4	[3, 6]	7	[7, 11]
Singapore	7	[6, 8]	1	[1, 1]	25	[21, 26]
Denmark	8	[7, 9]	9	[7, 10]	12	[8, 13]
Luxembourg	9	[8, 13]	21	[18, 23]	5	[5, 6]
Hong Kong (China)	10	[9, 16]	2	[2, 6]	24	[20, 28]
Ireland	11	[9, 14]	12	[10, 17]	11	[9, 13]
Canada	12	[11, 17]	8	[6, 9]	20	[18, 26]
Germany	13	[9, 14]	19	[17, 19]	8	[7, 9]
Norway	14	[12, 17]	14	[11, 18]	14	[12, 17]
Israel	15	[11, 20]	17	[11, 21]	13	[11, 16]
Korea, Rep.	16	[11, 17]	16	[11, 17]	15	[11, 15]
Australia	17	[14, 17]	10	[10, 12]	22	[19, 25]
New Zealand	18	[17, 19]	13	[12, 19]	18	[17, 22]
Iceland	19	[17, 22]	24	[23, 26]	9	[7, 15]
Austria	20	[18, 20]	18	[15, 20]	21	[19, 23]
Japan	21	[20, 23]	15	[13, 16]	27	[26, 31]
France	22	[20, 23]	20	[19, 21]	26	[20, 25]
Belgium	23	[23, 25]	22	[21, 23]	23	[21, 25]
Estonia	24	[21, 24]	23	[21, 25]	19	[15, 20]
Malta	25	[25, 27]	33	[30, 35]	10	[9, 13]
Czech Republic	26	[24, 26]	27	[26, 29]	17	[15, 18]
Spain	27	[26, 27]	26	[23, 26]	28	[27, 29]
Slovenia	28	[28, 29]	28	[27, 29]	31	[29, 31]
China	29	[28, 31]	45	[34, 53]	16	[14, 23]
Cyprus	30	[29, 35]	31	[29, 36]	34	[33, 34]
Italy	31	[30, 33]	32	[30, 33]	33	[32, 34]
Portugal	32	[30, 33]	29	[28, 32]	36	[35, 36]
Malaysia	33	[30, 34]	30	[26, 31]	35	[35, 38]
Latvia	34	[34, 35]	35	[35, 39]	32	[32, 33]
Hungary	35	[31, 35]	41	[37, 45]	29	[28, 31]
United Arab Emirates	36	[36, 50]	25	[23, 31]	68	[57, 95]
Slovakia	37	[36, 40]	43	[41, 48]	38	[37, 40]
Saudi Arabia	38	[37, 55]	39	[35, 45]	41	[41, 69]
Lithuania	39	[37, 42]	36	[34, 39]	52	[46, 53]
Mauritius	40	[39, 49]	42	[39, 58]	43	[43, 48]
Barbados	41	[39, 48]	38	[34, 48]	53	[48, 56]
Croatia	42	[38, 46]	50	[45, 53]	40	[39, 42]
Moldova, Rep.	43	[38, 47]	80	[74, 83]	30	[27, 30]
Bulgaria	44	[36, 45]	54	[48, 55]	37	[35, 38]
Poland	45	[41, 45]	40	[39, 43]	48	[45, 50]
Chile	46	[41, 47]	37	[33, 39]	54	[53, 60]
Qatar	47	[43, 59]	34	[32, 36]	69	[65, 83]
Thailand	48	[43, 49]	52	[42, 55]	49	[46, 50]
Russian Federation	49	[44, 51]	56	[45, 60]	45	[42, 49]
Greece	50	[48, 52]	44	[42, 51]	58	[56, 62]
Seychelles	51	[50, 65]	53	[47, 78]	56	[52, 63]
Panama	52	[49, 55]	64	[60, 72]	42	[39, 51]
South Africa	53	[49, 58]	47	[39, 56]	63	[60, 70]
Turkey	54	[49, 55]	78	[66, 83]	39	[37, 40]
Romania	55	[51, 58]	65	[59, 70]	44	[41, 52]
Mongolia	56	[36, 60]	51	[40, 54]	67	[39, 68]
Costa Rica	57	[54, 59]	66	[58, 72]	51	[50, 53]
Belarus	58	[48, 60]	70	[57, 80]	50	[41, 54]
Montenegro	59	[55, 66]	46	[44, 55]	74	[61, 79]
TFYR of Macedonia	60	[58, 63]	57	[54, 72]	66	[58, 68]
Brazil	61	[58, 68]	63	[54, 69]	64	[61, 71]
Bahrain	62	[61, 69]	48	[44, 53]	80	[75, 83]
Ukraine	63	[57, 67]	88	[72, 92]	46	[43, 50]
Jordan	64	[61, 70]	72	[66, 81]	57	[56, 65]
Armenia	65	[61, 69]	81	[71, 87]	55	[53, 56]
Mexico	66	[63, 66]	62	[55, 63]	70	[67, 71]
Serbia	67	[57, 67]	75	[64, 82]	59	[52, 61]
Colombia	68	[64, 71]	58	[49, 59]	77	[74, 78]
Kuwait	69	[69, 77]	79	[72, 84]	62	[59, 76]
Argentina	70	[65, 73]	83	[62, 89]	61	[61, 67]
Viet Nam	71	[67, 82]	100	[89, 103]	47	[44, 58]
Uruguay	72	[69, 74]	73	[68, 82]	72	[68, 72]

Table 4: GII 2014 and Input/Output Sub-Indices: Ranks and 90% confidence intervals *(continued)*

	GII 2014		Input Sub-Index		Output Sub-Index	
Country/Economy	Rank	Interval	Rank	Interval	Rank	Interval
Peru	73	[70, 84]	60	[56, 69]	85	[81, 104]
Georgia	74	[70, 77]	68	[61, 82]	75	[72, 78]
Oman	75	[75, 87]	59	[53, 65]	96	[93, 117]
India	76	[72, 78]	93	[84, 95]	65	[61, 69]
Lebanon	77	[72, 77]	61	[55, 71]	95	[79, 95]
Tunisia	78	[76, 82]	77	[60, 87]	87	[84, 91]
Kazakhstan	79	[78, 85]	69	[59, 72]	101	[97, 102]
Guyana	80	[77, 90]	92	[83, 111]	76	[63, 86]
Bosnia and Herzegovina	81	[79, 88]	82	[72, 89]	92	[84, 95]
Jamaica	82	[80, 92]	84	[75, 89]	91	[90, 100]
Dominican Republic	83	[81, 98]	101	[91, 128]	71	[69, 108]
Morocco	84	[78, 87]	89	[76, 98]	86	[78, 86]
Kenya	85	[83, 91]	103	[89, 110]	73	[69, 79]
Bhutan	86	[78, 136]	76	[72, 89]	102	[73, 140]
Indonesia	87	[80, 104]	117	[105, 124]	60	[59, 86]
Brunei Darussalam	88	[75, 101]	55	[50, 72]	124	[100, 128]
Paraguay	89	[63, 90]	99	[90, 101]	79	[45, 94]
Trinidad and Tobago	90	[84, 104]	86	[80, 91]	98	[97, 124]
Uganda	91	[86, 118]	98	[93, 109]	90	[85, 125]
Botswana	92	[80, 97]	67	[59, 77]	116	[101, 118]
Guatemala	93	[90, 104]	94	[93, 109]	97	[95, 110]
Albania	94	[86, 98]	71	[65, 84]	117	[91, 117]
Fiji	95	[77, 108]	49	[41, 74]	136	[92, 137]
Ghana	96	[89, 118]	106	[104, 116]	82	[75, 121]
Cabo Verde	97	[89, 102]	85	[78, 91]	114	[90, 116]
Senegal	98	[93, 112]	116	[108, 118]	78	[75, 113]
Egypt	99	[85, 113]	104	[90, 115]	89	[83, 115]
Philippines	100	[92, 101]	110	[102, 114]	84	[79, 85]
Azerbaijan	101	[98, 116]	91	[88, 108]	109	[108, 123]
Rwanda	102	[92, 111]	74	[66, 101]	128	[94, 128]
El Salvador	103	[96, 108]	97	[91, 102]	110	[108, 118]
Gambia	104	[96, 106]	111	[107, 125]	93	[74, 100]
Sri Lanka	105	[95, 117]	125	[113, 136]	81	[77, 87]
Cambodia	106	[96, 108]	113	[100, 122]	99	[95, 102]
Mozambique	107	[104, 125]	96	[88, 100]	115	[111, 138]
Namibia	108	[96, 117]	95	[88, 107]	119	[115, 123]
Burkina Faso	109	[104, 130]	112	[107, 126]	104	[102, 130]
Nigeria	110	[107, 127]	133	[131, 139]	83	[80, 103]
Bolivia, Plurinational St.	111	[99, 121]	115	[101, 127]	106	[104, 115]
Kyrgyzstan	112	[109, 129]	90	[82, 99]	131	[128, 140]
Malawi	113	[110, 136]	109	[106, 127]	108	[107, 135]
Cameroon	114	[106, 132]	127	[119, 129]	100	[98, 132]
Ecuador	115	[99, 115]	105	[96, 111]	113	[110, 117]
Côte d'Ivoire	116	[111, 126]	135	[132, 137]	88	[82, 108]
Lesotho	117	[102, 120]	87	[80, 97]	137	[124, 137]
Honduras	118	[109, 118]	102	[96, 104]	126	[123, 127]
Mali	119	[117, 139]	132	[130, 141]	103	[102, 130]
Iran, Islamic Rep.	120	[89, 122]	107	[91, 121]	125	[78, 124]
Zambia	121	[120, 135]	131	[123, 142]	105	[103, 131]
Venezuela, Bolivarian Rep.	122	[109, 140]	137	[120, 142]	94	[92, 122]
Tanzania, United Rep.	123	[121, 135]	120	[111, 126]	122	[120, 137]
Madagascar	124	[117, 126]	123	[117, 133]	121	[110, 122]
Nicaragua	125	[107, 132]	108	[89, 123]	130	[128, 133]
Ethiopia	126	[121, 139]	128	[124, 139]	118	[114, 134]
Swaziland	127	[121, 129]	119	[109, 133]	127	[117, 128]
Uzbekistan	128	[113, 131]	124	[111, 133]	123	[106, 129]
Bangladesh	129	[100, 132]	130	[113, 135]	120	[88, 121]
Zimbabwe	130	[123, 136]	136	[131, 142]	111	[99, 112]
Niger	131	[120, 141]	118	[112, 132]	134	[119, 141]
Benin	132	[110, 134]	129	[120, 132]	129	[89, 129]
Algeria	133	[126, 142]	122	[107, 127]	132	[130, 142]
Pakistan	134	[125, 136]	139	[138, 142]	107	[96, 107]
Angola	135	[128, 142]	138	[136, 141]	112	[108, 137]
Nepal	136	[130, 136]	121	[112, 134]	135	[125, 137]
Tajikistan	137	[97, 140]	114	[104, 129]	140	[88, 141]
Burundi	138	[135, 141]	126	[119, 129]	141	[138, 141]
Guinea	139	[137, 140]	140	[117, 141]	138	[135, 139]
Myanmar	140	[113, 141]	143	[117, 143]	133	[93, 134]
Yemen	141	[130, 141]	141	[119, 141]	139	[130, 140]
Togo	142	[111, 142]	134	[123, 137]	142	[88, 142]
Sudan	143	[143, 143]	142	[142, 143]	143	[143, 143]

Source: Saisana and Saltelli, European Commission Joint Research Centre, 2014.

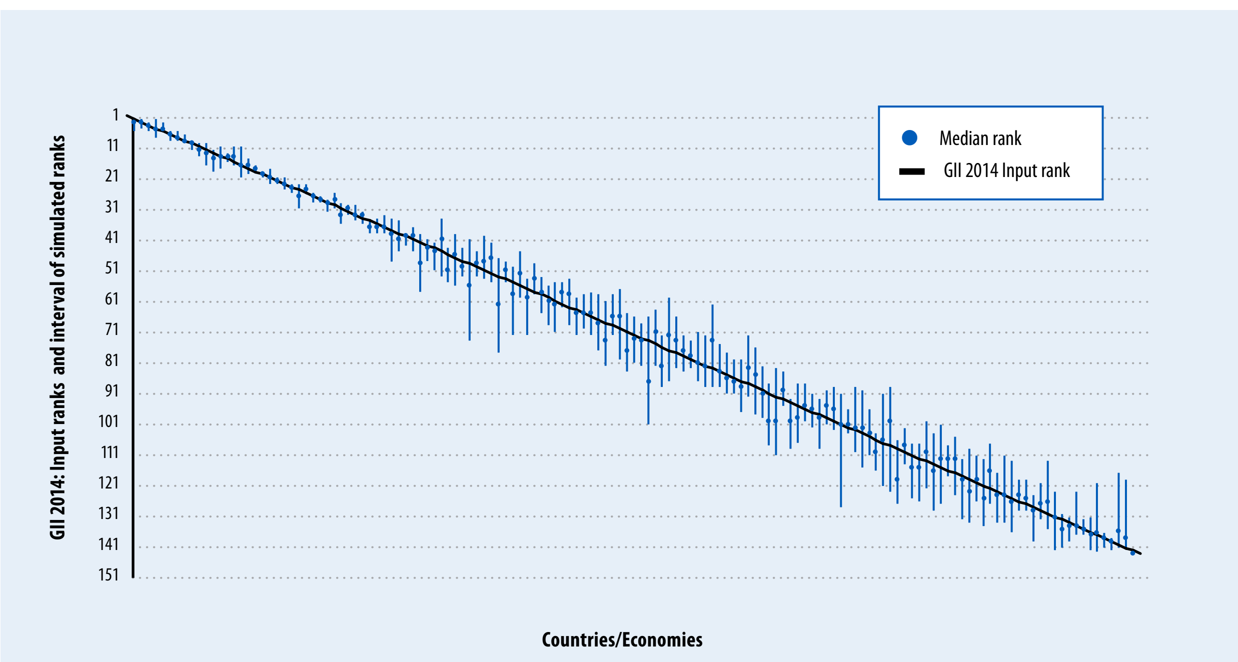
Figure 2a: Robustness analysis (GII rank vs. median rank, 90% confidence intervals)



Source: Saisana and Saltelli, European Commission Joint Research Centre, 2014.

Notes: The Spearman rank correlation between the median rank and the GII 2014 rank is 0.993. Median ranks and intervals are calculated for over 4,000 simulated scenarios combining random weights, imputed versus missing values, and geometric versus arithmetic averages at the pillar level.

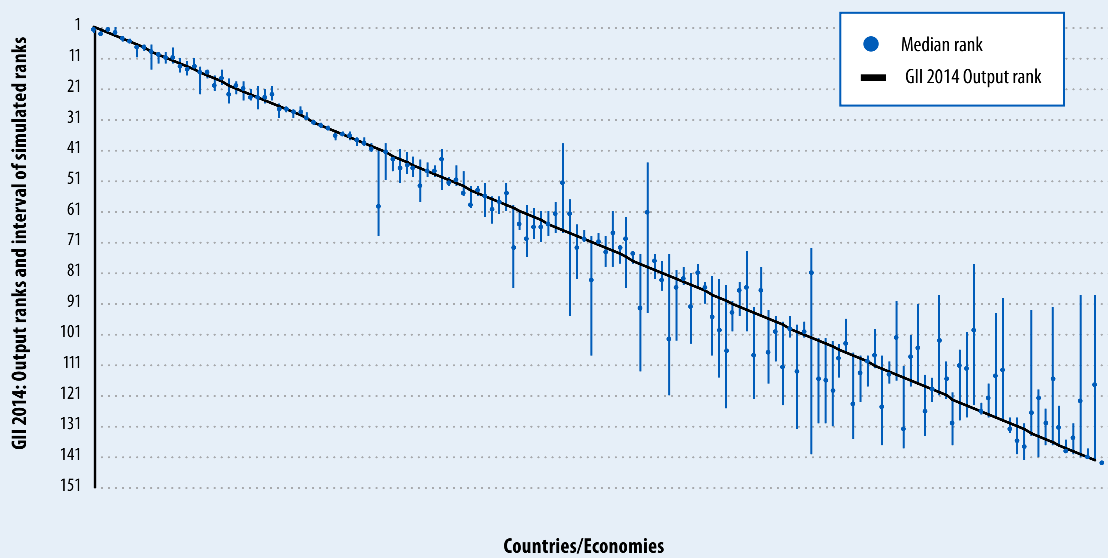
Figure 2b: Robustness analysis (Input rank vs. median rank, 90% confidence intervals)



Source: Saisana and Saltelli, European Commission Joint Research Centre, 2014.

Notes: The Spearman rank correlation between the median rank and the Innovation Input 2014 rank is 0.997. Median ranks and intervals are calculated for over 4,000 simulated scenarios combining random weights, imputed versus missing values, and geometric versus arithmetic averages at the pillar level.

Figure 2c: Robustness analysis (Output rank vs. median rank, 90% confidence intervals)



Source: Saisana and Saltelli, European Commission Joint Research Centre, 2014.

Note: The Spearman rank correlation between the median rank and the Innovation Output 2014 rank is 0.981. Median ranks and intervals are calculated for over 4,000 simulated scenarios combining random weights, imputation versus no imputation of missing values, and geometric versus arithmetic averages at the pillar level.

Index and Sub-Index country ranks together with the simulated median ranks and 90% confidence intervals in order to better appreciate the robustness of the results to the choice of weights and aggregation function and the impact of estimating missing data (where applicable).

Sensitivity analysis results

Complementary to the uncertainty analysis, sensitivity analysis has been used to identify which of the modelling assumptions have the highest impact on certain country ranks. Figure 3 plots the rankings of the GII and its sub-indices versus one-at-a-time changes of either the EM imputation method or the geometric aggregation formula, with random weights, with summary results included in Table 5. The most influential assumption is the choice of no imputation versus EM imputation, in particular for the Output Sub-Index, next for the GII, and least for the Input Sub-Index. This sensitivity is a result of data availability, which is less satisfactory in the case of the Output Sub-Index: although no economy has indicator coverage of less than 63% over the 54 variables in the Input Sub-Index, 38 economies have data coverage below this threshold over the 27 variables in the Output Sub-Index. This factor has impacted the uncertainty analysis as well, and has propagated from the Output Sub-Index to the estimation of the overall GII. The choice of the aggregation formula has a very limited impact on the country/economy ranks.

Our recommendation would be to consider country/economy ranks in the GII 2014 and in the Input and Output Sub-Indices not only at face value but also within the 90% confidence intervals in order to better appreciate to what degree a country/economy rank depends on the modelling choices.

Distance to the efficient frontier in the GII by data envelopment analysis

Several innovation-related policy issues at the national level entail an intricate balance between global priorities and economy-specific strategies. Comparing the multidimensional performance on innovation by subjecting economies to a fixed and common set of weights may prevent acceptance of an innovation index on the grounds that a given weighting scheme might not be fair to a particular economy. An appealing feature of the more recent data envelopment analysis (DEA) literature applied in real decision-making settings is that it

Figure 3a: Sensitivity analysis: Impact of modelling choices
(Imputation)

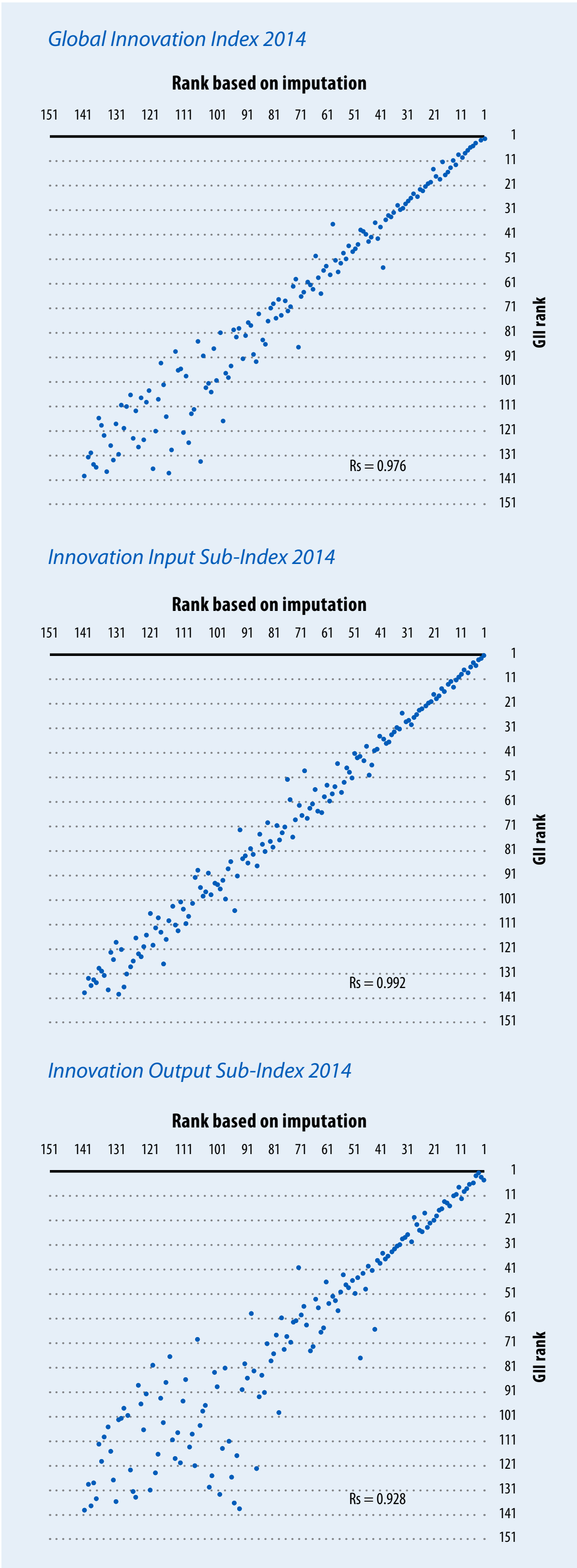
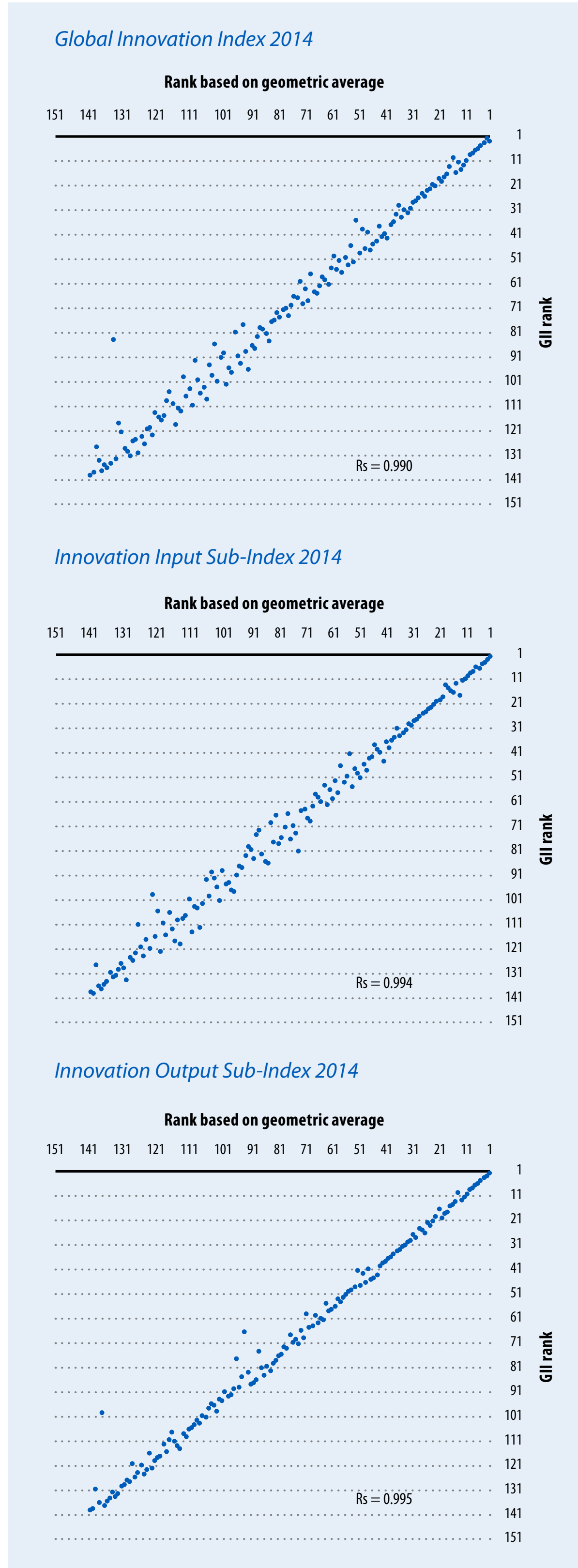


Figure 3b: Sensitivity analysis: Impact of modelling choices
(Geometric average)



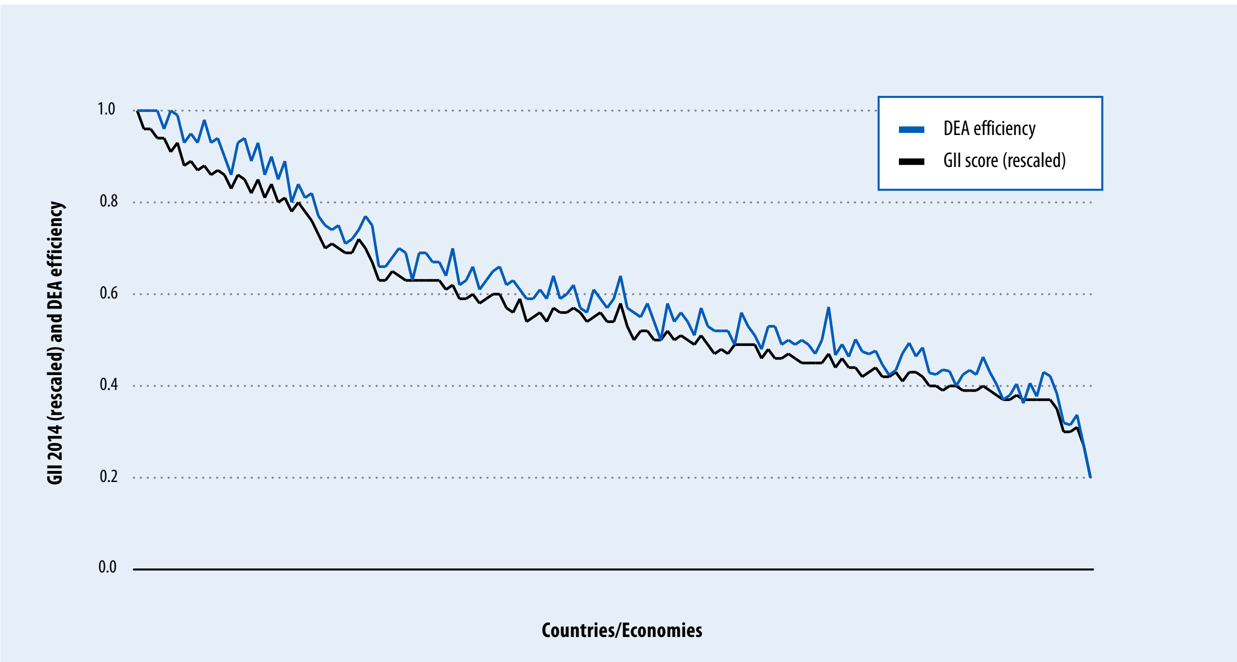
Source: Saisana and Saltelli, European Commission Joint Research Centre, 2014.
Note: Rs = Spearman rank correlation; imputation based on expectation-maximization algorithm.

Table 5: Sensitivity analysis: Impact of modelling choices on economies with most sensitive ranks

Index or Sub-Index	Uncertainty tested (pillar level only)	Number of economies that *improve* by 20 or more positions	Number of economies that *deteriorate* by 20 or more positions
GII	Geometric vs. arithmetic average	0	1
	EM imputation vs. no imputation of missing data	6	0
	Geometric average and EM imputation vs. arithmetic average and missing values	7	3
Input Sub-Index	Geometric vs. arithmetic average	0	0
	EM imputation vs. no imputation of missing data	0	0
	Geometric average and EM imputation vs. arithmetic average and missing values	1	1
Output Sub-Index	Geometric vs. arithmetic average	0	1
	EM imputation vs. no imputation of missing data	13	16
	Geometric average and EM imputation vs. arithmetic average and missing values	13	16

Source: Saisana and Saltelli, European Commission Joint Research Centre, 2014.

Figure 4: GII 2014 scores and DEA 'distance to the efficient frontier' scores



Note: For comparison purposes, we have rescaled the GII scores by dividing them with the best performer in the overall GII 2014.

Table 6: Pie shares and distance to the efficient frontier: Top 15 economies in the GII 2014

Counttry/Economy	DEA efficiency	Institutions	Human capital and research	Infrastructure	Market sophistication	Business sophistication	Knowledge and technology outputs	Creative outputs
United Kingdom	1.00	0.08	0.19	0.19	0.19	0.07	0.19	0.08
Switzerland	1.00	0.06	0.19	0.10	0.09	0.18	0.19	0.19
Singapore	1.00	0.07	0.20	0.20	0.12	0.20	0.17	0.05
Sweden	1.00	0.15	0.20	0.20	0.05	0.10	0.20	0.11
Finland	1.00	0.20	0.20	0.11	0.05	0.16	0.20	0.08
United States of America	0.99	0.20	0.20	0.06	0.20	0.09	0.20	0.05
Hong Kong (China)	0.98	0.20	0.05	0.20	0.20	0.15	0.05	0.15
Netherlands	0.96	0.20	0.06	0.20	0.05	0.20	0.09	0.20
Denmark	0.95	0.20	0.20	0.20	0.15	0.05	0.05	0.15
Canada	0.94	0.20	0.17	0.20	0.20	0.05	0.05	0.13
Ireland	0.93	0.20	0.08	0.05	0.20	0.20	0.20	0.07
Israel	0.93	0.05	0.20	0.07	0.20	0.20	0.20	0.08
Luxembourg	0.93	0.20	0.07	0.20	0.05	0.20	0.08	0.20
Germany	0.90	0.20	0.20	0.18	0.05	0.05	0.20	0.12
Iceland	0.86	0.20	0.20	0.20	0.09	0.05	0.06	0.20

Source: Saisana and Saltelli, European Commission Joint Research Centre, 2014.
Note: Pie shares are in absolute terms, bounded by 0.05 and 0.20.

can determine endogenous weights that maximize the overall score of each decision-making unit given a set of other observations.

In this section, the assumption of fixed pillar weights common to all economies is relaxed once more; this time economy-specific weights that maximize an economy's score are determined endogenously by DEA.[10] In theory, each economy is free to decide on the relative contribution of each pillar to its score, so as to achieve the best possible score in a computation that reflects its innovation strategy. In practice, the DEA method assigns a higher (lower) contribution to those pillars in which an economy is relatively strong (weak). Reasonable constraints on the weights are assumed to preclude the possibility of an economy achieving a perfect score by assigning a zero weight to weak pillars: for each economy, the share of each pillar score (i.e., the pillar score multiplied by the DEA weight over the total score) has upper and lower bounds of 5% and 20%, respectively. The DEA score is then measured as the weighted average of all seven pillar scores, where the weights are the economy-specific DEA weights compared to the best performance among all other economies with those same weights. The DEA score can be interpreted as a measure of the distance to the efficient frontier.

Table 6 presents the pie shares and DEA scores for the top 15 economies in the GII 2014. All pie shares are determined in accordance with a starting point that grants leeway to each economy when assigning shares while not violating the (relative) upper and lower bounds. The pie shares are quite diverse and reflect current national innovation strategies. This year, for example, Switzerland assigns 19% of its DEA score to *Creative outputs,* while the same pillar accounts for no more than 5% of Sweden's DEA score. More than half of the top 15 economies assign the maximum allowed (20%) to the first three Input pillars of the GII: *Institutions, Human capital and research,* and *Infrastructure.* Five economies—the United Kingdom, Switzerland, Singapore, Sweden, and Finland—reach a perfect DEA score of 1, and the United States of America and Hong Kong (China) are very close to the frontier. It is worth noting that the 15 economies that achieved the highest DEA scores are the same economies in the top 15 of the GII (except for Iceland, which ranks 19th in the GII). Figure 4 shows how closely related the DEA scores and GII 2014 scores are for all 143 economies (correlation of 0.994).

Conclusion

The JRC analysis suggests that the conceptualized multi-level structure of the GII 2014 with its 21 sub-pillars, 7 pillars, 2 sub-indices, and overall index is statistically sound and balanced: that is, each indicator and sub-pillar makes a similar contribution to the variation of its respective sub-pillar or pillar. The no-imputation choice of not treating missing values, common in relevant contexts and justified on the grounds of transparency and replicability, can at times have an undesirable impact on some country scores for the Innovation Output Sub-Index in particular, with the additional negative side effect that it may encourage countries not to report low data values. The choice of the GII team this year to use weights as scaling coefficients during the development of the index (the same choice that was made for the GII 2012 and 2013) constitutes a significant departure from the traditional vision of weights as a reflection of indicators' importance in a weighted average. Such a consideration will, it is hoped, also be made by other developers of composite indicators.

The strong correlations among the GII components are proven not to be a sign of redundancy of information in the GII. For more than 51.7% (up to 74.1%) of the 143 economies included in the GII 2014, the GII ranking and any of the seven pillar rankings differ by 10 positions or more. This demonstrates the added value of the GII ranking, which helps to highlight other components of innovation that do not emerge directly by looking into the seven pillars separately.

All published GII 2014 ranks lie within the simulated 90% confidence intervals that take into account the unavoidable uncertainties in the estimation of missing data, the weights (fixed vs. random), and the aggregation formula (arithmetic vs. geometric average) at the pillar level. For most economies, these intervals are narrow enough for meaningful inferences to be drawn: fewer than 10 positions for 81 of the 143 economies. Caution is needed for some countries with ranks that are highly sensitive to the methodological choices. The Output Sub-index is more sensitive to the methodological choices, mostly because of the estimation of missing data and the fact that this sub-index has only two pillars (with 0.68 correlation); hence changes to the imputation method, weights, or aggregation formula have a more notable impact on the country ranks. Nevertheless, country ranks, either in the GII 2014 or in the two sub-indices, can be considered representative of the many possible scenarios: 75% of the economies shift fewer than five positions with respect to the median rank in the GII (four and seven positions, respectively, in the Input and Output Sub-Indices).

The distance to the efficient frontier measure calculated with DEA scores could replace the Innovation Efficiency Ratio as a measure of efficiency, even if it is conceptually closer to the GII score than it is to the Efficiency Ratio. In fact, the 15 economies that achieved the highest DEA scores are the same economies in the top 15 of the GII (except for Iceland, which is ranked 19th in the GII).

All things considered, the JRC audit conducted herein shows the usefulness of the GII 2014 as a statistically sound benchmarking tool in reliably identifying strengths and weaknesses in national innovation practices around the world. We invite readers and users of the GII 2014 not to use this index as a standalone metric but to see it instead as a pointer back to the wealth of information gathered in the GII framework, which is a sound attempt to pave the way for better and more informed innovation policies worldwide.

Notes

1 OECD/EC JRC, 2008, p. 26.

2 The JRC analysis was based on the recommendations of the OECD/EC JRC (2008) *Handbook on Composite Indicators* and on more recent research from the JRC. The JRC audits on composite indicators are conducted upon request of the Index developers and are available at http://composite-indicators.jrc.ec.europa.eu/.

3 Groeneveld and Meeden (1984) set the criteria for absolute skewness above 1 and kurtosis above 3.5. The skewness criterion was relaxed to account for the small sample of 143 economies.

4 See Nunnally, 1978.

5 Saisana et al., 2005; Saisana et al., 2011.

6 With arithmetic average, the no-imputation choice is equivalent to replacing missing values with the average of the available (normalized) data within each sub-pillar.

7 The Expectation-Maximization (EM) algorithm (Little and Rubin, 2002) is an iterative procedure that finds the maximum likelihood estimates of the parameter vector by repeating two steps: (1) The expectation E-step: Given a set of parameter estimates, such as a mean vector and covariance matrix for a multivariate normal distribution, the E-step calculates the conditional expectation of the complete-data log likelihood given the observed data and the parameter estimates. (2) The maximization M-step: Given a complete-data log likelihood, the M-step finds the parameter estimates to maximize the complete-data log likelihood from the E-step. The two steps are iterated until the iterations converge.

8 Munda, 2008.

9 In the geometric average, pillars are multiplied as opposed to summed as they are in the arithmetic average. Pillar weights appear as exponents in the multiplication. All pillar scores were greater than zero, hence there was no reason to rescale them to avoid zero values that would have led to zero geometric averages.

10 The original question in the DEA literature concerned how to measure each unit's relative efficiency in production compared with a sample of peers, given observations on input and output quantities and, often, no reliable information on prices (Charnes and Cooper, 1985). A notable difference between the original DEA question and the one applied here is that no differentiation between inputs and outputs is made (Melyn and Moesen, 1991; Cherchye et al., 2008). To estimate DEA-based distance to the efficient frontier scores, we consider the m = 7 pillars in the GII 2014 for n = 143 economies, with y_{ij} the value of pillar j in economy i. The objective is to combine the pillar scores per economy into a single number, calculated as the weighted average of the m pillars, where w_j represents the weight of the jth pillar. In absence of reliable information about the true weights, the weights that maximize the DEA-based scores are endogenously determined. This gives the following linear programming problem for each country i:

$$Y_i = \max_{wij} \frac{\sum_{j=1}^{7} y_{ij} w_{ij}}{\max\limits_{y_c \in \{\text{dataset}\}} \sum_{j=1}^{7} y_{cj} w_{ij}} \qquad \text{(bounding constraint)}$$

subject to

$$w_{ij} \geqq 0, \qquad \text{(non-negativity constraint)}$$

where

$j = 1, \ldots, 7,$
$i = 1, \ldots, 143$

In this basic programming problem, the weights are non-negative and a country's score is between 0 (worst) and 1 (best).

References

Charnes, A., and W. W. Cooper. 1985. 'Preface to Topics in Data Envelopment Analysis'. *Annals of Operations Research* 2: 59–94.

Cherchye, L., W. Moesen, N. Rogge, T. Van Puyenbroeck, M. Saisana, A. Saltelli, R. Liska, and S. Tarantola. 2008. 'Creating Composite Indicators with DEA and Robustness Analysis: The Case of the Technology Achievement Index'. *Journal of Operational Research Society* 59: 239–51.

Groeneveld, R. A. and G. Meeden. 1984. 'Measuring Skewness and Kurtosis'. *The Statistician* 33: 391–99.

Little, R. J. A. and D. B. Rubin. 2002. *Statistical Analysis with Missing Data*. 2nd edition. Hoboken, NJ: John Wiley & Sons, Inc.

Melyn, W. and W. Moesen. 1991. 'Towards a Synthetic Indicator of Macroeconomic Performance: Unequal Weighting when Limited Information Is Available'. *Public Economics Research Paper* No. 17. Leuven: Centre for Economic Studies.

Munda, G. 2008. *Social Multi-Criteria Evaluation for a Sustainable Economy*. Berlin Heidelberg: Springer-Verlag.

Nunally, J. 1978. *Psychometric Theory*. New York: McGraw-Hill.

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., M. Saisana, and A. Saltelli. 2013. 'Ratings and Rankings: Voodoo or Science?' *Journal of the Royal Statistical Society* A 176 (3): 609–34.

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, 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.

Table 6: Pie shares and distance to the efficient frontier: Top 15 economies in the GII 2014

Counttry/Economy	DEA efficiency	Institutions	Human capital and research	Infrastructure	Market sophistication	Business sophistication	Knowledge and technology outputs	Creative outputs
United Kingdom	1.00	0.08	0.19	0.19	0.19	0.07	0.19	0.08
Switzerland	1.00	0.06	0.19	0.10	0.09	0.18	0.19	0.19
Singapore	1.00	0.07	0.20	0.20	0.12	0.20	0.17	0.05
Sweden	1.00	0.15	0.20	0.20	0.05	0.10	0.20	0.11
Finland	1.00	0.20	0.20	0.11	0.05	0.16	0.20	0.08
United States of America	0.99	0.20	0.20	0.06	0.20	0.09	0.20	0.05
Hong Kong (China)	0.98	0.20	0.05	0.20	0.20	0.15	0.05	0.15
Netherlands	0.96	0.20	0.06	0.20	0.05	0.20	0.09	0.20
Denmark	0.95	0.20	0.20	0.20	0.15	0.05	0.05	0.15
Canada	0.94	0.20	0.17	0.20	0.20	0.05	0.05	0.13
Ireland	0.93	0.20	0.08	0.05	0.20	0.20	0.20	0.07
Israel	0.93	0.05	0.20	0.07	0.20	0.20	0.20	0.08
Luxembourg	0.93	0.20	0.07	0.20	0.05	0.20	0.08	0.20
Germany	0.90	0.20	0.20	0.18	0.05	0.05	0.20	0.12
Iceland	0.86	0.20	0.20	0.20	0.09	0.05	0.06	0.20

Source: Saisana and Saltelli, European Commission Joint Research Centre, 2014.
Note: Pie shares are in absolute terms, bounded by 0.05 and 0.20.

can determine endogenous weights that maximize the overall score of each decision-making unit given a set of other observations.

In this section, the assumption of fixed pillar weights common to all economies is relaxed once more; this time economy-specific weights that maximize an economy's score are determined endogenously by DEA.[10] In theory, each economy is free to decide on the relative contribution of each pillar to its score, so as to achieve the best possible score in a computation that reflects its innovation strategy. In practice, the DEA method assigns a higher (lower) contribution to those pillars in which an economy is relatively strong (weak). Reasonable constraints on the weights are assumed to preclude the possibility of an economy achieving a perfect score by assigning a zero weight to weak pillars: for each economy, the share of each pillar score (i.e., the pillar score multiplied by the DEA weight over the total score) has upper and lower bounds of 5% and 20%, respectively. The DEA score is then measured as the weighted average of all seven pillar scores, where the weights are the economy-specific DEA weights compared to the best performance among all other economies with those same weights. The DEA score can be interpreted as a measure of the distance to the efficient frontier.

Table 6 presents the pie shares and DEA scores for the top 15 economies in the GII 2014. All pie shares are determined in accordance with a starting point that grants leeway to each economy when assigning shares while not violating the (relative) upper and lower bounds. The pie shares are quite diverse and reflect current national innovation strategies. This year, for example, Switzerland assigns 19% of its DEA score to *Creative outputs,* while the same pillar accounts for no more than 5% of Sweden's DEA score. More than half of the top 15 economies assign the maximum allowed (20%) to the first three Input pillars of the GII: *Institutions, Human capital and research,* and *Infrastructure.* Five economies—the United Kingdom, Switzerland, Singapore, Sweden, and Finland—reach a perfect DEA score of 1, and the United States of America and Hong Kong (China) are very close to the frontier. It is worth noting that the 15 economies that achieved the highest DEA scores are the same economies in the top 15 of the GII (except for Iceland, which ranks 19th in the GII). Figure 4 shows how closely related the DEA scores and GII 2014 scores are for all 143 economies (correlation of 0.994).

Conclusion

The JRC analysis suggests that the conceptualized multi-level structure of the GII 2014 with its 21 sub-pillars, 7 pillars, 2 sub-indices, and overall index is statistically sound and balanced: that is, each indicator and sub-pillar makes a similar contribution to the variation of its respective sub-pillar or pillar. The no-imputation choice of not treating missing values, common in relevant contexts and justified on the grounds of transparency and replicability, can at times have an undesirable impact on some country scores for the Innovation Output Sub-Index in particular, with the additional negative side effect that it may encourage countries not to report low data values. The choice of the GII team this year to use weights as scaling coefficients during the development of the index (the same choice that was made for the GII 2012 and 2013) constitutes a significant departure from the traditional vision of weights as a reflection of indicators' importance in a weighted average. Such a consideration will, it is hoped, also be made by other developers of composite indicators.

The strong correlations among the GII components are proven not to be a sign of redundancy of information in the GII. For more than 51.7% (up to 74.1%) of the 143 economies included in the GII 2014, the GII ranking and any of the seven pillar rankings differ by 10 positions or more. This demonstrates the added value of the GII ranking, which helps to highlight other components of innovation that do not emerge directly by looking into the seven pillars separately.

All published GII 2014 ranks lie within the simulated 90% confidence intervals that take into account the unavoidable uncertainties in the estimation of missing data, the weights (fixed vs. random), and the aggregation formula (arithmetic vs. geometric average) at the pillar level. For most economies, these intervals are narrow enough for meaningful inferences to be drawn: fewer than 10 positions for 81 of the 143 economies. Caution is needed for some countries with ranks that are highly sensitive to the methodological choices. The Output Sub-index is more sensitive to the methodological choices, mostly because of the estimation of missing data and the fact that this sub-index has only two pillars (with 0.68 correlation); hence changes to the imputation method, weights, or aggregation formula have a more notable impact on the country ranks. Nevertheless, country ranks, either in the GII 2014 or in the two sub-indices, can be considered representative of the many possible scenarios: 75% of the economies shift fewer than five positions with respect to the median rank in the GII (four and seven positions, respectively, in the Input and Output Sub-Indices).

The distance to the efficient frontier measure calculated with DEA scores could replace the Innovation Efficiency Ratio as a measure of efficiency, even if it is conceptually closer to the GII score than it is to the Efficiency Ratio. In fact, the 15 economies that achieved the highest DEA scores are the same economies in the top 15 of the GII (except for Iceland, which is ranked 19th in the GII).

All things considered, the JRC audit conducted herein shows the usefulness of the GII 2014 as a statistically sound benchmarking tool in reliably identifying strengths and weaknesses in national innovation practices around the world. We invite readers and users of the GII 2014 not to use this index as a standalone metric but to see it instead as a pointer back to the wealth of information gathered in the GII framework, which is a sound attempt to pave the way for better and more informed innovation policies worldwide.

Notes

1 OECD/EC JRC, 2008, p. 26.

2 The JRC analysis was based on the recommendations of the OECD/EC JRC (2008) *Handbook on Composite Indicators* and on more recent research from the JRC. The JRC audits on composite indicators are conducted upon request of the Index developers and are available at http://composite-indicators.jrc.ec.europa.eu/.

3 Groeneveld and Meeden (1984) set the criteria for absolute skewness above 1 and kurtosis above 3.5. The skewness criterion was relaxed to account for the small sample of 143 economies.

4 See Nunnally, 1978.

5 Saisana et al., 2005; Saisana et al., 2011.

6 With arithmetic average, the no-imputation choice is equivalent to replacing missing values with the average of the available (normalized) data within each sub-pillar.

7 The Expectation-Maximization (EM) algorithm (Little and Rubin, 2002) is an iterative procedure that finds the maximum likelihood estimates of the parameter vector by repeating two steps: (1) The expectation E-step: Given a set of parameter estimates, such as a mean vector and covariance matrix for a multivariate normal distribution, the E-step calculates the conditional expectation of the complete-data log likelihood given the observed data and the parameter estimates. (2) The maximization M-step: Given a complete-data log likelihood, the M-step finds the parameter estimates to maximize the complete-data log likelihood from the E-step. The two steps are iterated until the iterations converge.

8 Munda, 2008.

9 In the geometric average, pillars are multiplied as opposed to summed as they are in the arithmetic average. Pillar weights appear as exponents in the multiplication. All pillar scores were greater than zero, hence there was no reason to rescale them to avoid zero values that would have led to zero geometric averages.

10 The original question in the DEA literature concerned how to measure each unit's relative efficiency in production compared with a sample of peers, given observations on input and output quantities and, often, no reliable information on prices (Charnes and Cooper, 1985). A notable difference between the original DEA question and the one applied here is that no differentiation between inputs and outputs is made (Melyn and Moesen, 1991; Cherchye et al., 2008). To estimate DEA-based distance to the efficient frontier scores, we consider the m = 7 pillars in the GII 2014 for n = 143 economies, with y_{ij} the value of pillar j in economy i. The objective is to combine the pillar scores per economy into a single number, calculated as the weighted average of the m pillars, where w_i represents the weight of the jth pillar. In absence of reliable information about the true weights, the weights that maximize the DEA-based scores are endogenously determined. This gives the following linear programming problem for each country i:

$$Y_i = \max_{w_{ij}} \frac{\sum_{j=1}^{7} y_{ij} w_{ij}}{\max_{y_c \in \{\text{dataset}\}} \sum_{j=1}^{7} y_{cj} w_{ij}} \qquad \text{(bounding constraint)}$$

subject to

$$w_{ij} \geqq 0, \qquad \text{(non-negativity constraint)}$$

where

$j = 1, \ldots, 7,$
$i = 1, \ldots, 143$

In this basic programming problem, the weights are non-negative and a country's score is between 0 (worst) and 1 (best).

References

Charnes, A., and W. W. Cooper. 1985. 'Preface to Topics in Data Envelopment Analysis'. *Annals of Operations Research* 2: 59–94.

Cherchye, L., W. Moesen, N. Rogge, T. Van Puyenbroeck, M. Saisana, A. Saltelli, R. Liska, and S. Tarantola. 2008. 'Creating Composite Indicators with DEA and Robustness Analysis: The Case of the Technology Achievement Index'. *Journal of Operational Research Society* 59: 239–51.

Groeneveld, R. A. and G. Meeden. 1984. 'Measuring Skewness and Kurtosis'. *The Statistician* 33: 391–99.

Little, R. J. A. and D. B. Rubin. 2002. *Statistical Analysis with Missing Data.* 2nd edition. Hoboken, NJ: John Wiley & Sons, Inc.

Melyn, W. and W. Moesen. 1991. 'Towards a Synthetic Indicator of Macroeconomic Performance: Unequal Weighting when Limited Information Is Available'. *Public Economics Research Paper* No. 17. Leuven: Centre for Economic Studies.

Munda, G. 2008. *Social Multi-Criteria Evaluation for a Sustainable Economy.* Berlin Heidelberg: Springer-Verlag.

Nunally, J. 1978. *Psychometric Theory.* New York: McGraw-Hill.

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., M. Saisana, and A. Saltelli. 2013. 'Ratings and Rankings: Voodoo or Science?' *Journal of the Royal Statistical Society* A 176 (3): 609–34.

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, 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.