APPENDIX V

JOINT RESEARCH CENTRE (JRC) STATISTICAL AUDIT OF THE 2019 GLOBAL INNOVATION INDEX

Michaela Saisana, **Marcos Domínguez-Torreiro**, **Daniel Vértesy**, and **Marcos Alvarez**,
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 12th edition, the Global Innovation Index (GII) 2019 considers these conceptual challenges in Chapter 1 and deals with practical challenges—related to data quality and methodological choices—by grouping economy-level data over 129 economies; and across 80 indicators into 21 sub-pillars, 7 pillars, 2 sub-indices and, finally, an overall index. This appendix offers detailed insights into the practical issues related to the construction of the GII, analysing 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 (COIN) at the Joint Research Centre (JRC) in Ispra has been invited for the ninth consecutive year to audit the GII. As in previous editions, the present JRC-COIN audit focuses on the statistical soundness of the multi-level structure of the index as well as on the impact of key modeling assumptions on the results.[2] The independent statistical assessment of the GII provided by the JRC-COIN 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 the innovation field.

As in past GII reports, the JRC-COIN analysis complements the economy 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. Finally, the JRC-COIN 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-COIN in April/May 2019. Fine-tuning suggestions were taken into account in the final computation of the rankings in an iterative process with the JRC-COIN aimed at setting the foundation for a balanced index. The entire process followed four steps (Figure A-V.1).

Step 1: conceptual consistency

Eighty indicators were selected for their relevance to a specific innovation pillar based on literature review, expert opinion, economy coverage, and timeliness. To represent a fair picture of economy differences, indicators were scaled either at source or by the GII team, as appropriate, and where needed. For example, Expenditure on education (indicator 2.1.1) is expressed as a percentage of GDP, while Government funding per pupil at secondary level (indicator 2.1.2), is expressed as a percentage of GDP per capita.

Step 2: data checks

The data, which were most recently released within the period 2008 to 2018, were used for each economy: 78% of the available data refer to 2017 or more recent years. The exception are data values for six economies: Argentina, Lebanon, Trinidad and Tobago, Pakistan, Ghana, and Madagascar, on Printing & other media, % manufacturing (indicator 7.2.4) that refer to the period 2002 to 2007. The JRC-COIN recommendation was to offer an explanation behind the choice to use data that may not reflect recent advances in the relevant field in these economies (Appendix III). In past editions, until 2015, economies were included if data availability was at least 60% across all variables in the GII framework. More stringent criterion were adopted in 2016, following the JRC-COIN recommendation in past GII audits, where economies were only included if data availability was at least 66% within each of the two sub-indices (i.e., 35 out of 53 variables within the Input Sub-Index and 18 out of the 27 variables in the Output Sub-Index) and where at least two of the three sub-pillars in each pillar could be computed. These

FIGURE A-V.1

Conceptual and statistical coherence in the GII 2019 framework

STEP 4. QUALITATIVE REVIEW

Internal qualitative review (INSEAD, WIPO, and Cornell University)

External qualitative review (JRC-COIN, international experts)

STEP 3. STATISTICAL COHERENCE

Treatment of pairs of highly collinear variables as a single indicator

Assessment of grouping indicators into sub-pillars, pillars, sub-indices, and 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 (78% of available data refer to 2017 and 2018)

Availability requirements per economy: coverage ≥66% for the Input and the Output Sub-Indices, separately and data availability for at least two sub-pillars per pillar

Check for reporting errors (interquartile range)

Outlier identification (skewness and kurtosis) and treatment (winsorisation or logarithmic transformation)

Direct contact with data providers

STEP 1. CONCEPTUAL CONSISTENCY

Compatibility with existing literature on innovation and pillar definition

Use of scaling factors (denominators) per indicator to represent a fair picture of country differences (e.g., GDP, population)

Source: European Commission, Joint Research Centre, 2019.

criterion aim to ensure that economy scores for the GII and for the two Input and Output Sub-Indices are not particularly sensitive to missing values (as was the case for the Output Sub-Index scores of several economies in past editions).
In practice, data availability for all economies included in the GII 2019 is good: 80% of data is available for 87% of the economies (equivalent to 112 economies out of 129). 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. Since 2011, and decided jointly with the JRC-COIN, values were treated if the indicators had absolute skewness greater than 2.0 and kurtosis greater than 3.5.
In 2017, and after having analyzed data in the GII 2011 to the GII 2017, a less stringent criterion were adopted. An indicator was only treated if the absolute skewness was greater than 2.25 and kurtosis greater than 3.5.[3] These indicators were treated either by winsorization or by natural logarithm (in cases of more than five outliers; Appendix IV: Technical Notes). In 2018, an exceptional behaviour for FDI net outflows (indicator 6.3.4) was observed (Chapter 1, Annex 3, JRC Audit, 2018) and from 2018 on, it was recommended to adjust the GII rule for the treatment of outliers as follows:

(a) for indicators with absolute skewness greater than 2.25 and kurtosis greater than 3.5, apply either winsorization or the natural logarithm (in case of more than five outliers);
(b) for indicators with absolute skewness of less than 2.25 and kurtosis greater than 10.0, produce scatterplots to identify potentially problematic values that need to be considered as outliers and treated accordingly.

Step 3: statistical coherence

Weights as scaling coefficients

Jointly decided between the JRC-COIN and the GII team in 2012, weights of 0.5 or 1.0 were 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, W. et al. (2017) and Paruolo, P. 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 80 indicators and two sub-pillars—7.2 Creative goods and services and 7.3 Creation of online content—were assigned half weights, while all other indicators and sub-pillars were assigned a weight of 1.0. In past GII editions, despite this weighting adjustment, a small number of indicators (seven in the GII 2017 edition) were found to be non-influential in the GII framework, implying that they could not explain at least 9% of economy variation in the respective sub-pillar scores.[4] This year, as it was the case also in 2018, all 80 indicators are found to be sufficiently influential in the GII framework, which is worthy highlighting as a very positive feature of this year's 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 55% (pillar 4: Market sophistication) up to 83% (pillar 1: Institutions) of the total variance in the three underlying sub-pillars. 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. (Table A-V.1).

The five input pillars share a single statistical dimension that summarizes 82% of the total variance, and the five loadings (correlation coefficients) of these pillars are very similar to each other (0.84–0.93). 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.[5]

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 (0.94 to 0.96).

Finally, an important part of the analysis relates to clarifying the importance of the Input and Output Sub-Indices with respect to variation in the GII scores. The GII is built as a 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 the Input or the Output 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 2019 framework, and that the GII has a balanced structure at each aggregation level. Furthermore, this year, all 80 indicators are found to be sufficiently influential in the GII framework, namely each indicator explains at least 9% of countries variation in the respective sub-pillar scores, which is worthy highlighting as a very positive feature of this year's GII framework.[6]

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

TABLE A-V.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 & technology outputs	Creative outputs
Innovation Input Sub-index	1.1. Political environment	0.95	0.81	0.88	0.73	0.82	0.73	0.82
	1.2. Regulatory environment	0.92	0.70	0.72	0.63	0.74	0.64	0.74
	1.3. Business environment	0.86	0.69	0.70	0.63	0.68	0.65	0.61
	2.1. Education	0.60	0.81	0.61	0.50	0.57	0.55	0.55
	2.2. Tertiary education	0.62	0.81	0.69	0.52	0.50	0.52	0.56
	2.3. Research and development (R&D)	0.77	0.88	0.76	0.69	0.88	0.86	0.74
	3.1. Information and communication technologies (ICTs)	0.81	0.84	0.94	0.71	0.74	0.71	0.78
	3.2. General infrastructure	0.56	0.54	0.70	0.48	0.50	0.50	0.48
	3.3. Ecological sustainability	0.64	0.56	0.75	0.44	0.61	0.56	0.69
	4.1. Credit	0.70	0.62	0.61	0.88	0.62	0.55	0.62
	4.2. Investment	0.35	0.26	0.21	0.63	0.28	0.23	0.21
	4.3. Trade, competition, and market scale	0.53	0.68	0.71	0.68	0.61	0.66	0.59
	5.1. Knowledge workers	0.79	0.83	0.79	0.69	0.89	0.78	0.75
	5.2. Innovation linkages	0.63	0.57	0.53	0.52	0.81	0.67	0.66
	5.3. Knowledge absorption	0.65	0.64	0.62	0.49	0.85	0.80	0.65
Innovation Output Sub-index	6.1. Knowledge creation	0.71	0.81	0.69	0.65	0.84	0.90	0.79
	6.2. Knowledge impact	0.54	0.60	0.59	0.47	0.57	0.80	0.60
	6.3. Knowledge diffusion	0.65	0.65	0.64	0.54	0.82	0.88	0.66
	7.1. Intangible assets	0.62	0.61	0.68	0.53	0.63	0.66	0.88
	7.2. Creative goods and services	0.67	0.61	0.69	0.59	0.65	0.64	0.82
	7.3. Online creativity	0.80	0.72	0.73	0.59	0.82	0.76	0.84

Source: European Commission Joint Research Centre, 2019.

TABLE A-V.2

Statistical coherence in the GII: correlations between sub-pillars and pillars

Rank differences (positions)	Innovation Input Sub-Index: Institutions %	Innovation Input Sub-Index: Human capital and research %	Innovation Input Sub-Index: Infrastructure %	Innovation Input Sub-Index: Market sophistication %	Innovation Input Sub-Index: Business sophistication %	Innovation Output Sub-Index: Knowledge & technology outputs %	Innovation Output Sub-Index: Creative outputs %
More than 30	12.4%	10.1%	10.1%	24.0%	11.6%	10.1%	7.0%
20-29	13.2%	13.2%	10.1%	17.8%	14.0%	12.4%	10.1%
10-19	28.7%	30.2%	24.0%	30.2%	20.2%	25.6%	21.7%
10 or more*	54.3%	53.5%	44.2%	72.1%	45.7%	48.1%	38.8%
5-9	22.5%	24.0%	29.5%	7.0%	19.4%	25.6%	27.1%
Less than 5	20.9%	22.5%	25.6%	17.1%	31.0%	22.5%	30.2%
Same rank	2.3%	0.0%	0.8%	3.9%	3.9%	3.9%	3.9%
Total**	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Pearson correlation coefficient with the GII	0.90	0.90	0.89	0.78	0.92	0.93	0.92

Source: European Commission Joint Research Centre, 2019.
Notes: *This column is the sum of the prior three rows. **This column is the sum of all white rows.

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-COIN confirm that this is not the case. In fact, for more than 44% (up to 72%) of the 129 economies included in the GII 2019, the GII ranking and any of the seven pillar rankings differ by 10 positions or more (Table A-V.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, economy-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 economy 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 modeling assumptions on the GII results

An important part of the GII statistical audit is to check the effect of varying assumptions inside plausible ranges. Modeling 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 modeling choices is accounted for in the robustness assessment carried out by the JRC-COIN. 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 2019 individual economy rankings.

As suggested in the relevant literature on composite indicators,[7] the robustness assessment was based on Monte Carlo simulation and multi-modeling 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 modeling 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 centered in the reference values. The ranges of simulated weights were defined by considering 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 between 10% and 30%; the limit values for the two output pillars are between 40% and 60%. (Table A-V.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 12 editions of the GII, the index-developing 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 economy 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-COIN estimated missing data using the Expectation Maximization (EM) algorithm that was applied within each GII pillar.[8]

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.[9] To assess the impact of this compensability issue, the JRC-COIN 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.[10]

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 (Table A-V.3 for a summary of the uncertainties considered).

Uncertainty analysis results

The main results of the robustness analysis are shown in Figure A-V.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 inn ascending order (best to worst) according to their reference rank (black line), the dot being the median rank over the simulations.

TABLE A-V.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		

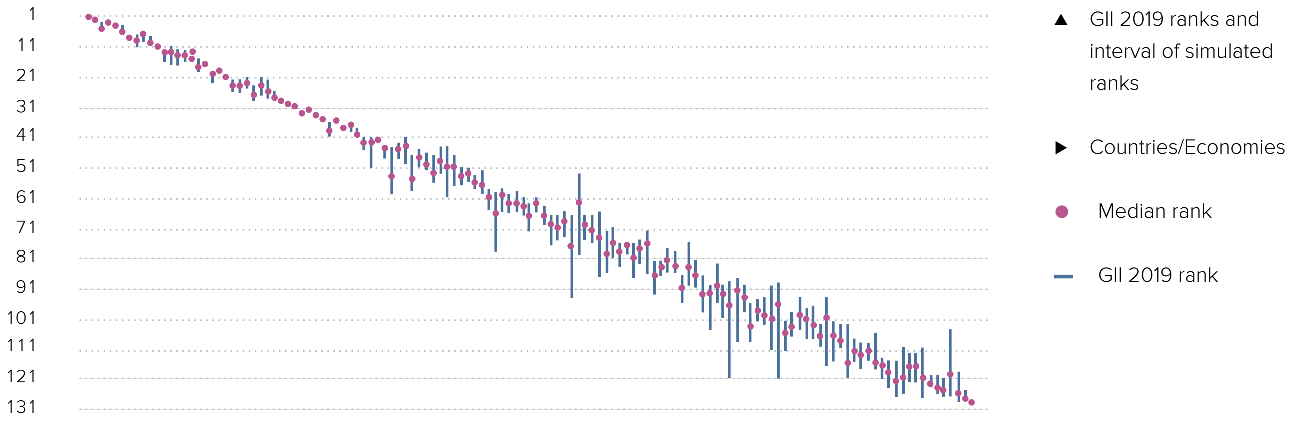
GII Sub-Index	Pillar	Reference value for the weight	Distribution assigned for robustness analysis
Innovation Input	Institutions	0.2	U[0.1,0.3]
	Human capital and research	0.2	U[0.1,0.3]
	Infrastructure	0.2	U[0.1,0.3]
	Market sophistication	0.2	U[0.1,0.3]
	Business sophistication	0.2	U[0.1,0.3]
Innovation 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, 2019.

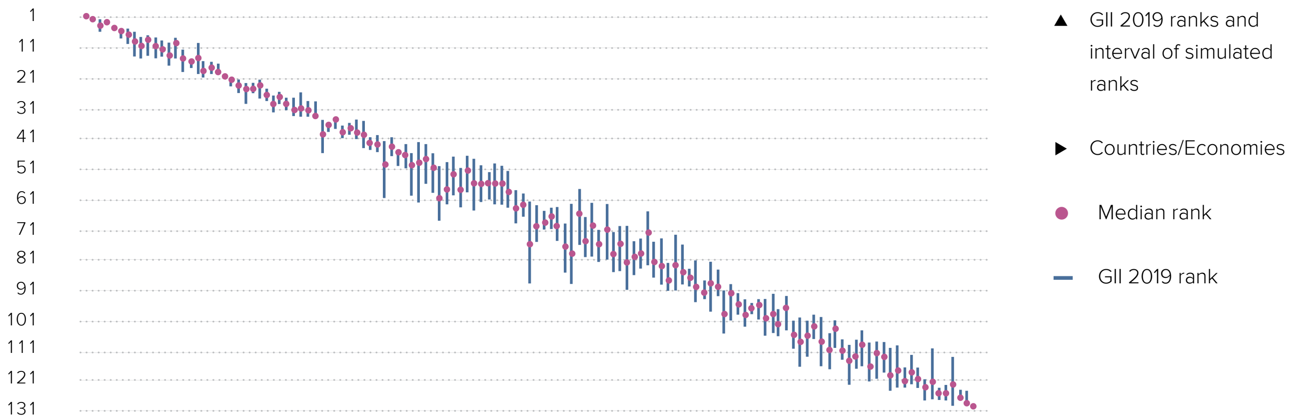
FIGURE A-V.2

Robustness analysis of the GII and Input and Output Sub-Indices

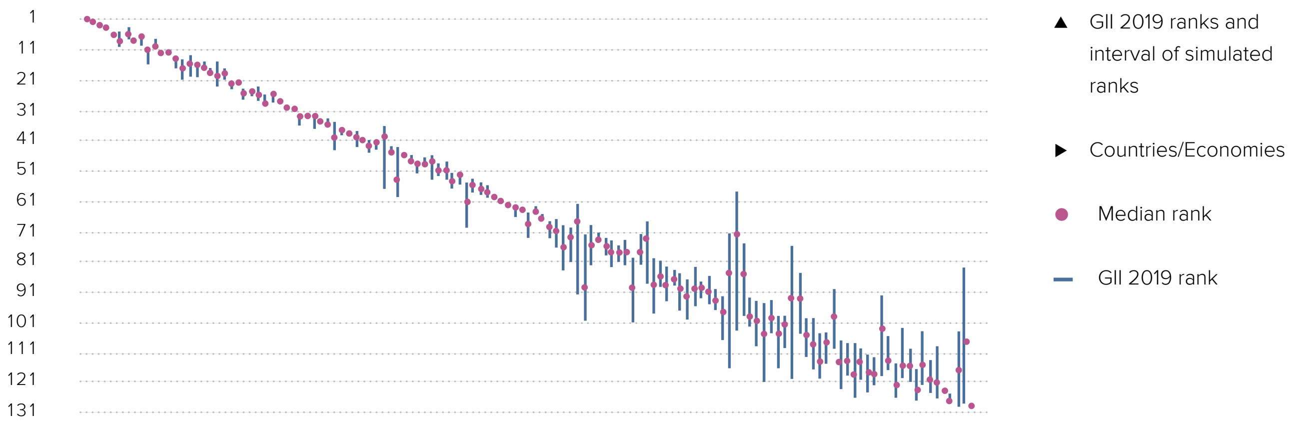
GII rank vs. median rank, 90% confidence intervals



Input rank vs. median rank, 90% confidence intervals



Output rank vs. median rank, 90% confidence



Source: European Commission Joint Research Centre, 2019.
Notes: Median ranks and intervals are calculated over 4,000 simulated scenarios combining simulated 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 GII 2019 rank is 0.997; between the median rank and Innovation Input 2019 rank it is 0.997; and between the median rank and the Innovation Output 2019 rank it is 0.992.

All published GII 2019 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 98 of the 129 economies. However, it is also true that ranks for a few economies vary significantly with changes in weights and aggregation formula and because of the estimation of missing data. Nine economies, Brunei Darussalam, Belarus, Panama, Rwanda, Paraguay, Tajikistan, Namibia, El Salvador, and Togo have 90% confidence interval widths of 20 positions (up to 32 positions in the case of Rwanda and Namibia). Consequently, their GII ranks—between the 71st (Brunei Darussalam) and 126th position (Togo) in the GII classification—should be interpreted cautiously and certainly not taken at face value. This is a remarkable improvement compared to GII versions until 2016, where more than 40 economies had confidence interval widths of more than 20 positions. The improvement in the confidence that one can attach to the GII 2019 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 66% data availability within each of the two sub-indices. Some caution is also warranted in the Input Sub-Index for 3 economies— Panama, Bosnia and Herzegovina, and Rwanda —that have 90% confidence interval widths over 20 (up to 27 for Rwanda). The Output Sub-Index is slightly more sensitive to the methodological choices: 13 economies, Mongolia, Belarus, Panama, Mauritius, Lebanon, Trinidad and Tobago, Paraguay, the United Republic of Tanzania, Namibia, El Salvador, Ethiopia, Togo, and the Niger, have 90% confidence interval widths over 20 (up to 46 for 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 economy ranks in the Innovation Output Sub-Index.

Although ranks for a few economies, in the GII 2019 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 A-V.4 reports the GII 2019 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).

Emphasizing the identification of and relation between input and output indicators seems irresistible from a policy perspective since doing so may possibly shed light on the effectiveness of innovation systems and policies. Yet, last year's statistical audit concluded that innovation efficiency ratios, calculated as ratios of indices, have to be approached with care. The reason was that the simulated 90% confidence intervals for most economies were too wide for meaningful inferences to be drawn: there was a shift of more than 20 positions for 50% of the economies. Hence, whilst propagating the uncertainty in the two GII sub-indices over to their sum the GII had a modest impact to the rankings, this same uncertainty propagation over to their ratio had a very high impact on the economy ranks. This is not a challenge specific to the GII framework per se but a statistical property that comes with ratios of composite indicators. In this present audit, the JRC-COIN complements the GII team for having opted to drop the Efficiency Ratio in this year's publication, drawing instead policy inference on the Input-Output performance in a similar way as per the plot of GII scores against the economies' level of economic development and commenting on those pairs/groups of economies that have similar Innovation Input level but very different Innovation Output level, and vice versa.

Sensitivity analysis results

Complementary to the uncertainty analysis, sensitivity analysis has been used to identify which of the modeling assumptions have the highest impact on certain country ranks. Table A-V.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 the imputation of missing data, or the aggregation formula. Depending on the combination of the choices made, only nine economies, Belarus, Paraguay, Namibia, El Salvador, Togo, the Niger, Brunei Darussalam, Rwanda, the United Republic of Tanzania, shift rank by 20 positions or more.

All in all, the published GII 2019 ranks are reliable and for most economies the simulated 90% confidence intervals are narrow enough for meaningful inferences to be drawn. Nevertheless, the readers of the GII 2019 report should consider economy ranks in the GII 2019 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 an economy's rank depends on the modeling choices. These confidence intervals have to be taken into account also when comparing economy rank changes from one year to another at the GII or Innovation Sub-indices level in order to avoid drawing erroneous conclusions on economies' ascent or descent in the overall classifications. Since 2016, following the JRC-COIN 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 2019 has led to a net increase in the reliability of economy ranks for the GII and the two sub-indices. Furthermore, the adoption in 2017 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.

TABLE A-V.4

GII 2019 and Input/Output Sub-Indices: ranks and 90% confidence intervals

	GII 2019		Input Sub-Index		Output Sub-Index	
	Rank	Interval	Rank	Interval	Rank	Interval
Switzerland	1	[1, 1]	2	[2, 3]	1	[1, 1]
Sweden	2	[2, 2]	4	[2, 4]	3	[3, 3]
United States of America	3	[3, 6]	3	[2, 6]	6	[5, 10]
Netherlands	4	[3, 5]	11	[8, 15]	2	[2, 2]
United Kingdom	5	[3, 5]	6	[6, 8]	4	[4, 5]
Finland	6	[4, 6]	7	[5, 10]	7	[4, 8]
Denmark	7	[7, 9]	5	[4, 6]	12	[11, 13]
Singapore	8	[7, 11]	1	[1, 2]	15	[14, 21]
Germany	9	[7, 9]	12	[9, 14]	9	[7, 9]
Israel	10	[8, 10]	17	[10, 20]	8	[7, 9]
Republic of Korea	11	[10, 12]	10	[7, 14]	13	[12, 13]
Ireland	12	[12, 16]	20	[17, 20]	10	[10, 16]
Hong Kong, China	13	[11, 17]	8	[6, 14]	16	[13, 20]
China	14	[12, 17]	26	[22, 28]	5	[5, 6]
Japan	15	[12, 16]	14	[8, 15]	17	[15, 20]
France	16	[14, 16]	16	[15, 18]	14	[14, 17]
Canada	17	[15, 19]	9	[8, 15]	22	[21, 24]
Luxembourg	18	[16, 18]	23	[22, 26]	11	[8, 11]
Norway	19	[19, 23]	13	[10, 17]	27	[26, 29]
Iceland	20	[18, 20]	22	[22, 24]	18	[15, 18]
Austria	21	[20, 21]	19	[16, 20]	25	[24, 26]
Australia	22	[22, 26]	15	[12, 19]	31	[30, 31]
Belgium	23	[22, 26]	21	[20, 21]	24	[24, 27]
Estonia	24	[21, 25]	27	[25, 29]	19	[17, 20]
New Zealand	25	[24, 29]	18	[16, 21]	32	[32, 36]
Czech Republic	26	[21, 27]	29	[26, 30]	21	[17, 21]
Malta	27	[22, 28]	32	[26, 34]	20	[15, 23]
Cyprus	28	[26, 29]	28	[27, 33]	23	[22, 23]
Spain	29	[28, 29]	25	[23, 26]	28	[25, 28]
Italy	30	[30, 30]	30	[28, 32]	29	[28, 29]
Slovenia	31	[31, 32]	33	[29, 34]	30	[30, 31]
Portugal	32	[32, 33]	31	[28, 34]	35	[34, 36]
Hungary	33	[31, 33]	39	[36, 40]	26	[23, 28]
Latvia	34	[34, 35]	36	[36, 39]	34	[32, 37]
Malaysia	35	[34, 36]	34	[29, 34]	39	[38, 40]
United Arab Emirates	36	[36, 41]	24	[23, 30]	58	[55, 59]
Slovakia	37	[36, 37]	42	[41, 45]	33	[32, 34]
Lithuania	38	[37, 39]	38	[37, 41]	40	[38, 43]
Poland	39	[37, 39]	37	[35, 38]	41	[40, 42]
Bulgaria	40	[38, 40]	45	[41, 47]	38	[37, 39]
Greece	41	[41, 45]	40	[35, 41]	54	[52, 57]
Viet Nam	42	[41, 51]	63	[58, 69]	37	[35, 44]
Thailand	43	[41, 43]	47	[43, 51]	43	[42, 44]
Croatia	44	[44, 48]	46	[45, 50]	52	[49, 53]
Montenegro	45	[44, 60]	55	[51, 64]	46	[43, 60]
Russian Federation	46	[43, 48]	41	[36, 44]	59	[56, 60]
Ukraine	47	[41, 50]	82	[65, 83]	36	[34, 37]
Georgia	48	[47, 59]	44	[42, 61]	60	[59, 60]
Turkey	49	[45, 51]	56	[47, 59]	49	[48, 52]
Romania	50	[46, 52]	54	[47, 60]	53	[48, 54]
Chile	51	[47, 56]	43	[40, 46]	62	[61, 63]
India	52	[44, 53]	61	[50, 63]	51	[46, 54]
Mongolia	53	[44, 61]	73	[67, 80]	44	[36, 57]
Philippines	54	[47, 57]	76	[63, 81]	42	[41, 45]
Costa Rica	55	[51, 57]	68	[64, 71]	48	[46, 49]
Mexico	56	[51, 56]	59	[52, 61]	55	[53, 55]
Serbia	57	[54, 58]	62	[52, 64]	57	[54, 58]
Republic of Moldova	58	[52, 60]	81	[74, 84]	45	[43, 46]
North Macedonia	59	[58, 65]	52	[50, 68]	63	[63, 66]
Kuwait	60	[59, 79]	75	[71, 82]	56	[55, 70]
Iran (Islamic Republic of)	61	[58, 66]	86	[73, 91]	47	[45, 47]
Uruguay	62	[60, 66]	66	[63, 75]	61	[61, 62]
South Africa	63	[59, 66]	51	[46, 59]	68	[68, 73]
Armenia	64	[61, 67]	85	[82, 91]	50	[47, 50]
Qatar	65	[63, 72]	53	[49, 63]	70	[69, 84]

CONTINUED

TABLE A-V.4

GII 2019 and Input/Output Sub-Indices: ranks and 90% confidence intervals, continued

	GII 2019		Input Sub-Index		Output Sub-Index	
	Rank	Interval	Rank	Interval	Rank	Interval
Brazil	66	[61, 66]	60	[49, 63]	67	[66, 68]
Colombia	67	[64, 70]	58	[50, 62]	76	[73, 79]
Saudi Arabia	68	[67, 77]	49	[42, 62]	85	[83, 94]
Peru	69	[67, 75]	48	[46, 60]	86	[84, 89]
Tunisia	70	[66, 74]	74	[62, 80]	65	[65, 73]
Brunei Darussalam	71	[67, 94]	35	[35, 46]	120	[110, 121]
Belarus	72	[53, 80]	50	[43, 56]	95	[58, 104]
Argentina	73	[67, 75]	72	[58, 76]	75	[72, 75]
Morocco	74	[67, 76]	83	[75, 87]	66	[63, 66]
Panama	75	[66, 87]	79	[70, 91]	72	[62, 92]
Bosnia and Herzegovina	76	[72, 86]	71	[63, 89]	79	[74, 82]
Kenya	77	[71, 81]	89	[81, 95]	64	[64, 65]
Bahrain	78	[76, 84]	69	[64, 75]	87	[85, 97]
Kazakhstan	79	[76, 80]	64	[59, 67]	92	[90, 97]
Oman	80	[76, 88]	57	[48, 65]	101	[99, 116]
Jamaica	81	[75, 83]	84	[74, 89]	69	[67, 76]
Mauritius	82	[72, 86]	67	[65, 71]	96	[75, 99]
Albania	83	[82, 93]	70	[69, 85]	93	[92, 107]
Azerbaijan	84	[82, 87]	77	[72, 85]	90	[88, 93]
Indonesia	85	[78, 86]	87	[76, 89]	78	[76, 81]
Jordan	86	[79, 86]	91	[82, 98]	71	[70, 81]
Dominican Republic	87	[87, 96]	90	[88, 94]	88	[87, 100]
Lebanon	88	[76, 90]	92	[84, 93]	82	[68, 89]
Sri Lanka	89	[82, 91]	94	[89, 101]	77	[74, 83]
Kyrgyzstan	90	[87, 99]	78	[70, 85]	111	[108, 119]
Trinidad and Tobago	91	[90, 105]	88	[84, 90]	99	[95, 121]
Egypt	92	[83, 96]	106	[99, 107]	74	[69, 82]
Botswana	93	[90, 101]	80	[75, 86]	117	[106, 117]
Rwanda	94	[89, 121]	65	[62, 89]	123	[114, 124]
Paraguay	95	[88, 109]	95	[92, 99]	94	[72, 116]
Senegal	96	[90, 99]	103	[101, 110]	81	[72, 82]
United Republic of Tanzania	97	[96, 109]	115	[108, 120]	73	[72, 101]
Cambodia	98	[95, 102]	104	[100, 116]	84	[81, 89]
Ecuador	99	[94, 103]	98	[94, 101]	98	[94, 109]
Tajikistan	100	[90, 112]	107	[100, 116]	83	[80, 98]
Namibia	101	[89, 121]	99	[94, 106]	103	[76, 120]
Uganda	102	[102, 112]	96	[94, 103]	107	[105, 120]
Côte d'Ivoire	103	[99, 107]	110	[107, 114]	91	[86, 95]
Honduras	104	[94, 105]	101	[97, 106]	104	[85, 104]
Pakistan	105	[98, 108]	113	[104, 116]	89	[83, 96]
Ghana	106	[97, 108]	109	[101, 110]	97	[93, 103]
Guatemala	107	[103, 110]	105	[101, 113]	102	[99, 110]
El Salvador	108	[94, 117]	97	[95, 99]	116	[92, 119]
Nepal	109	[102, 115]	93	[91, 105]	119	[103, 120]
Bolivia (Plurinational State of)	110	[103, 111]	102	[93, 104]	113	[110, 120]
Ethiopia	111	[103, 121]	124	[123, 127]	80	[80, 101]
Mali	112	[108, 116]	120	[112, 122]	100	[94, 105]
Algeria	113	[109, 118]	100	[92, 105]	118	[115, 126]
Nigeria	114	[109, 115]	116	[108, 118]	105	[100, 112]
Cameroon	115	[106, 118]	112	[107, 117]	106	[100, 117]
Bangladesh	116	[114, 121]	117	[110, 124]	108	[105, 115]
Burkina Faso	117	[115, 124]	111	[109, 122]	115	[113, 122]
Malawi	118	[115, 127]	119	[116, 123]	112	[108, 126]
Mozambique	119	[111, 126]	118	[109, 123]	114	[112, 124]
Nicaragua	120	[113, 122]	108	[105, 117]	122	[104, 122]
Madagascar	121	[113, 122]	122	[120, 127]	109	[90, 109]
Zimbabwe	122	[111, 127]	123	[110, 127]	110	[107, 123]
Benin	123	[120, 124]	114	[108, 121]	125	[123, 125]
Zambia	124	[120, 126]	126	[113, 129]	121	[117, 127]
Guinea	125	[121, 127]	127	[124, 127]	124	[109, 126]
Togo	126	[105, 127]	121	[116, 123]	128	[83, 128]
Niger	127	[119, 129]	125	[122, 127]	127	[104, 129]
Burundi	128	[125, 128]	128	[124, 129]	126	[125, 128]
Yemen	129	[128, 129]	129	[128, 129]	129	[128, 129]

Source: European Commission Joint Research Centre, 2019.
Notes: Confidence intervals are calculated over 4,000 simulated scenarios combining simulated weights, imputation versus no imputation of missing values, and geometric versus arithmetic average at the pillar level.

TABLE A-V.5

Sensitivity analysis: impact of modeling choices on countries with most sensitive ranks

Index or Sub-Index	Uncertainty tested (pillar level only)	Spearman rank correlation between the two series	Number of economies that improve		Number of economies that deteriorate	
			by more than 20 positions	between 10 and 20 positions	by more than 20 positions	between 10 and 20 positions
GII	Geometric vs. arithmetic average	0.991	0	1	2[3]	2
	EM imputation vs. no imputation of missing data	0.992	0	4	0	5
	Geometric average and EM imputation vs. arithmetic average and missing values	0.989	0	5	0	7
Input Sub-Index	Geometric vs. arithmetic average	0.996	0	1	0	2
	EM imputation vs. no imputation of missing data	0.993	0	2	0	3
	Geometric average and EM imputation vs. arithmetic average and missing values	0.990	0	3	1[4]	6
Output Sub-Index	Geometric vs. arithmetic average	0.996	0	0	1[5]	3
	EM imputation vs. no imputation of missing data	0.969	5[1]	8	1[6]	11
	Geometric average and EM imputation vs. arithmetic average and missing values	0.969	4[2]	9	1[7]	15

Source: European Commission Joint Research Centre, 2019.
Notes:
1 Belarus, Paraguay, Namibia, El Salvador, Togo
2 Belarus, El Salvador, Togo, the Niger
3 Brunei Darussalam, Rwanda
4 Rwanda
5 Paraguay
6 United Republic of Tanzania
7 United Republic of Tanzania

Efficiency frontier in the GII by Data Envelopment Analysis

Is there a way to benchmark economies' multi-dimensional performance on innovation without imposing a fixed and common set of weights that may not be fair to a particular economy?

Several innovation-related policy issues at the national level entail an intricate balance between global priorities and economy-specific strategies. Comparing the multi-dimensional performance on innovation by subjecting economies 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 economy. 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 segment, the assumption of fixed pillar weights common to all economies is relaxed once more; this time economy-specific weights that maximize an economies' global innovation score are determined endogenously by DEA.[11] In theory, each economy is free to decide on the relative contribution of each innovation 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 an economy is relatively strong (weak). Reasonable constraints on the weights are applied to preclude the possibility of an economy achieving a perfect score by assigning a zero weight to weak pillars: for each economy, 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 innovation pillar scores, where the weights are the economy-specific DEA weights, compared to the best performance among all other economies with those same weights. The DEA score can be interpreted as a measure of the "distance to the efficient frontier".

Table A-V.6 presents the pie shares and DEA scores for the top 25 economies in the GII 2019, next to the GII 2019 ranks. All pie shares are in accordance with the starting point of granting leeway to each economy 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 economies' comparative advantage in certain GII pillars vis-à-vis

TABLE A-V.6

Pie shares (absolute terms) and efficiency scores for the top 25 economies in the GII 2019

	Input pillars					Output pillars					
	Institutions	Human capital and research	Infrastructure	Market sophistication	Business sophistication	Knowledge and technology outputs	Creative outputs	Efficient frontier score (DEA)	Efficient frontier rank DEA)	GII rank	Difference from GII rank
Switzerland	0.07	0.17	0.12	0.06	0.19	0.19	0.19	1.00	1	1	0
Sweden	0.20	0.20	0.20	0.05	0.20	0.05	0.10	0.99	2	2	0
United States of America	0.20	0.20	0.10	0.20	0.20	0.05	0.05	0.97	3	3	0
Netherlands	0.20	0.05	0.20	0.05	0.20	0.10	0.20	0.93	8	4	-4
United Kingdom	0.20	0.20	0.20	0.20	0.05	0.05	0.10	0.96	5	5	0
Finland	0.20	0.20	0.20	0.05	0.20	0.05	0.10	0.95	6	6	0
Denmark	0.20	0.20	0.20	0.20	0.10	0.05	0.05	0.95	6	7	1
Singapore	0.20	0.20	0.20	0.20	0.10	0.05	0.05	0.97	3	8	5
Germany	0.20	0.20	0.20	0.10	0.05	0.05	0.20	0.91	10	9	-1
Israel	0.20	0.20	0.10	0.20	0.20	0.05	0.05	0.89	12	10	-2
Republic of Korea	0.20	0.20	0.20	0.20	0.10	0.05	0.05	0.91	10	11	1
Ireland	0.20	0.05	0.20	0.10	0.20	0.20	0.05	0.86	17	12	-5
Hong Kong, China	0.20	0.05	0.20	0.20	0.10	0.05	0.20	0.92	9	13	4
China	0.05	0.05	0.20	0.20	0.20	0.10	0.20	0.83	22	14	-8
Japan	0.20	0.10	0.20	0.20	0.20	0.05	0.05	0.89	12	15	3
France	0.20	0.20	0.20	0.20	0.05	0.05	0.10	0.88	15	16	1
Canada	0.20	0.20	0.20	0.20	0.10	0.05	0.05	0.89	12	17	5
Luxembourg	0.20	0.05	0.20	0.10	0.20	0.05	0.20	0.85	19	18	-1
Norway	0.20	0.20	0.20	0.20	0.05	0.05	0.10	0.87	16	19	3
Iceland	0.20	0.10	0.20	0.20	0.05	0.05	0.20	0.83	22	20	-2
Austria	0.20	0.20	0.20	0.10	0.20	0.05	0.05	0.85	19	21	2
Australia	0.20	0.20	0.20	0.20	0.05	0.05	0.10	0.86	17	22	5
Belgium	0.20	0.20	0.20	0.20	0.10	0.05	0.05	0.82	24	23	-1
Estonia	0.20	0.10	0.20	0.20	0.05	0.05	0.20	0.80	25	24	-1
New Zealand	0.20	0.20	0.20	0.20	0.05	0.05	0.10	0.84	21	25	4

Source: European Commission, Joint Research Centre, 2019.
Notes: Pie shares are in absolute terms, bounded by 0.05 and 0.20 for all seven innovation pillars. In the GII 2019, 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. Darker colors represent higher contribution of those pillars to the overall DEA score as a result of a country's stronger performance in those pillars, which may help to evidence economy-specific strategies.

all other economies and all pillars. For example, Switzerland is the only economy this year that obtains a perfect DEA score of 1.00, followed closely by Sweden (DEA score of 0.99). In the case of Switzerland this is achieved by assigning 17 to 19% of its DEA score to a mix of input and output pillars, namely Human capital and research, Business sophistication, Knowledge and technology outputs, and Creative outputs. Instead, merely 6% to 12% of Switzerland's DEA score comes from three input pillars, namely Institutions, Infrastructure, and Market sophistication. Using a different mix, Sweden would assign 20% of its DEA score to four input pillars—Institutions, Human capital and research, Infrastructure, and Business sophistication—while merely 5 to 10% of its DEA score comes from the two output pillars capturing Knowledge and technology outputs, and Creative Outputs, and from the input pillar measuring Market sophistication. Switzerland and Sweden are closely followed by the United States of America, Singapore, United Kingdom, Finland, and Denmark, who score between 0.95 (Denmark) and 0.97 (the United States of America and Singapore) in terms of efficiency. Figure A-V.3 shows how close the DEA scores and the GII 2019 scores are for all 129 economies (Pearson correlation of 0.993).

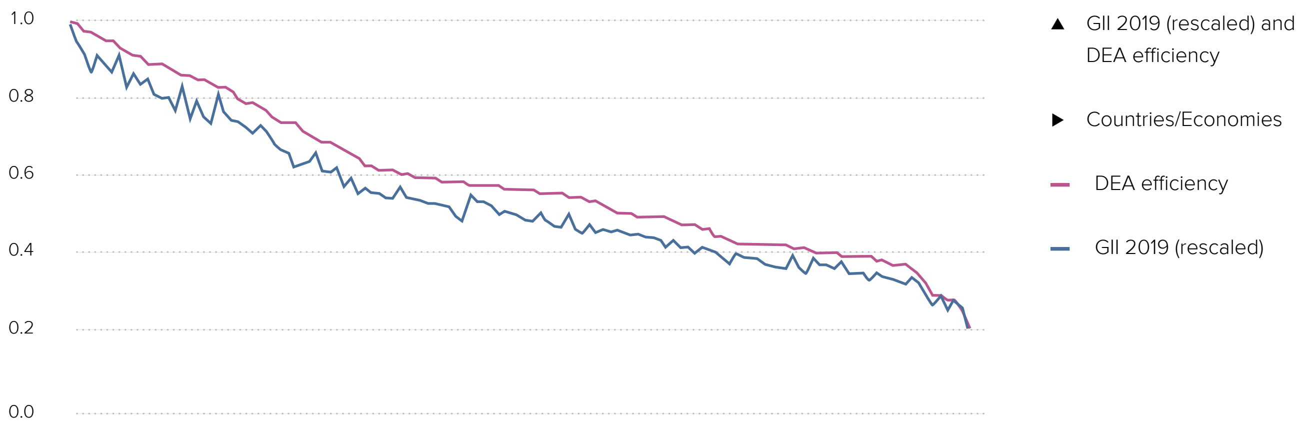
Conclusion

The JRC-COIN analysis suggests that the conceptualized multi-level structure of the GII 2019—with its 80 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. This year, the refinements made by the developing team have helped to enhance the already strong statistical coherence in the GII framework where for all 80 indicators their capacity to distinguish economies' performance is maintained at the sub-pillar level or higher.

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 economy scores, with the additional negative side-effect that it may encourage economies not to report low data values. The adoption, since 2016, by the GII team of a more

FIGURE A-V.3

GII 2019 scores and DEA "distance to the efficient frontier" scores



Source: European Commission Joint Research Centre, 2019
Note: For comparison purposes, the GII scores were rescaled by dividing them with the best performer (Switzerland) in the overall GII 2019.

stringent data coverage threshold (at least 66% for the input- and output-related indicators, separately) has notably improved the confidence in the economy 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 index development 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 44% (up to 72%) of the 129 economies included in the GII 2019, 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 considering the GII pillars, sub-pillars, and individual indicators on their own merit. By doing so, economy-specific strengths and bottlenecks in innovation can be identified and serve as an input for evidence-based policy making.

All published GII 2019 ranks lie within the simulated 90% confidence intervals that consider 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 economies these intervals are narrow enough for meaningful inferences to be drawn: the intervals comprise fewer than 10 positions for 76% (98 out of 129) of the economies. Some caution is needed mainly for nine countries—Brunei Darussalam, Belarus, Panama, Rwanda, Paraguay, Tajikistan, Namibia, El Salvador, Togo—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. Economy ranks, either in the GII 2019 or in the two sub-indices, can be considered representative of the many possible scenarios: 75% of economies 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-COIN audit findings confirm that the GII 2019 meets international quality standards for statistical soundness, which indicates that the GII index is a reliable benchmarking tool for innovation practices at the economy level around the world.

Finally, the "distance to the efficient frontier" measure calculated with Data Envelopment Analysis can be used as a measure of efficiency, and a suitable approach to benchmark economies' multidimensional performance on innovation without imposing a fixed and common set of weights that may not be fair to particular economy. The choice of the GII team to abandon the efficiency ratio (ratio of Output to Input Sub-index) is particularly applaudable. In fact, ratios of composite indicators (Output to

Input Sub-Index in this case) come with much higher uncertainty than the sum of the components (Input plus Output Sub-Index, equivalent to the GII). For this reason, developers and users of indices alike need to take efficiency ratios of this nature with great care. The GII should not be the ultimate and definitive ranking of economies with respect to innovation. On the contrary, the GII best represents an ongoing attempt by Cornell University, 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, matured over 12 years of constant refinements, to pave the way for better and more informed innovation policies worldwide.

Notes:

1 OECD/EC JRC, 2008.

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 and https://composite-indicators.jrc.ec.europa.eu

3 Groeneveld, R.A., et al., 1984: set the criteria for absolute skewness above 1 and kurtosis above 3.5. The skewness criterion was relaxed in the GII case after having conducted ad-hoc tests in the GII 2008-2018 timeseries.

4 An indicator can explain 9% of the economy's variation in the GII sub-pillar scores if the Pearson correlation coefficient between the two series is 0.3.

5 Nunnally, 1978.

6 See footnote 4.

7 Saisana et al., 2005; Saisana et al., 2011; Vértesy, 2016; Vértesy et al., 2016; Montalto et al., 2019.

8 The Expectation-Maximization (EM) algorithm (Little, R.J., et al., 2002; Schneider, T., 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.

9 Munda, 2008.

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

11 A question that arises from the GII approach is whether there is a way to benchmark economies' multi-dimensional performance on innovation without imposing a fixed and common set of weights that may not be fair to an economy. 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, A. et al., 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, L. et al., 2008; Melyn, W. et al., 1991). To estimate DEA-based distance to the efficient frontier scores, we consider the $m = 7$ pillars in the GII 2019 for $n = 129$ economies, with y_{ij} the value of pillar j in economy i. The objective is to combine the pillar scores per economy 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 economy j:

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

Subject to

$$w_{ij} \geq 0 \text{ , where } j = 1,\ldots,7 \text{ , } i = 1,\ldots,129 \qquad \text{(non-negativity constraint)}$$

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

References:

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

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

Charnes, A. & Cooper, W. W. (1985). Preface to Topics in Data Envelopment Analysis. *Annals of Operations Research* 2: 59–94.

Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T., Saisana, M. et al. (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. & Meeden, G. (1984). Measuring Skewness and Kurtosis. *The Statistician* 33: 391–99.

Little, R.J. A., & Rubin, D.B. (2002). *Statistical Analysis with Missing Data*. 2nd edition. Hoboken, NJ: John Wiley & Sons, Inc.

Melyn, W. & Moesen, W. (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.

Montalto, V., Tacao Moura, C.J., Langedijk, S., & Saisana, M. (2019). Culture counts: An empirical approach to measure the cultural and creative vitality of European cities. *Cities* 89: 167-85.

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

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

Saisana, M., Saltelli, A., & Tarantola, S. (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., Ratto, M., Andres, T., Campolongo, F., Cariboni et al. (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.

Vértesy, D. (2016, July). A Critical Assessment of Quality and Validity of Composite Indicators of Innovation. Paper presented at the OECD Blue Sky III Forum on Science and Innovation Indicators. Ghent, 19-21 Sept 2016.

Vertesy, D., & Deiss, R. (2016). The Innovation Output Indicator 2016. Methodology Update. EUR 27880. European Commission, Joint Research Centre.