Statistical Tests on the Global Innovation Index

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

The assessment of conceptual and statistical coherence of the Global Innovation Index (GII) and the estimation of the impact of modelling assumptions on a country's performance are necessary steps to ensure the transparency and reliability of the GII and enable policy makers to derive more accurate and meaningful conclusions and potentially guide choices on priority setting and policy formulation. Modelling the versatile concepts underlying innovation at national scale around the globe, as attempted in the 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 second consecutive year by INSEAD and the World Intellectual Property Organization (WIPO) to audit the GII along two main issues: the conceptual and statistical coherence of the structure, and the impact of key modelling assumptions on the GII 2012 scores and ranks.[1]

Conceptual and statistical coherence in the GII framework

An earlier version of the GII model was assessed by the JRC in March 2012. Fine-tuning suggestions were made and taken into account in the final version of the GII model. In this way, the development of the 2012 GII moved from a one-way design process to an iterative process with the JRC with a view to set the foundation for a balanced index. This section will consider these refinements and provide an additional assessment of the conceptual/statistical coherence in the final GII model. 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 (based on literature review and expert opinion) and timeliness. To represent a fair picture of country differences, indicators were scaled (by GDP, population, total goods, or others), as appropriate and where needed, either at the source or by the GII team.

Step 2: Data checks

The most recently released data were used for each country with a cut-off at year 2001. 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. These two criteria were jointly decided by the JRC and the GII team as suitable for the dataset already at hand from the GII 2011. Data values outside the 2.0 interquartile range[2] were checked for reporting errors. Potentially problematic indicators that could bias the overall results were identified as those having skewness (absolute) > 2 and kurtosis > 3.5[3] and were treated either by winsorisation (country values distorting the indicator distribution were assigned the next highest value, up to the level where skewness and kurtosis entered within the specified ranges) or by taking the natural logarithm (in case of more than five outliers).

Step 3: Statistical coherence

Only two cases of strong collinearity (i.e., Pearson correlation coefficients greater than ~ 0.92) were spotted within the same sub-pillar: these involved variables 1.2.1 with 1.2.2, Regulatory quality and Rule of Law; and finally 3.2.1 with 3.2.2 Electricity output and consumption.[4] This issue was dealt with by treating them as a single indicator (by assigning half weight to each normalized score). Besides these four variables, 17 more variables in the GII 2012 framework of 84 variables were assigned half weight in order to arrive at sub-pillar scores that were balanced in the underlying variables. For the same reason, two sub-pillars—7.2 and 7.3, Creative goods and services and Online creativity—were assigned half weight, while all other sub-pillars were assigned a weight of 1.0. These 0.5 or 1.0 weights were jointly decided between the JRC and the GII team, as scaling coefficients and not as importance coefficients. The aim was to attain a balance between

Figure 1: Conceptual and statistical coherence in the GII 2012 framework

Step 4. Qualitative review
- Internal qualitative review (INSEAD, WIPO)
- External qualitative review (JRC, international experts)

Step 3. Statistical coherence
- Assessment of grouping sub-pillars to pillars, to Sub-Indices, and to overall GII
- Use of weights as scaling coefficients to ensure statistical coherence
- Assessment of arithmetic average assumption

Step 2. Data check
- Availability requirements per country: coverage > 63% and at least two sub-pillars per pillar
- Check for reporting errors (interquartile range)
- Outlier treatment (skewness and kurtosis)

Step 1. Conceptual consistency
- Compatibility with exisiting literature on innovation and pillar definition
- Scaling factors per indicator to represent a fair picture of country differences (e.g., GDP, population, total exports)

Source: Saisana and Philippas, European Commission Joint Research Centre, 2012.

the contribution of variables to their respective sub-pillars and also a balance of the sub-pillars to their respective pillars. Paruolo et al. (2012) show that nominal weights in weighted arithmetic averages are not a measure of variable importance, although weights are assigned so as to reflect some stated target importance and they are communicated as such. In weighted averages, the ratio of two nominal weights gives the rate of substitutability between the two individual variables, and hence can be used to reveal the target relative importance of individual indicators. This target importance can then be compared with ex-post measures of variables' importance, such as the Karl Pearson's 'correlation ratio'.

Principal component analysis confirms the presence of a single latent dimension in the first six pillars (one component with an eigenvalue greater than 1.0) that captures between 57% (business sophistication) and 80% (institutions) of the total variance in the three underlying sub-pillars. For the seventh pillar (creative outputs), two principal components have eigenvalues greater than 1.0; nevertheless, the first component captures 56% of the variance of the three underlying sub-pillars. Further, results confirm the expectation that the sub-pillars are more correlated to their own pillar than to any other.

The five pillars in the Innovation Input Sub-index also share a single latent dimension that captures 80% of the total variance. The five loadings are very similar to each other, which suggests that building the Input Sub-index as a simple average (equal weights) of the five pillars is statistically supported by the data. This analysis could not be carried out on the Innovation Output Sub-index given that it is made of only two pillars[5]—Knowledge and technology outputs and Creative outputs, which are both correlated strongly with the Output Sub-index (Pearson correlation coefficients 0.92 and 0.90, respectively). This latter implies that also the Output Sub-index is well balanced in its two pillars.

Finally, building the GII as the simple average of the Input and Output Sub-index 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 conceptual grouping of sub-pillars into pillars, sub-indices, and in an overall GII is statistically coherent, has a balanced structure (i.e., not dominated by any pillar or sub-pillar), and

Table 1: 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, 2012.

gives further justification for the use of simple averages at the various levels of aggregation.

Step 4: Qualitative review

Finally, the GII results, including overall country classification and relative performance in terms of Innovation Input or Output, were evaluated by the GII team and the JRC to verify that the overall results are, to a great extent, consistent with current evidence, existing research or prevailing theory.

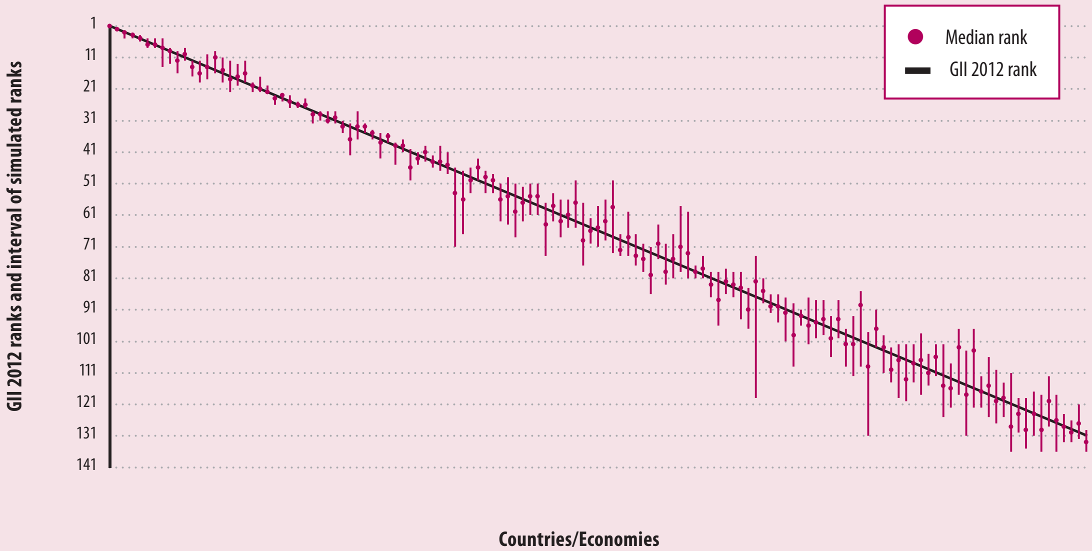
Notwithstanding these statistical tests and the positive outcomes on the statistical coherence of the GII structure, it is important to mention that the GII model is and has to be open for future improvements as better data, more comprehensive surveys and assessments, and new relevant research studies become available.

Impact of modelling assumptions on the GII results

Every country score on the overall GII and its two Innovation Sub-Indices depends on choices: the seven-pillar structure, the selected variables, the estimation or not of missing data, the normalization of the variables, the weights assigned to them, and the aggregation method, among other elements. Some of these choices are based on the opinion of experts in the field (e.g., the selection of variables and equal weights within pillars) or common practice (e.g., min-max method to normalize the variables in 0 to 100 scale), driven by statistical analysis (e.g., treating outliers) or simplicity (e.g., no estimation of missing data). The aim of the robustness analysis is to assess to what extent these choices might impact the GII results. We have dealt with these uncertainties in order to check their simultaneous and joint influence with a view to fully acknowledging their implications. In the present analysis, the data are assumed to be error-free since INSEAD already undertook a double-check control of potential outliers and eventual errors and typos were corrected during this phase (see Step 2 in Figure 1).

The robustness assessment of the GII was based on the combination of a Monte Carlo experiment and a multi-modelling approach. 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 has been driven by two opposite needs: on the one hand, to ensure a wide enough interval to have meaningful robustness checks; on the other hand, to respect the rationale of the GII that the Input Sub-Index (five pillars) and the Output Sub-Index (two pillars) are placed on equal footing when building the overall GII. Given these considerations, limit values of uncertainty intervals have been defined as shown in Table 1. The multi-modelling approach involved combinations of the remaining two key assumptions on the 'no imputation' of missing data and the aggregation formula at the pillar level. The GII developing team, for transparency and replicability, opted not to estimate missing

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



Source: Saisana and Philippas, European Commission Joint Research Centre, 2012.

Note: The Spearman rank correlation between the median rank and the GII 2012 rank is 0.996. 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.

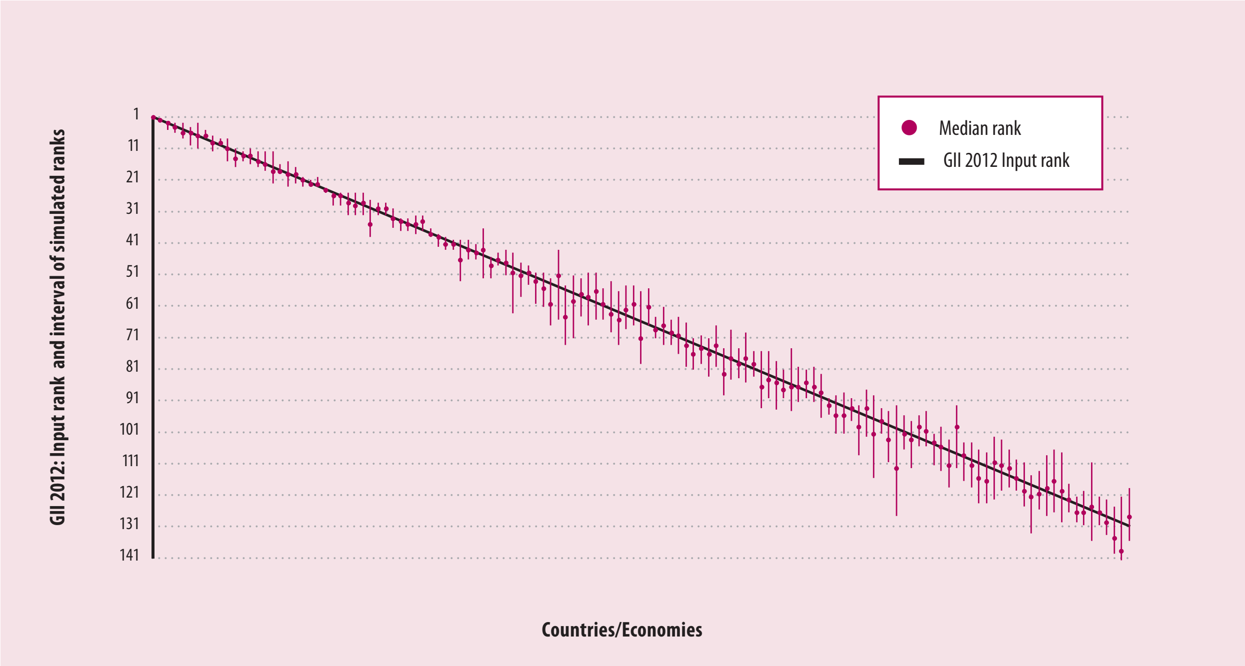
data and instead calculated sub-pillar and pillar scores using only available information per country. The "no imputation" choice, which is common in relevant contexts, might discourage countries from reporting low data values.[7] To overcome this limitation, we opted to use the Expectation Maximization (EM) algorithm.[8] Regarding the GII assumption on the aggregation function (arithmetic average), and despite that it received statistical support in the previous section, decision-theory practitioners have challenged this type of aggregation because of inherent theoretical inconsistencies and because of the fully compensatory nature, in which a comparative high advantage on few variables can compensate a comparative disadvantage on many variables.[9] Hence, we considered the geometric average instead,[10] which is a partially compensatory approach and further 'motivates' countries to improve in the dimensions of innovation where they perform poorly, as opposed to in any dimension (which is instead done under the arithmetic average).

Consequently, we tested four models based on the combination of 'no imputation' versus EM, or arithmetic versus geometric average. Combined with the 1,000 simulations per model to account for the uncertainty in the weights at the pillar level, we carried out altogether 4,000 simulations for the GII, and an equal number of simulations for either the Innovation Input or the Innovation Output Sub-index (see Table 1 for a summary of the uncertainties considered in the GII 2012).

Uncertainty analysis results

The main results of the robustness analysis are shown in Figure 2 with median ranks and intervals computed across the 4,000 Monte Carlo simulations for the overall GII, and the two Innovation 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. GII ranks are rather robust: the median rank is close to the reference rank (less than four positions for 75% of the countries). Results for the Input Sub-Index are more robust (75% of the countries shift less than 3 positions), while the Output Sub-Index is more sensitive to the methodological choices (75% of the countries shift less than 6 positions).The fact that the Output Sub-Index is more sensitive to methodological changes is twofold: there are only two pillars and they are only moderately associated to each other (Pearson correlation coefficient: 0.65). However, it cannot be ruled out altogether that the correlation

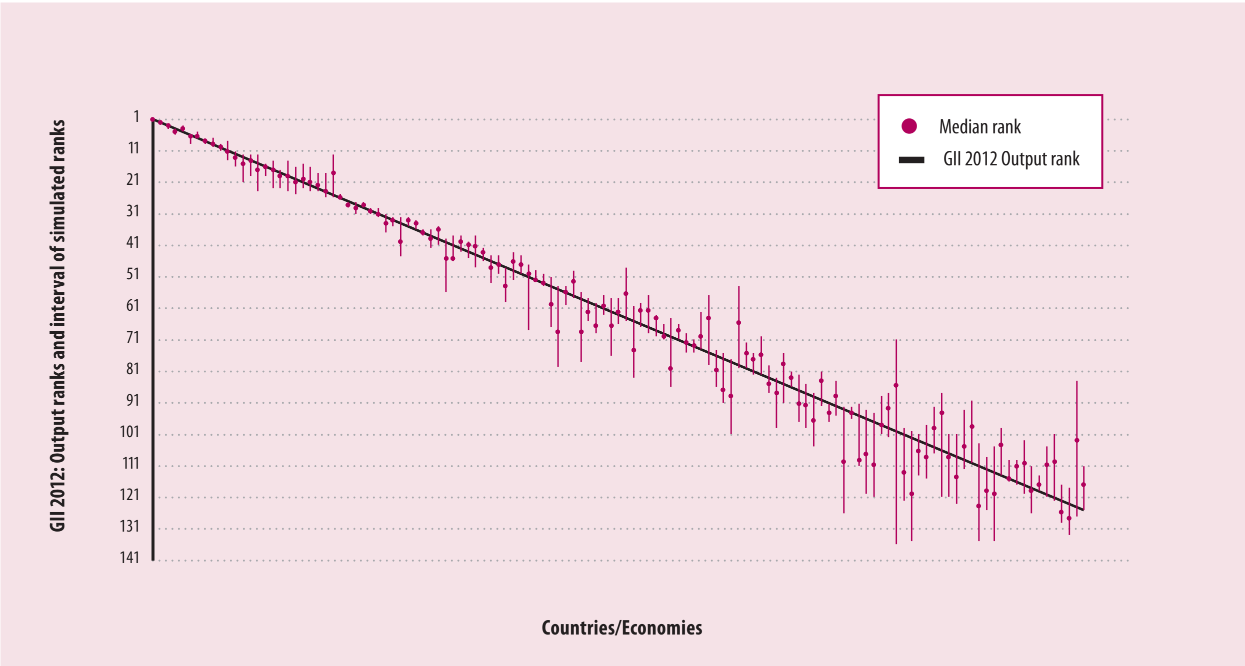
Figure 2b: Robustness analysis (Input rank vs. median rank, 90% confidence intervals)



Source: Saisana and Philippas, European Commission Joint Research Centre, 2012.

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.

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



Source: Saisana and Philippas, European Commission Joint Research Centre, 2012.

Note: The Spearman rank correlation between the median rank and the Output rank is 0.988. 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 2: GII 2012 and Input and Output Sub-Indices: Ranks and 90% confidence intervals

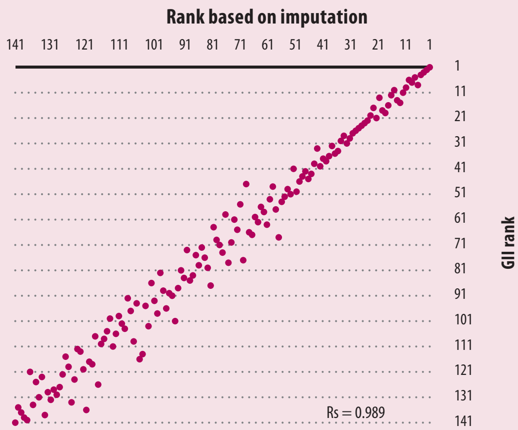
	GII 2012		Input Sub-Index		Output Sub-Index	
Country/Economy	Rank	Interval	Rank	Interval	Rank	Interval
Switzerland	1	[1, 1]	4	[3, 6]	1	[1, 1]
Sweden	2	[2, 2]	3	[3, 5]	2	[2, 2]
Singapore	3	[3, 5]	1	[1, 1]	11	[8, 14]
Finland	4	[3, 5]	6	[4, 10]	5	[3, 5]
United Kingdom	5	[4, 6]	5	[3, 8]	6	[6, 9]
Netherlands	6	[5, 8]	15	[12, 17]	3	[3, 4]
Denmark	7	[5, 8]	8	[5, 8]	9	[7, 10]
Hong Kong SAR, China	8	[5, 14]	2	[2, 2]	25	[12, 26]
Ireland	9	[8, 13]	7	[3, 11]	14	[12, 19]
United States	10	[9, 16]	9	[7, 12]	16	[15, 19]
Luxembourg	11	[8, 12]	14	[11, 16]	10	[9, 11]
Canada	12	[12, 17]	10	[8, 10]	20	[16, 25]
New Zealand	13	[12, 19]	12	[11, 17]	15	[12, 24]
Norway	14	[10, 18]	11	[8, 15]	17	[14, 23]
Germany	15	[9, 16]	23	[20, 23]	7	[5, 7]
Malta	16	[11, 19]	27	[25, 32]	4	[4, 6]
Israel	17	[12, 22]	17	[12, 22]	13	[12, 21]
Iceland	18	[13, 20]	19	[15, 23]	12	[11, 16]
Estonia	19	[12, 19]	24	[24, 25]	8	[8, 9]
Belgium	20	[19, 22]	20	[17, 22]	18	[17, 23]
Korea, Rep.	21	[17, 22]	16	[12, 18]	24	[18, 26]
Austria	22	[20, 22]	21	[20, 23]	21	[15, 23]
Australia	23	[23, 26]	13	[12, 15]	31	[29, 32]
France	24	[23, 25]	22	[21, 23]	26	[25, 26]
Japan	25	[23, 27]	18	[16, 19]	28	[27, 31]
Slovenia	26	[25, 27]	32	[28, 32]	22	[16, 24]
Czech Republic	27	[24, 27]	31	[28, 32]	23	[18, 24]
Cyprus	28	[28, 32]	25	[25, 29]	32	[31, 37]
Spain	29	[28, 31]	26	[25, 29]	35	[32, 35]
Latvia	30	[28, 32]	36	[32, 38]	27	[27, 29]
Hungary	31	[28, 32]	37	[32, 37]	29	[27, 29]
Malaysia	32	[31, 35]	29	[25, 32]	38	[36, 42]
Qatar	33	[32, 42]	30	[27, 39]	41	[38, 46]
China	34	[28, 37]	55	[43, 65]	19	[14, 24]
Portugal	35	[32, 35]	33	[30, 36]	33	[32, 35]
Italy	36	[34, 37]	34	[33, 37]	39	[35, 41]
United Arab Emirates	37	[35, 43]	28	[25, 32]	51	[47, 68]
Lithuania	38	[35, 38]	38	[36, 39]	37	[36, 38]
Chile	39	[38, 45]	43	[40, 46]	34	[32, 45]
Slovakia	40	[37, 41]	40	[39, 43]	43	[40, 45]
Bahrain	41	[40, 50]	35	[33, 37]	60	[59, 69]
Croatia	42	[41, 45]	44	[41, 46]	45	[42, 46]
Bulgaria	43	[39, 44]	47	[44, 48]	42	[38, 43]
Poland	44	[42, 46]	41	[40, 43]	50	[44, 50]
Montenegro	45	[39, 47]	48	[43, 51]	44	[38, 48]
Serbia	46	[41, 48]	65	[54, 67]	36	[33, 36]
Oman	47	[46, 71]	42	[40, 53]	55	[54, 80]
Saudi Arabia	48	[47, 67]	39	[38, 42]	70	[64, 86]
Mauritius	49	[46, 54]	49	[44, 63]	48	[48, 59]
Moldova	50	[43, 50]	79	[74, 84]	30	[29, 31]
Russian Federation	51	[47, 54]	60	[50, 65]	49	[43, 52]
Romania	52	[48, 54]	51	[48, 54]	57	[49, 58]
Brunei Darussalam	53	[51, 63]	46	[45, 52]	69	[66, 71]
South Africa	54	[49, 64]	45	[36, 52]	73	[71, 75]
Kuwait	55	[51, 68]	61	[55, 65]	54	[51, 67]
Jordan	56	[52, 63]	72	[66, 80]	46	[44, 53]
Thailand	57	[51, 61]	59	[50, 67]	56	[54, 60]
Brazil	58	[51, 61]	69	[61, 73]	52	[49, 52]
Tunisia	59	[57, 74]	64	[54, 68]	58	[56, 78]
Costa Rica	60	[54, 63]	71	[64, 76]	53	[50, 54]
Lebanon	61	[56, 68]	62	[53, 69]	63	[58, 66]
Macedonia, FYR	62	[56, 65]	52	[50, 60]	71	[66, 71]
Ukraine	63	[50, 65]	78	[68, 84]	47	[44, 50]
India	64	[57, 77]	96	[87, 102]	40	[39, 56]
Colombia	65	[62, 70]	58	[52, 64]	72	[69, 75]
Greece	66	[58, 71]	50	[47, 58]	82	[70, 82]
Uruguay	67	[56, 69]	68	[64, 71]	67	[57, 69]
Mongolia	68	[50, 73]	53	[50, 61]	79	[54, 80]
Armenia	69	[67, 74]	73	[71, 81]	68	[63, 70]
Argentina	70	[60, 74]	76	[67, 80]	66	[59, 67]
Georgia	71	[67, 77]	63	[56, 73]	81	[75, 82]

Table 2: GII 2012 and Input and Output Sub-Indices: Ranks and 90% confidence intervals *(cont'd.)*

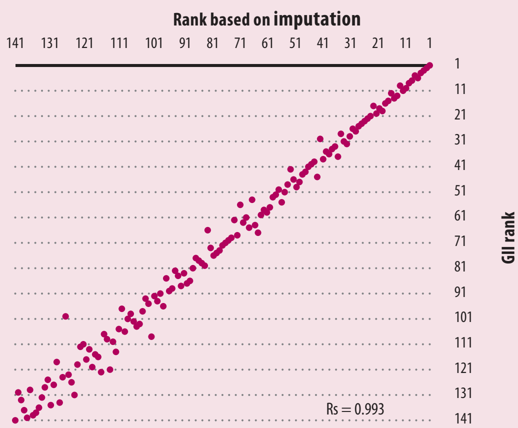
	GII 2012		Input Sub-Index		Output Sub-Index	
Country/Economy	Rank	Interval	Rank	Interval	Rank	Interval
Bosnia and Herzegovina	72	[70, 79]	66	[56, 79]	80	[72, 80]
Namibia	73	[71, 86]	56	[54, 73]	87	[82, 97]
Turkey	74	[64, 75]	81	[75, 83]	61	[57, 63]
Peru	75	[70, 83]	57	[51, 71]	88	[85, 99]
Vietnam	76	[67, 81]	83	[75, 90]	59	[58, 65]
Guyana	77	[58, 79]	86	[74, 94]	64	[48, 65]
Belarus	78	[60, 81]	80	[69, 85]	75	[57, 79]
Mexico	79	[77, 81]	70	[65, 73]	86	[81, 86]
Belize	80	[74, 81]	87	[80, 91]	74	[62, 74]
Trinidad and Tobago	81	[79, 87]	74	[70, 79]	84	[83, 99]
Swaziland	82	[79, 96]	99	[94, 110]	65	[60, 83]
Kazakhstan	83	[78, 86]	67	[55, 67]	105	[92, 107]
Paraguay	84	[79, 87]	103	[95, 105]	62	[58, 76]
Botswana	85	[79, 94]	54	[52, 67]	121	[101, 122]
Dominican Republic	86	[84, 97]	93	[88, 101]	77	[75, 91]
Panama	87	[74, 119]	75	[71, 83]	100	[71, 136]
Morocco	88	[81, 89]	88	[81, 90]	90	[81, 92]
Azerbaijan	89	[86, 92]	85	[81, 90]	94	[92, 96]
Albania	90	[86, 95]	82	[75, 93]	98	[89, 101]
Jamaica	91	[87, 101]	77	[74, 89]	107	[101, 121]
Ghana	92	[89, 109]	91	[90, 95]	93	[92, 126]
El Salvador	93	[91, 96]	94	[90, 97]	91	[91, 97]
Sri Lanka	94	[87, 102]	115	[106, 118]	76	[74, 86]
Philippines	95	[88, 100]	106	[99, 113]	83	[79, 88]
Kenya	96	[88, 99]	89	[80, 91]	114	[99, 114]
Senegal	97	[93, 106]	114	[103, 119]	78	[77, 101]
Ecuador	98	[88, 100]	109	[104, 116]	85	[75, 91]
Guatemala	99	[97, 109]	98	[93, 101]	101	[99, 122]
Indonesia	100	[93, 112]	113	[102, 122]	89	[88, 105]
Fiji	101	[85, 109]	84	[78, 94]	124	[84, 127]
Rwanda	102	[98, 131]	95	[92, 108]	113	[105, 135]
Egypt	103	[91, 103]	104	[94, 105]	99	[88, 102]
Iran	104	[99, 111]	97	[89, 115]	117	[103, 120]
Nicaragua	105	[103, 114]	102	[97, 112]	119	[114, 119]
Gabon	106	[102, 119]	112	[107, 123]	106	[88, 121]
Zambia	107	[102, 120]	122	[107, 129]	96	[93, 120]
Tajikistan	108	[102, 114]	111	[106, 123]	109	[93, 112]
Kyrgyzstan	109	[99, 118]	90	[82, 96]	131	[127, 132]
Mozambique	110	[105, 115]	107	[103, 118]	115	[109, 116]
Honduras	111	[102, 112]	105	[101, 111]	116	[109, 117]
Bangladesh	112	[102, 125]	118	[114, 133]	104	[98, 115]
Nepal	113	[108, 122]	127	[121, 130]	95	[91, 111]
Bolivia	114	[97, 118]	108	[92, 112]	120	[105, 121]
Zimbabwe	115	[104, 131]	130	[121, 141]	92	[84, 95]
Lesotho	116	[97, 122]	92	[89, 101]	133	[107, 133]
Uganda	117	[112, 122]	121	[106, 126]	112	[108, 125]
Venezuela	118	[106, 125]	126	[110, 135]	103	[101, 114]
Mali	119	[110, 127]	131	[118, 135]	97	[94, 121]
Malawi	120	[114, 125]	110	[104, 118]	122	[117, 129]
Cameroon	121	[111, 136]	125	[119, 130]	111	[104, 135]
Burkina Faso	122	[119, 130]	120	[108, 127]	123	[118, 133]
Nigeria	123	[119, 135]	134	[120, 138]	102	[100, 135]
Algeria	124	[117, 131]	101	[95, 106]	134	[133, 137]
Benin	125	[118, 136]	132	[128, 140]	108	[101, 123]
Madagascar	126	[112, 128]	116	[109, 119]	126	[116, 129]
Uzbekistan	127	[118, 136]	100	[92, 127]	137	[119, 138]
Tanzania	128	[124, 133]	117	[110, 124]	129	[124, 137]
Cambodia	129	[126, 133]	119	[113, 125]	132	[130, 132]
Gambia, The	130	[121, 132]	128	[122, 133]	125	[111, 125]
Ethiopia	131	[129, 136]	124	[121, 129]	128	[126, 135]
Syria	132	[124, 133]	123	[117, 126]	130	[126, 130]
Pakistan	133	[115, 134]	140	[134, 141]	110	[90, 110]
Côte d'Ivoire	134	[124, 136]	139	[131, 139]	118	[111, 126]
Angola	135	[132, 141]	133	[128, 139]	127	[123, 141]
Togo	136	[112, 138]	v135	[128, 139]	136	[99, 136]
Burundi	137	[135, 139]	137	[131, 140]	135	[132, 139]
Lao PDR	138	[125, 139]	129	[124, 139]	139	[113, 139]
Yemen	139	[137, 140]	138	[129, 139]	138	[137, 139]
Niger	140	[137, 140]	136	[132, 139]	140	[131, 140]
Sudan	141	[140, 141]	141	[139, 141]	141	[140, 141]

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

Global Innovation Index 2012



Innovation Input Sub-Index 2012



Innovation Output Sub-Index 2012

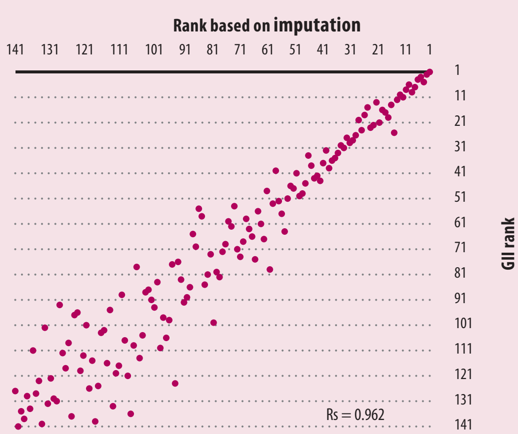
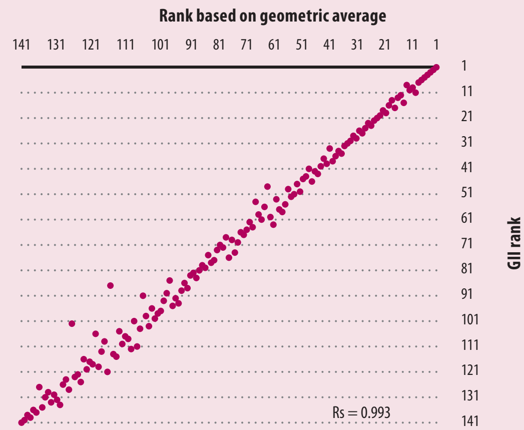
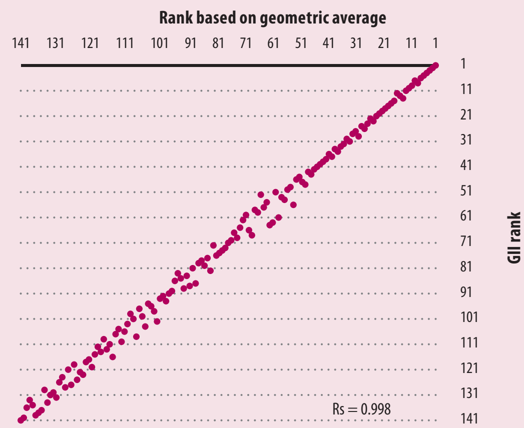


Figure 3b: Sensitivity analysis: Impact of modelling choices
(Geometric average)

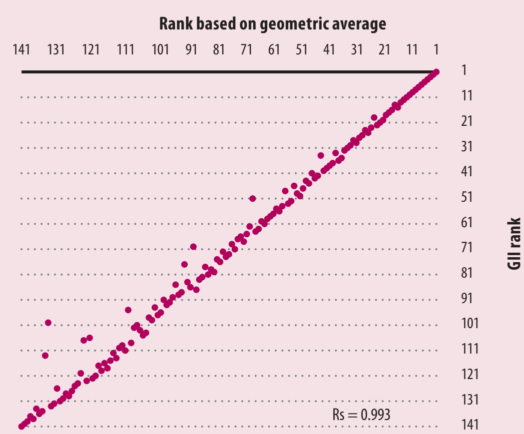
Global Innovation Index 2012



Innovation Input Sub-Index 2012



Innovation Output Sub-Index 2012



Source: Saisana and Philippas, European Commission Joint Research Centre, 2012.
Note: Rs = Spearman rank correlation; imputation based on expectation-maximization algorithm.

could improve as data become available, as suggested by theory. In fact, between 2011 and 2012 the association between these two output pillars increased from 0.51 to 0.65.The currently observed moderate correlation might be caused by (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, particularly on 7.2 (expenditure on recreation and culture, exports of creative goods and services as proxies for creative outputs). For an in depth discussion of these results, the reader is referred to Saisana and Philippas, 2012.

For transparency, Table 2 reports the original country ranks and the 90% interval for the simulated rank for the GII, the Input Sub-Index, and the Output Sub-Index. Our intention is to be explicit about on which countries the simulated interval either does not include the reference rank or is too wide to allow for a reasonable inference. Overall, all country ranks in the GII or any of the Innovation Sub-Indices lay within the simulated intervals. Simulated intervals are narrow enough for most countries (less than 10 positions) to allow for meaningful inferences to be drawn.

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. Detailed results are available in the main JRC assessment report, but the main conclusion is that the impact of the imputation alone is noteworthy for some countries, although it may be moderated when considering a geometric aggregation and a variation in the weights for the pillars. Figure 3 plots the reference GII ranks (and the two sub-indices) versus one-at-a-time changes of either the imputation method or the aggregation formula.

These plots show that the most influential assumption is the choice of no imputation versus EM imputation in particular for the Output Sub-Index, then for the GII and least for the Input Sub-index. For example, in one case a country does not shift position if a geometric aggregation is applied, although it is found to lose 24 positions in the Output ranking if EM imputation is applied. If both assumptions are changed (and weights remain at the reference values), the impact of the imputation would be moderated. This sensitivity is the result of data availability. Although all countries have data coverage above 70% in the Input variables, 21 countries have data coverage below 65% in the Output variables, which explains the impact of imputation on these countries ranks. Sensitivity analysis, by assessing the impact of the modelling choices, has given more transparency in the entire process and can help to appreciate the GII results with respect to the assumptions made during the development phase. Sensitive ranks usually concern countries with poor data coverage on the Innovation Output Sub-Index, and to a more limited extent on the Innovation Input Sub-Index—an impact that propagates to the estimation of the overall GII. For an in depth discussion of these results, the reader is referred to Saisana and Philippas, 2012.

The recommendation for the future would be to apply the 63% criterion for data availability within each of the two sub-indices so as to avoid drawing a better picture for countries with poor data quality on one of the two sub-indices, in particular on the Innovation Output Sub-Index. For this year, drawing upon the analysis made by the JRC, the recommendation is to consider country ranks in the GII 2012 and in the Input and Output Sub-Indices not only at face value but also within the ranges simulated by uncertainty analysis in order to better appreciate to what degree a country rank depends on the methodological choices made during the development of the GII 2012.

Conclusion

The JRC analysis suggests that the conceptualized multi-level structure of the GII 2012 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 is 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 and justified on grounds of transparency and replicability, can at times have an undesirable impact on aggregate scores, with the additional negative side-effect that it may discourage countries from reporting low data values. Finally, the GII team's choice this year to use weights as scaling coefficients during the development of the index constitutes a significant departure from the traditional vision of weights as a reflection of indicators' importance in a weighted average. Such a consideration will hopefully be made also by other developers of composite indicators.

Overall, GII country ranks are in most cases fairly robust (less than three positions shift for 94 out of 141 countries) to methodological assumptions related to the estimation of missing data, weighting and aggregation formula. Consequently, inferences can be drawn for most countries in the GII, although some caution may be needed for a few countries. Note that perfect robustness would have been undesirable because this would have implied that the GII components are perfectly correlated and hence redundant. The JRC analysis suggests that the GII 2012 and its Innovation Input and Output Sub-Indices are fairly robust to the methodological choices without being redundant.

Notes

1 The JRC analysis was based on the recommendations of the OECD/EC JRC *Handbook on Constructing Composite Indicators* (2008) 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 The 'interquartile range' is the difference between the upper (75% of values) and the lower (25% of values) quartiles.

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 (130 countries).

4 High collinearity can be problematic when analysing the statistical coherence of a framework and may result in aggregate scores that are dominated by the highly collinear indicators.

5 Principal Components Analysis requires at three least pillars (variables in general).

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

7 Note that here 'no imputation' is equivalent to replacing missing values with the average of the available data within each sub-pillar.

8 The Expectation-Maximization (EM) algorithm 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. See Little and Rubin, 2002.

9 Munda, 2008.

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, hence there was no reason to rescale them to avoid zero values that would have led to zero geometric averages.

References

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. Hobboken, NJ: John Wiley & Sons.

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

OECD/EC JRC. 2008. *Handbook on Constructing Composite Indicators: Methodology and User Guide*. Paris: OECD.

Paruolo, P., M. Saisana, and A. Saltelli. 2012. 'Ratings and Rankings: Voodoo or Science?' *Journal of the Royal Statistical Society – A* (in print). A draft version is available at http://arxiv.org/abs/1104.3009.

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.