Joint Research Centre Statistical Audit of the 2013 Global Innovation Index

Michaela Saisana and **Dionisis Th. Philippas**, European Commission Joint Research Centre (Ispra, Italy)

Modelling versatile concepts underlying innovation at the national scale around the globe, as attempted in the Global Innovation Index (GII), raises practical challenges related to the quality of data and the combination of these into a single number. The Econometrics and Applied Statistics Unit at the European Commission Joint Research Centre (JRC) in Ispra (Italy) was invited for a third consecutive year to audit the GII because of the adjustments made to the list of indicators included in the GII framework (see Annex 2 for more details).

The JRC assessment of the 2013 GII focused on two main issues: the conceptual and statistical coherence of the structure, and the impact of key modelling assumptions on the GII scores and ranks.[1] These are necessary steps to ensure the transparency and reliability of the GII, to enable policy makers to derive more accurate and meaningful conclusions, and to potentially guide choices on priority setting and policy formulation.

As in the previous two 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, for the first time this year, the JRC analysis includes both 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 (DEA).

Conceptual and statistical coherence in the GII framework

An earlier version of the GII model was assessed by the JRC in April 2013. Fine-tuning suggestions were taken into account in the final computation of the rankings in an iterative process with the JRC, aiming to set the foundation for a balanced index. The entire process followed four steps (see Figure 1):

Step 1: Conceptual consistency

Candidate 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

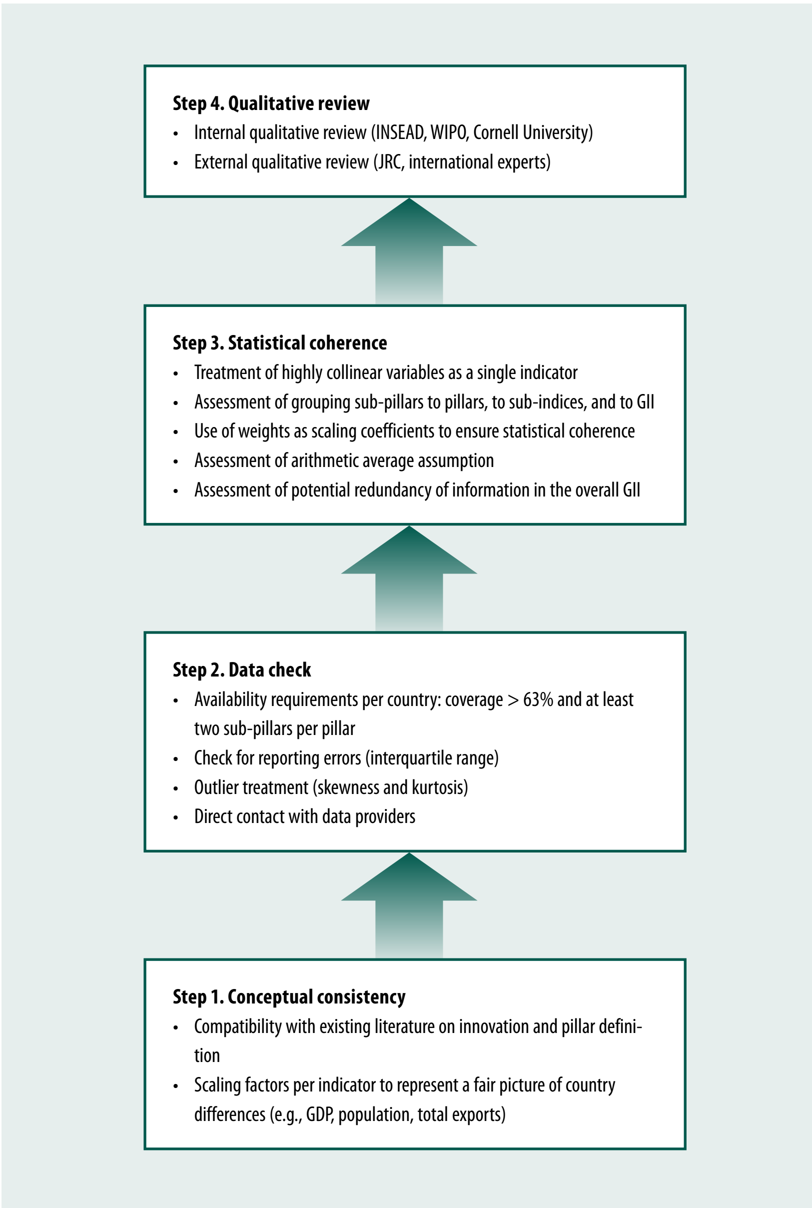
The most recently released data were used for each country with a cut-off at year 2003. Countries were included if data availability was at least 63% (i.e., 54 out of 84 variables) and at least two of the three sub-pillars in each pillar could be computed. Potentially problematic indicators that could bias the overall results were identified as those having absolute skewness greater than 2 and kurtosis greater than 3.5.[2] These indicators were treated either by winsorisation or by taking the natural logarithm (in case of more than five outliers). These criteria were decided jointly with the JRC back in 2011 (see Appendix IV, Technical Notes, for details).

Step 3: Statistical coherence

Weights as 'scaling coefficients'

Weights of 0.5 or 1.0 were jointly decided between the JRC and the GII team as 'scaling coefficients' and not as 'importance coefficients', with the aim of arriving at sub-pillar and pillar scores that were balanced in their underlying components (with balanced contributions of indicators/sub-pillars to the variance of their respective sub-pillars/pillars). Paruolo, Saisana, and Saltelli (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's 'correlation ratio'. As a result of this analysis, 23 out of 84 indicators and three sub-pillars—6.1 Knowledge creation, 7.2 Creative goods and services, and 7.3 Online creativity—were assigned half weights, while all

Figure 1: Conceptual and statistical coherence in the GII 2013 framework



Source: Saisana and Philippas, European Commission Joint Research Centre, 2013.

other indicators and sub-pillars were assigned a weight of 1.0.[3]

Principal component analysis

Principal component analysis confirms the presence of a single latent dimension in each of the seven pillars (one component with eigenvalue greater than 1.0) that captures between 63% (pillars 5 and 6) up to 83% (pillar 1) of the total variance in the three underlying sub-pillars.[4] These results reveal that the adjustments made to the 2013 GII framework led to a further improvement of its statistical coherence.[5] Furthermore, results confirm the expectation that the sub-pillars are more correlated with their own pillar than with any other. It is interesting to note that sub-pillar 6.1 Knowledge creation has the same degree of correlation (0.76) with its own pillar 6 Knowledge and technology outputs than with pillar 2 Human capital and research, a confirmation of the link between human capital and the creation of knowledge.

The five pillars in the Innovation Input Sub-index also share a single latent dimension that captures 82% of the total variance. The five loadings are very similar to each other; thereafter, building the Input Sub-Index as a simple average (equal weights) of the five pillars is statistically supported by the data. The two output pillars, Knowledge and technology outputs and Creative outputs, are moderately correlated with each other (0.60), but they are both strongly correlated with the Innovation Output Sub-Index (0.88), implying that that sub-index is also well balanced in its two pillars.

Last, 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 roughly 0.90. So far, results show that the grouping of sub-pillars into pillars, sub-indices, and the GII is statistically coherent, and that the GII has a balanced structure justifying the various levels of aggregation.

Assessing potential redundancy of information in the GII

As discussed, the Input and Output Sub-Indices correlate well with each other and with the overall GII. However, the information summarized by the GII is not redundant. In fact, one way in which the GII helps to highlight other components of innovation is by pinpointing the differences in rankings that emerge from a comparison between

Table 1: 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 30	19.7	13.4	10.6	20.4	18.3	25.4	17.6
20 to 29	13.4	20.4	15.5	14.1	16.2	15.5	14.8
10 to 19	20.4	24.6	29.6	27.5	20.4	19.0	29.6
5 to 9	26.1	19.0	19.7	20.4	24.6	21.1	16.2
Less than 5	20.4	22.5	22.5	15.5	17.6	16.2	19.0
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 Philippas, European Commission Joint Research Centre, 2013.

the GII and each of the seven pillars (see Table 1). Of the 142 countries included in the GII 2013, for more than 53.5% (up to 62.0%) of the countries, the GII ranking and any of the seven pillar rankings differ by 10 positions or more.

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, or prevailing theory.

Notwithstanding these statistical tests and the positive outcomes on the statistical coherence of the GII structure, it is important to mention 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 country score on the GII and its two sub-indices depends on modelling choices: the seven-pillar structure, selected indicators, imputation or not of missing data, normalization, weights, aggregation method, among other elements. These choices are based on expert opinion (e.g., selection of indicators), 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 is aimed at assessing the simultaneous and joint impact of these modelling choices on the rankings. The data are assumed to be error-free, since potential outliers and eventual errors and typos were corrected during the computation phase (see Step 2 in Figure 1).

The robustness assessment of the GII was based on a combination of a Monte Carlo experiment and a multi-modelling approach that dealt with three issues: pillar weights, missing data, and the aggregation formula. This type of assessment aims to respond to eventual criticism that the country scores associated with aggregate measures are generally not calculated under conditions of certainty, even if they are frequently presented as such.[6]

The Monte Carlo simulation related to the issue of weighting and comprised 1,000 runs, each corresponding to a different set of weights 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 opposite needs: (1) to ensure a wide enough interval to have meaningful robustness checks, and (2) to respect the rationale of the GII that places on an equal footing the Input Sub-Index and the Output Sub-Index. 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 2).[7]

The GII developing team, for transparency and replicability, opted to not estimate missing data. The 'no imputation' choice, which is common in similar contexts, might encourage countries not to report low data values.[8] To overcome this limitation, the JRC opted to impute missing data using the Expectation Maximization (EM) algorithm.[9]

Regarding the aggregation formula, decision-theory practitioners have challenged the use of simple arithmetic averages because of their fully compensatory nature, in which a comparative high advantage on a

Table 2: 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 Philippas, European Commission Joint Research Centre, 2013.

few indicators can compensate a comparative disadvantage on many indicators (Munda, 2008). Despite receiving statistical support in the previous section, the geometric average was considered instead,[10] which is a partially compensatory approach that rewards economies with balanced profiles and motivates them to improve in the dimensions in which they perform poorly, and not just in *any* dimension.

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 2 for a summary of the uncertainties considered in the GII 2013).

Uncertainty analysis results

The main results of the robustness analysis are shown in Figures 2a, 2b, and 2c 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. Error bars represent, for each country, the 90% interval across all simulations. Table 3 reports the published rankings and the 90% confidence intervals. It can be verified that all but five country ranks lie within the simulated intervals, and that these are narrow enough for most countries (less than 10 positions) to allow meaningful inferences to be drawn.

GII ranks are rather robust: the median rank is close to the reference rank (six or fewer positions away) for 75% of the countries. Results for the Input Sub-Index are relatively more robust (75% of the countries shift fewer than three positions) for two main reasons: the high correlations between the five Input pillars (the average bivariate Pearson correlation coefficient of 0.82) and the very good data coverage (only 1 of the 142 countries has an indicator coverage below 63% of the 57 variables included in the Input Sub-Index).

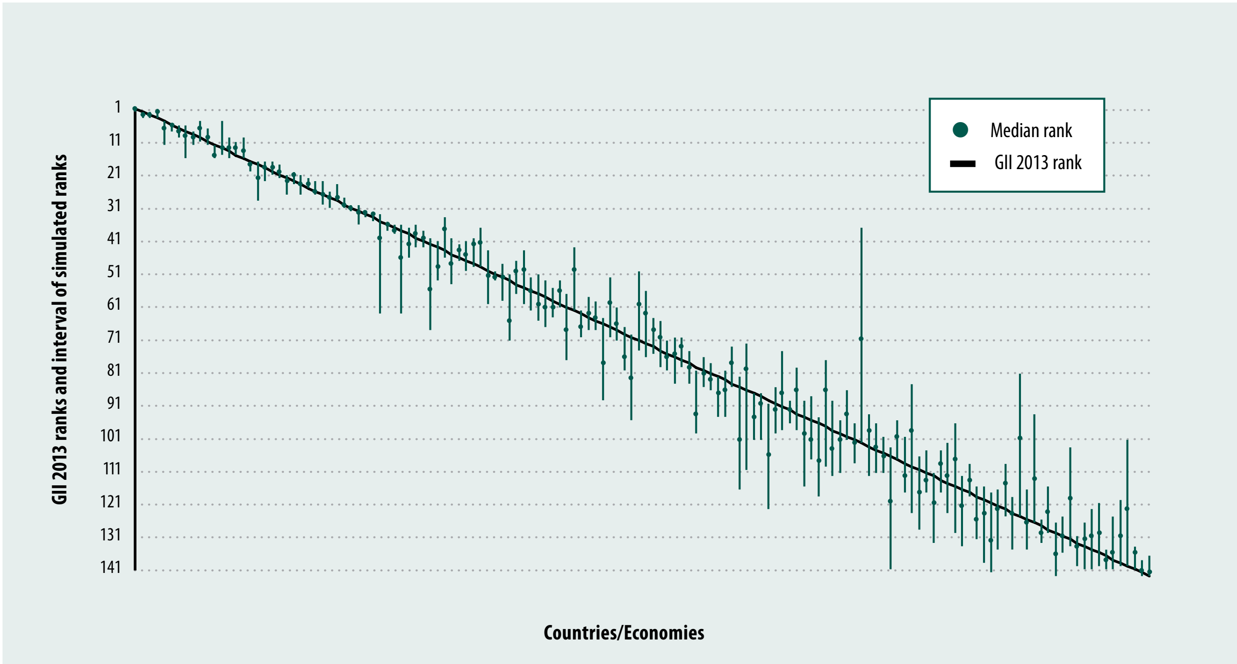
In contrast, the Output Sub-Index is more sensitive to the methodological choices (one-fourth of the countries shift more than 10 positions) for the same two reasons: there are only two pillars that are moderately correlated (0.60) and the data coverage is less satisfactory (15 countries have an indicator coverage of less than 63% of the 27 variables included in the Output Sub-Index). However, it cannot be ruled out altogether that the correlation between the two Output pillars could improve as data become available, as suggested by theory. The currently observed moderate correlation might be the result of (1) the fact that missing values are particularly distorting; (2) the use of count and not value variables; (3) the use of proxies due to the lack of statistics.

Sensitivity analysis results

Complementary to the uncertainty analysis, sensitivity analysis has been used to identify which of the modelling assumptions have the greatest impact on certain country ranks. Figure 3 plots the rankings of the GII and 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 4. Figure 4 presents the box plots of ranking shifts with respect to the original ranking resulting from random weights only.

The most influential assumption is the choice of no imputation versus EM imputation, particularly

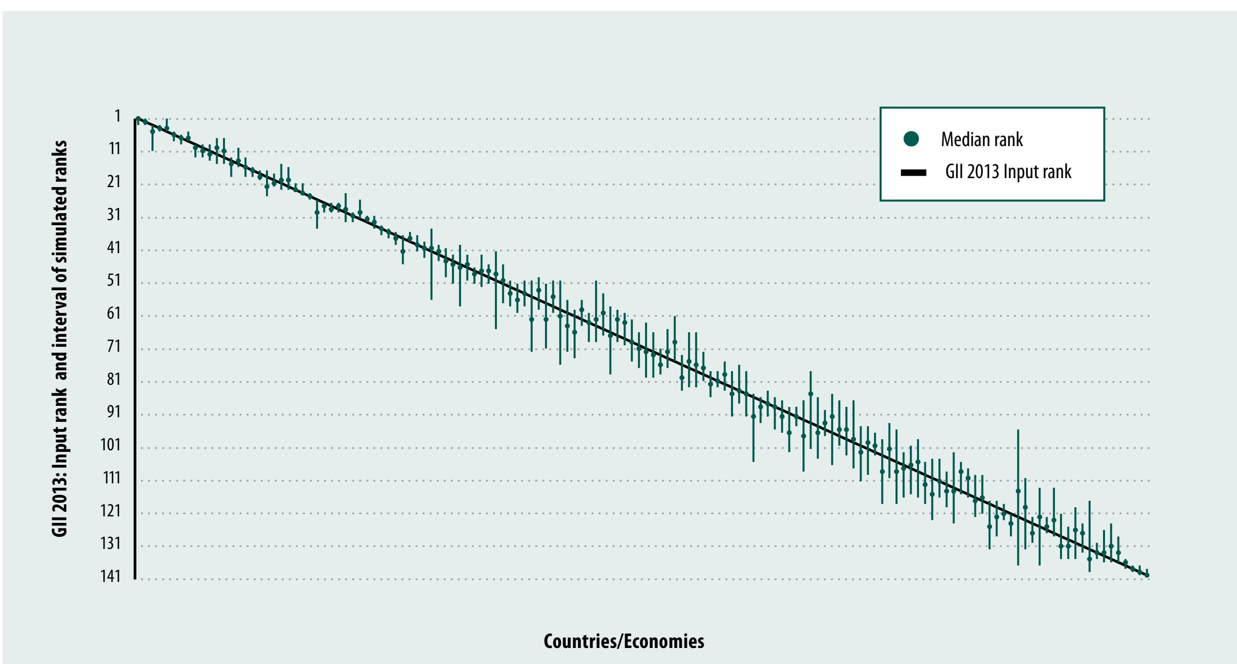
Figure 2a: Robustness analysis (GII rank vs. median rank, 90% confidence intervals)



Source: Saisana and Philippas, European Commission Joint Research Centre, 2013.

Note: The Spearman rank correlation between the median rank and the GII 2013 rank is 0.987. Median ranks and intervals are calculated over 4,000 simulated scenarios combining different sets of weights, imputed versus non imputed (missing) values and geometric versus arithmetic average at the pillar level.

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



Source: Saisana and Philippas, European Commission Joint Research Centre, 2013.

Note: The Spearman rank correlation between the median rank and the Input rank is 0.998. Median ranks and intervals are calculated over 4,000 simulated scenarios combining different sets of weights, imputed versus non imputed (missing) values and geometric versus arithmetic average at the pillar level.

Table 3: GII 2013 and Input/Output Sub-Indices: Ranks and 90% confidence intervals

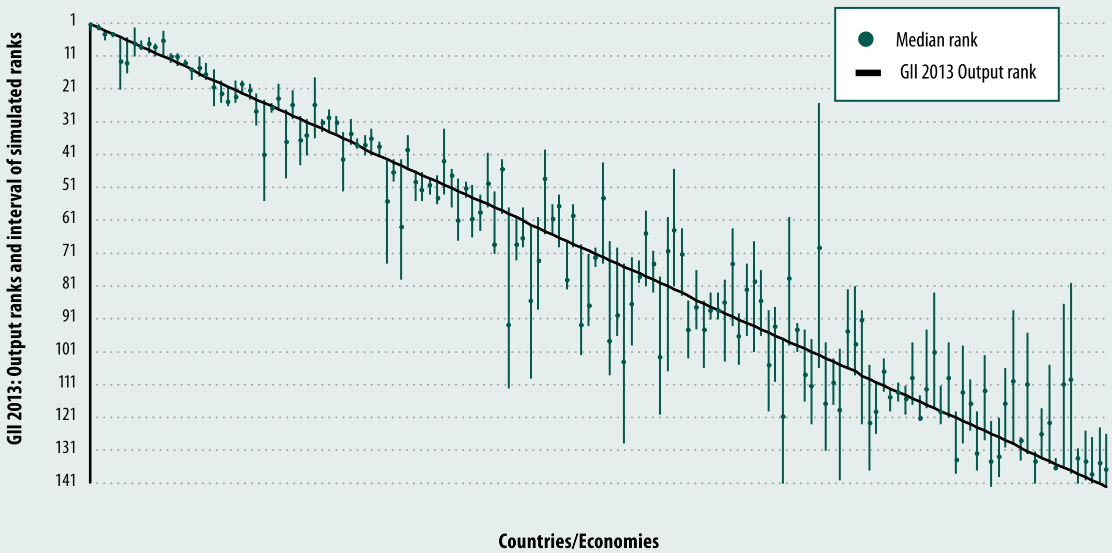
	GII 2013		Input Sub-Index		Output Sub-Index	
Country/Economy	Rank	Interval	Rank	Interval	Rank	Interval
Switzerland	1	[1, 2]	7	[5, 11]	1	[1, 3]
Sweden	2	[2, 4]	5	[3, 5]	3	[3, 6]
United Kingdom	3	[2, 4]	4	[3, 6]	4	[4, 5]
Netherlands	4	[1, 4]	10	[9, 13]	2	[1, 2]
United States of America	5	[5, 12]	3	[3, 13]	12	[10, 13]
Finland	6	[6, 8]	6	[4, 10]	8	[6, 9]
Hong Kong (China)	7	[6, 10]	2	[1, 3]	15	[15, 18]
Singapore	8	[6, 16]	1	[1, 2]	18	[15, 26]
Denmark	9	[8, 12]	8	[5, 8]	14	[12, 14]
Ireland	10	[5, 11]	12	[6, 15]	11	[3, 11]
Canada	11	[7, 12]	9	[6, 13]	13	[10, 14]
Luxembourg	12	[12, 16]	18	[15, 21]	6	[5, 16]
Iceland	13	[5, 15]	21	[15, 23]	7	[2, 11]
Israel	14	[10, 16]	19	[16, 25]	9	[5, 10]
Germany	15	[11, 15]	20	[18, 22]	10	[7, 11]
Norway	16	[10, 16]	13	[7, 15]	16	[11, 17]
New Zealand	17	[17, 20]	15	[11, 17]	19	[18, 25]
Korea, Rep.	18	[17, 29]	16	[11, 21]	24	[22, 32]
Australia	19	[17, 23]	11	[9, 14]	32	[17, 36]
France	20	[17, 21]	23	[21, 24]	17	[13, 18]
Belgium	21	[18, 22]	22	[16, 24]	22	[18, 22]
Japan	22	[21, 27]	14	[12, 19]	33	[30, 34]
Austria	23	[20, 24]	17	[16, 20]	27	[19, 27]
Malta	24	[21, 27]	34	[30, 36]	5	[5, 21]
Estonia	25	[22, 25]	25	[24, 26]	21	[18, 25]
Spain	26	[23, 27]	24	[21, 25]	35	[29, 35]
Cyprus	27	[23, 30]	30	[24, 33]	20	[20, 26]
Czech Republic	28	[26, 31]	27	[26, 31]	26	[25, 28]
Italy	29	[24, 29]	28	[26, 31]	29	[21, 30]
Slovenia	30	[28, 31]	29	[27, 30]	34	[27, 34]
Hungary	31	[30, 32]	36	[35, 40]	23	[19, 24]
Malaysia	32	[30, 36]	32	[26, 33]	30	[29, 44]
Latvia	33	[32, 34]	33	[29, 33]	37	[30, 38]
Portugal	34	[33, 35]	31	[29, 34]	39	[35, 41]
China	35	[33, 63]	46	[39, 58]	25	[24, 55]
Slovakia	36	[35, 38]	37	[36, 41]	45	[35, 45]
Croatia	37	[36, 39]	43	[40, 45]	41	[37, 41]
United Arab Emirates	38	[36, 63]	26	[26, 36]	81	[60, 107]
Costa Rica	39	[37, 46]	66	[55, 70]	31	[30, 41]
Lithuania	40	[36, 43]	35	[34, 38]	56	[40, 57]
Bulgaria	41	[38, 43]	50	[46, 53]	38	[36, 39]
Saudi Arabia	42	[40, 68]	44	[40, 52]	44	[42, 79]
Qatar	43	[41, 53]	38	[37, 45]	52	[48, 67]
Montenegro	44	[34, 46]	40	[35, 43]	50	[33, 53]
Moldova, Rep.	45	[40, 54]	76	[63, 77]	28	[27, 48]
Chile	46	[42, 47]	41	[40, 45]	48	[48, 53]
Barbados	47	[41, 50]	42	[32, 59]	49	[47, 56]
Romania	48	[40, 49]	55	[51, 60]	40	[33, 41]
Poland	49	[37, 49]	39	[36, 40]	64	[39, 65]
Kuwait	50	[44, 60]	74	[66, 78]	36	[34, 52]
TFYR of Macedonia	51	[50, 53]	48	[47, 55]	66	[53, 69]
Uruguay	52	[48, 59]	64	[58, 72]	46	[46, 55]
Mauritius	53	[51, 71]	60	[47, 78]	57	[52, 71]
Serbia	54	[47, 57]	63	[56, 67]	51	[45, 57]
Greece	55	[44, 60]	45	[42, 53]	82	[45, 81]
Argentina	56	[52, 62]	78	[66, 84]	43	[42, 49]
Thailand	57	[51, 65]	57	[49, 62]	61	[57, 69]
South Africa	58	[53, 67]	51	[41, 68]	71	[69, 75]
Armenia	59	[55, 64]	71	[66, 79]	47	[46, 55]
Colombia	60	[53, 61]	59	[51, 62]	65	[56, 65]
Jordan	61	[57, 77]	61	[56, 76]	63	[60, 88]
Russian Federation	62	[43, 62]	52	[46, 60]	72	[43, 74]
Mexico	63	[62, 70]	68	[60, 70]	60	[60, 73]
Brazil	64	[58, 68]	67	[58, 80]	68	[56, 69]
Bosnia and Herzegovina	65	[59, 68]	58	[51, 71]	78	[58, 81]
India	66	[64, 89]	87	[87, 106]	42	[42, 74]
Bahrain	67	[52, 70]	47	[44, 52]	90	[63, 93]
Turkey	68	[61, 71]	81	[78, 87]	53	[49, 54]
Peru	69	[67, 80]	70	[61, 79]	70	[71, 93]
Tunisia	70	[69, 95]	80	[71, 83]	59	[57, 112]
Ukraine	71	[50, 74]	83	[75, 85]	58	[42, 59]

Table 3: GII 2013 and Input/Output Sub-Indices: Ranks and 90% confidence intervals *(continued)*

	GII 2013		Input Sub-Index		Output Sub-Index	
Country/Economy	Rank	Interval	Rank	Interval	Rank	Interval
Mongolia	72	[56, 76]	49	[44, 54]	93	[67, 101]
Georgia	73	[64, 75]	62	[58, 78]	83	[63, 84]
Brunei Darussalam	74	[65, 79]	54	[46, 61]	89	[79, 104]
Lebanon	75	[71, 80]	56	[51, 76]	88	[83, 91]
Viet Nam	76	[70, 84]	89	[85, 96]	54	[50, 66]
Belarus	77	[70, 79]	75	[65, 80]	79	[70, 83]
Guyana	78	[74, 84]	94	[87, 113]	55	[53, 64]
Dominican Republic	79	[80, 99]	93	[90, 101]	69	[68, 102]
Oman	80	[76, 85]	53	[51, 63]	111	[103, 113]
Trinidad and Tobago	81	[78, 86]	82	[78, 84]	87	[83, 91]
Jamaica	82	[82, 94]	85	[77, 92]	84	[85, 103]
Ecuador	83	[80, 94]	100	[90, 107]	67	[67, 82]
Kazakhstan	84	[73, 85]	69	[61, 71]	106	[82, 106]
Indonesia	85	[82, 116]	115	[104, 125]	62	[62, 109]
Panama	86	[72, 110]	73	[65, 82]	108	[88, 123]
Guatemala	87	[87, 101]	91	[88, 102]	91	[89, 105]
El Salvador	88	[87, 101]	88	[86, 98]	96	[87, 110]
Uganda	89	[90, 122]	109	[103, 117]	75	[74, 129]
Philippines	90	[85, 99]	108	[103, 118]	77	[73, 80]
Botswana	91	[74, 98]	65	[51, 71]	125	[102, 128]
Morocco	92	[89, 96]	90	[86, 101]	99	[92, 101]
Albania	93	[79, 98]	77	[72, 84]	118	[83, 118]
Ghana	94	[89, 115]	99	[89, 105]	95	[88, 119]
Bolivia, Plurinational St.	95	[88, 107]	106	[95, 116]	86	[85, 106]
Senegal	96	[94, 118]	116	[107, 117]	80	[78, 120]
Fiji	97	[77, 109]	72	[60, 83]	129	[88, 129]
Sri Lanka	98	[89, 112]	118	[110, 125]	76	[72, 99]
Kenya	99	[95, 111]	98	[87, 108]	100	[94, 116]
Paraguay	100	[86, 101]	104	[100, 105]	94	[76, 96]
Tajikistan	101	[96, 108]	113	[109, 126]	85	[76, 94]
Belize	102	[37, 102]	95	[79, 103]	102	[25, 106]
Cape Verde	103	[93, 112]	84	[78, 94]	122	[99, 129]
Swaziland	104	[96, 111]	124	[99, 140]	74	[69, 96]
Azerbaijan	105	[100, 111]	92	[90, 99]	114	[111, 120]
Mali	106	[103, 140]	132	[128, 137]	73	[67, 108]
Honduras	107	[95, 107]	96	[88, 99]	115	[98, 117]
Egypt	108	[100, 117]	101	[88, 112]	112	[112, 119]
Namibia	109	[84, 123]	79	[63, 84]	134	[105, 135]
Cambodia	110	[106, 128]	120	[118, 129]	101	[97, 123]
Gabon	111	[104, 117]	117	[107, 118]	104	[103, 117]
Rwanda	112	[111, 132]	102	[88, 114]	121	[119, 138]
Iran, Islamic Rep.	113	[104, 117]	107	[97, 122]	120	[98, 121]
Venezuela, Bolivarian Rep.	114	[102, 123]	134	[117, 141]	92	[74, 96]
Nicaragua	115	[96, 129]	103	[89, 112]	128	[106, 130]
Burkina Faso	116	[112, 133]	119	[107, 125]	109	[105, 137]
Kyrgyzstan	117	[108, 118]	97	[90, 101]	133	[118, 133]
Zambia	118	[115, 131]	128	[120, 139]	103	[98, 131]
Malawi	119	[115, 138]	125	[115, 134]	105	[100, 140]
Nigeria	120	[117, 141]	137	[133, 138]	97	[97, 141]
Mozambique	121	[116, 134]	111	[102, 119]	124	[119, 137]
Gambia	122	[108, 124]	127	[122, 135]	107	[81, 108]
Tanzania, United Rep.	123	[118, 134]	110	[104, 118]	127	[121, 139]
Lesotho	124	[81, 124]	86	[74, 95]	136	[86, 136]
Cameroon	125	[116, 134]	131	[123, 133]	110	[110, 126]
Guinea	126	[93, 126]	139	[134, 141]	98	[60, 99]
Benin	127	[125, 132]	121	[117, 128]	130	[127, 134]
Nepal	128	[115, 129]	129	[123, 129]	123	[109, 125]
Ethiopia	129	[126, 142]	126	[123, 133]	126	[122, 142]
Bangladesh	130	[124, 135]	135	[132, 137]	119	[111, 123]
Niger	131	[103, 133]	130	[111, 131]	131	[95, 132]
Zimbabwe	132	[130, 139]	138	[132, 142]	116	[114, 122]
Uzbekistan	133	[126, 140]	114	[106, 127]	138	[130, 141]
Syrian Arab Republic	134	[122, 140]	105	[99, 117]	140	[127, 141]
Angola	135	[120, 139]	140	[137, 141]	117	[94, 118]
Côte d'Ivoire	136	[134, 140]	133	[126, 134]	132	[131, 141]
Pakistan	137	[124, 140]	142	[140, 142]	113	[110, 116]
Algeria	138	[119, 139]	112	[105, 118]	141	[124, 141]
Togo	139	[101, 139]	122	[119, 127]	137	[80, 138]
Madagascar	140	[133, 140]	123	[119, 130]	135	[133, 137]
Sudan	141	[137, 142]	136	[124, 140]	142	[126, 142]
Yemen	142	[136, 142]	141	[137, 142]	139	[125, 140]

Source: Saisana and Philippas, European Commission Joint Research Centre, 2013.

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



Source: Saisana and Philippas, European Commission Joint Research Centre, 2013.

Note: The Spearman rank correlation between the median rank and the Output rank is 0.964. Median ranks and intervals are calculated over 4,000 simulated scenarios combining different sets of weights, imputed versus non imputed (missing) values and geometric versus arithmetic average at the pillar level.

for the Output Sub-Index, then for the GII, and least for the Input Sub-index. For example, in one case, a country improves by three positions in the Output Sub-Index ranking if a geometric aggregation is applied, although it is found to improve by 36 positions if EM imputation is applied. If both assumptions are changed with fixed (equal) pillar weights, the impact of the imputation is moderated (to a 19-position improvement). This sensitivity is the result of data availability, a factor that impacted the uncertainty analysis as well and that propagates from the Output Sub-Index to the estimation of the overall GII.

A recommendation for the future would be to apply the 63% criterion for data availability within each of the two sub-indices. For this year, drawing upon the analysis made by the JRC, the recommendation is to consider country ranks in the GII 2013 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 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 country-specific strategies. Comparing the multi-dimensional performance on innovation by subjecting countries 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 country. An appealing feature of the more recent DEA literature applied in real decision-making settings is that it allows for the determination of 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 countries is relaxed once more; this time country-specific weights that maximize a country's score are determined endogenously by DEA.[11] In theory, each country 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 a country is relatively strong (weak). Reasonable constraints on the weights are assumed to preclude the possibility of a country achieving a perfect score by assigning a zero weight to weak pillars: for each country, the share of

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

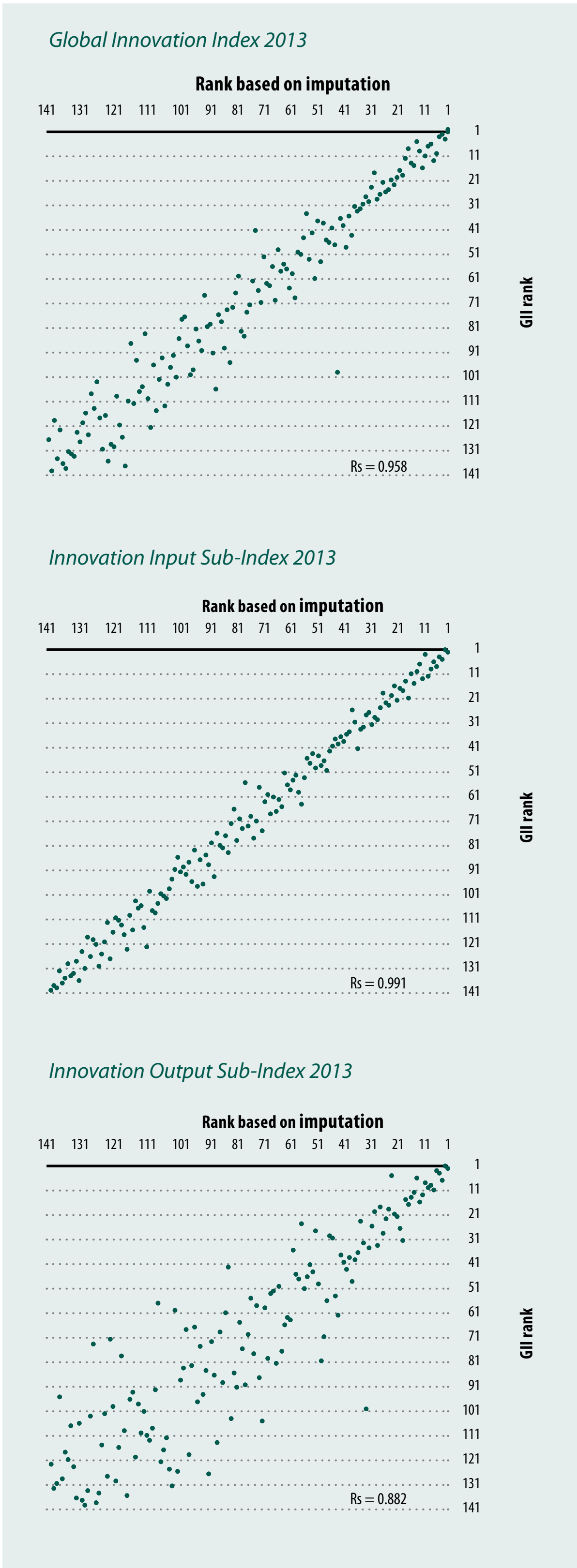
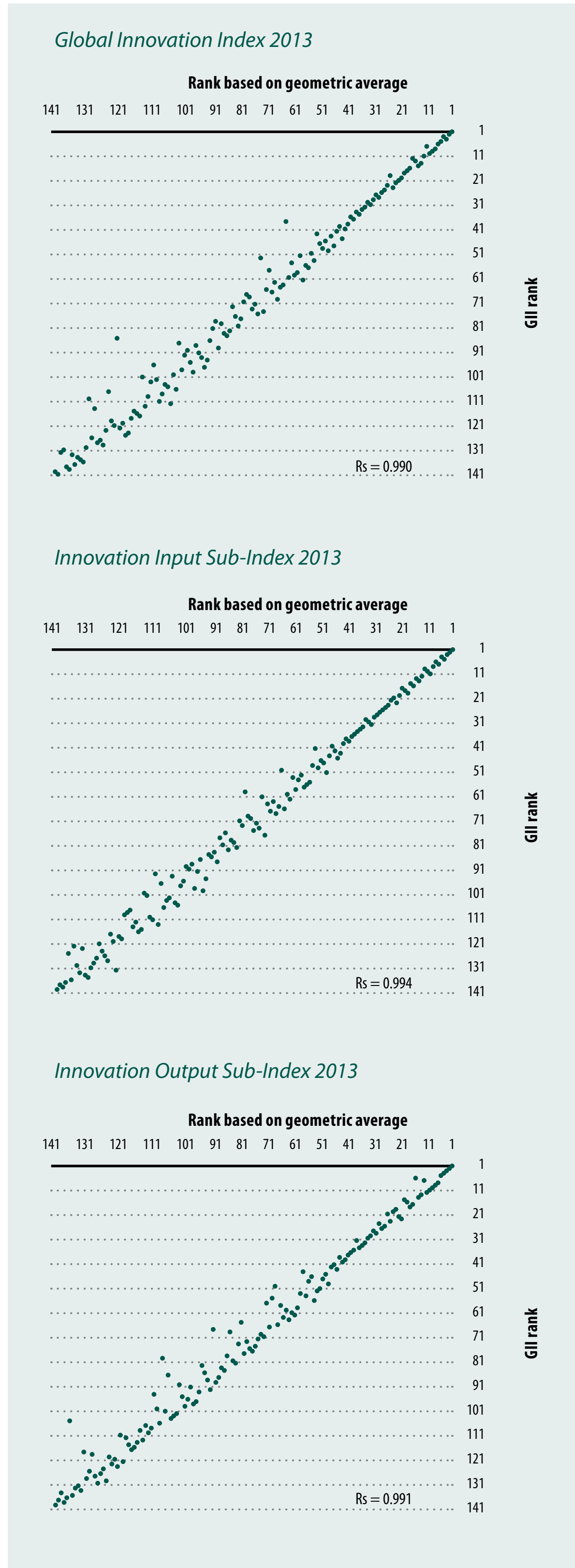


Figure 3b: Sensitivity analysis: Impact of modelling choices
(Geometric average)



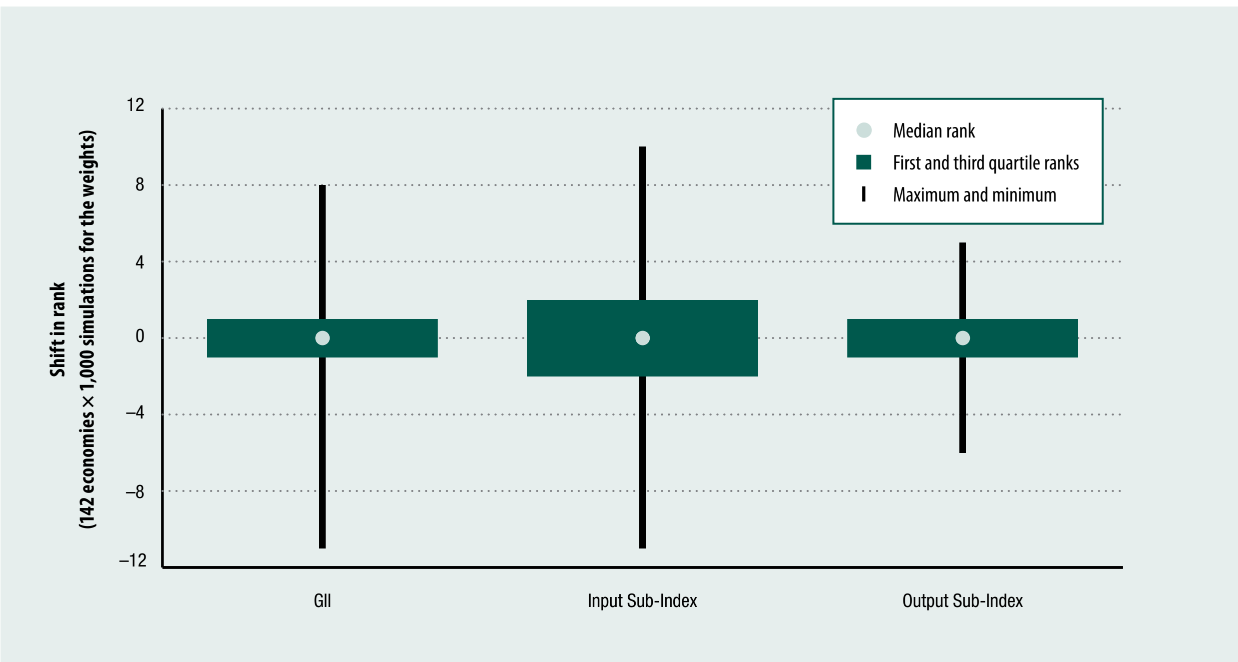
Source: Saisana and Philippas, European Commission Joint Research Centre, 2013.
Note: Rs = Spearman rank correlation; imputation based on expectation-maximization algorithm.

Table 4: Sensitivity analysis: Impact of modelling choices on countries with the most sensitive ranks

Index or Sub-Index	Uncertainty tested (pillar level only)	Number of countries that *improve* by 20 or more positions	Number of countries that *deteriorate* by 20 or more positions
GII	Geometric vs. arithmetic average	0	2
	EM imputation vs. no imputation of missing data	6	7
	Geometric average and EM imputation vs. arithmetic average and missing values	2	0
Input Sub-Index	Geometric vs. arithmetic average	0	0
	EM imputation vs. no imputation of missing data	1	0
	Geometric average and EM imputation vs. arithmetic average and missing values	0	0
Output Sub-Index	Geometric vs. arithmetic average	0	2
	EM imputation vs. no imputation of missing data	19	19
	Geometric average and EM imputation vs. arithmetic average and missing values	4	7

Source: Saisana and Philippas, European Commission Joint Research Centre, 2013.

Figure 4: Sensitivity analysis: Impact of random vs. fixed weights on the GII, Input, and Output Sub-Indices



Source: Saisana and Philippas, European Commission Joint Research Centre, 2013.

Table 5: Pie shares and distance to the efficient frontier: Top 10 economies in the GII 2013

Economy	DEA efficiency	Institutions	Human capital and research	Infrastructure	Market sophistication	Business sophistication	Knowledge and technology outputs	Creative outputs
Switzerland	1.00	0.06	0.18	0.11	0.08	0.19	0.19	0.19
Singapore	1.00	0.12	0.19	0.19	0.10	0.20	0.14	0.05
Hong Kong (China)	1.00	0.20	0.05	0.20	0.20	0.19	0.05	0.12
Sweden	1.00	0.20	0.20	0.20	0.17	0.05	0.13	0.05
United States of America	0.99	0.12	0.20	0.05	0.20	0.18	0.20	0.05
United Kingdom	0.99	0.20	0.20	0.20	0.20	0.05	0.06	0.09
Finland	0.98	0.20	0.20	0.20	0.05	0.11	0.19	0.05
Denmark	0.96	0.20	0.20	0.20	0.20	0.05	0.06	0.09
Ireland	0.95	0.20	0.20	0.05	0.20	0.12	0.18	0.05
Netherlands	0.95	0.20	0.12	0.20	0.05	0.20	0.05	0.18

Source: Saisana and Philippas, European Commission Joint Research Centre, 2013.
Note: The 10 economies that achieved the highest DEA scores are the same economies in the top 10 in the GII. Pie shares are in absolute terms, bounded by 0.05 and 0.20.

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 country-specific DEA weights, compared with the best performance among all other countries with those same weights. The DEA score can be interpreted as a measure of the 'distance to the efficient frontier'.

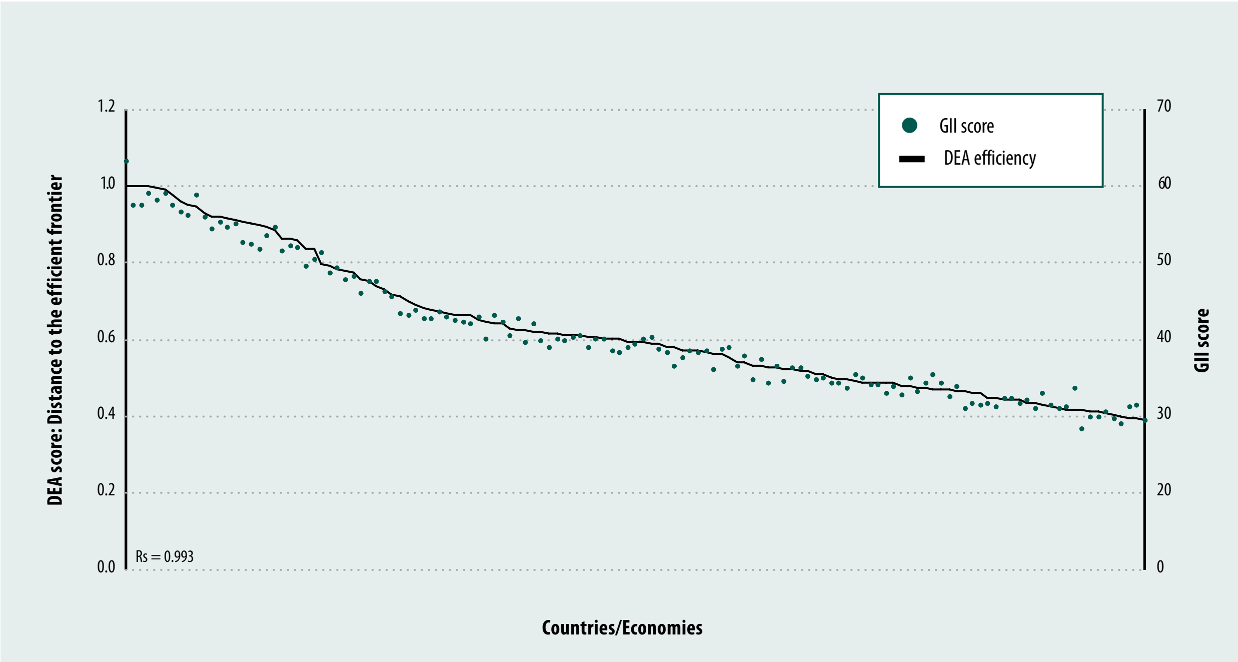
Table 5 presents the pie shares and DEA scores for the top 10 economies next to their GII scores. All pie shares are determined in accordance with a starting point that grants leeway to each country when assigning shares while not violating the (relative) upper and lower bounds. The pie shares are quite diverse, reflecting the different national innovation strategies. 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. Four of the top 10 economies assign the maximum allowed, 20%, to Institutions, Human capital and research, and Infrastructure. Four economies—Switzerland, Sweden, Hong Kong (China), and Singapore—reach a perfect DEA score of 1. Figure 5 shows how close the DEA scores and the GII 2013 scores are for all 142 economies (correlation of 0.993).[12]

Conclusion

The JRC analysis suggests that the conceptualized multi-level structure of the GII 2013 is statistically coherent and balanced (i.e., not dominated by any pillar or sub-pillar). Furthermore, the analysis has offered statistical justification for the weights and the use of arithmetic averaging at the various levels of aggregation. Together with other fine-tuning suggestions made in the sections above, a key recommendation for future years is to apply the data coverage criterion for countries' inclusion not at the overall GII level, as currently done, but within each of the two Innovation Sub-Indices. Furthermore, the 'no imputation' choice for not treating missing values, common in relevant contexts, as justified on grounds of transparency and replicability, can at times have undesirable impact on aggregate scores, with the additional negative side-effect that it may encourage countries not to report low data values. Finally, this year's choice of the GII team to use weights as scaling coefficients during the development of the index (as in the GII 2012) constitutes a significant departure from the traditional vision of weights as a reflection of indicators' importance in a weighted average. It is hoped that such a consideration will also be made by other developers of composite indicators. The 'distance to the efficient frontier' measure calculated with DEA scores could substitute for the Innovation Efficiency Ratio as a measure of efficiency, even if it is conceptually closer to the GII score than to the Efficiency Ratio.

Overall, the country/economy ranks of the GII and its sub-indices are fairly robust to methodological assumptions related to the estimation of missing data, weighting, and aggregation formula, without being redundant (four or fewer position shifts for 88 out of 142 countries).

Figure 5: GII 2013 scores and DEA 'distance to the efficient frontier' scores



Source: Saisana and Philippas, European Commission Joint Research Centre, 2013.

Consequently, inferences can be drawn for most economies in the GII, although some caution may be needed for a few. Note that perfect robustness would have been undesirable as this would have implied that the GII components are perfectly correlated and hence redundant, which is not the case for the GII 2013.

Notes

1 The JRC analysis was based on the recommendations of the OECD (2008) *Handbook on Composite Indicators,* and on more recent research from the JRC. The JRC auditing studies of composite indicators are available at http://composite-indicators.jrc.ec.europa.eu/; all audits were carried upon request of the index developers.

2 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 (142 countries).

3 When analyzing the statistical coherence of a framework, highly collinear indicators may dominate the aggregate scores. This problem is also taken care of by weights taken as 'scaling coefficients'. Only four cases of strong collinearity (i.e., Pearson correlation coefficients greater than ~ 0.92) were spotted within the same sub-pillar: 1.2.1 with 1.2.2, 3.1.1 with 3.1.2, 3.2.1 with 3.2.2, and 7.1.3 with 7.1.4. Indicators 1.2.1, 1.2.2, 3.2.1, and 3.2.2 were assigned half weights because of their high correlation with the sub-pillar score; while 3.1.1, 3.1.2, 7.1.3, and 7.1.4 were not treated, this was found not to bias the results of the respective sub-pillars 3.1 and 7.1.

4 Principal component analysis was applied to the GII dataset after treating pairs of highly collinear variables as a single indicator.

5 In GII 2012, the first principal component captured from 57% (Business sophistication) up to 80% (Institutions) of the total variance in the three underlying sub-pillars, while for the seventh pillar (Creative outputs) two principal components with eigenvalues greater than 1.0 were identified (in that case, the first component captured 56% of the variance of the three underlying sub-pillars).

6 Saisana, Saltelli, and Tarantola, 2005; Saisana et al., 2011.

7 The prior ranges are then rescaled to unity sum leading to posterior ranges of 5%–15% for the input pillar weights and 20%–30% for the output pillar weights. The ratio of the sum of the five Input pillar weights to the sum of the two pillar weights ranges between 0.77 and 1.39.

8 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.

9 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.

10 In the geometric average, pillars are multiplied as opposed to summed in the arithmetic average. Pillar weights appear as exponents in the multiplication. All pillar scores were greater than 1.0, so there was no reason to rescale them to avoid zero values that would have led to zero geometric averages.

11 The original question in the DEA-literature was how to measure each unit's relative efficiency in production compared to 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 the DEA-based distance to the efficient frontier scores, we consider the $m = 7$ pillars in the GII 2013 for $n = 142$ countries, with y_j the value of pillar j in country i. The objective is to combine the pillar scores per country into a single number, calculated as the weighted average of the m pillars, where w_i represents the weight of the ith pillar. In the 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 j:

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

Subject to

$$w_{ij} \geqq 0, \text{ where } j = 1, ..., 7, \; i = 1, ..., 142 \qquad \text{(non-negativity constraint)}$$

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

12 Of these, only Switzerland achieved a 1.0 score in the Innovation Efficiency Ratio, calculated as the ratio of the Output Sub-Index over the Input Sub-Index. The Efficiency Ratio and the DEA score embody very different concepts of efficiency, leading to completely different results and insights. A high score in the Innovation Efficiency Ratio is obtained by scoring higher on the Output Sub-Index than on the Input Sub-Index, irrespective of the actual scores in these two Sub-Indices. A high score in the DEA score can be obtained by having comparative advantages on several GII pillars (irrespective of these being input or output pillars). The DEA scores are therefore closer to the GII scores than to the Innovation Efficiency Ratio.

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.

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. doi: 0964–1998/13/176000

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.