Joint Research Centre Statistical Audit of the 2016 Global Innovation Index

Michaela Saisana, Marcos Domínguez-Torreiro, and Daniel Vertesy, 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. In its ninth edition, the 2016 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 across 82 indicators 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 a 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 European Commission's Competence Centre on Composite Indicators and Scoreboards at the Joint Research Centre (JRC) in Ispra has been invited for the sixth 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 independent statistical assessment of the GII 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 2016. 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

Eighty-two 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 2006–15 were used for each economy. Almost 75% of the available data refer to 2014 or more recent years. In past editions, countries were included if data availability was at least 60% across all variables in the GII framework. A more stringent criterion was adopted this year, following the JRC recommendation of past GII audits. This year countries were included if data availability was at least 60% within each of the two sub-indices (i.e., 33 out of 55 variables within the Input Sub-Index and 16 out of the 27 variables in the Output Sub-Index) and at least two of the three sub-pillars in each pillar could be computed. This more stringent criterion for a country's inclusion in the GII is introduced this year to ensure

Figure 1: Conceptual and statistical coherence in the GII 2016 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 indicators into sub-pillars, to pillars, to sub-indices, and to the 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 2014–2015)
- Availability requirements per country: coverage > 60% for the Input and the Output Sub-Indices 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, Domínquez-Torreiro, and Vertesy, European Commission Joint Research Centre, 2016.

that country scores for the GII and for the two Input and Output Sub-Indices are not particularly sensitive to the missing values (as was the case for the Output Sub-Index scores of several countries in past editions). 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 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) and Becker et al. (2016) 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 82 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 seven indicators with Pearson correlation coefficients less than 0.3 with the respective sub-pillars, some further reflection is needed because they seem to behave as 'noise' at all aggregation levels in the GII framework despite the fact that their inclusion was based on conceptual grounds or practical experience. This applies to 2.2.2 Graduates in science and engineering; 3.2.3 Gross capital formation; 3.3.1 GDP per unit of energy use; 4.1.3 Microfinance institutions' gross loan portfolio; 5.2.3 GERD financed by abroad; 5.3.4 Foreign

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	0.77	0.84	0.70	0.74	0.69	0.80
	Regulatory environment	0.92	0.63	0.67	0.58	0.65	0.60	0.67
	Business environment	0.90	0.71	0.75	0.68	0.62	0.68	0.71
	Education	0.52	0.75	0.54	0.40	0.48	0.51	0.52
	Tertiary education	0.65	0.79	0.75	0.57	0.46	0.52	0.57
	Research and development (R&D)	0.69	0.89	0.77	0.78	0.81	0.84	0.74
	Information and communication technologies (ICTs)	0.77	0.83	0.94	0.70	0.64	0.69	0.76
	General infrastructure	0.62	0.62	0.74	0.55	0.58	0.56	0.54
	Ecological sustainability	0.62	0.60	0.74	0.52	0.51	0.52	0.63
	Credit	0.65	0.60	0.58	0.85	0.59	0.52	0.57
	Investment	0.48	0.50	0.42	0.76	0.52	0.49	0.40
	Trade, competition, & market scale	0.51	0.66	0.71	0.71	0.50	0.64	0.60
	Knowledge workers	0.63	0.80	0.68	0.67	0.85	0.73	0.67
	Innovation linkages	0.53	0.40	0.42	0.43	0.72	0.49	0.49
	Knowledge absorption	0.60	0.56	0.56	0.54	0.82	0.71	0.59
OUTPUT	Knowledge creation	0.63	0.79	0.63	0.66	0.75	0.88	0.77
	Knowledge impact	0.51	0.51	0.56	0.44	0.50	0.73	0.59
	Knowledge diffusion	0.52	0.54	0.51	0.52	0.64	0.73	0.50
	Intangible assets	0.62	0.61	0.67	0.54	0.56	0.65	0.89
	Creative goods and services	0.67	0.63	0.64	0.54	0.61	0.69	0.84
	Online creativity	0.81	0.78	0.77	0.67	0.76	0.77	0.88

Source: Saisana, Domínguez-Torreiro, and Vertesy, European Commission Joint Research Centre, 2016.

direct investment net inflows; and 6.2.1 Growth rate of GDP per person engaged.

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 60% (pillar 4: Market sophistication) up to 84% (pillar 1: Institutions) of the total variance in the three underlying sub-pillars. These results reveal that the adjustments made to the 2016 GII framework have left unaffected 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 0.70 (see Table 1).

The five input pillars share a single statistical dimension that summarizes 76% 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 (0.80); 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 0.97; the two sub-indices have a correlation of 0.88. Thus far, results show that the grouping of sub-pillars into pillars, sub-indices, and the GII 2016 is statistically coherent, and that the GII has a balanced structure at each aggregation level. The only recommendation for next year relates to a careful evaluation of the seven indicators discussed above—2.2.2 Graduates in science and engineering; 3.2.3 Gross capital formation; 3.3.1 GDP per unit

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	12.5	10.2	7.8	21.1	21.9	10.9	4.7
20–29	16.4	14.8	12.5	16.4	10.2	10.2	11.7
10–19	21.9	23.4	35.9	25.0	21.9	30.5	15.6
10 or more*	**50.8**	**48.4**	**56.3**	**62.5**	**53.9**	**51.6**	**32.0**
5–9	28.1	22.7	16.4	16.4	23.4	19.5	32.0
Less than 5	18.0	25.8	24.2	20.3	20.3	21.9	32.8
Same rank	3.1	3.1	3.1	0.8	2.3	7.0	3.1
Total†	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
Pearson correlation coefficient with the GII	0.88	0.90	0.89	0.81	0.86	0.92	0.93

Source: Saisana, Domínguez-Torreiro, and Vertesy, European Commission Joint Research Centre, 2016.
* This column is the sum of the prior three rows.
† This column is the sum of all white rows.

of energy use; 4.1.3 Microfinance institutions' gross loan portfolio; 5.2.3 GERD financed by abroad; 5.3.4 Foreign direct investment net inflows; and 6.2.1 Growth rate of GDP per person engaged. Because their information content is lost in the aggregation at the pillar level or higher (sub-index and overall GII), the recommendation is either to increase the weight attached to these indicators so that their information is not lost in the aggregation or to replace them with some more suitable indicators that are better proxies of the conditions they are intended to capture.

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. The tests conducted by the JRC indicate that this is not the case. In fact, for more than 32% (up to 62.5%) of the 128 economies included in the GII 2016, 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.

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

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 pillar level		**Arithmetic average**	**Geometric average**
III. Uncertainty intervals for the GII pillar 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, Dominguez-Torreiro, and Vertesy, European Commission Joint Research Centre, 2016.

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 2016 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.

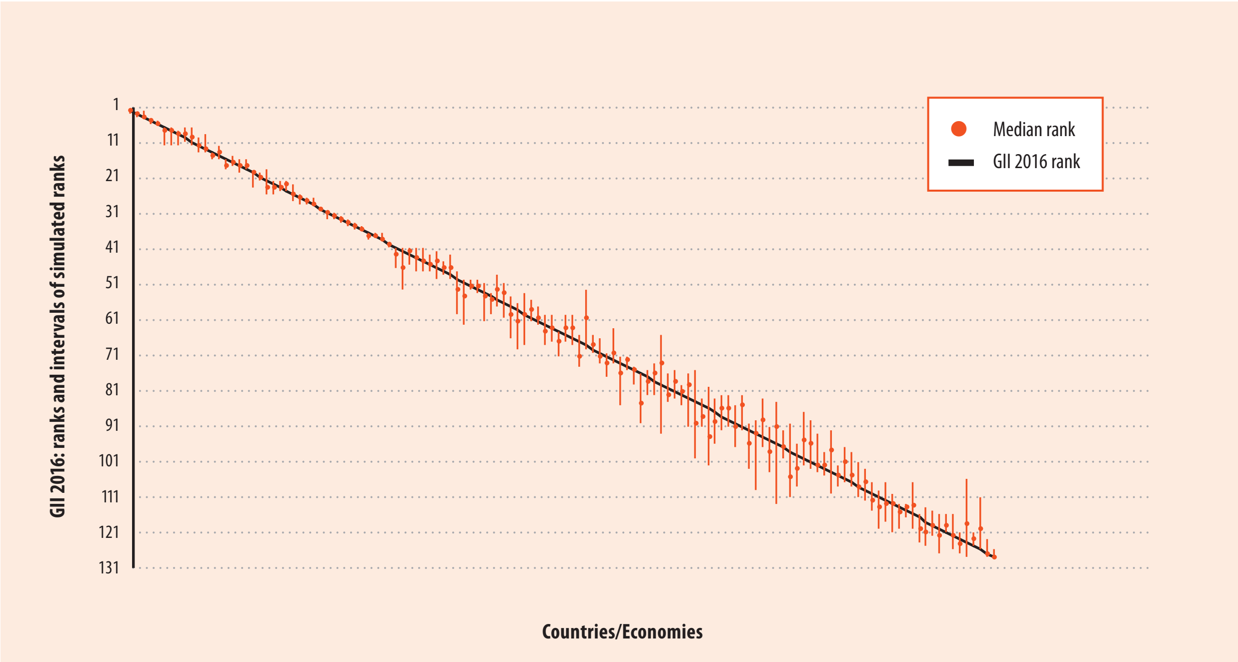
Monte Carlo simulation comprised 1,000 runs of different sets of weights for the seven pillars in the GII. The weights were assigned to the pillars based on uniform continuous distributions centred in the reference values. The ranges of simulated weights were defined by taking into account both the need 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 'no imputation' choice is equivalent to replacing an indicator's missing value for a given country with the respective sub-pillar score. 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 for a comparative disadvantage on many indicators.[7] For example, one may argue that Ireland and Iceland, despite their similar performance at the Innovation Output Sub-Index—both close to 55.5 points (rank 5th and 6th, respectively) are very different if one considers how these countries perform within the sub-index. Ireland ranks 3rd in Knowledge and technology outputs and 10th in Creative outputs, while Iceland is much more diverse: the country ranks 22nd in Knowledge and technology outputs, but it notably improves its overall position in the Output Sub-Index thanks to its 1st place position in Creative outputs. To assess the impact of this compensability issue, the JRC relaxed the 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

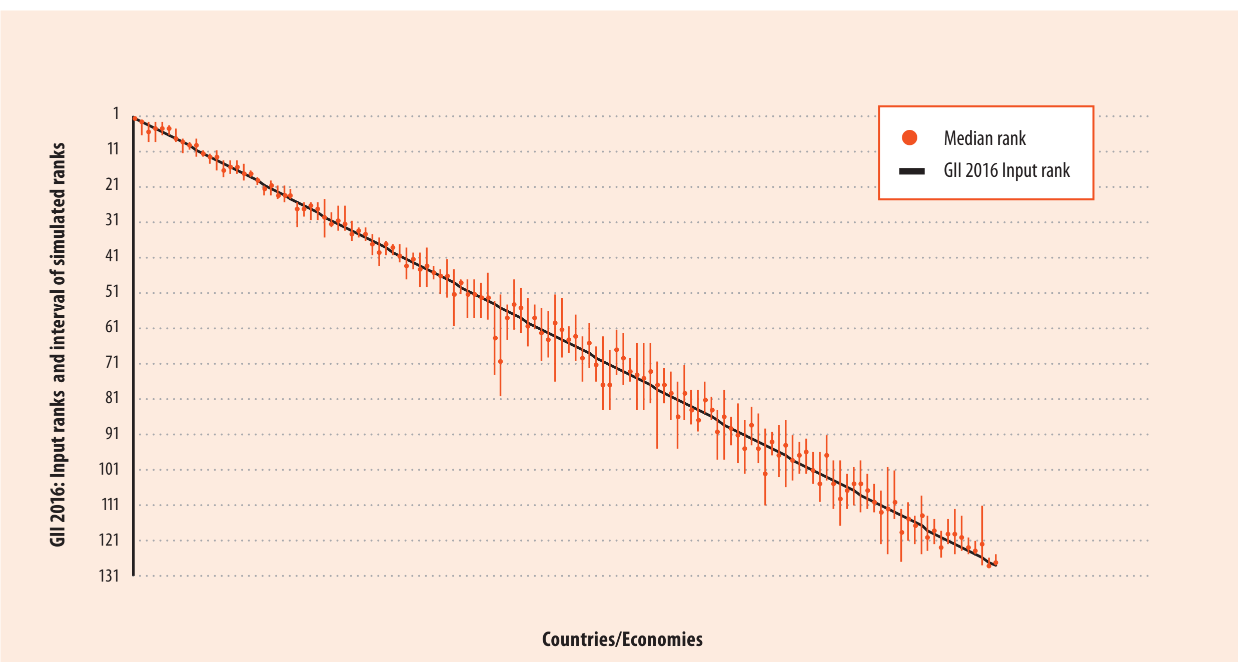
Figure 2a: Robustness analysis (GII rank vs. median rank, 90% confidence intervals)



Source: Saisana, Domínguez-Torreiro, and Vertesy, European Commission Joint Research Centre, 2016.

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 2016 rank is 0.997.

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



Source: Saisana, Domínguez-Torreiro, and Vertesy, European Commission Joint Research Centre, 2016.

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 2016 rank is 0.997.

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

Source: Saisana, Domínguez-Torreiro, and Vertesy, European Commission Joint Research Centre, 2016.

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 2016 rank is 0.992.

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 over the simulations.

All published GII 2016 ranks lay within the simulated 90% confidence intervals, and for most economies these intervals are narrow enough for meaningful inferences to be drawn: there is a shift of fewer than 10 positions (roughly plus or minus 5 positions) for 93 of the 128 economies. However, it is also true that six economy ranks vary significantly with changes in weights and aggregation formula and, where applicable, they also vary because of the estimation of missing data. These six economies—Belarus, Mozambique, Tajikistan, Bhutan, Malawi, and Niger, in rank order—have 90% confidence interval widths between 20 and 29, hence their GII ranks should be interpreted cautiously and certainly not taken at face value. This is a remarkable improvement compared to last year's GII, where confidence interval widths for 32 economies lay between 20 and 29, for another 7 economies between 30 and 39, and for 2 countries the widths were 40 or greater. This improvement in the confidence one can attach to the GII 2016 ranks is the direct result of the developers' choice to adopt a more stringent criterion for a country's inclusion, which requires at least 60% data availability within each of the two sub-indices. Some caution is also warranted in the Input Sub-Index for 8 economies—Kuwait, Oman, Jordan, Rwanda, Bosnia and Herzegovina, Cambodia, Bhutan, and Venezuela—that have 90% confidence interval widths over 20 (up to 29 for Rwanda). The Output Sub-Index is slightly more sensitive to the methodological choices: 14 countries—Kuwait, Oman, Belarus, Rwanda, Mozambique, Tajikistan, Namibia, Paraguay, Malawi, Ecuador, Honduras, Nepal, Niger, and Togo—have 90% confidence interval widths over 20 (up to 44

Table 4: GII 2016 and Input/Output Sub-Indices: Ranks and 90% confidence intervals

	GII 2016		Input Sub-Index		Output Sub-Index	
Country/Economy	Rank	Interval	Rank	Interval	Rank	Interval
Switzerland	1	[1, 2]	6	[3, 6]	1	[1, 2]
Sweden	2	[2, 3]	5	[2, 6]	2	[2, 3]
United Kingdom	3	[1, 3]	7	[4, 8]	4	[1, 5]
United States of America	4	[4, 5]	3	[2, 8]	7	[6, 11]
Finland	5	[4, 5]	4	[2, 8]	10	[8, 10]
Singapore	6	[6, 11]	1	[1, 1]	20	[17, 22]
Ireland	7	[6, 11]	16	[13, 17]	5	[4, 8]
Denmark	8	[7, 11]	8	[7, 11]	13	[12, 13]
Netherlands	9	[6, 10]	12	[11, 14]	9	[7, 10]
Germany	10	[6, 11]	18	[16, 18]	8	[5, 8]
Korea, Republic of	11	[9, 13]	13	[10, 16]	11	[10, 12]
Luxembourg	12	[8, 13]	23	[20, 25]	3	[2, 4]
Iceland	13	[13, 15]	24	[21, 24]	6	[5, 11]
Hong Kong (China)	14	[11, 15]	2	[2, 6]	25	[22, 25]
Canada	15	[15, 18]	10	[7, 12]	23	[23, 26]
Japan	16	[14, 17]	9	[8, 10]	24	[21, 25]
New Zealand	17	[15, 19]	14	[13, 18]	17	[16, 20]
France	18	[15, 18]	15	[13, 17]	19	[15, 20]
Australia	19	[19, 23]	11	[10, 12]	27	[27, 31]
Austria	20	[19, 21]	19	[18, 20]	22	[18, 22]
Israel	21	[18, 25]	21	[19, 23]	16	[16, 24]
Norway	22	[21, 25]	17	[14, 19]	26	[25, 28]
Belgium	23	[21, 24]	20	[20, 23]	18	[17, 22]
Estonia	24	[21, 24]	27	[25, 30]	14	[13, 14]
China	25	[22, 27]	29	[24, 35]	15	[13, 15]
Malta	26	[25, 28]	35	[32, 36]	12	[11, 18]
Czech Republic	27	[26, 28]	26	[25, 29]	21	[18, 22]
Spain	28	[26, 28]	22	[20, 24]	28	[27, 30]
Italy	29	[29, 29]	28	[25, 30]	31	[30, 31]
Portugal	30	[30, 32]	30	[28, 32]	32	[29, 32]
Cyprus	31	[30, 32]	33	[30, 36]	29	[26, 30]
Slovenia	32	[31, 33]	31	[26, 33]	33	[33, 34]
Hungary	33	[32, 34]	38	[36, 40]	30	[26, 32]
Latvia	34	[33, 35]	36	[34, 40]	34	[33, 34]
Malaysia	35	[34, 35]	32	[26, 33]	39	[38, 39]
Lithuania	36	[36, 38]	34	[32, 35]	41	[40, 46]
Slovakia	37	[36, 37]	42	[39, 44]	38	[37, 38]
Bulgaria	38	[36, 38]	49	[47, 51]	35	[35, 36]
Poland	39	[39, 40]	39	[37, 40]	46	[45, 46]
Greece	40	[40, 46]	37	[35, 43]	49	[48, 53]
United Arab Emirates	41	[40, 52]	25	[25, 32]	75	[68, 78]
Turkey	42	[40, 45]	59	[52, 66]	37	[37, 39]
Russian Federation	43	[40, 47]	44	[38, 49]	47	[47, 51]
Chile	44	[40, 47]	40	[37, 42]	53	[51, 55]
Costa Rica	45	[42, 47]	50	[47, 58]	44	[40, 45]
Moldova, Republic of	46	[41, 49]	74	[69, 76]	36	[35, 36]
Croatia	47	[44, 48]	45	[43, 47]	48	[47, 51]
Romania	48	[42, 49]	52	[48, 58]	45	[42, 46]
Saudi Arabia	49	[47, 59]	43	[39, 49]	54	[53, 70]
Qatar	50	[49, 62]	41	[38, 47]	58	[58, 77]
Montenegro	51	[49, 52]	46	[44, 51]	52	[52, 54]
Thailand	52	[49, 53]	57	[47, 61]	50	[48, 52]
Mauritius	53	[50, 61]	48	[44, 60]	68	[55, 69]
South Africa	54	[53, 59]	47	[41, 52]	71	[66, 71]
Mongolia	55	[48, 57]	66	[57, 70]	51	[40, 51]
Ukraine	56	[50, 60]	76	[65, 84]	40	[40, 44]
Bahrain	57	[54, 66]	51	[47, 58]	67	[63, 69]
TFYR of Macedonia	58	[56, 69]	62	[56, 69]	55	[53, 63]
Viet Nam	59	[53, 68]	79	[72, 82]	42	[41, 50]
Armenia	60	[55, 61]	80	[73, 87]	43	[43, 45]
Mexico	61	[57, 62]	60	[54, 61]	62	[59, 64]
Uruguay	62	[59, 68]	61	[55, 70]	66	[60, 70]
Colombia	63	[59, 67]	53	[45, 58]	74	[70, 74]
Georgia	64	[64, 71]	67	[63, 76]	60	[58, 65]

Table 4: GII 2016 and Input/Output Sub-Indices: Ranks and 90% confidence intervals *(continued)*

	GII 2016		Input Sub-Index		Output Sub-Index	
Country/Economy	Rank	Interval	Rank	Interval	Rank	Interval
Serbia	65	[59, 67]	68	[59, 72]	63	[57, 64]
India	66	[59, 68]	72	[61, 74]	59	[55, 67]
Kuwait	67	[65, 74]	78	[70, 95]	56	[52, 74]
Panama	68	[52, 69]	73	[62, 77]	61	[47, 62]
Brazil	69	[65, 70]	58	[49, 62]	79	[74, 80]
Lebanon	70	[66, 73]	85	[76, 85]	57	[57, 60]
Peru	71	[70, 77]	56	[54, 64]	87	[87, 95]
Morocco	72	[63, 73]	75	[65, 84]	70	[56, 72]
Oman	73	[71, 85]	63	[51, 76]	86	[85, 114]
Philippines	74	[71, 75]	86	[80, 87]	64	[57, 69]
Kazakhstan	75	[74, 79]	65	[60, 68]	90	[87, 99]
Dominican Republic	76	[76, 90]	84	[78, 90]	82	[80, 98]
Tunisia	77	[75, 82]	82	[71, 87]	84	[78, 84]
Iran, Islamic Republic of	78	[73, 82]	90	[82, 99]	72	[59, 73]
Belarus	79	[65, 93]	64	[52, 69]	103	[67, 110]
Kenya	80	[76, 84]	97	[87, 106]	65	[62, 76]
Argentina	81	[75, 83]	77	[65, 82]	89	[82, 89]
Jordan	82	[79, 85]	88	[78, 98]	77	[76, 81]
Rwanda	83	[76, 91]	55	[51, 80]	114	[79, 114]
Mozambique	84	[75, 100]	96	[88, 103]	73	[70, 100]
Azerbaijan	85	[83, 91]	81	[76, 95]	94	[90, 97]
Tajikistan	86	[80, 102]	102	[96, 110]	69	[65, 92]
Bosnia and Herzegovina	87	[83, 96]	70	[63, 84]	104	[102, 113]
Indonesia	88	[82, 92]	99	[93, 102]	76	[74, 77]
Jamaica	89	[82, 91]	83	[78, 88]	99	[82, 100]
Botswana	90	[85, 97]	69	[66, 76]	111	[108, 118]
Sri Lanka	91	[82, 92]	98	[91, 104]	78	[72, 79]
Albania	92	[90, 103]	71	[67, 84]	115	[113, 118]
Namibia	93	[89, 109]	87	[84, 98]	101	[94, 116]
Paraguay	94	[83, 97]	95	[90, 97]	88	[73, 101]
Cambodia	95	[91, 104]	94	[89, 111]	95	[90, 96]
Bhutan	96	[84, 113]	54	[53, 74]	122	[107, 126]
Guatemala	97	[92, 100]	101	[96, 105]	93	[88, 99]
Malawi	98	[90, 111]	110	[105, 113]	83	[80, 106]
Uganda	99	[96, 108]	91	[87, 102]	105	[100, 116]
Ecuador	100	[87, 102]	100	[92, 102]	97	[79, 103]
Honduras	101	[89, 103]	93	[85, 99]	106	[84, 107]
Ghana	102	[93, 104]	103	[91, 104]	100	[96, 112]
Kyrgyzstan	103	[98, 105]	92	[83, 94]	109	[106, 119]
El Salvador	104	[92, 110]	89	[86, 94]	110	[101, 118]
Tanzania, United Republic of	105	[102, 108]	117	[108, 125]	80	[78, 90]
Senegal	106	[97, 107]	109	[102, 112]	96	[86, 97]
Egypt	107	[98, 108]	107	[100, 111]	98	[95, 105]
Côte d'Ivoire	108	[101, 111]	121	[115, 122]	81	[78, 95]
Bolivia, Plurinational State of	109	[103, 112]	108	[98, 113]	102	[100, 108]
Ethiopia	110	[107, 115]	124	[120, 125]	85	[84, 95]
Madagascar	111	[109, 120]	120	[118, 126]	91	[90, 102]
Mali	112	[107, 118]	119	[115, 122]	92	[87, 104]
Algeria	113	[110, 121]	104	[98, 112]	116	[114, 125]
Nigeria	114	[113, 120]	122	[112, 125]	107	[104, 111]
Nepal	115	[113, 117]	116	[114, 122]	112	[92, 112]
Nicaragua	116	[107, 120]	106	[102, 112]	120	[106, 120]
Bangladesh	117	[116, 124]	115	[110, 121]	117	[115, 124]
Cameroon	118	[114, 125]	118	[114, 124]	113	[110, 121]
Pakistan	119	[115, 122]	123	[114, 124]	108	[106, 116]
Venezuela, Bolivarian Republic of	120	[116, 127]	112	[100, 125]	119	[119, 128]
Benin	121	[116, 121]	111	[107, 122]	121	[109, 121]
Burkina Faso	122	[116, 126]	105	[98, 117]	127	[121, 128]
Burundi	123	[121, 127]	114	[112, 127]	123	[122, 127]
Niger	124	[106, 128]	113	[101, 115]	125	[106, 128]
Zambia	125	[121, 125]	126	[111, 128]	118	[117, 125]
Togo	126	[111, 126]	125	[121, 125]	126	[82, 126]
Guinea	127	[123, 128]	127	[126, 128]	124	[120, 125]
Yemen	128	[126, 128]	128	[125, 128]	128	[125, 128]

Source: Saisana, Domínguez-Torreiro, and Vertesy, European Commission Joint Research Centre, 2016.

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	0	0
	Geometric average and EM imputation vs. arithmetic average and missing values	0	0
Input Sub-Index	Geometric vs. arithmetic average	0	0
	EM imputation vs. no imputation of missing data	19	16
	Geometric average and EM imputation vs. arithmetic average and missing values	0	1
Output Sub-Index	Geometric vs. arithmetic average	0	0
	EM imputation vs. no imputation of missing data	17	19
	Geometric average and EM imputation vs. arithmetic average and missing values	4	3

Source: Saisana, Domínguez-Torreiro, and Vertesy, European Commission Joint Research Centre, 2016.

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 in the Innovation Output Sub-Index.

Although a few economy ranks, in the GII 2016 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 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 three positions with respect to the median rank in the GII (three and four positions in the Input and Output Sub-Index, respectively). Note that in the past GII 2015, 75% of the economies included were 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-Indices, respectively). This result further confirms that the developers' choice to require higher data availability for a country's inclusion in this year's GII has led to more reliable country ranks for the GII and the two sub-indices.

For full transparency and information, Table 4 reports the GII 2016 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).

Note: 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 2016 rank is 0.992.

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. As in past versions of the GII, the most influential assumption is the choice of no imputation versus EM imputation. Yet, unlike past editions, the decision as to whether to impute or not missing data has the same influence on both the Input and the Output Sub-Index (note that in past GII editions the Output Sub-Index was found to be much more sensitive to the estimation of missing data than the Input Sub-Index). The GII is found not to be heavily influenced by the imputation of missing data. The choice of the aggregation formula does not have a pronounced impact on the economies' ranks; if the geometric averaging across the pillars is used instead of an arithmetic averaging, then merely four countries—Belarus, Albania, Namibia, and Bhutan, in rank order—would decline by more than 10 positions (up to 26 for Bhutan), while no economy would improve by 10 positions or more.

All in all, the published GII 2016 ranks are reliable and for the vast majority of countries the simulated 90% confidence intervals are narrow

Table 6: Pie shares (absolute terms) and efficiency scores for the top 25 economies in the GII 2016

	Input pillars					Output pillars						
Country/Economy	Institutions	Human capital and research	Infrastructure	Market sophistication	Business sophistication	Knowledge and technology outputs	Creative outputs	Efficient frontier rank (DEA)	GII rank	Difference	Efficiency ratio rank	Difference from GII rank
Switzerland	0.08	0.19	0.07	0.10	0.17	0.19	0.19	1	1	0	5	–4
Sweden	0.14	0.20	0.20	0.05	0.20	0.16	0.05	1	2	1	10	–8
Singapore	0.15	0.19	0.20	0.17	0.18	0.05	0.05	1	6	5	78	–72
United Kingdom	0.10	0.20	0.20	0.20	0.05	0.05	0.20	4	3	–1	14	–11
United States of America	0.20	0.05	0.20	0.20	0.18	0.05	0.12	5	4	–1	25	–21
Finland	0.20	0.20	0.18	0.05	0.20	0.05	0.12	6	5	–1	32	–27
Denmark	0.20	0.20	0.20	0.20	0.05	0.05	0.10	7	8	1	34	–26
Hong Kong (China)	0.20	0.20	0.20	0.20	0.07	0.05	0.08	8	14	6	83	–69
Netherlands	0.20	0.10	0.20	0.05	0.20	0.05	0.20	9	9	0	20	–11
Ireland	0.20	0.05	0.20	0.05	0.20	0.10	0.20	10	7	–3	8	–1
Korea, Rep.	0.05	0.20	0.20	0.20	0.17	0.13	0.05	10	11	1	24	–13
Germany	0.20	0.20	0.20	0.10	0.05	0.05	0.20	12	10	–2	9	1
Canada	0.20	0.20	0.20	0.20	0.06	0.05	0.09	12	15	3	57	–42
Japan	0.20	0.20	0.20	0.20	0.06	0.09	0.05	12	16	4	65	–49
Australia	0.20	0.20	0.20	0.20	0.05	0.05	0.10	12	19	7	73	–54
Luxembourg	0.20	0.05	0.20	0.05	0.20	0.10	0.20	16	12	–4	1	11
New Zealand	0.20	0.20	0.20	0.20	0.05	0.05	0.10	16	17	1	40	–23
France	0.20	0.20	0.20	0.20	0.05	0.05	0.10	18	18	0	44	–26
Iceland	0.20	0.10	0.20	0.05	0.20	0.05	0.20	19	13	–6	3	10
Austria	0.20	0.20	0.20	0.05	0.18	0.05	0.12	19	20	1	43	–23
Norway	0.20	0.20	0.20	0.05	0.19	0.05	0.11	21	22	1	55	–33
Israel	0.05	0.20	0.20	0.19	0.20	0.05	0.11	22	21	–1	23	–2
Belgium	0.20	0.20	0.20	0.05	0.18	0.05	0.12	22	23	1	27	–4
Estonia	0.20	0.05	0.20	0.10	0.20	0.05	0.20	25	24	–1	6	18
China	0.05	0.10	0.20	0.20	0.20	0.20	0.05	25	25	0	7	18

Source: Saisana, Domínguez-Torreiro, and Vertesy, European Commission Joint Research Centre, 2016.
Notes: Pie shares are in absolute terms, bounded by 0.05 and 0.20. In the GII 2016, however, the five input pillars each have a fixed weight of 0.10; the two output pillars each have a fixed weight of 0.25.

enough for meaningful inferences to be drawn. Nevertheless, the readers of the GII 2016 report should consider country ranks in the GII 2016 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 the degree to which a country's rank depends on the modelling choices. This year, following the JRC recommendation from past GII audits, the developers' choice to apply the 60% indicator coverage threshold separately to the Input and the Output Sub-Indices has led to a net increase in the reliability of country ranks for the GII and the two sub-indices.

Distance to the efficiency 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 grounds that a given weighting scheme might not be fair to a particular country. An appealing feature of the more recent Data Envelopment Analysis (DEA) literature applied in real decision-making settings is to determine endogenous weights that maximize the overall score of each decision-making unit given a set of other observations.

In this section, the assumption of fixed pillar weights common to all countries is relaxed once more; this time country-specific weights that maximize a country's score are determined endogenously by DEA.[9] 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 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 to 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'.

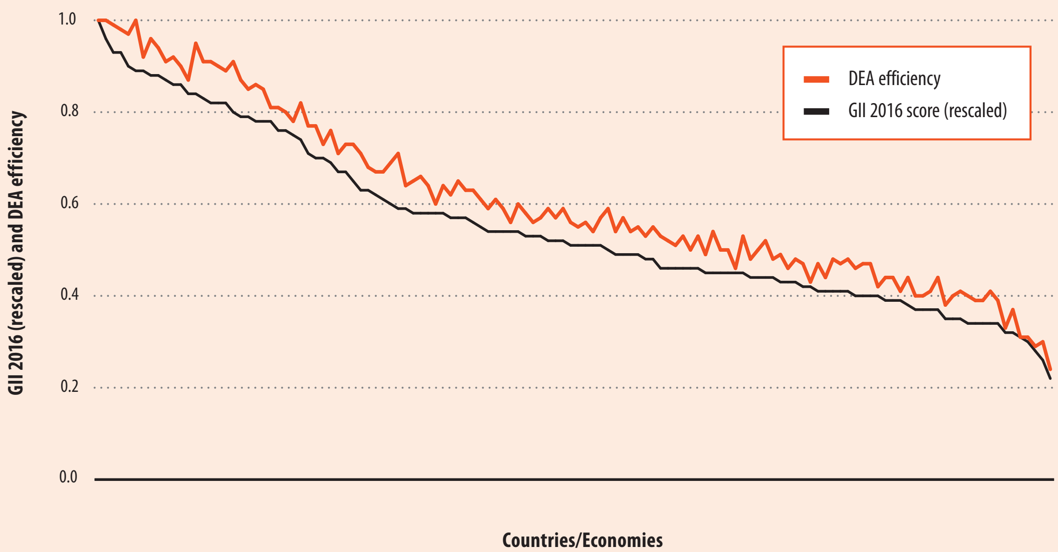
Table 6 presents the pie shares and DEA scores for the top 25 countries in the GII 2016, next to the GII 2016 ranks and efficiency ratio ranks. All pie shares are in accordance with the starting point of granting leeway to each country when assigning shares, while not violating the (relative) upper and lower bounds. The pie shares are quite diverse, reflecting the different national innovation strategies. These pie shares can also be seen to reflect countries comparative advantage in certain GII pillars vis-à-vis all other countries and all pillars. For example, Sweden obtains a perfect DEA score of 1 by assigning 20% of its DEA score to *Human capital and research, Infrastructure,* and *Business sophistication,* while merely 5% of its DEA score comes from *Market sophistication* and *Creative outputs*. Instead, countries including the United Kingdom (UK), the United States of America (USA), and Denmark would assign 20% of their DEA scores to Market sophistication. Three countries—Switzerland, Sweden, and Singapore—reach a perfect DEA score of 1. These countries are closely followed by the UK, the USA, Finland, Denmark, and Hong Kong (China), which score between 0.95 and 0.99 in terms of efficiency. Figure 3 shows how close the DEA scores and the GII 2016 scores are for all 128 economies (correlation of 0.98).[10] Note that, by construction, the version of the DEA used herein is closer to the GII than to the efficiency ratio calculated as the Output Sub-Index score divided by the Input Sub-Index score (which has a correlation of 0.59).

Conclusions

The JRC analysis suggests that the conceptualized multi-level structure of the GII 2016—with its 82 indicators, 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. Nevertheless, a careful reflection by the GII team is needed for seven out of the 82 indicators because their capacity to distinguish countries' performance is lost in the aggregation at the pillar level or higher. Six indicators related to the inputs of innovation—2.2.2 Graduates in science and engineering; 3.2.3 Gross capital formation; 3.3.1 GDP per unit of energy use; 4.1.3 Microfinance institutions' gross loan portfolio; 5.2.3 GERD financed by abroad; 5.3.4 Foreign direct investment net inflows—and one indicator related to the outputs of innovation, 6.2.1 Growth rate of GDP per person engaged, need to be reviewed because their statistical relevance to the GII framework is very weak, unlike their strong conceptual relevance. 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, with the additional negative side-effect that it may encourage countries not to report low data values. This year's adoption by the GII team of a more stringent data coverage threshold (at least 60% for the input- and output-related indicators, separately) has notably improved the confidence in the country ranks for the GII and the two sub-indices. Additionally, the choice of the GII team, which has been followed since 2012, to use weights as scaling coefficients during the development of the index constitutes a significant departure from the traditional, yet erroneous, 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 to avoid situations where bias sneaks in when least expected.

The strong correlations between the GII components are proven not to be a sign of redundancy of information in the GII. For more than 32% (up to 62.5%) of the 128 economies included in the GII 2016, 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

Figure 3: GII 2016 scores and DEA 'distance to the efficient frontier' scores



Source: Saisana, Domínguez-Torreiro, and Vertesy, European Commission Joint Research Centre, 2016.
Note: For comparison purposes, we have rescaled the GII scores by dividing them with the best performer in the overall GII 2016.

seven pillars separately. At the same time, this finding 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 in innovation can be identified and serve as an input for evidence-based policy making.

All published GII 2016 ranks lie within the simulated 90% confidence intervals that take into account the unavoidable uncertainties in the estimation of missing data, the weights (fixed vs. random), and the aggregation formula (arithmetic vs. geometric average) at the pillar level. For the vast majority of countries these intervals are narrow enough for meaningful inferences to be drawn: the intervals comprise fewer than 10 positions for 93 of the 128 economies. Some caution is needed merely for six countries with ranks that are highly sensitive to the methodological choices. The Input and Output Sub-Indices have the same modest degree of sensitivity to the methodological choices related to the imputation method, weights, or aggregation formula. Country ranks, either in the GII 2016 or in the two sub-indexes, can be considered representative of the many possible scenarios: 75% of the countries shift fewer than three positions with respect to the median rank in the GII (three and four positions, respectively, in the Input and Output Sub-Indices).

All things considered, the present JRC audit findings suggest that the GII 2016 meets international quality standards for statistical soundness, indicating that the GII index is a reliable benchmarking tool for innovation practices at the country level around the world.

That said, the GII should not be seen as the ultimate and definitive ranking of countries with respect to innovation. On the contrary, the GII best represents an ongoing attempt by Cornell University, the business school INSEAD, and the World Intellectual Property Organization to find metrics and approaches that better capture the richness of innovation, continuously adapting the GII framework to reflect the improved availability of statistics and the theoretical advances in the field. In any case, the GII 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; Schneider, 2001) 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.

9 A question that arises from the GII approach is whether there is a way to benchmark countries' multi-dimensional performance on innovation without imposing a fixed and common set of weights that may not be fair to a particular country. 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 (Cherchye et al., 2008; Melyn and Moesen, 1991). To estimate DEA-based distance to the efficient frontier scores, we consider the $m = 7$ pillars in the GII 2016 for $n = 128$ countries, with y_{ij} 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 *i*-th pillar. In absence of reliable information about the true weights, the weights that maximize the DEA-based scores are endogenously determined. This gives the following linear programming problem for each country *j*:

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

subject to

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

where

$j = 1, \ldots, 7,$
$i = 1, \ldots, 128$

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

10 Of these, only Luxembourg 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 Ration and the DEA score embed very different concepts of efficiency leading to completely different results and insights. A high score in the Innovation Efficiency Ratio is obtained by scoring more on the Output Sub-Index than on the Input Sub-Index, irrespective of the actual scores in these two sub-indices. Instead, 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

Becker, W., P. Paruolo, M. Saisana, and A. Saltelli. 2016. 'Weights and Importance in Composite Indicators: Mind the Gap'. In *Handbook of Uncertainty Quantification,* eds. R. Ghanem, D. Higdon, and H. Owhadi. Cham, Switzerland: Springer International Publishing. Chapter 4.

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* 17. Leuven: Centre for Economic Studies.

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.

Schneider, T. 2001. 'Analysis of Incomplete Climate Data: Estimation of Mean Values and Covariance Matrices and Imputation of Missing Values'. *Journal of Climate* 14 (5): 853–71.