Joint Research Centre (JRC) statistical audit of the 2021 Global Innovation Index

Joint Research Centre (JRC) statistical audit of the 2021 Global Innovation Index

Michaela Saisana, **Ana Rita Neves**, **Valentina Montalto**,
Giulio Caperna and Giacomo Damioli,
European Commission, JRC, Ispra, Italy

Conceptual and practical challenges are inevitable when trying to understand and model the fundamentals of innovation at the national level worldwide. Now in its 14th edition, the Global Innovation Index (GII) 2021 considers these conceptual challenges and deals with practical challenges – related to data quality and methodological choices – by grouping economy-level data for 132 economies across 81 indicators into 21 sub-pillars, 7 pillars, 2 sub-indices and, finally, an overall index. This appendix offers detailed insights into the practical issues relating to the construction of the GII, analyzing 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" (OECD/EC JRC, 2008: 26). 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 to audit the GII for the 11th consecutive year. As in previous editions, the present JRC-COIN audit focuses on the statistical soundness of the multilevel structure of the index as well as on the impact of key modeling assumptions on the results.[1] The independent statistical assessment of the GII provided by the JRC-COIN guarantees the transparency and reliability of the index for both policymakers 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 rankings 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 efficiency frontier" of innovation by using data envelopment analysis.

Conceptual and statistical coherence in the GII framework

The GII model was assessed by the JRC-COIN in June 2021. Fine-tuning suggestions were taken into account in the final computation of the rankings during an iterative process with the JRC-COIN aiming to set the foundations for a balanced index. The entire process followed four steps, as shown in Figure 1.

Step 1: Conceptual consistency

A total of 81 indicators were selected for their relevance to specific innovation pillars, based on literature review, expert opinion, economy coverage and timeliness. To present 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 used for each economy were those most recently released within the period 2010 to 2020: 71.4 percent of the available data refer to 2019 or more recent years. The JRC-COIN recommendation was to offer an explanation for the reasoning behind the decision to use data that may not reflect recent advances in the relevant field in these economies. In past editions, up to 2015, economies were included in the GII if sufficient data were available for at least 60 percent of all variables within the GII framework. More stringent criteria were adopted in 2016, following the JRC-COIN recommendation in past GII audits, with the result that economies were only included if data availability reached at least 66 percent 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 if at least two of the three sub-pillars in each pillar could be computed. These criteria aim to ensure that economy scores for the GII and for the two Input and Output Sub-Indices are not overly 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 2021 is very good: 80 percent of data is available for 88 percent of the economies (equivalent to 116 economies out of 132). Potentially problematic indicators that could bias the

Figure 1
Conceptual and statistical coherence in the GII 2021 framework

Step 1
Conceptual consistency

- Compatibility with existing literature on innovation and pillar definition
- Use of scaling factors per indicator to present a fair picture of economy differences (e.g., GDP, population)

Step 2
Data checks

- Check for data recency
  (71.4 percent of available data refer to 2019 and 2020)
- Inclusion requirements per economy:
  availability of ≥66 percent 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 (winsorization or logarithmic transformation)
- Direct contact with data providers

Step 3
Statistical coherence

- Treatment of pairs of highly collinear variables as a single indicator
- Assessment of grouping of 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 4
Qualitative review

- Internal qualitative review (by WIPO in partnership with the Portulans Institute, our Corporate and Academic Network partners as well as our Advisory Board members).
- External qualitative review (JRC-COIN, international experts)

Source: European Commission, Joint Research Centre, 2021.

overall results were identified on the basis of two measures related to the shape of the data distributions: skewness and kurtosis. Since 2011, a joint decision of the GII team and JRC-COIN determined that values would be treated if the indicators had absolute skewness greater than 2.0 and kurtosis greater than 3.5.[2] In 2017, having analyzed data in the GIIs between 2011 and 2017, less stringent criteria were adopted. An indicator was only 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 natural logarithm (in cases of more than five outliers; see Appendix I). In 2018, an exceptional behavior for FDI net outflows (indicator 6.3.4 at the time) was observed (Annex 3, JRC Audit, GII 2018) and, from 2018 on, it was recommended that the GII rule for the treatment of outliers be amended 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 cases of more than five outliers);

(a) 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

The JRC-COIN and GII team made the joint decision in 2012 that 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 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, 27 out of 81 indicators and two sub-pillars – 7.2 Creative goods and services and 7.3 Online creativity – were assigned a weight of 0.5, while all other indicators and sub-pillars were assigned a weight of 1.0. Despite this weighting adjustment, only two indicators (5.3.4 FDI net inflows and 6.2.1 Labor productivity growth) were found to be non-influential in the GII framework, meaning that they could not explain at least 9 percent of a given economy's variation in the respective sub-pillar scores.[3] These two indicators also remain non-influential at both the sub-index and the index

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
Innovation Input Sub-Index	1.1. Political environment	0.94	0.80	0.86	0.68	0.80	0.72	0.80
	1.2. Regulatory environment	0.92	0.67	0.70	0.59	0.67	0.59	0.69
	1.3. Business environment	0.85	0.69	0.70	0.61	0.66	0.68	0.60
	2.1. Education	0.61	0.82	0.66	0.56	0.58	0.58	0.60
	2.2. Tertiary education	0.66	0.82	0.74	0.55	0.56	0.58	0.61
	2.3. Research and development (R&D)	0.75	0.89	0.77	0.71	0.89	0.88	0.78
	3.1. Information and communication technologies (ICTs)	0.80	0.86	0.94	0.69	0.74	0.74	0.77
	3.2. General infrastructure	0.54	0.54	0.68	0.37	0.52	0.48	0.46
	3.3. Ecological sustainability	0.66	0.63	0.80	0.49	0.64	0.65	0.67
	4.1. Credit	0.63	0.58	0.52	0.83	0.56	0.48	0.58
	4.2. Investment	0.45	0.38	0.33	0.74	0.46	0.37	0.46
	4.3. Trade, competition, and market scale	0.45	0.68	0.62	0.63	0.60	0.70	0.56
	5.1. Knowledge workers	0.75	0.83	0.77	0.70	0.93	0.82	0.79
	5.2. Innovation linkages	0.72	0.71	0.66	0.63	0.88	0.76	0.76
	5.3. Knowledge absorption	0.64	0.68	0.67	0.61	0.88	0.80	0.76
Innovation Output Sub-Index	6.1. Knowledge creation	0.68	0.83	0.69	0.68	0.84	0.90	0.79
	6.2. Knowledge impact	0.63	0.70	0.74	0.59	0.67	0.85	0.64
	6.3. Knowledge diffusion	0.61	0.67	0.65	0.54	0.79	0.89	0.66
	7.1. Intangible assets	0.59	0.65	0.62	0.59	0.69	0.66	0.90
	7.2. Creative goods and services	0.63	0.64	0.68	0.65	0.73	0.68	0.80
	7.3. Online creativity	0.82	0.77	0.78	0.64	0.81	0.72	0.83

Source: European Commission, Joint Research Centre, 2021.

Table 2
Distribution of differences between pillar and GII rankings

Rank differences (positions)	Innovation Input Sub-Index					Innovation Output Sub-Index	
	Institutions (%)	Human capital and research (%)	Infrastructure (%)	Market sophistication (%)	Business sophistication (%)	Knowledge and technology outputs (%)	Creative outputs (%)
More than 30	15.9%	8.3%	8.3%	24.2%	9.8%	7.6%	6.1%
20–29	16.7%	9.1%	14.4%	20.5%	15.2%	12.1%	12.9%
10–19	22.0%	31.1%	37.1%	20.5%	23.5%	25.8%	24.2%
10 or more*	54.5%	48.5%	59.8%	65.2%	48.5%	45.5%	43.2%
5–9	18.9%	25.8%	18.2%	15.9%	22.7%	27.3%	28.0%
Less than 5	24.2%	22.0%	19.7%	16.7%	27.3%	22.0%	25.8%
Same rank	2.3%	3.8%	2.3%	2.3%	1.5%	5.3%	3.0%
Total**	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Spearman rank correlation coefficient with the GII	0.85	0.92	0.91	0.80	0.89	0.91	0.93

Source: European Commission, Joint Research Centre, 2021.

Notes: * This column is the sum of the previous three rows.

** This column is the sum of all white rows.

level. This means that there is almost no relationship between a country's level of innovation and its FDI net inflows or Labor productivity growth, which calls either for better formulation of those indicators or for better proxies for those concepts. However, the fact remains that almost all of the indicators were found to be sufficiently influential in the GII framework.

Principal component analysis and reliability item analysis

Principal component analysis (PCA) was used to assess the extent to which 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 approximately 56 percent (pillar 4: Market sophistication) and up to 81 percent (pillar 1: Institutions) of the total variance in the three underlying sub-pillars. Furthermore, results confirm the expectation that the sub-pillars are more closely correlated with their own pillar than with any other pillar and that all correlation coefficients are close to or greater than 0.70 (Table 1).

The five input pillars share a single statistical dimension that summarizes 82 percent of the total variance, and the five loadings (correlation coefficients) of these pillars are very similar to each other (0.83–0.94). 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. Consequently, the reliability of the Input Sub-Index, measured by Cronbach's alpha value, is very high at 0.94 – well above the 0.70 threshold for a reliable aggregate (Nunally, 1978).

The two output pillars – Knowledge and technology outputs and Creative outputs – are strongly correlated with each other (0.80); they are also both strongly correlated with the Innovation Output Sub-Index (0.94 to 0.95).

Finally, a vital 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 the output-related pillars' weight of 2/7. 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 0.98 (and the two sub-indices have a correlation of 0.92), which suggests that the sub-indices are effectively placed on an 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 2021 framework, and that the GII has a balanced structure at each aggregation level. Furthermore, this year, all but two of the 81 indicators are found to be sufficiently influential in the GII framework – that is, each indicator explains at least 9 percent of countries' variation in their respective sub-pillar scores, which is worth highlighting as a very positive feature of this year's GII framework.[4] The only recommendation for a possible refinement to the GII framework relates to two indicators – 5.3.4 FDI net inflows and 6.2.1 Labor productivity growth – which seem to bear little relation to any of the GII indicators or to the overall sub-indices and GII index. In spite of expectations to the contrary, an economy's innovation level is almost independent of the FDI net inflows and Labor productivity growth in the country.

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-COIN confirm that this is not the case. In fact, for more than 43 percent (up to 65 percent) of the 132 economies included in the GII 2021, the GII ranking and any of the 7 pillar rankings differ by 10 positions or more (Table 2). This is a desirable outcome because it demonstrates the added value of the GII ranking, which helps to highlight other aspects of innovation that are not immediately apparent from analysis of the seven pillars individually. This result highlights the value of taking due account of the merits of each of the GII pillars, sub-pillars and indicators individually. By doing so, economy-specific strengths and bottlenecks in terms of 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 suggest the use of a minimum–maximum normalization approach in the [0–100] range, the treatment of outliers is driven by statistical analysis, and simplicity and parsimony criteria 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 2021 individual economy rankings.

As suggested in the relevant literature on composite indicators,[5] the robustness assessment was based on Monte Carlo simulation and multi-modeling approaches, applied to “error-free” data where potential outliers, eventual errors and typos have already been corrected in

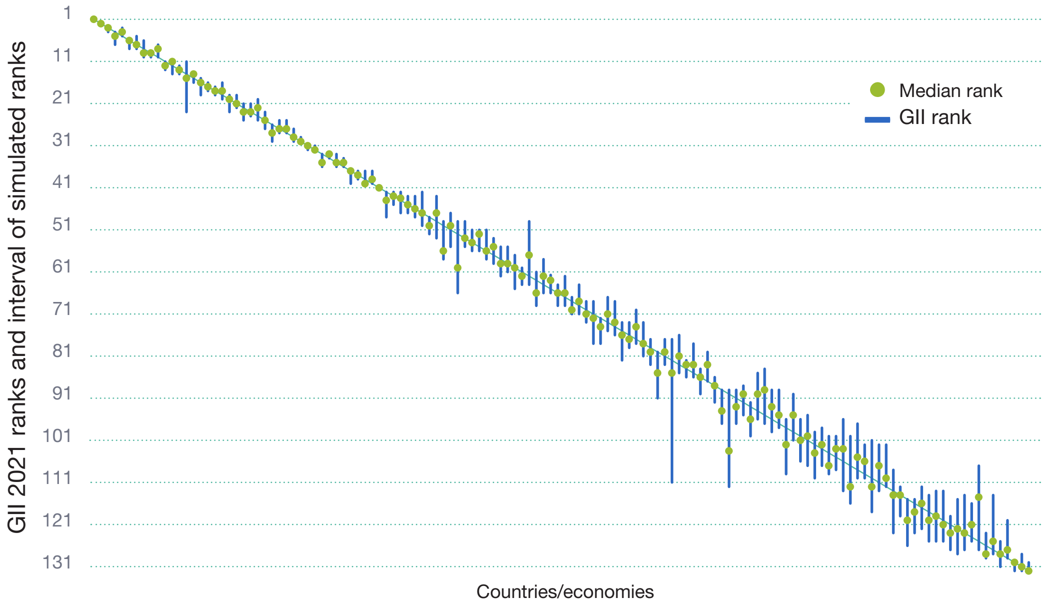
Table 3
Uncertainty parameters: Missing values, aggregation and weights

			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, 2021.

Figure 2
Robustness analysis of the GII, Input and Output Sub-Indices

GII rank vs. median rank, 90 percent confidence intervals



Input rank vs. median rank, 90 percent confidence intervals

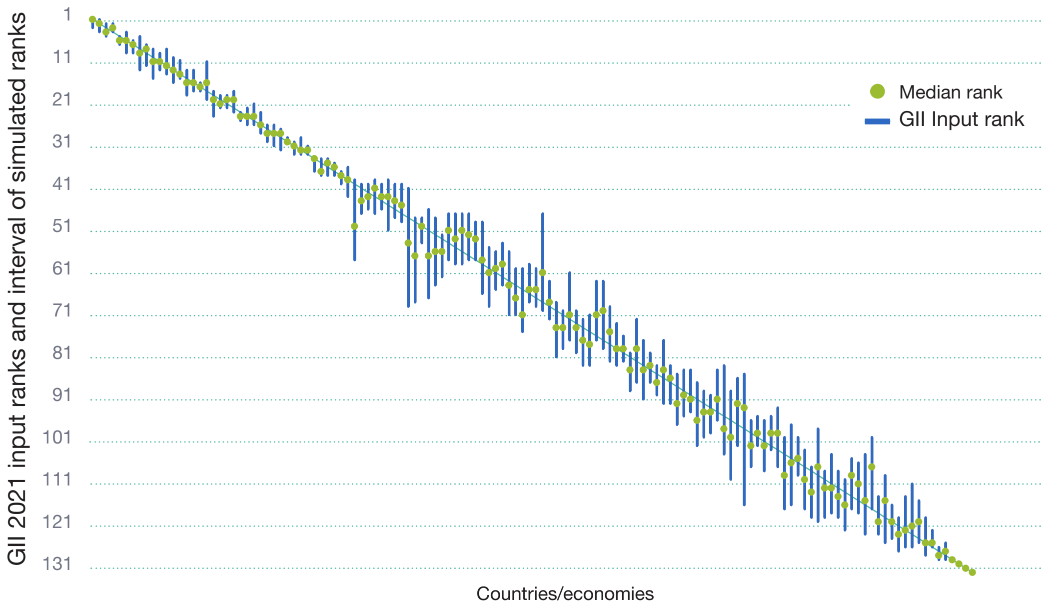
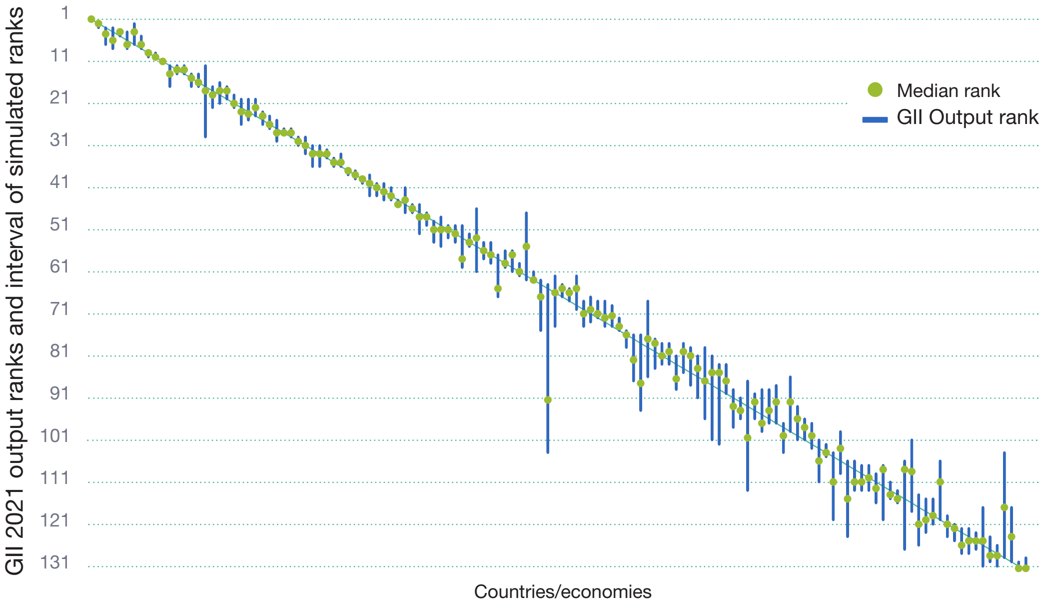


Figure 2
Robustness analysis of the GII, Input and Output Sub-Indices (continued)

Output rank vs. median rank, 90 percent confidence intervals



Source: European Commission, Joint Research Centre, 2021.
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 2021 rank is 0.998; between the median rank and Innovation Input 2021 rank is 0.997; and between the median rank and the Innovation Output 2021 rank is 0.995.

a preliminary stage. In particular, the three key modeling issues considered in the assessment of the GII were the treatment of missing data, pillar weights and the aggregation formula used at the pillar level.

The 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 an equal footing. As a result of these considerations, the limit values of uncertainty for the five input pillars are between 10 percent and 30 percent; the limit values for the two output pillars are between 40 percent and 60 percent (Table 3).

For transparency and replicability purposes, the GII team 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. However, this is not the case for the GII. After 14 editions, the GII team has not encountered any strategy of deliberate no-reporting for the indicators used. The consequence of not imputing missing values in an arithmetic average 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 not imputing missing values, the JRC-COIN estimated missing data using the expectation–maximization (EM) algorithm that was applied within each GII pillar and then compared it to the non-imputation approach (Table 5).[6]

Regarding the aggregation formula, decision-theory practitioners challenge the use of simple arithmetic averages because of their fully compensatory nature, in which a high comparative advantage on a few indicators can compensate for a comparative disadvantage on many indicators (Munda, 2008). To assess the impact of this 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.[7]

Four models were tested, based on the combination of no imputation versus EM imputation and arithmetic versus geometric average, using 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 3 provides 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 percent confidence intervals computed across the 4,000 Monte Carlo simulations for the GII and the two sub-indices. The figure orders economies in ascending order (best to worst) according to their reference rank (black line), with the dot representing the median rank over the simulations.

All published GII 2021 ranks lie within the simulated 90 percent confidence intervals and for most economies these intervals are sufficiently narrow to allow meaningful inferences to be drawn: there is a shift of fewer than 10 positions for 106 of the 132 economies. However, it is also true that a few economies experience significant changes in rank with variations in weights and aggregation formula and because of the estimation of missing data. Two economies – Brunei Darussalam and the United Republic of Tanzania – have 90 percent confidence interval widths over 20 positions (34 and 23 positions, respectively). Consequently, their GII ranks – between the 82nd (Brunei Darussalam) and 90th position (United Republic of Tanzania) in the GII classification – should be interpreted cautiously and certainly not taken at face value. This is a remarkable improvement compared to GII versions up to 2016, when more than 40 economies had confidence interval widths of more than 20 positions. The improvement in the confidence that can be placed in the GII 2021 ranking is the direct result of the decision to adopt a more stringent criterion for an economy's inclusion since 2016, which now requires at least 66 percent data availability within each of the two sub-indices. Some caution is also warranted in regard to the Input Sub-Index for seven economies – Mauritius, Brunei Darussalam, Belarus, the Plurinational State of Bolivia, Cabo Verde, Botswana and Algeria – that have 90 percent confidence interval widths of more than 20 positions (up to 31 for Botswana). A similar degree of caution is also needed in the Output Sub-Index for four economies – the United Republic of Tanzania, Malawi, Brunei Darussalam and Togo – that have 90 percent confidence interval widths of more than 20 positions (up to 40 for Tanzania). Compared to the GII 2019, the higher data availability in the Output Sub-Index this year has led to a much lower number of countries with very wide intervals (4 compared to 13 in the GII 2019 edition), which is a noteworthy improvement.

Although ranks for a few economies, in the GII 2021 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 to be representative of the plurality of scenarios simulated in this audit. Taking the median rank as the benchmark for an economy's expected rank in the realm of the GII's unavoidable methodological uncertainties, 75 percent 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-Indices.

In order to offer full transparency and complete information, Table 4 reports the GII 2021 Index and Input and Output Sub-Indices' economy ranks together with the simulated 90 percent confidence intervals to allow a better appreciation of the robustness of the results to the choice of weights and aggregation formula and the impact of estimating missing data (where applicable).

Table 4
GII 2021 and Input/Output Sub-Indices: Ranks and 90 percent confidence intervals

	GII 2021		Input Sub-Index		Output Sub-Index	
	Rank	Interval	Rank	Interval	Rank	Interval
Switzerland	1	[1, 1]	4	[2, 4]	1	[1, 1]
Sweden	2	[2, 2]	2	[1, 4]	2	[2, 3]
United States	3	[3, 4]	3	[2, 5]	4	[3, 8]
United Kingdom	4	[4, 7]	7	[6, 9]	6	[4, 8]
Republic of Korea	5	[3, 5]	9	[7, 12]	5	[4, 5]
Netherlands	6	[6, 8]	12	[8, 14]	3	[3, 7]
Finland	7	[5, 8]	6	[4, 9]	9	[9, 10]
Singapore	8	[6, 10]	1	[1, 3]	13	[12, 14]
Denmark	9	[9, 10]	5	[5, 7]	11	[11, 11]
Germany	10	[7, 10]	14	[11, 15]	8	[5, 8]
France	11	[11, 13]	17	[16, 18]	10	[9, 10]
China	12	[11, 14]	25	[21, 26]	7	[2, 7]
Japan	13	[12, 14]	11	[9, 13]	14	[12, 14]
Hong Kong, China	14	[11, 23]	10	[8, 15]	17	[12, 29]

Table 4
GII 2021 and Input/Output Sub-Indices: Ranks and 90 percent confidence intervals (continued)

	GII 2021		Input Sub-Index		Output Sub-Index	
	Rank	Interval	Rank	Interval	Rank	Interval
Israel	15	[14, 16]	18	[11, 20]	12	[12, 17]
Canada	16	[15, 19]	8	[5, 13]	23	[20, 25]
Iceland	17	[16, 18]	20	[19, 22]	16	[14, 17]
Austria	18	[17, 19]	16	[13, 18]	24	[20, 24]
Ireland	19	[16, 20]	22	[18, 23]	19	[16, 21]
Norway	20	[19, 23]	13	[10, 16]	28	[27, 28]
Estonia	21	[19, 22]	24	[22, 26]	20	[17, 20]
Belgium	22	[21, 25]	21	[19, 22]	26	[24, 27]
Luxembourg	23	[21, 24]	26	[23, 28]	18	[17, 22]
Czech Republic	24	[20, 25]	30	[29, 30]	15	[14, 17]
Australia	25	[23, 27]	15	[13, 19]	33	[31, 36]
New Zealand	26	[26, 30]	19	[18, 24]	32	[31, 36]
Malta	27	[25, 28]	29	[27, 32]	22	[20, 26]
Cyprus	28	[25, 28]	31	[30, 33]	21	[19, 22]
Italy	29	[27, 30]	33	[31, 33]	25	[23, 26]
Spain	30	[29, 30]	28	[26, 31]	29	[27, 29]
Portugal	31	[31, 32]	32	[29, 33]	30	[29, 31]
Slovenia	32	[31, 32]	27	[26, 30]	36	[33, 36]
United Arab Emirates	33	[33, 36]	23	[23, 25]	47	[45, 52]
Hungary	34	[33, 34]	34	[34, 37]	31	[29, 33]
Bulgaria	35	[33, 36]	46	[40, 48]	27	[25, 30]
Malaysia	36	[34, 36]	36	[34, 38]	34	[32, 34]
Slovakia	37	[37, 40]	42	[40, 46]	35	[34, 36]
Latvia	38	[37, 39]	38	[37, 40]	39	[39, 40]
Lithuania	39	[37, 40]	35	[34, 38]	43	[41, 44]
Poland	40	[37, 40]	37	[35, 38]	42	[40, 44]
Turkey	41	[41, 41]	45	[39, 51]	41	[40, 43]
Croatia	42	[42, 48]	41	[40, 47]	48	[47, 50]
Thailand	43	[42, 45]	47	[40, 49]	46	[45, 47]
Viet Nam	44	[42, 47]	60	[55, 69]	38	[37, 39]
Russian Federation	45	[43, 47]	43	[39, 47]	52	[50, 54]
India	46	[43, 48]	57	[47, 58]	45	[41, 47]
Greece	47	[42, 50]	39	[36, 43]	60	[56, 61]
Romania	48	[48, 52]	54	[47, 58]	50	[48, 55]
Ukraine	49	[43, 53]	76	[63, 77]	37	[37, 38]
Montenegro	50	[49, 58]	53	[52, 62]	53	[50, 60]
Philippines	51	[47, 55]	72	[61, 77]	40	[38, 43]
Mauritius	52	[49, 66]	48	[41, 69]	58	[57, 67]
Chile	53	[49, 55]	44	[40, 46]	61	[59, 62]
Serbia	54	[51, 56]	50	[48, 54]	57	[54, 59]
Mexico	55	[51, 56]	62	[54, 64]	51	[50, 53]
Costa Rica	56	[51, 58]	66	[59, 68]	49	[49, 54]
Brazil	57	[53, 59]	56	[47, 59]	59	[56, 60]
Mongolia	58	[55, 62]	65	[60, 75]	55	[46, 61]
North Macedonia	59	[55, 61]	40	[39, 58]	69	[62, 70]
Iran (Islamic Republic of)	60	[57, 65]	86	[77, 92]	44	[44, 45]
South Africa	61	[60, 64]	55	[47, 59]	68	[65, 68]
Belarus	62	[49, 64]	68	[47, 70]	62	[47, 63]
Georgia	63	[61, 69]	49	[48, 68]	74	[69, 74]
Republic of Moldova	64	[58, 66]	80	[76, 82]	54	[52, 55]
Uruguay	65	[62, 66]	69	[63, 72]	63	[61, 63]
Saudi Arabia	66	[64, 69]	59	[49, 66]	72	[68, 72]
Colombia	67	[62, 69]	58	[49, 58]	75	[72, 75]
Qatar	68	[67, 71]	64	[60, 71]	70	[68, 74]
Armenia	69	[64, 71]	85	[83, 90]	56	[54, 58]
Peru	70	[68, 73]	52	[48, 64]	82	[78, 83]
Tunisia	71	[68, 78]	78	[69, 82]	64	[63, 75]
Kuwait	72	[72, 78]	73	[70, 80]	73	[68, 74]
Argentina	73	[67, 75]	77	[63, 79]	71	[67, 73]
Jamaica	74	[68, 76]	82	[72, 87]	66	[62, 74]
Bosnia and Herzegovina	75	[73, 82]	70	[68, 81]	80	[77, 84]
Oman	76	[73, 79]	67	[60, 69]	90	[83, 90]
Morocco	77	[70, 78]	84	[80, 87]	67	[64, 67]
Bahrain	78	[73, 81]	63	[56, 71]	99	[86, 99]
Kazakhstan	79	[77, 83]	61	[56, 65]	101	[96, 101]

Table 4
GII 2021 and Input/Output Sub-Indices: Ranks and 90 percent confidence intervals (continued)

	GII 2021		Input Sub-Index		Output Sub-Index	
	Rank	Interval	Rank	Interval	Rank	Interval
Azerbaijan	80	[80, 91]	74	[72, 83]	91	[89, 98]
Jordan	81	[77, 83]	79	[73, 83]	81	[78, 83]
Brunei Darussalam	82	[77, 111]	51	[46, 67]	115	[106, 127]
Panama	83	[76, 85]	83	[77, 91]	79	[68, 86]
Albania	84	[82, 86]	71	[70, 79]	92	[91, 96]
Kenya	85	[78, 86]	89	[84, 95]	76	[75, 79]
Uzbekistan	86	[84, 90]	75	[71, 83]	100	[93, 101]
Indonesia	87	[80, 87]	87	[83, 92]	84	[78, 85]
Paraguay	88	[86, 92]	90	[84, 94]	87	[79, 96]
Cabo Verde	89	[89, 97]	96	[89, 110]	88	[81, 101]
United Republic of Tanzania	90	[89, 112]	120	[116, 124]	65	[64, 104]
Ecuador	91	[89, 97]	92	[89, 100]	94	[90, 96]
Lebanon	92	[88, 95]	94	[84, 96]	97	[88, 97]
Dominican Republic	93	[92, 100]	93	[90, 99]	98	[97, 104]
Egypt	94	[85, 96]	102	[95, 103]	86	[81, 91]
Sri Lanka	95	[84, 97]	103	[93, 107]	85	[79, 88]
El Salvador	96	[89, 99]	100	[95, 102]	89	[83, 102]
Trinidad and Tobago	97	[89, 98]	97	[86, 102]	95	[89, 99]
Kyrgyzstan	98	[96, 109]	81	[80, 89]	119	[115, 121]
Pakistan	99	[90, 101]	117	[100, 117]	77	[76, 87]
Namibia	100	[96, 106]	88	[85, 97]	110	[107, 113]
Guatemala	101	[95, 107]	112	[108, 119]	83	[81, 89]
Rwanda	102	[99, 110]	91	[87, 102]	108	[106, 113]
Tajikistan	103	[98, 107]	104	[100, 117]	96	[89, 97]
Bolivia (Plurinational State of)	104	[100, 109]	95	[83, 104]	111	[109, 116]
Senegal	105	[100, 108]	105	[97, 116]	102	[97, 103]
Botswana	106	[96, 113]	98	[85, 116]	109	[107, 113]
Malawi	107	[100, 116]	118	[114, 123]	93	[87, 113]
Honduras	108	[97, 110]	101	[96, 108]	106	[99, 109]
Cambodia	109	[102, 110]	106	[100, 109]	104	[102, 105]
Madagascar	110	[101, 118]	127	[126, 129]	78	[76, 94]
Nepal	111	[102, 113]	99	[96, 107]	116	[101, 118]
Ghana	112	[102, 112]	114	[105, 117]	103	[101, 111]
Zimbabwe	113	[108, 123]	116	[104, 123]	105	[104, 120]
Côte d'Ivoire	114	[112, 119]	107	[103, 117]	121	[119, 124]
Burkina Faso	115	[115, 126]	108	[107, 119]	123	[122, 128]
Bangladesh	116	[115, 123]	121	[119, 127]	113	[111, 115]
Lao People's Democratic Republic	117	[112, 122]	123	[111, 126]	112	[107, 120]
Nigeria	118	[114, 125]	115	[106, 118]	124	[122, 128]
Uganda	119	[113, 125]	119	[109, 125]	122	[121, 125]
Algeria	120	[113, 125]	109	[98, 120]	128	[126, 131]
Zambia	121	[119, 127]	111	[104, 118]	127	[124, 130]
Mozambique	122	[115, 128]	122	[114, 126]	118	[115, 123]
Cameroon	123	[114, 127]	124	[115, 125]	117	[114, 126]
Mali	124	[116, 125]	126	[122, 126]	114	[113, 116]
Togo	125	[107, 127]	110	[108, 119]	129	[104, 129]
Ethiopia	126	[123, 129]	129	[128, 129]	107	[106, 124]
Myanmar	127	[114, 128]	128	[125, 129]	120	[106, 120]
Benin	128	[125, 131]	113	[110, 122]	132	[129, 132]
Niger	129	[120, 129]	125	[119, 128]	130	[117, 130]
Guinea	130	[130, 132]	130	[130, 132]	126	[117, 131]
Yemen	131	[128, 132]	132	[130, 132]	125	[123, 127]
Angola	132	[130, 132]	131	[130, 132]	131	[130, 132]

Source: European Commission, Joint Research Centre, 2021.

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.

Emphasizing the identification of and relationship between input and output indicators may seem irresistible from a policymaking perspective, since doing so has the potential to shed light on the effectiveness of innovation systems and policies. However, the 2018 statistical audit concluded that innovation efficiency ratios, calculated as ratios of indices, have to be approached with care. The reason for advising caution was that the simulated 90 percent confidence intervals for most economies were too wide to allow meaningful inferences to be drawn: there was a shift of more than 20 positions for 50 percent of the economies. Hence, while propagating the uncertainty in the two GII sub-indices over to their sum (the GII) had a modest impact on the rankings, applying the same uncertainty propagation to their ratio had a very high impact on the economy rankings. This challenge is not specific to the GII framework per se but is a statistical property that comes with ratios of composite indicators. In this present audit, the JRC-COIN commends the GII team's decision to drop the efficiency ratio from the 2019 edition onwards and instead to draw policy inferences from scrutiny of the Input–Output performance, as per the plot of GII scores against the economies' level of economic development, and comment 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 5 summarizes the impact of changes in 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-Indices are found to be heavily influenced by the imputation of missing data, or by the aggregation formula. Depending on the combination of the choices made in Table 5, only two economies – Togo and the United Republic of Tanzania – shift rank by more than 20 positions.

All in all, the published GII 2021 ranks are reliable and, for most economies, the simulated 90 percent confidence intervals are narrow enough to allow meaningful inferences to be drawn. Nevertheless, the readers of the GII 2021 report should consider economy ranks in the GII 2021 and in the Input and Output Sub-Indices not only at face value but also within the 90 percent confidence intervals in order to better appreciate the degree to which an economy's rank depends on the modeling choices. These confidence intervals also have to be taken into account when comparing economy rank changes from one year to another at the GII or Innovation Sub-Index level in order to avoid drawing erroneous conclusions on economies' rise or fall in the overall classifications. Since 2016, following the JRC-COIN recommendation in past GII audits, the developers' decision to apply the 66 percent indicator coverage threshold separately to the Input and Output Sub-Indices in the GII 2021 has led to a net increase in the reliability of economy ranks for both the GII and the two sub-indices. Furthermore, the adoption in 2017 of less stringent criteria for the skewness and kurtosis (greater than 2.25 in absolute value and greater than 3.5, respectively) has not introduced any bias into the estimates.

Table 5
Sensitivity analysis: Impact of modeling choices on countries with most sensitive ranks

		Number of countries that improve			Number of countries that deteriorate	
Index or sub-index	**Uncertainty tested (pillar level only)**	Spearman rank correlation between the two series	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	0.994	0	0	1**	1
	EM imputation vs. no imputation of missing data	0.995	0	2	1***	0
	Geometric average and EM imputation vs. arithmetic average and missing values	0.992	0	5	1***	3
Input Sub-Index	Geometric vs. arithmetic average	0.996	0	0	0	2
	EM imputation vs. no imputation of missing data	0.994	0	1	0	2
	Geometric average and EM imputation vs. arithmetic average and missing values	0.991	0	3	0	5
Output Sub-Index	Geometric vs. arithmetic average	0.997	0	0	0	4
	EM imputation vs. no imputation of missing data	0.987	1*	5	1****	6
	Geometric average and EM imputation vs. arithmetic average and missing values	0.987	1*	4	1****	5

Source: European Commission, Joint Research Centre, 2021.

Notes:

* The United Republic of Tanzania (down from 65th to 104th in the Output Sub-Index).

** Brunei Darussalam (down from 82nd to 111th in the GII).

*** The United Republic of Tanzania (down from 90th to 111th in the GII).

**** Togo (up from 129th to 105th in the Output Sub-Index).

Efficiency frontier in the GII by data envelopment analysis

Is there a way to benchmark economies' multidimensional 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 multidimensional performance on innovation by subjecting all economies to a fixed and common set of weights may prevent acceptance of an innovation index on the 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 the determination of 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 and this time, economy-specific weights that maximize an economy's global innovation score are determined endogenously by DEA.[8] 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 are applied to the weights 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 percent and 20 percent, 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

Table 6

Pie shares (absolute terms) and efficiency scores for the top 25 economies in the GII 2021

	Input pillars					Output pillars					
	Institutions	Human capital and research	Infrastructure	Market sophistication	Business sophistication	Knowledge and technology outputs	Creative outputs	Efficiency frontier score (DEA)	Efficiency frontier rank (DEA)	GII rank	Difference from GII rank
Switzerland	0.09	0.14	0.13	0.12	0.19	0.14	0.19	1.00	1	1	0
Sweden	0.18	0.19	0.20	0.05	0.20	0.13	0.05	1.00	1	2	1
United States	0.20	0.18	0.05	0.20	0.20	0.12	0.05	0.99	3	3	0
Singapore	0.20	0.17	0.13	0.20	0.20	0.05	0.05	0.98	4	8	4
United Kingdom	0.20	0.20	0.20	0.20	0.05	0.05	0.10	0.96	5	4	–1
Republic of Korea	0.14	0.20	0.20	0.05	0.20	0.05	0.16	0.95	6	5	–1
Finland	0.20	0.20	0.20	0.05	0.13	0.17	0.05	0.95	6	7	1
Denmark	0.20	0.20	0.20	0.20	0.10	0.05	0.05	0.95	6	9	3
Netherlands	0.20	0.15	0.20	0.05	0.20	0.05	0.15	0.93	9	6	–3
Germany	0.20	0.20	0.20	0.05	0.16	0.05	0.14	0.91	10	10	0
Japan	0.20	0.10	0.20	0.20	0.20	0.05	0.05	0.90	11	13	2
Hong Kong, China	0.20	0.10	0.20	0.20	0.05	0.05	0.20	0.90	11	14	3
Canada	0.20	0.20	0.20	0.20	0.10	0.05	0.05	0.90	11	16	5
France	0.20	0.20	0.20	0.10	0.05	0.05	0.20	0.89	14	11	–3
Israel	0.20	0.10	0.05	0.20	0.20	0.20	0.05	0.87	15	15	0
Austria	0.20	0.20	0.20	0.14	0.16	0.05	0.05	0.87	15	18	3
Norway	0.20	0.20	0.20	0.20	0.10	0.05	0.05	0.87	15	20	5
China	0.05	0.10	0.20	0.20	0.20	0.20	0.05	0.86	18	12	–6
Australia	0.20	0.20	0.20	0.20	0.10	0.05	0.05	0.85	19	25	6
Iceland	0.20	0.20	0.20	0.05	0.10	0.05	0.20	0.84	20	17	–3
Ireland	0.20	0.10	0.20	0.05	0.20	0.20	0.05	0.83	21	19	–2
Estonia	0.20	0.10	0.20	0.20	0.05	0.05	0.20	0.83	21	21	0
Belgium	0.20	0.20	0.20	0.14	0.16	0.05	0.05	0.83	21	22	1
Luxembourg	0.20	0.05	0.20	0.10	0.20	0.05	0.20	0.82	25	23	–2
Czech Republic	0.20	0.20	0.20	0.05	0.10	0.20	0.05	0.78	27	24	–3

Source: European Commission, Joint Research Centre, 2021.

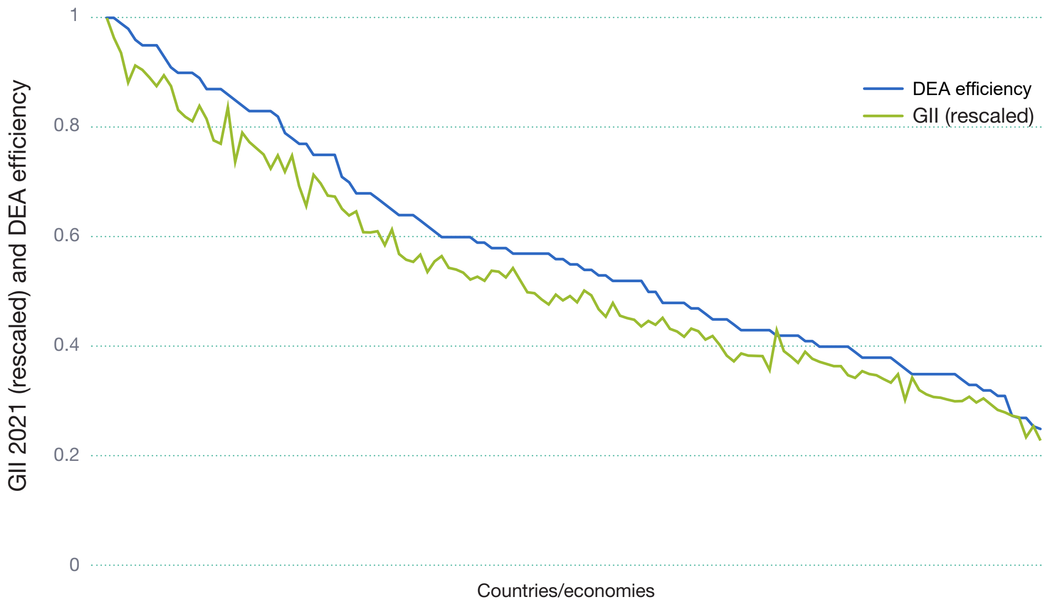
Notes: Pie shares are in absolute terms, bounded by 0.05 and 0.20 for all seven innovation pillars. In the GII 2021, however, the five input pillars each have a fixed weight of 0.10 while the two output pillars each have a fixed weight of 0.25. Darker colors represent a higher contribution of those pillars to the overall DEA score as a result of a country's stronger performance in those pillars, which may help to provide evidence for economy-specific strategies. Countries are ordered by their GII 2021 ranking.

weights. The DEA score can be interpreted as a measure of the "distance to the efficiency frontier."

Table 6 presents the pie shares and DEA scores for the top 25 economies in the GII 2021, next to the GII 2021 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 different economies' comparative advantage in certain GII pillars vis-à-vis all other economies and all pillars. For example, this year, Switzerland and Sweden are the only economies to obtain a perfect DEA score of 1.00, followed closely by the United States of America and Singapore (with DEA scores of 0.99 and 0.98, respectively). In the case of Switzerland, this is achieved by assigning 19 percent of its DEA score to a combination of input and output pillars, namely Business sophistication and Creative outputs, while 9 percent to 14 percent of Switzerland's DEA score comes from the remaining pillars. Using a different approach, Sweden has assigned 18–20 percent of its DEA score to four input pillars – Institutions, Human capital and research, Infrastructure and Business sophistication – while just 5 to 13 percent 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 (0.99) and Singapore (0.98) in terms of efficiency. Figure 3 shows how close the DEA scores and the GII 2021 scores are for all 132 economies (Pearson correlation of 0.994).

Figure 3

GII 2021 scores and DEA "distance to the efficiency frontier" scores



Source: European Commission, Joint Research Centre, 2021.

Notes: For comparison purposes, the GII scores were rescaled by dividing them by the result of the best performer in the overall GII 2021 (Switzerland).

Conclusion

The JRC-COIN analysis suggests that the conceptualized multilevel structure of the GII 2021 – with its 81 indicators, 21 sub-pillars, 7 pillars and 2 sub-indices comprising the overall index – is statistically sound and balanced: that is, each sub-pillar makes a similar contribution to the variation of its respective pillar. The refinements made by the developing team have helped to enhance the already strong statistical coherence in the GII framework, in which the capacity of the 81 indicators to distinguish economies' performance is maintained at the sub-pillar level or higher in all but two cases.

The decision not to impute missing values, which is common in comparable contexts and justified on the grounds of transparency and replicability, can at times have an undesirable impact on some economies' scores, with the additional negative side-effect that it might encourage economies not to report low data values. The GII team's adoption, in 2016, of a more stringent data coverage threshold (at least 66 percent data availability for each of the input- and output-related indicators) has notably improved confidence in the economy ranking for the GII and the two sub-indices.

Additionally, the GII team's decision, 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 an approach will be adopted 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 43 percent (up to 65 percent) of the 132 economies included in the GII 2021, the GII ranking and the rankings of any of the 7 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 are not immediately apparent from an analysis of the seven pillars separately. At the same time, this finding points to the value of duly considering merits of the GII pillars, sub-pillars and their constituent indicators individually. By doing so, economy-specific strengths and bottlenecks in innovation can be identified and serve as an input for evidence-based policymaking.

All published GII 2021 ranks lie within the simulated 90 percent confidence intervals that consider the unavoidable uncertainties inherent 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 80 percent (106 out of 132) of the economies. Some caution is needed, mainly for two countries – Brunei Darussalam and the United Republic of Tanzania – which have GII rankings that are highly sensitive to the methodological choices. The Input and Output Sub-Indices have the same modest degree of sensitivity to the methodological choices relating to the imputation method, weights or aggregation formula. Economy ranks, either in the GII 2021 or in the two sub-indices, can be considered to be representative of the many possible scenarios: 75 percent of economies shift fewer than three positions with respect to the median rank in the GII or either of the Input or Output Sub-Indices.

All things considered, the present JRC-COIN audit findings confirm that the GII 2021 meets international quality standards for statistical soundness, which indicates that the GII is a reliable benchmarking tool for innovation practices at the economy level around the world.

Finally, the "distance to the efficiency frontier" measure calculated using data envelopment analysis can be used both as a measure of efficiency and as a suitable approach to benchmarking economies' multidimensional performance on innovation without imposing a fixed and common set of weights that may not be fair to a particular economy. The decision made by the GII team to abandon the efficiency ratio (ratio of Output to Input Sub-Index) is particularly laudable. 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 approach efficiency ratios of this nature with great care. The GII should not be considered as the ultimate and definitive ranking of economies with respect to innovation. On the contrary, the GII best represents an ongoing attempt to find metrics and approaches that capture the richness of innovation more effectively, 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, based on the principle of transparency, matured over 14 years of constant refinements, to pave the way for better and more informed innovation policies worldwide.

Notes

1 The JRC analysis was based on the recommendations of the OECD/EC JRC (2008) *Handbook on Constructing Composite Indicators* and on more recent research from the JRC. The JRC audits on composite indicators are conducted at the request of the index developers and are available at https://knowledge4policy.ec.europa.eu/composite-indicators_en and https://composite-indicators.jrc.ec.europa.eu.

2 Groeneveld and Meeden (1984) set the criteria for absolute skewness above 1 and for kurtosis above 3.5. The skewness criterion was relaxed in the GII case after ad hoc tests were conducted in the GII 2008–2018 time series.

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

4 See note 3.

5 See Saisana *et al*., 2005; Saisana *et al*., 2011; Vertesy, 2016; Vertesy and Deiss, 2016; and Montalto *et al*., 2019.

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:

(a) The expectation step (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.

(b) The maximization step (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 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.

8 A question that arises from the GII approach is whether there is a way to benchmark economies' multidimensional performance on innovation without imposing a fixed and common set of weights that might not be fair to a particular 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 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 efficiency frontier scores, we consider the $m = 7$ pillars in the GII 2021 for $n = 132$ 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 wj represents the weight of the j-th pillar. In the 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 i:

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

Subject to

$w_{ij} \geq 0$, where, $j=1,\ldots,7$, $i = 1,\ldots,132$ (*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

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, 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. 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.

Montalto, V., C.J. Tacao Moura, S. Langedijk and M. Saisana (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., 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., 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.

Saisana, M., B. D'Hombres and A. Saltelli (2011). Rickety numbers: Volatility of university rankings and policy implications. *Research Policy*, 40, 165–77.

Schneider, T. (2001). Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. *Journal of Climate*, 14, 853–71.

Vertesy, 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. and R. Deiss (2016). *The Innovation Output Indicator 2016. Methodology Update*. EUR 27880. Luxembourg: European Commission, Joint Research Centre.