Joint Research Centre Statistical Audit of the 2015 Global Innovation Index

Michaela Saisana and **Marcos Domínguez-Torreiro,** European Commission, Joint Research Centre (JRC), Ispra, Italy

Conceptual and practical challenges are inevitable when trying to understand and model the fundamentals of innovation at the national level worldwide. The Global Innovation Index (GII) considers these conceptual challenges in Chapter 1 and deals with practical challenges—related to data quality and methodological choices—by grouping country-level data into 21 sub-pillars, 7 pillars, 2 sub-indices, and, finally, an overall index. The object of this annex is to offer a detailed insight into the practical issues related to the construction of the index, analysing in-depth the statistical soundness of the calculations and assumptions made to arrive at the final index rankings. Notwithstanding, statistical soundness should be regarded as a necessary but not sufficient condition for a sound GII, since the correlations underpinning the majority of the statistical analyses carried out herein 'need not necessarily represent the real influence of the individual indicators on the phenomenon being measured'.[1] Consequently, the development of the GII must be nurtured by a dynamic iterative dialogue between the principles of statistical and conceptual soundness or, to put it another way, between the theoretical understanding of innovation and the empirical observations of the data underlying the variables.

The Econometrics and Applied Statistics at the European Commission Joint Research Centre (JRC) in Ispra has been invited for the fifth consecutive year to audit the GII. As in previous editions, the present JRC audit will focus on the statistical soundness of the multi-level structure of the index as well as on the impact of key modelling assumptions on the results.[2] The external qualitative check provided by the JRC guarantees the transparency and reliability of the index for both policy makers and other stakeholders, thus facilitating more accurate priority setting and policy formulation in this particular field.

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 the added value of 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-May 2015. Fine-tuning suggestions were taken into account in the final computation of the rankings in an iterative process with the JRC aimed at setting the foundation for a balanced index. The entire process followed four steps (see Figure 1).

Step 1: Conceptual consistency

Seventy-nine 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 within the period 2004–14 were used for each economy. Almost 75% of the available data refer to 2013 or more recent years. Countries were included if data availability was at least 60% (i.e., 47 out of 79 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 case of more than five outliers). These criteria were decided jointly with the JRC back in 2011 (see Appendix IV Technical Notes in this report 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 in 2012 to be scaling coefficients and not importance

Figure 1: Conceptual and statistical coherence in the GII 2015 framework

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

Step 3. Statistical coherence
- Treatment of highly collinear variables as a single indicator
- Assessment of grouping sub-pillars to pillars, to sub-indices, and to GII
- Use of weights as scaling coefficients to ensure statistical coherence
- Assessment of arithmetic average assumption
- Assessment of potential redundancy of information in the overall GII

Step 2. Data checks
- Check for data recency (almost 75% of available data refer to 2013–2015)
- Availability requirements per country: coverage > 60% and at least two sub-pillars per pillar
- Check for reporting errors (interquartile range)
- Outlier treatment (skewness and kurtosis)
- Direct contact with data providers

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

Source: Saisana and Domínguez-Torreiro, European Commission Joint Research Centre, 2015.

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 79 indicators and two sub-pillars—7.2 Creative goods and services and 7.3 Creation of online content—were assigned half weight while all other indicators and sub-pillars were assigned a weight of 1.0. Nevertheless, for five indicators with Pearson correlation coefficients that have respective sub-pillar scores below 0.5, some further reflection is needed because they seem to behave as 'noise' at all aggregation levels in the GII framework. This applies to 5.2.3 GERD financed by abroad; 5.3.4 Foreign direct investment, net inflows; 6.2.1 Growth rate of GDP per person engaged; 6.2.2 New business density; and 7.2.4 Printing and publishing output.

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 61% (pillar 4: Market sophistication) up to 85% (pillar 1: Institutions) of the total variance in the three underlying sub-pillars. These results reveal that the adjustments made to the 2015 GII framework have further improved the already good 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 pillar and that all coefficients are greater than 0.75 (see Table 1).

The five input pillars share a single statistical dimension that

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	0.94	0.75	0.79	0.69	0.75	0.67	0.80
	Regulatory environment	0.92	0.64	0.65	0.64	0.63	0.59	0.71
	Business environment	0.89	0.73	0.76	0.74	0.71	0.69	0.76
	Education	0.64	0.82	0.63	0.52	0.67	0.64	0.67
	Tertiary education	0.56	0.79	0.60	0.39	0.45	0.51	0.59
	Research and development (R&D)	0.71	0.89	0.79	0.70	0.80	0.87	0.80
	Information and communication technologies (ICTs)	0.76	0.80	0.95	0.70	0.73	0.75	0.80
	General infrastructure	0.60	0.61	0.78	0.54	0.55	0.58	0.59
	Ecological sustainability	0.67	0.68	0.83	0.65	0.65	0.69	0.75
	Credit	0.44	0.33	0.40	0.75	0.46	0.40	0.45
	Investment	0.67	0.67	0.68	0.80	0.70	0.71	0.68
	Trade and competition	0.60	0.53	0.62	0.79	0.60	0.59	0.71
	Knowledge workers	0.62	0.75	0.72	0.65	0.86	0.77	0.74
	Innovation linkages	0.72	0.65	0.65	0.70	0.76	0.67	0.74
	Knowledge absorption	0.48	0.44	0.41	0.47	0.79	0.57	0.51
OUTPUT	Knowledge creation	0.58	0.78	0.64	0.57	0.69	0.85	0.69
	Knowledge impact	0.53	0.67	0.72	0.63	0.63	0.83	0.69
	Knowledge diffusion	0.64	0.62	0.61	0.65	0.81	0.83	0.71
	Intangible assets	0.68	0.66	0.70	0.73	0.72	0.71	0.90
	Creative goods and services	0.71	0.77	0.76	0.72	0.76	0.76	0.87
	Online creativity	0.81	0.85	0.81	0.69	0.78	0.77	0.90

Source: Saisana and Dominguez-Torreiro, European Commission Joint Research Centre, 2015.

summarizes 81% of the total variance, and the five loadings (correlation coefficients) of these pillars are very similar to each other. 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 0.95—well above the 0.70 threshold for a reliable aggregate.[4]

The two output pillars—Knowledge and technology outputs and Creative outputs—are strongly correlated to each other (0.83); they are also both strongly correlated with the Innovation Output Sub-index (0.95). 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 Sub-Index and Output Sub-Index is also statistically justifiable because the Pearson correlation coefficient of either sub-index with the overall GII is 0.98; the two sub-indices have a correlation of 0.93. Thus far, results show that the grouping of sub-pillars into pillars, sub-indices, and the GII 2015 is statistically coherent, and that the GII has a balanced structure at each aggregation level.

Added value of 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 high statistical reliability of the five input pillars—may be interpreted by some as a sign of redundancy of information in the GII. However, this is not the case here. In fact, for more than 50.4% (up to 69.5%) of the 141 economies included in the GII 2015, 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 aspects of innovation that do not emerge directly by looking into the seven pillars separately. At the same time, this result points to the value of duly taking into account the GII pillars, sub-pillars, and individual indicators on their own merit. By doing so, country-specific strengths and bottlenecks on innovation can be identified and serve as an input for evidence-based policymaking.

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 30	15.6	14.2	12.1	27.0	25.5	17.7	7.8
20–29	12.1	17.0	15.6	14.2	10.6	14.9	12.8
10–19	24.1	22.7	26.2	28.4	24.8	27.0	29.8
10 or more*	**51.8**	**53.9**	**53.9**	**69.5**	**61.0**	**59.6**	**50.4**
5–9	25.5	25.5	22.7	14.9	17.7	17.0	22.0
Less than 5	19.1	19.9	20.6	14.2	19.1	22.0	24.8
Same rank	3.5	0.7	2.8	1.4	2.1	1.4	2.8
Total†	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**

Source: Saisana and Dominguez-Torreiro, European Commission Joint Research Centre, 2015.
* This column is the sum of the prior three rows.
† This column is the sum of all white rows.

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 are, 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, 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.

The impact of modelling assumptions on the GII results

Setting up an underlying structure for the index based on a battery of pillars; choosing the individual variables to be used as indicators; deciding whether or not to impute missing data; selecting the normalization approach to be applied, the weights to be assigned, the rule of aggregation to be implemented, and other elements of the index are all modelling assumptions with a direct impact on the GII scores and rankings. The rationale for these choices is manifold. For instance, expert opinion is behind the selection of the individual indicators, common practice suggests the use of a min-max normalization approach in the [0–100] range, the treatment of outliers is driven by statistical analysis, and simplicity and parsimony criteria seem to advocate for not imputing missing data. The unavoidable uncertainty stemming from the above-mentioned modelling choices is accounted for in the robustness assessment carried out by the JRC. More precisely, the methodology applied herein allows for the joint and simultaneous analysis of the impact of such choices on the national scores, resulting in error estimates and confidence intervals calculated for the GII 2015 individual country rankings.

As suggested in the relevant literature on composite indicators,[5] the robustness assessment was based on Monte Carlo simulation and multi-modelling approaches, applied to 'error-free' data where potential outliers and eventual errors and typos have already been corrected in a preliminary stage. In particular, the three key modelling issues considered in the assessment of the GII were the pillar weights, the treatment of missing data, and the aggregation formula used.

Monte Carlo simulation comprised 1,000 runs of different sets of weights for the seven pillars in the GII. The weights were assigned to the pillars based on uniform continuous distributions centred in the reference values. The ranges of simulated weights were defined by taking into account both the need for a wide enough interval to allow for meaningful robustness checks and the need to respect the underlying principle of the GII that the Input and the Output Sub-Indices should be placed on equal footings. As a result of these considerations, the limit values of uncertainty for the five input pillars are 10%–30%; the limit values for the two output pillars are 40%–60% (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. In fact, with arithmetic average, the

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 Domínguez-Torreiro, European Commission Joint Research Centre, 2015.

'no imputation' choice is equivalent to replacing missing values with the average of the available (normalized) data within each sub-pillar. To overcome this limitation, the JRC estimated missing data using the Expectation Maximization (EM) algorithm.[6]

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 few indicators can compensate a comparative disadvantage on many indicators.[7] The JRC relaxed this strong perfect substitutability assumption inherent in the arithmetic average and considered instead the geometric average, which is a partially compensatory approach that rewards economies with balanced profiles and motivates economies to improve in the GII pillars in which they perform poorly, and not just in *any* GII pillar.[8]

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).

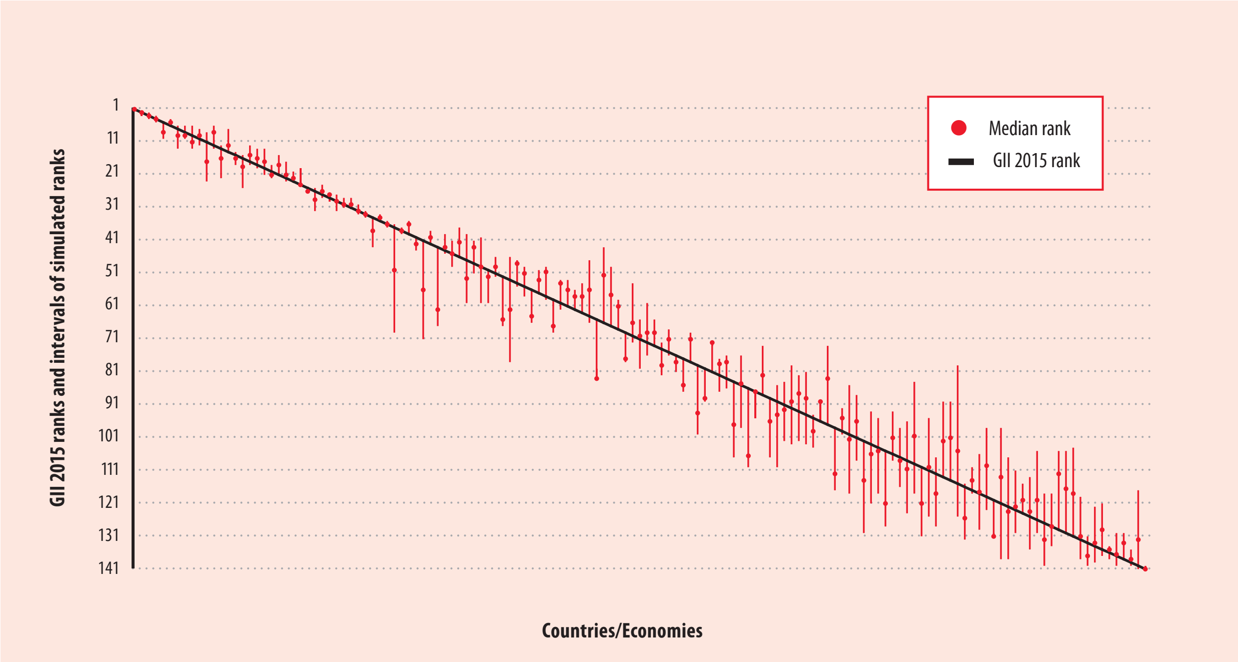
Uncertainty analysis results

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. The figure orders economies from best to worst according to their reference rank (black line), the dot being the median rank.

All published GII 2015 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 80 of the 141 economies. However, it is also true that some economy ranks vary significantly with changes in weights and aggregation formula and, where applicable, they also vary because of the estimation of missing data. Indeed, 32 economies have 90% confidence interval widths between 20 and 29. Confidence interval widths for another 7 economies lie between 30 and 39 (Montenegro, Uganda, Uzbekistan, Belarus, Barbados, Egypt, Kyrgyzstan), and for 2 countries (Bhutan and Fiji) the widths are 40 or greater. For these economies the GII ranks should be interpreted cautiously and certainly not taken at face value. Some caution is also warranted in the Input Sub-Index for 37 economies that have 90% confidence interval widths over 20 (up to 53 for Bosnia and Herzegovina). The Output Sub-Index is slightly more sensitive to the methodological choices: 48 countries have 90% confidence interval widths over 20 (up to 48 for Togo). This sensitivity is mostly the consequence of the estimation of missing data and the fact that there are only two pillars: this means that changes to the imputation method, weights, or aggregation formula have a more notable impact on the country ranks.

Although some economy ranks, in the GII 2015 overall or in the two sub-indices, appear to be sensitive to the methodological choices, the published rankings for the vast majority can be considered as representative of the plurality of scenarios simulated herein. Taking

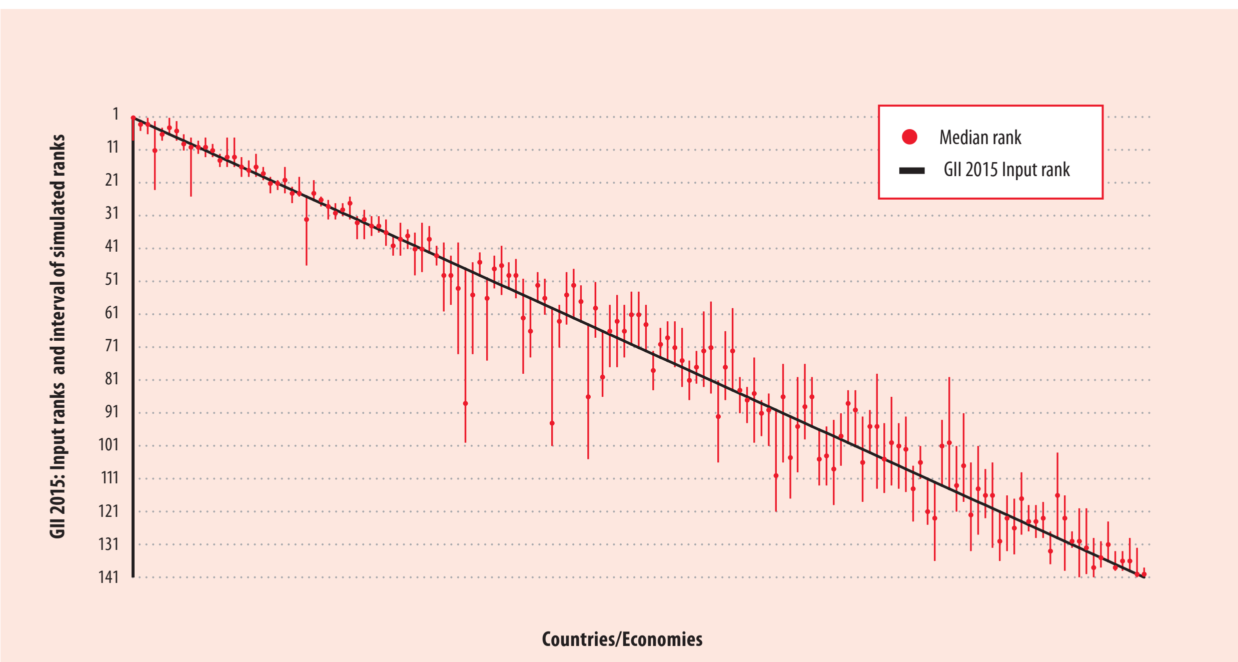
Figure 2a: Robustness analysis (GII rank vs. median rank, 90% confidence intervals)



Source: Saisana and Domínguez-Torreiro, European Commission Joint Research Centre, 2015.

Notes: Median ranks and intervals are calculated over 4,000 simulated scenarios combining random weights, imputed versus missing values, and geometric versus arithmetic average at the pillar level. The Spearman rank correlation between the median rank and the GII 2015 rank is 0.986.

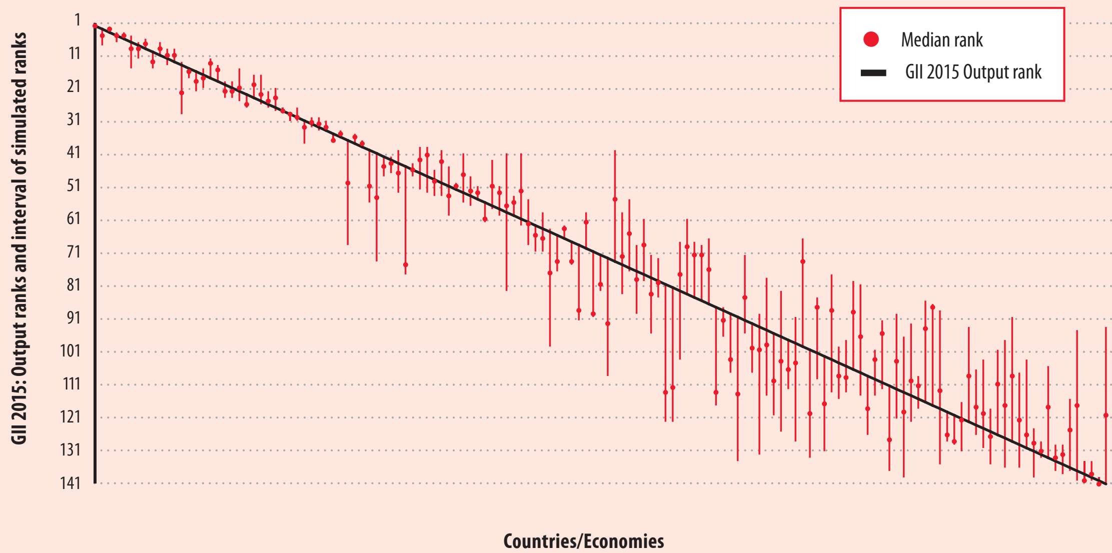
Figure 2b: Robustness analysis (Input rank vs. median rank, 90% confidence intervals)



Source: Saisana and Domínguez-Torreiro, European Commission Joint Research Centre, 2015.

Notes: Median ranks and intervals are calculated over 4,000 simulated scenarios combining random weights, imputed versus missing values, and geometric versus arithmetic average at the pillar level. The Spearman rank correlation between the median rank and the Innovation Input 2015 rank is 0.983.

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



Source: Saisana and Domínguez-Torreiro, European Commission Joint Research Centre, 2015.

Notes: Median ranks and intervals are calculated over 4,000 simulated scenarios combining random weights, imputation versus no imputation of missing values, and geometric versus arithmetic average at the pillar level. The Spearman rank correlation between the median rank and the Innovation Output 2015 rank is 0.966.

the median rank as the yardstick for an economy's expected rank in the realm of the GII's unavoidable methodological uncertainties, 75% of the economies are found to shift fewer than seven positions with respect to the median rank in the GII (seven and eleven positions in the Input and Output Sub-Index, respectively).

For full transparency and information, Table 4 reports the GII 2015 Index and Input and Output Sub-Indices economy ranks together with the simulated 90% confidence intervals in order to better appreciate the robustness of the results to the choice of weights, of the aggregation formula 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. Table 5 summarizes the impact of one-at-a-time changes of either the EM imputation method or the geometric aggregation formula, with random weights. The most influential assumption is the choice of no imputation versus EM imputation; this is particularly influential for the Output Sub-Index, then for the GII, and least for the Input Sub-Index. This sensitivity is the result of the data availability, which is less satisfactory in the case of the Output Sub-Index: 29 countries have data coverage well below the 60% threshold over the 27 variables in the Output Sub-Index. Instead, data coverage is satisfactory in the case of the Input Sub-Index (all economies have indicator coverage more than 65% over the 52 variables). This factor has affected 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 economies' ranks.

Our recommendation would be to consider country ranks in the GII 2015 and in the Input and Output Sub-Indexes not only at face value but also within the 90% confidence intervals in order to better appreciate to what degree a country's rank depends on the modelling choices. Furthermore, the 60% indicator coverage threshold needs to be applied separately to the Input and the Output Sub-Indices.

Table 4: GII 2015 and Input/Output Sub-Indices: Ranks and 90% confidence intervals

	GII 2015		Input Sub-Index		Output Sub-Index	
Country/Economy	Rank	Interval	Rank	Interval	Rank	Interval
Switzerland	1	[1, 1]	2	[2, 5]	1	[1, 1]
United Kingdom	2	[2, 3]	6	[1, 6]	5	[3, 5]
Sweden	3	[2, 4]	7	[2, 8]	4	[3, 6]
Netherlands	4	[3, 5]	11	[7, 13]	3	[2, 3]
United States of America	5	[5, 10]	5	[4, 8]	9	[9, 14]
Finland	6	[4, 6]	3	[1, 6]	10	[6, 10]
Singapore	7	[6, 13]	1	[1, 8]	20	[18, 23]
Ireland	8	[6, 10]	14	[7, 16]	7	[6, 11]
Luxembourg	9	[6, 13]	20	[19, 24]	2	[2, 7]
Denmark	10	[7, 12]	8	[6, 11]	12	[8, 12]
Hong Kong (China)	11	[8, 23]	4	[2, 23]	19	[18, 23]
Germany	12	[6, 13]	18	[12, 19]	8	[5, 8]
Iceland	13	[12, 22]	23	[22, 27]	6	[4, 14]
Korea, Republic of	14	[7, 14]	15	[7, 16]	11	[8, 13]
New Zealand	15	[14, 18]	13	[12, 16]	15	[15, 21]
Canada	16	[15, 25]	9	[7, 25]	22	[22, 26]
Australia	17	[12, 18]	10	[8, 12]	24	[16, 25]
Austria	18	[13, 19]	19	[16, 20]	18	[13, 18]
Japan	19	[13, 21]	12	[9, 13]	26	[20, 27]
Norway	20	[18, 22]	16	[13, 19]	25	[21, 26]
France	21	[15, 21]	17	[14, 19]	23	[16, 23]
Israel	22	[17, 23]	22	[16, 24]	16	[14, 20]
Estonia	23	[20, 23]	26	[20, 26]	14	[14, 17]
Czech Republic	24	[19, 24]	27	[25, 28]	17	[11, 17]
Belgium	25	[25, 26]	21	[20, 23]	28	[28, 30]
Malta	26	[25, 32]	33	[29, 38]	13	[12, 28]
Spain	27	[24, 28]	24	[19, 25]	29	[26, 30]
Slovenia	28	[27, 29]	30	[27, 31]	27	[26, 27]
China	29	[27, 32]	41	[33, 48]	21	[14, 24]
Portugal	30	[28, 31]	28	[26, 32]	33	[30, 33]
Italy	31	[28, 31]	29	[27, 32]	32	[29, 33]
Malaysia	32	[30, 33]	31	[25, 32]	34	[34, 36]
Latvia	33	[32, 34]	34	[31, 37]	30	[30, 37]
Cyprus	34	[34, 43]	32	[31, 38]	43	[39, 52]
Hungary	35	[33, 35]	42	[34, 42]	37	[34, 37]
Slovakia	36	[35, 37]	37	[37, 43]	38	[36, 38]
Barbados	37	[36, 69]	46	[39, 73]	36	[36, 68]
Lithuania	38	[37, 39]	35	[31, 36]	42	[41, 46]
Bulgaria	39	[35, 39]	49	[42, 49]	35	[33, 35]
Croatia	40	[40, 44]	43	[40, 46]	41	[41, 47]
Montenegro	41	[41, 71]	50	[50, 75]	40	[40, 73]
Chile	42	[38, 42]	36	[32, 40]	48	[45, 53]
Saudi Arabia	43	[43, 67]	45	[43, 58]	44	[44, 77]
Moldova, Republic of	44	[39, 45]	74	[65, 74]	31	[29, 32]
Greece	45	[41, 49]	38	[33, 43]	57	[50, 59]
Poland	46	[37, 46]	39	[35, 41]	56	[42, 57]
United Arab Emirates	47	[39, 60]	25	[25, 46]	99	[66, 99]
Russian Federation	48	[41, 51]	52	[40, 55]	49	[39, 53]
Mauritius	49	[40, 60]	44	[39, 60]	60	[40, 62]
Qatar	50	[50, 60]	40	[36, 49]	62	[62, 70]
Costa Rica	51	[46, 52]	61	[48, 64]	45	[43, 47]
Viet Nam	52	[52, 67]	78	[75, 87]	39	[39, 55]
Belarus	53	[46, 78]	55	[50, 79]	58	[40, 82]
Romania	54	[47, 55]	57	[48, 57]	52	[40, 55]
Thailand	55	[49, 58]	62	[47, 62]	50	[44, 59]
TFYR Macedonia	56	[56, 66]	56	[56, 74]	55	[55, 61]
Mexico	57	[50, 57]	58	[50, 61]	54	[50, 54]
Turkey	58	[49, 59]	71	[54, 71]	46	[38, 51]
Bahrain	59	[59, 69]	48	[45, 73]	65	[65, 76]
South Africa	60	[53, 62]	54	[44, 56]	61	[54, 68]
Armenia	61	[53, 61]	69	[58, 74]	51	[49, 51]
Panama	62	[56, 62]	72	[58, 72]	53	[47, 56]
Serbia	63	[54, 63]	70	[54, 70]	59	[53, 59]
Ukraine	64	[47, 66]	84	[59, 84]	47	[38, 52]
Seychelles	65	[65, 80]	59	[59, 101]	64	[63, 99]
Mongolia	66	[43, 66]	53	[45, 53]	73	[39, 73]
Colombia	67	[49, 67]	51	[43, 53]	75	[54, 76]
Uruguay	68	[59, 68]	63	[52, 63]	66	[62, 66]
Oman	69	[69, 78]	68	[55, 77]	68	[68, 91]
Brazil	70	[54, 72]	65	[51, 68]	74	[58, 83]
Peru	71	[65, 80]	60	[58, 71]	82	[67, 103]

Table 4: GII 2015 and Input/Output Sub-Indices: Ranks and 90% confidence intervals *(continued)*

	GII 2015		Input Sub-Index		Output Sub-Index	
Country/Economy	Rank	Interval	Rank	Interval	Rank	Interval
Argentina	72	[60, 76]	81	[57, 85]	63	[58, 70]
Georgia	73	[65, 73]	67	[58, 77]	86	[66, 86]
Lebanon	74	[72, 82]	77	[65, 83]	76	[68, 89]
Jordan	75	[68, 76]	80	[60, 83]	67	[67, 74]
Tunisia	76	[76, 81]	83	[66, 87]	71	[71, 82]
Kuwait	77	[77, 87]	87	[74, 100]	70	[70, 90]
Morocco	78	[69, 78]	76	[63, 77]	84	[67, 84]
Bosnia and Herzegovina	79	[79, 100]	47	[47, 100]	122	[93, 122]
Trinidad and Tobago	80	[80, 90]	86	[83, 94]	88	[87, 96]
India	81	[73, 81]	100	[84, 100]	69	[58, 69]
Kazakhstan	82	[77, 87]	75	[63, 75]	107	[80, 114]
Philippines	83	[76, 86]	101	[84, 101]	77	[60, 79]
Senegal	84	[84, 107]	110	[101, 111]	72	[72, 108]
Sri Lanka	85	[76, 98]	104	[79, 114]	79	[72, 84]
Guyana	86	[86, 110]	90	[90, 121]	93	[89, 132]
Albania	87	[87, 95]	73	[72, 84]	112	[89, 121]
Paraguay	88	[73, 88]	103	[90, 103]	83	[60, 83]
Dominican Republic	89	[89, 107]	88	[87, 98]	98	[90, 128]
Botswana	90	[85, 110]	79	[72, 82]	108	[108, 126]
Cambodia	91	[84, 102]	96	[96, 113]	91	[71, 95]
Kenya	92	[79, 103]	113	[93, 113]	78	[71, 95]
Azerbaijan	93	[82, 102]	89	[85, 101]	103	[77, 113]
Rwanda	94	[81, 103]	66	[66, 86]	128	[90, 128]
Mozambique	95	[94, 104]	93	[80, 109]	97	[97, 112]
Jamaica	96	[96, 96]	85	[80, 91]	110	[91, 112]
Indonesia	97	[73, 97]	114	[80, 114]	85	[68, 85]
Malawi	98	[98, 117]	111	[111, 125]	89	[89, 107]
El Salvador	99	[92, 100]	95	[80, 95]	104	[99, 115]
Egypt	100	[85, 118]	108	[92, 115]	96	[82, 125]
Guatemala	101	[88, 108]	107	[94, 114]	101	[84, 109]
Burkina Faso	102	[102, 130]	109	[107, 124]	100	[100, 133]
Cabo Verde	103	[93, 121]	92	[92, 117]	114	[92, 122]
Bolivia, Plurinational State of	104	[95, 118]	118	[101, 128]	92	[92, 107]
Mali	105	[105, 128]	125	[119, 127]	81	[81, 122]
Iran, Islamic Republic of	106	[93, 108]	106	[86, 113]	105	[97, 113]
Namibia	107	[98, 116]	91	[76, 106]	119	[119, 128]
Ghana	108	[96, 124]	116	[91, 118]	102	[102, 131]
Kyrgyzstan	109	[84, 118]	94	[76, 99]	118	[88, 135]
Cameroon	110	[110, 131]	126	[119, 129]	90	[90, 134]
Uganda	111	[95, 125]	102	[92, 118]	113	[96, 139]
Gambia	112	[107, 128]	121	[121, 136]	94	[78, 114]
Honduras	113	[90, 113]	105	[94, 115]	116	[85, 116]
Tajikistan	114	[90, 114]	115	[101, 121]	106	[79, 106]
Fiji	115	[79, 125]	64	[64, 105]	137	[94, 140]
Côte d'Ivoire	116	[115, 132]	131	[127, 132]	87	[87, 117]
Tanzania, United Republic of	117	[110, 118]	124	[109, 128]	95	[95, 120]
Lesotho	118	[106, 129]	97	[95, 113]	125	[118, 135]
Ecuador	119	[98, 123]	99	[91, 107]	124	[103, 130]
Angola	120	[120, 130]	137	[133, 139]	80	[80, 122]
Bhutan	121	[98, 138]	82	[81, 106]	135	[129, 138]
Uzbekistan	122	[107, 138]	112	[112, 136]	127	[97, 136]
Swaziland	123	[111, 130]	98	[93, 119]	132	[128, 133]
Zambia	124	[115, 124]	130	[116, 139]	115	[108, 118]
Madagascar	125	[113, 135]	123	[117, 134]	123	[106, 126]
Algeria	126	[105, 130]	119	[108, 123]	129	[107, 136]
Ethiopia	127	[118, 140]	132	[120, 141]	111	[111, 137]
Nigeria	128	[118, 134]	135	[130, 138]	109	[100, 114]
Bangladesh	129	[105, 129]	129	[103, 129]	126	[100, 126]
Nicaragua	130	[105, 130]	120	[106, 132]	133	[105, 133]
Pakistan	131	[104, 131]	136	[124, 136]	117	[86, 117]
Venezuela, Bolivarian Republic of	132	[119, 138]	133	[120, 140]	121	[116, 131]
Zimbabwe	133	[131, 140]	134	[129, 141]	120	[120, 129]
Niger	134	[122, 139]	117	[106, 133]	139	[134, 140]
Nepal	135	[121, 137]	127	[118, 129]	136	[115, 137]
Burundi	136	[134, 138]	122	[116, 133]	138	[134, 140]
Yemen	137	[130, 140]	138	[133, 139]	131	[124, 139]
Myanmar	138	[130, 138]	139	[129, 139]	130	[103, 130]
Guinea	139	[135, 140]	140	[132, 141]	134	[129, 137]
Togo	140	[117, 141]	128	[127, 137]	141	[93, 141]
Sudan	141	[140, 141]	141	[138, 141]	140	[139, 141]

Source: Saisana and Domínguez-Torreiro, European Commission Joint Research Centre, 2015.

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	8	2
	Geometric average and EM imputation vs. arithmetic average and missing values	8	4
Input Sub-Index	Geometric vs. arithmetic average	1	1
	EM imputation vs. no imputation of missing data	1	1
	Geometric average and EM imputation vs. arithmetic average and missing values	1	1
Output Sub-Index	Geometric vs. arithmetic average	2	1
	EM imputation vs. no imputation of missing data	15	18
	Geometric average and EM imputation vs. arithmetic average and missing values	15	18

Source: Saisana and Domínguez-Torreiro, European Commission Joint Research Centre, 2015.

Conclusion

The JRC analysis suggests that the conceptualized multi-level structure of the GII 2015—with its 21 sub-pillars, 7 pillars, 2 sub-indices, up to an overall index—is statistically sound and balanced: that is, each sub-pillar makes a similar contribution to the variation of its respective pillar. 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 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 has been made since 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 be made also by other developers of composite indicators.

The strong correlations between the GII components are proven not to be a sign of redundancy of information in the GII. For more than 50.4% (up to 69.5%) of the 141 economies included in the GII 2015, the GII ranking and the rankings of any of the seven pillars 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 2015 ranks lay 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 countries these intervals are narrow enough for meaningful inferences to be drawn: the intervals comprise fewer than 10 positions for 80 of the 141 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; sensitivity is mostly the consequence of the estimation of missing data and the fact that there are only two pillars; 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 2015 or in the two sub-indexes, can be considered representative of the many possible scenarios: 75% of the countries shift fewer than seven positions with respect to the median rank in the GII (seven and eleven positions, respectively, in the Input and Output Sub-Indices).

All things considered, the present JRC audit endorses the statistical soundness and reliability of the GII index as a benchmarking tool for innovation practices at the country level around the world. Needless to say, the usefulness of the GII index as a standalone policy evaluation tool should be enhanced by simultaneously reading and reflecting on the wealth of information on innovation issues gathered and disseminated within the overall GII framework, which in any case should be regarded as 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 https://ec.europa.eu/jrc/en/coin.

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 (141 economies).

4 Nunnally, 1978.

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

6 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.

7 Munda, 2008.

8 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 zero, 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. Hoboken, NJ: John Wiley & Sons, Inc.

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

Nunnally, 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.