Joint Research Centre Statistical Audit of the 2017 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 10th edition, the 2017 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 over 127 countries and across 81 indicators into 21 sub-pillars, 7 pillars, 2 sub-indices and, finally, an overall index. This annex offers detailed insights into the practical issues related to the construction of the GII, analysing in depth the statistical soundness of the calculations and assumptions made to arrive at the final index rankings. 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 European Commission's Competence Centre on Composite Indicators and Scoreboards at the Joint Research Centre (JRC) in Ispra has been invited for the seventh consecutive year to audit the GII. As in previous editions, the present JRC audit focuses 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 2017. 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).

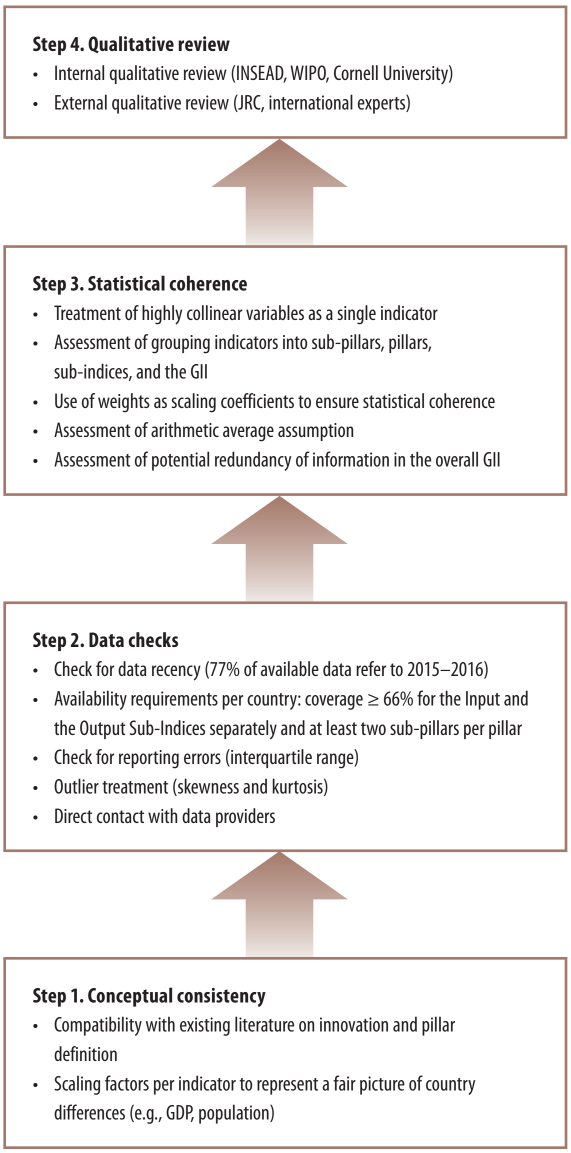
Step 1: Conceptual consistency

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

Step 2: Data checks

The most recently released data within the period 2006–16 were used for each economy: 77% of the available data refer to 2015 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. That is, countries were included if data availability was at least 66% within each of the two sub-indices (i.e., 36 out of 54 variables within the Input Sub-Index and 18 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 was introduced this year in order to ensure that country scores for the GII and for the two Input and Output Sub-Indices are

Figure 1: Conceptual and statistical coherence in the GII 2017 framework



Source: European Commission Joint Research Centre, 2017.

not particularly sensitive to the missing values (as it was the case for the Output Sub-Index scores of several countries in past editions). In practice, data availability for all countries included in the GII 2017 is very good: 80% data availability for 84% (107 out of 127) of the countries. Potentially problematic indicators that could bias the overall results were identified on the basis of two measures related to the shape of the distributions: skewness and kurtosis. In past editions since 2011, values were treated if the indicators had absolute skewness greater than 2.0 and kurtosis greater than 3.5.[3] These criteria were decided jointly with the JRC back in 2011. This year and after having analysed data in GII 2011–GII 2017, a less stringent criterion was adopted: an indicator was treated if the absolute skewness was greater than 2.25 and kurtosis greater than 3.5. These indicators were treated either by winsorization or by taking the natural logarithm (in case of more than five outliers; 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). Becker et al. (2017) and Paruolo et al. (2013) show that, in weighted arithmetic averages, the ratio of two nominal weights gives the rate of substitutability between 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, 35 out of 81 indicators and two sub-pillars—7.2 Creative goods and services and 7.3 Online creativity—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 be non-influential (i.e., they behave as 'noise') at all aggregation levels in the GII 2017 framework, despite the fact that their inclusion was based on conceptual grounds or practical experience. This applies to 2.1.2 Government expenditure on education per pupil, secondary; 2.2.2 Graduates in science and engineering; 3.2.3 Gross capital formation; 5.2.3 GERD financed by abroad,

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	0.76	0.85	0.69	0.71	0.68	0.77
	Regulatory environment	0.93	0.68	0.71	0.60	0.68	0.60	0.67
	Business environment	0.89	0.73	0.77	0.69	0.66	0.66	0.71
	Education	0.57	0.78	0.56	0.45	0.50	0.51	0.55
	Tertiary education	0.67	0.80	0.73	0.58	0.48	0.54	0.60
	Research and development (R&D)	0.69	0.88	0.76	0.74	0.83	0.85	0.74
	Information and communication technologies (ICTs)	0.80	0.85	0.94	0.75	0.68	0.72	0.82
	General infrastructure	0.57	0.53	0.69	0.47	0.49	0.56	0.47
	Ecological sustainability	0.65	0.59	0.77	0.54	0.55	0.53	0.67
	Credit	0.63	0.58	0.58	0.87	0.53	0.56	0.60
	Investment	0.53	0.47	0.42	0.71	0.52	0.44	0.42
	Trade, competition, & market scale	0.48	0.66	0.73	0.71	0.52	0.62	0.62
	Knowledge workers	0.69	0.79	0.72	0.67	0.86	0.72	0.67
	Innovation linkages	0.52	0.42	0.40	0.38	0.74	0.51	0.45
	Knowledge absorption	0.56	0.60	0.58	0.55	0.81	0.77	0.61
OUTPUT	Knowledge creation	0.62	0.79	0.64	0.65	0.78	0.89	0.76
	Knowledge impact	0.50	0.55	0.61	0.48	0.54	0.76	0.62
	Knowledge diffusion	0.59	0.60	0.61	0.58	0.69	0.80	0.59
	Intangible assets	0.61	0.63	0.70	0.59	0.55	0.67	0.91
	Creative goods and services	0.69	0.66	0.69	0.61	0.68	0.71	0.85
	Online creativity	0.82	0.80	0.82	0.71	0.75	0.79	0.88

Source: European Commission Joint Research Centre, 2017..

5.3.4 Foreign direct investment net inflows; 6.2.1 Growth rate of GDP per person engaged; and 7.2.4 Printing and publishing output. For two out of the seven indicators listed above—2.1.2 and 7.2.4—this is the first time that they are found to be non-influential at all in the GII framework. Instead, the remaining five indicators were found to be non-influential also in the GII 2016. On the other hand, two indicators that were found to be non-influential last year—3.3.1 GDP per unit of energy use and 4.1.3 Microfinance institutions' gross loan portfolio—are instead found to be influential in this year's framework. It is suggested that the GII development team carefully assess how these variables behave in the coming releases of the index. If the 'noisy' behaviour persists, these variables could eventually be removed from the GII framework.

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 close to 60% (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 modest adjustments made to the 2017 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 correlation coefficients are close to or greater than 0.70. (see Table 1).

The five input pillars share a single statistical dimension that summarizes 80% of the total variance, and the five loadings (correlation coefficients) of these pillars are very similar to each other (0.86–0.92). 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 0.94—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 (0.81); they are also both strongly correlated with the Innovation Output Sub-Index (0.95). This result suggests that the

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	14.8%	9.4%	3.9%	21.9%	17.2%	9.4%	3.1%
20–29	15.6%	14.8%	14.1%	10.2%	12.5%	11.7%	8.6%
10–19	23.4%	21.9%	28.1%	28.9%	18.8%	26.6%	30.5%
10 or more*	**53.9%**	**46.1%**	**46.1%**	**60.9%**	**48.4%**	**47.7%**	**42.2%**
5–9	21.1%	23.4%	25.8%	16.4%	22.7%	23.4%	19.5%
Less than 5	21.9%	26.6%	23.4%	18.8%	25.0%	25.8%	32.0%
Same rank	2.3%	3.1%	3.9%	3.1%	3.1%	2.3%	5.5%
Total†	**99.2%**	**99.2%**	**99.2%**	**99.2%**	**99.2%**	**99.2%**	**99.2%**
Pearson correlation coefficient with the GII	0.88	0.90	0.89	0.81	0.86	0.92	0.93

Source: European Commission Joint Research Centre, 2017.

* This column is the sum of the prior three rows.
† This column is the sum of all white rows.

Output Sub-Index is also well balanced in its two pillars. Furthermore, 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 0.89.

Finally, an important part of the analysis relates to clarifying the importance of the Input and Output Sub-Indices with respect to the variation of the GII scores. The GII is built as the simple arithmetic average of the five Input sub-pillars and the two Output sub-pillars, which implies that the Input-related pillars have a weight of 5/7 versus a weight of 2/7 for the Output-related pillars. Yet this does not imply that the Input aspect is more important than the Output aspect in determining the variation of the GII scores. In fact, the Pearson correlation coefficient of either sub-index with the overall GII is 0.97 (and the two sub-indices have a correlation of 0.89), which suggests that the sub-indices are effectively placed on equal footing.

Overall, the tests so far show that the grouping of variables into sub-pillars, pillars, and an overall index is statistically coherent in the GII 2017 framework, and that the GII has a balanced structure at e`ach aggregation level.

The only recommendation for next year relates to a careful reflection of the seven indicators discussed above—2.1.2 Government expenditure on education per pupil, secondary; 2.2.2 Graduates in science and engineering; 3.2.3 Gross capital formation; 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; and 7.2.4 Printing and publishing output—because their information content is lost in the aggregation at the pillar level or higher (sub-index and overall GII). For five out of the seven indicators (2.2.2, 3.2.3, 5.2.3, 5.3.4, 6.2.1) this was also the case in last year's audit.

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 confirm that this is not the case. In fact, for more than 42% (up to 61%) of the 127 economies included in the GII 2017, 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 policy making.

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

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: European Commission Joint Research Centre, 2017.

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

Modelling assumptions with a direct impact on the GII scores and rankings relate to:

- 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 (and how) or not to impute missing data,
- deciding whether (and how) or not to treat outliers,
- selecting the normalization approach to be applied,
- choosing the weights to be assigned, and
- deciding on the aggregation rule to be implemented.

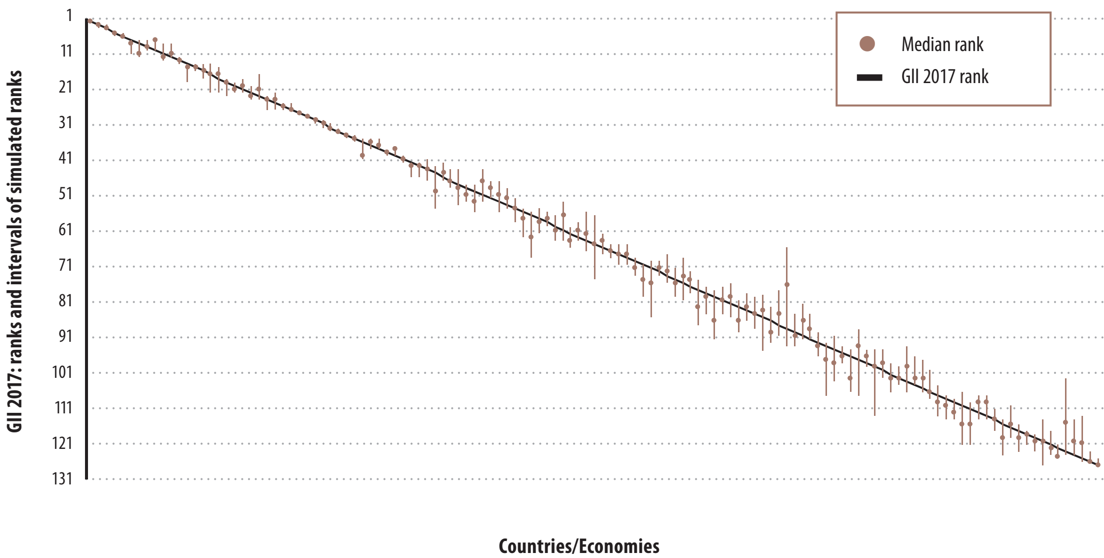
The rationale for these choices is manifold. For instance, expert opinion coupled with statistical analysis is behind the selection of the individual indicators, common practice and ease of interpretation suggests the use of a min-max normalization approach in the [0–100] range, the treatment of outliers 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 aggregate scores, resulting in error estimates and confidence intervals calculated for the GII 2017 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 treatment of missing data, the pillar weights, and the aggregation formula used at the pillar level.

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. Yet this is not the case for the GII. After 10 editions of the GII, the index-developing

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



Source: European Commission Joint Research Centre, 2017.

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 2017 rank is 0.997.

team has not encountered any intentional no-reporting strategy. The consequence of the 'no imputation' choice in an arithmetic average is that it is equivalent to replacing an indicator's missing value for a given country with the respective sub-pillar score. Hence, the available data (indicators) in the incomplete pillar may dominate, sometimes biasing the ranks up or down. To test the impact of the 'no imputation' choice, the JRC estimated missing data using the Expectation Maximization (EM) algorithm.[6]

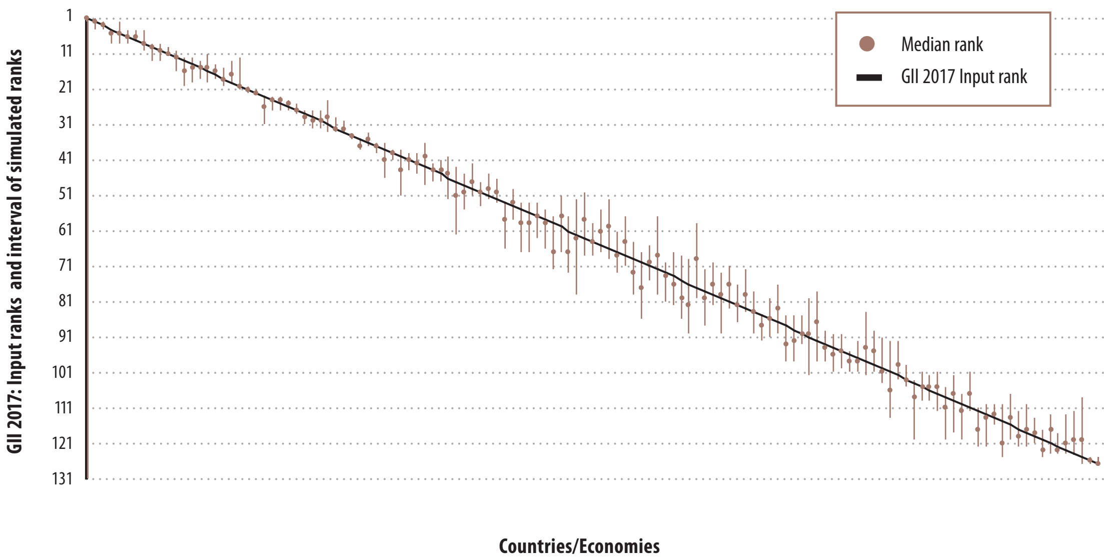
Regarding the aggregation formula, decision-theory practitioners challenge the use of simple arithmetic averages because of their fully compensatory nature, in which a comparative high advantage on a few indicators can compensate a comparative disadvantage on many indicators.[7] For example, one may argue that the United Kingdom and Germany, despite their similar performance at the Innovation Output Sub-Index—both close to 53.5 points (rank 6th and 7th respectively)—are very different if one considers how these countries perform within the sub-index. Germany ranks 8th in Knowledge and technology outputs and 7th in Creative outputs, while the United Kingdom is much more diverse: the country ranks 13th position in Knowledge and technology outputs, but it notably improves its overall position in the Output Sub-Index thanks to its 4th rank 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 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

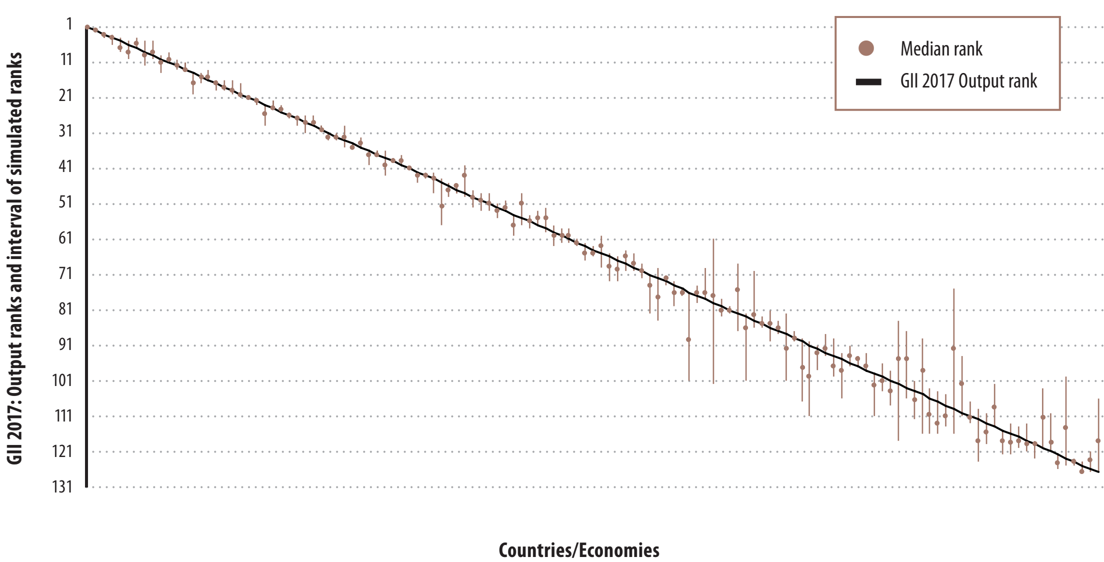
Figure 2b: Robustness analysis (Input rank vs. median rank, 90% confidence intervals)



Source: European Commission Joint Research Centre, 2017.

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 2017 rank is 0.997.

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



Source: European Commission Joint Research Centre, 2017.

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 2017 rank is 0.995.

Table 4: GII 2017 and Input/Output Sub-Indices: Ranks and 90% confidence intervals

	GII 2017		Input Sub-Index		Output Sub-Index	
Country/Economy	Rank	Interval	Rank	Interval	Rank	Interval
Switzerland	1	[1, 1]	3	[2, 4]	1	[1, 1]
Sweden	2	[2, 3]	2	[1, 4]	3	[3, 4]
Netherlands	3	[2, 3]	9	[8, 13]	2	[2, 2]
United States of America	4	[4, 5]	5	[2, 8]	5	[4, 8]
United Kingdom	5	[4, 5]	7	[4, 7]	6	[5, 10]
Denmark	6	[6, 10]	6	[4, 8]	12	[10, 13]
Singapore	7	[6, 11]	1	[1, 2]	17	[16, 19]
Finland	8	[6, 9]	4	[4, 8]	13	[11, 13]
Germany	9	[6, 9]	17	[14, 18]	7	[4, 7]
Ireland	10	[7, 12]	19	[13, 19]	8	[5, 12]
Korea, Republic of	11	[7, 11]	16	[11, 19]	9	[5, 10]
Luxembourg	12	[11, 13]	24	[23, 27]	4	[3, 6]
Iceland	13	[13, 18]	21	[20, 22]	10	[9, 14]
Japan	14	[13, 15]	11	[9, 11]	20	[17, 21]
France	15	[13, 17]	15	[13, 18]	18	[16, 19]
Hong Kong (China)	16	[13, 21]	8	[4, 10]	25	[23, 25]
Israel	17	[14, 21]	20	[12, 21]	14	[14, 20]
Canada	18	[17, 22]	10	[8, 13]	23	[23, 29]
Norway	19	[18, 21]	14	[12, 19]	22	[22, 23]
Austria	20	[17, 21]	18	[15, 20]	21	[20, 21]
New Zealand	21	[19, 23]	13	[12, 20]	24	[22, 24]
China	22	[16, 23]	31	[24, 33]	11	[8, 11]
Australia	23	[22, 26]	12	[10, 16]	30	[29, 30]
Czech Republic	24	[21, 26]	27	[25, 28]	16	[13, 16]
Estonia	25	[24, 26]	26	[24, 27]	19	[16, 20]
Malta	26	[24, 26]	28	[27, 31]	15	[14, 17]
Belgium	27	[27, 27]	22	[21, 22]	27	[26, 29]
Spain	28	[28, 28]	25	[23, 27]	26	[25, 27]
Italy	29	[29, 30]	29	[27, 32]	29	[26, 29]
Cyprus	30	[29, 31]	32	[29, 33]	28	[26, 31]
Portugal	31	[30, 32]	33	[30, 33]	31	[31, 33]
Slovenia	32	[31, 32]	30	[27, 32]	34	[34, 35]
Latvia	33	[33, 34]	35	[35, 38]	33	[29, 35]
Slovakia	34	[33, 35]	39	[38, 41]	35	[32, 35]
United Arab Emirates	35	[34, 40]	23	[23, 31]	56	[54, 58]
Bulgaria	36	[34, 37]	45	[41, 47]	32	[31, 33]
Malaysia	37	[34, 37]	36	[33, 37]	39	[38, 39]
Poland	38	[38, 39]	37	[36, 39]	41	[40, 41]
Hungary	39	[37, 39]	41	[39, 44]	37	[36, 37]
Lithuania	40	[39, 41]	34	[34, 35]	49	[47, 52]
Croatia	41	[41, 45]	44	[42, 47]	46	[45, 49]
Romania	42	[41, 45]	51	[45, 52]	44	[42, 48]
Turkey	43	[40, 46]	68	[57, 71]	36	[36, 40]
Greece	44	[42, 54]	38	[36, 46]	59	[57, 63]
Russian Federation	45	[41, 46]	43	[36, 48]	51	[48, 53]
Chile	46	[43, 48]	42	[39, 45]	53	[50, 53]
Viet Nam	47	[43, 53]	71	[65, 75]	38	[36, 43]
Montenegro	48	[47, 52]	50	[47, 54]	52	[51, 55]
Qatar	49	[47, 55]	48	[45, 55]	54	[54, 60]
Ukraine	50	[43, 52]	77	[59, 80]	40	[37, 40]
Thailand	51	[46, 51]	65	[55, 67]	43	[42, 44]
Mongolia	52	[46, 55]	67	[61, 73]	48	[40, 49]
Costa Rica	53	[48, 54]	57	[53, 63]	50	[48, 53]
Moldova, Republic of	54	[51, 58]	73	[70, 81]	42	[42, 45]
Saudi Arabia	55	[54, 62]	46	[40, 52]	66	[65, 73]
Kuwait	56	[55, 68]	80	[73, 90]	45	[44, 57]
South Africa	57	[54, 61]	49	[42, 51]	69	[65, 70]
Mexico	58	[55, 59]	54	[49, 58]	60	[58, 62]
Armenia	59	[56, 63]	82	[76, 87]	47	[46, 48]
India	60	[52, 63]	66	[52, 69]	58	[52, 59]
TFYR of Macedonia	61	[59, 65]	53	[53, 66]	63	[62, 67]
Serbia	62	[58, 63]	58	[55, 66]	61	[58, 62]
Panama	63	[55, 66]	74	[67, 82]	55	[48, 57]
Mauritius	64	[56, 74]	47	[43, 62]	82	[68, 87]

Table 4: GII 2017 and Input/Output Sub-Indices: Ranks and 90% confidence intervals *(continued)*

	GII 2017		Input Sub-Index		Output Sub-Index	
Country/Economy	Rank	Interval	Rank	Interval	Rank	Interval
Colombia	65	[61, 67]	52	[46, 53]	75	[75, 77]
Bahrain	66	[64, 68]	55	[53, 67]	67	[66, 73]
Uruguay	67	[64, 70]	61	[57, 73]	64	[63, 66]
Georgia	68	[64, 70]	69	[64, 79]	62	[61, 63]
Brazil	69	[68, 73]	60	[51, 67]	80	[78, 83]
Peru	70	[70, 79]	56	[53, 67]	85	[83, 86]
Brunei Darussalam	71	[69, 85]	40	[38, 51]	110	[94, 111]
Morocco	72	[69, 73]	79	[70, 82]	68	[64, 70]
Philippines	73	[67, 75]	83	[72, 84]	65	[60, 69]
Tunisia	74	[71, 79]	81	[71, 83]	71	[71, 82]
Iran, Islamic Republic of	75	[68, 80]	98	[84, 102]	57	[53, 57]
Argentina	76	[72, 78]	72	[57, 79]	81	[80, 82]
Oman	77	[74, 87]	62	[52, 79]	90	[89, 107]
Kazakhstan	78	[76, 84]	64	[59, 68]	93	[88, 94]
Dominican Republic	79	[77, 91]	88	[85, 98]	72	[69, 84]
Kenya	80	[76, 84]	91	[80, 102]	70	[68, 72]
Lebanon	81	[75, 85]	87	[76, 90]	78	[69, 78]
Azerbaijan	82	[80, 89]	78	[72, 88]	89	[87, 90]
Jordan	83	[78, 86]	92	[78, 98]	74	[73, 80]
Jamaica	84	[79, 88]	84	[78, 90]	84	[70, 86]
Paraguay	85	[79, 94]	90	[85, 93]	79	[61, 102]
Bosnia and Herzegovina	86	[82, 92]	75	[68, 86]	96	[91, 97]
Indonesia	87	[77, 90]	99	[89, 101]	73	[71, 74]
Belarus	88	[65, 93]	63	[50, 68]	109	[75, 116]
Botswana	89	[84, 93]	59	[57, 72]	111	[107, 113]
Sri Lanka	90	[81, 91]	94	[90, 101]	77	[74, 79]
Trinidad and Tobago	91	[84, 91]	85	[85, 92]	86	[81, 90]
Ecuador	92	[89, 96]	95	[90, 100]	83	[82, 101]
Albania	93	[92, 107]	70	[67, 86]	115	[115, 122]
Tajikistan	94	[90, 103]	100	[91, 104]	88	[82, 101]
Kyrgyzstan	95	[93, 98]	86	[80, 91]	104	[101, 112]
Tanzania, United Republic of	96	[94, 106]	109	[102, 118]	76	[76, 101]
Namibia	97	[88, 107]	89	[85, 98]	102	[84, 118]
Guatemala	98	[94, 99]	97	[92, 101]	92	[91, 98]
Rwanda	99	[94, 113]	76	[69, 90]	121	[110, 121]
Senegal	100	[94, 102]	102	[92, 103]	98	[93, 98]
Cambodia	101	[98, 106]	104	[103, 120]	87	[84, 88]
Uganda	102	[99, 104]	93	[89, 98]	106	[103, 116]
El Salvador	103	[93, 106]	96	[95, 101]	105	[89, 116]
Honduras	104	[96, 104]	103	[99, 105]	103	[87, 106]
Egypt	105	[97, 106]	106	[102, 109]	97	[94, 97]
Bolivia, Plurinational State of	106	[100, 108]	107	[101, 112]	99	[99, 111]
Mozambique	107	[104, 113]	114	[110, 116]	100	[96, 104]
Algeria	108	[107, 114]	105	[101, 109]	117	[114, 120]
Nepal	109	[108, 114]	108	[105, 120]	114	[102, 114]
Ethiopia	110	[106, 121]	122	[118, 124]	91	[90, 111]
Madagascar	111	[109, 121]	120	[117, 125]	95	[93, 106]
Côte d'Ivoire	112	[107, 114]	121	[113, 124]	94	[89, 100]
Pakistan	113	[107, 114]	116	[107, 120]	101	[98, 108]
Bangladesh	114	[111, 117]	113	[110, 122]	108	[105, 114]
Malawi	115	[114, 124]	112	[111, 122]	112	[109, 124]
Benin	116	[110, 119]	110	[107, 120]	120	[103, 120]
Cameroon	117	[115, 123]	117	[112, 122]	113	[110, 119]
Mali	118	[117, 121]	123	[113, 124]	107	[104, 116]
Nigeria	119	[118, 123]	118	[110, 122]	119	[118, 123]
Burkina Faso	120	[114, 127]	101	[92, 114]	126	[121, 127]
Zimbabwe	121	[117, 124]	124	[112, 124]	116	[113, 122]
Burundi	122	[121, 125]	115	[110, 125]	122	[122, 126]
Niger	123	[102, 124]	111	[101, 112]	123	[100, 125]
Zambia	124	[114, 124]	125	[108, 127]	118	[113, 121]
Togo	125	[113, 126]	119	[114, 121]	127	[106, 127]
Guinea	126	[123, 126]	126	[125, 127]	124	[123, 125]
Yemen	127	[125, 127]	127	[125, 127]	125	[124, 127]

Source: European Commission Joint Research Centre, 2017.

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*		Number of economies that *deteriorate*	
		by 20 or more positions	between 10 and 19 positions	by 20 or more positions	between 10 and 19 positions
GII	Geometric vs. arithmetic average	0	1	0	3
	EM imputation vs. no imputation of missing data	0	3	0	3
	Geometric average and EM imputation vs. arithmetic average and missing values	1 (Belarus)	3	0	3
Input Sub-Index	Geometric vs. arithmetic average	0	0	0	1
	EM imputation vs. no imputation of missing data	0	2	0	2
	Geometric average and EM imputation vs. arithmetic average and missing values	0	5	0	7
Output Sub-Index	Geometric vs. arithmetic average	0	0	0	3
	EM imputation vs. no imputation of missing data	1 (Belarus)	10	1 (Tanzania, U. Rep.)	7
	Geometric average and EM imputation vs. arithmetic average and missing values	1 (Belarus)	9	1 (Tanzania, U. Rep.)	7

Source: European Commission Joint Research Centre, 2017.

according to their reference rank (black line), the dot being the median rank over the simulations.

All published GII 2017 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 for 105 of the 127 economies. However, it is also true that merely two country ranks vary significantly with changes in weights and aggregation formula and because of the estimation of missing data. These two countries—Niger and Belarus—have 90% confidence interval widths of 22 and 28, respectively; hence their GII ranks should be interpreted cautiously and certainly not taken at face value. This is a remarkable improvement compared to the GII 2015, where confidence interval widths for 32 economies lay between 20 and 29, for another 7 economies between 30 and 39, and for 2 economies the widths were 40 or greater. This improvement in the confidence one can attach to the GII 2017 ranks is the direct result of the developers' choice since 2016 to adopt a more stringent criterion for an economy's inclusion, which requires at least 62% data availability within each of the two sub-indices. Some caution is also warranted in the Input Sub-Index for 7 economies—Ukraine, Argentina, Oman, Kenya, Jordan, Rwanda, and Burkina Faso—that have 90% confidence interval widths over 20 (up to 27 for Oman). The Output Sub-Index is slightly more sensitive to the methodological choices: 8 countries—Paraguay, Belarus, the United Republic of Tanzania, Namibia, El Salvador, Ethiopia, Niger, and Togo—have 90% confidence interval widths over 20 (up to 41 for Paraguay and Belarus). 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.

Although a few economy ranks, in the GII 2017 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, or in the Input and Output Sub-Index.

For full transparency and information, Table 4 reports the GII 2017 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 changes of the EM imputation method and/or the geometric aggregation formula, with fixed weights at their reference values (as in the original GII). Similar to last year's results, this year neither the GII nor the Input or Output Sub-Index are found to be heavily influenced by

Table 6: Pie shares (absolute terms) and efficiency scores for the top 25 economies in the GII 2017

	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	0.09	0.17	0.08	0.09	0.19	0.19	0.19	1	1	0	2	–1
Sweden	0.20	0.20	0.20	0.10	0.20	0.05	0.05	2	2	0	12	–10
Netherlands	0.20	0.05	0.20	0.05	0.20	0.10	0.20	5	3	–2	4	–1
United States of America	0.20	0.20	0.20	0.20	0.10	0.05	0.05	5	4	–1	21	–17
United Kingdom	0.20	0.20	0.20	0.20	0.05	0.05	0.10	4	5	1	20	–15
Denmark	0.20	0.20	0.20	0.20	0.05	0.05	0.10	5	6	1	34	–28
Singapore	0.20	0.20	0.20	0.20	0.10	0.05	0.05	2	7	5	63	–56
Finland	0.20	0.20	0.20	0.10	0.20	0.05	0.05	5	8	3	37	–29
Germany	0.20	0.20	0.20	0.10	0.05	0.05	0.20	9	9	0	7	2
Ireland	0.20	0.20	0.20	0.10	0.20	0.05	0.05	14	10	–4	6	4
Korea, Republic of	0.20	0.20	0.20	0.20	0.10	0.05	0.05	11	11	0	14	–3
Luxembourg	0.20	0.05	0.20	0.20	0.10	0.05	0.20	16	12	-4	1	11
Iceland	0.20	0.10	0.20	0.05	0.20	0.05	0.20	19	13	-6	5	8
Japan	0.20	0.20	0.20	0.20	0.10	0.05	0.05	9	14	5	49	–35
France	0.20	0.20	0.20	0.20	0.05	0.05	0.10	14	15	1	35	–20
Hong Kong (China)	0.20	0.10	0.20	0.20	0.20	0.05	0.05	11	16	5	73	–57
Israel	0.10	0.20	0.20	0.20	0.20	0.05	0.05	19	17	–2	23	–6
Canada	0.20	0.20	0.20	0.20	0.10	0.05	0.05	11	18	7	59	–41
Norway	0.20	0.20	0.20	0.20	0.10	0.05	0.05	19	19	0	51	–32
Austria	0.20	0.20	0.20	0.10	0.20	0.05	0.05	19	20	1	41	–21
New Zealand	0.20	0.20	0.20	0.20	0.05	0.05	0.10	16	21	5	56	–35
China	0.05	0.10	0.20	0.20	0.20	0.20	0.05	23	22	–1	3	19
Australia	0.20	0.20	0.20	0.20	0.05	0.05	0.10	16	23	7	76	–53
Czech Republic	0.20	0.20	0.20	0.10	0.05	0.05	0.20	28	24	–4	13	11
Estonia	0.20	0.05	0.20	0.20	0.10	0.05	0.20	23	25	2	19	6

Source: European Commission Joint Research Centre, 2017.

Notes: Pie shares are in absolute terms, bounded by 0.05 and 0.20. In the GII 2017, 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.

the imputation of missing data or the aggregation formula. Depending on the combination of the choices made, only Belarus or the United Republic of Tanzania can change rank by 20 positions or more.

All in all, the published GII 2017 ranks are reliable and for the vast majority of countries the simulated 90% confidence intervals are narrow enough for meaningful inferences to be drawn. Nevertheless, the readers of the GII 2017 report should consider country ranks in the GII 2017 and in the Input and Output Sub-Indices not only at face value but also within the 90% confidence intervals in order to better appreciate to what degree a country's rank depends on the modelling choices. Since 2016, following the JRC recommendation in past GII audits, the developers' choice to apply the 66% indicator coverage threshold separately to the Input and Output Sub-Indices in the GII 2017 has led to a net increase in the reliability of country ranks for the GII and the two sub-indices. Furthermore, the adoption of less stringent criterion for the skewness and kurtosis (greater than 2.25 in absolute value and greater than 3.5, respectively) has not introduced any bias in the estimates.

Efficiency frontier in the GII by Data Envelopment Analysis

Is there 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?

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

Figure 3: GII 2017 scores and DEA 'distance to the efficient frontier' scores

Source: European Commission Joint Research Centre, 2017.

Note: For comparison purposes, we have rescaled the GII scores by dividing them with the best performer in the overall GII 2017.

weighting scheme might not be fair to a particular country. An appealing feature of the 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 applied 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 2017, next to the GII 2017 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, Switzerland is the only country this year that obtains a perfect DEA score of 1 by assigning 19% of its DEA score to Business sophistication, Knowledge and technology outputs, and Creative outputs, while merely 8% to 9% of its DEA score comes from Institutions, Infrastructure, and Market sophistication. Instead, countries including the United States of America, the United Kingdom, Denmark, and Singapore would assign 20% of their DEA scores to Market sophistication. Only Switzerland reaches a perfect DEA score of 1, closely followed by Sweden, the Netherlands, the United States of America, the United

Kingdom, Denmark, Singapore, and Finland, which score between 0.96 (Finland) and 0.99 (Sweden) in terms of efficiency. Figure 3 shows how close the DEA scores and the GII 2017 scores are for all 127 economies (correlation of 0.99). [10] Note that by construction, the version of 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 (with a correlation of 0.63).

Conclusion

The JRC analysis suggests that the conceptualized multi-level structure of the GII 2017—with its 81 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 81 indicators because their capacity to distinguish countries' performance is lost in the aggregation at the pillar level or higher. Five indicators related to the inputs of innovation— 2.1.2 Government expenditure on education per pupil, secondary; 2.2.2 Graduates in science and engineering; 3.2.3 Gross capital formation; 5.2.3 GERD financed by abroad; 5.3.4 Foreign direct investment net inflows—and two indicators related to the outputs of innovation, 6.2.1 Growth rate of GDP per person engaged and 7.2.4 Printing and publishing output, 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. The adoption, since 2016, by the GII team of a more stringent data coverage threshold (at least 66% 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 was made in 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 42.2% (up to 60.9%) of the 127 economies included in the GII 2017, 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. 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 2017 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. simulated), 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 83% (105 out of 127) of the economies. Some caution is needed mainly for two countries—Belarus and Niger—with ranks that are highly sensitive to the methodological choices. The Input and the 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 2017 or in the two sub-indices, 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 or either of the Input and Output Sub-Indices.

All things considered, the present JRC audit findings confirm that the GII 2017 meets international quality standards for statistical soundness, which indicates that the GII index is a reliable benchmarking tool for innovation practices at the country level around the world.

Finally, the 'distance to the efficient frontier' measure calculated with Data Envelopment Analysis could complement the Innovation Efficiency Ratio as a measure of efficiency, even if it is conceptually closer to the GII score than to the efficiency ratio.

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 (127 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 2017 for $n = 127$ 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:

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

subject to

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

where

$j = 1, \ldots, 7,$
$i = 1, \ldots, 127$

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

10 Instead, only Switzerland achieved a 1.0 score in the Innovation Efficiency Ratio, calculated as the ratio of the Output Sub-Index over the Input Sub-Index. The Efficiency Ratio and the DEA score 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 and related reading

Barbosa, N. and A. P. Faria. 2011. 'Innovation across Europe: How important are institutional differences'. *Research Policy* 40: 1157–69.

Becker, W., M. Saisana, P. Paruolo, and I. Vandecasteele. 2017. 'Weights and Importance in Composite Indicators: Closing the Gap. *Ecological Indicators* 80: 12–22.

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.

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

OECD/EC JRC (Organisation for Economic Co-operation and Development/European Commission, Joint Research Centre). 2008. *Handbook on Constructing Composite Indicators: Methodology and User Guide.* Paris: OECD.

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

Saisana, M., B. D'Hombres, and A. Saltelli. 2011. 'Rickety Numbers: Volatility of University Rankings and Policy Implications'. *Research Policy* 40: 165–77.

Saisana, M., A. Saltelli, and S. Tarantola. 2005. 'Uncertainty and Sensitivity Analysis Techniques as Tools for the Analysis and Validation of Composite Indicators'. *Journal of the Royal Statistical Society* A 168 (2): 307–23.

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. 2008. *Global Sensitivity Analysis: The Primer.* Chichester, England: John Wiley & Sons.

Schneider, T. 2001. 'Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. *Journal of Climate* 14, 853–871.