Joint Research Centre (JRC) statistical audit of the 2022 Global Innovation Index

Michaela Saisana, **Michela Bello**, **Giulio Caperna**, **Ana Rita Neves** and
Carlos Tacao Moura, 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 15th edition, the *Global Innovation Index* (*GII*) *2022*, considers these conceptual challenges and deals with practical issues – related to data quality and methodological choices – by grouping economy-level data for 132 economies across 81 indicators into 21 sub-pillars, seven pillars, two sub-indices and, finally, an overall index. This appendix offers detailed insights into the practical issues as they relate to the construction of the GII, analyzing the statistical soundness of the calculations and assumptions used 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 observation 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, Italy, has been invited to audit the GII for a twelfth 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, through the use of data envelopment analysis, the JRC-COIN analysis includes an assessment of the added value of the GII together with a measure of the "distance to the efficiency frontier" of innovation.

Figure 1 Conceptual and statistical coherence in the GII 2022 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
  (74 percent of available data refer to 2020 or after)
- 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, the GII Corporate and Academic Network partners, as well as GII Advisory Board members)
- external qualitative review (JRC-COIN, international experts)

Source: European Commission, Joint Research Centre, 2022.

Conceptual and statistical coherence within the GII framework

The GII model was assessed by the JRC-COIN in June 2022. 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. This four-step process is outlined in its entirety in Figure 1.

Step 1: Conceptual consistency

A total of 81 indicators were selected for their relevance to specific innovation pillars, based on a 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 2011 to 2021: 74 percent of the available data refer to 2020 or after. The JRC-COIN recommended that an explanation ought to be given of the reasoning behind the decision to use data that may not reflect recent advances in the relevant fields in the economies covered. In previous 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 a 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 previous editions). In practice, data availability for all economies included in the GII 2022 is very good: 80 percent of data is available for 83 percent of the economies covered (equivalent to 110 economies out of 132). Potentially problematic indicators that could bias the 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 by the GII team and the JRC-COIN determined that values would be treated if an indicator had absolute skewness greater than 2.0 and kurtosis greater than 3.5.[2] In 2017, having analyzed data in the GIIs compiled 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. Such 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 by foreign direct investment (FDI) net outflows (indicator 6.3.4 at the time) was observed (Annex 3, JRC Audit, GII 2018) and, from 2018 onward, 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);
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

The JRC-COIN and GII team jointly decided 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 a variable's importance, such as the non-linear Pearson correlation ratio. As a result of this analysis, 8 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, five indicators (3.2.3 Gross capital formation, 4.1.3 Loans from microfinance institutions, 5.3.4 FDI net inflows, 6.2.1 Labor productivity growth, and 7.2.4 Printing and other media) were found to be non-influential in this year's GII framework, meaning that they could not explain at least 9 percent of economies' overall variation in the respective sub-pillar scores.[3] These five indicators also remain non-influential at both the sub-index and the index level. This means there is almost no relationship between a country's level of innovation and its Gross capital formation, Loans from microfinance institutions, FDI net inflows, Labor productivity growth, or Printing and other media, which calls either for better formulation of these indicators or for better metrics for those concepts. That said, the other 76 out of 81 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 62 percent (pillar 4: Market sophistication) and up to 83 percent (pillar 5: Business sophistication) of the total variance in the three underlying sub-pillars. Furthermore, results confirm 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 83 percent of the total variance, and the five loadings (correlation coefficients) of these pillars are very similar to each other (0.87–0.94). This similarity suggests that the five pillars make a roughly equal contribution to the variation of the Innovation Input Sub-Index scores, as envisaged by the development 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 (0.86); they are also both strongly correlated with the Innovation Output Sub-Index (both 0.96).

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 0.97 (and the two sub-indices have a correlation of 0.91), which suggests that the sub-indices are effectively placed on an equal footing.

Overall, tests so far show that the grouping of variables into sub-pillars, pillars and an overall index is statistically coherent within the GII 2022 framework, and that the GII has a balanced structure at each aggregation level. Furthermore, this year, all but five of the 81 indicators have been 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.[4] The only recommended possible refinement to the GII framework relates to five indicators – 3.2.3 Gross capital formation, 4.1.3 Loans from microfinance institutions, 5.3.4 FDI net inflows, 6.2.1 Labor productivity growth, and 7.2.4 Printing and other media – 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 Gross capital formation, Loans from microfinance institutions, FDI net inflows, Labor productivity growth, and Printing and other media within the country concerned.

Table 1 Statistical coherence in the GII: correlations between sub-pillars and pillars

	Sub-pillar	Institutions	Human capital and research	Infra-structure	Market sophistica-tion	Business sophistica-tion	Knowledge and technology outputs	Creative outputs
Innovation Input Sub-Index	1.1 Political environment	0.92	0.79	0.83	0.66	0.81	0.72	0.71
	1.2 Regulatory environment	0.87	0.73	0.71	0.58	0.71	0.63	0.61
	1.3 Business environment	0.82	0.48	0.47	0.45	0.51	0.39	0.40
	2.1 Education	0.58	0.80	0.71	0.55	0.63	0.63	0.62
	2.2 Tertiary education	0.62	0.82	0.74	0.61	0.61	0.60	0.61
	2.3 Research and development (R&D)	0.69	0.89	0.75	0.77	0.88	0.87	0.82
	3.1 Information and communication technologies (ICTs)	0.68	0.82	0.93	0.71	0.75	0.76	0.75
	3.2 General infrastructure	0.70	0.77	0.83	0.66	0.72	0.70	0.65
	3.3 Ecological sustainability	0.49	0.54	0.72	0.48	0.60	0.60	0.59
	4.1 Credit	0.61	0.64	0.64	0.80	0.59	0.58	0.65
	4.2 Investment	0.57	0.55	0.49	0.81	0.59	0.54	0.60
	4.3 Trade, diversification, and market scale	0.43	0.71	0.70	0.76	0.64	0.73	0.69
	5.1 Knowledge workers	0.69	0.87	0.81	0.70	0.94	0.86	0.80
	5.2 Innovation linkages	0.78	0.78	0.73	0.70	0.89	0.81	0.77
	5.3 Knowledge absorption	0.66	0.72	0.73	0.66	0.88	0.79	0.78
Innovation Output Sub-Index	6.1 Knowledge creation	0.62	0.84	0.71	0.71	0.86	0.91	0.83
	6.2 Knowledge impact	0.60	0.76	0.82	0.70	0.76	0.88	0.76
	6.3 Knowledge diffusion	0.59	0.74	0.76	0.65	0.82	0.92	0.74
	7.1 Intangible assets	0.48	0.72	0.66	0.69	0.69	0.73	0.92
	7.2 Creative goods and services	0.62	0.63	0.71	0.65	0.74	0.70	0.74
	7.3 Online creativity	0.74	0.78	0.74	0.67	0.85	0.77	0.80

Source: European Commission, Joint Research Centre, 2022.

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 a redundancy of information within the GII. The tests conducted by the JRC-COIN confirm this is not the case. In fact, for more than 34 percent (up to 67 percent) of the 132 economies included in the GII 2022, the GII ranking and any of the seven 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 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 a basis for evidence-based policymaking.

Table 2 Distribution of differences between pillar and GII rankings

	Innovation Input Sub-Index					Innovation Output Sub-Index	
Rank differences (positions)	**Institutions (%)**	**Human capital and research (%)**	**Infrastructure (%)**	**Market sophistication (%)**	**Business sophistication (%)**	**Knowledge and technology outputs (%)**	**Creative outputs (%)**
More than 30	25.8	9.1	3.8	12.9	4.5	3.8	5.3
20–29	18.9	11.4	17.4	19.7	14.4	5.3	10.6
10–19	22.7	21.2	28.0	22.0	28.8	25.0	32.6
10 or more*	67.4	41.7	49.2	54.5	47.7	34.1	48.5
5–9	18.2	26.5	25.8	23.5	26.5	25.8	26.5
Less than 5	12.9	24.2	23.5	18.9	24.2	36.4	21.2
Same rank	1.5	7.6	1.5	3.0	1.5	3.8	3.8
Total**	100	100	100	100	100	100	100
Spearman rank correlation coefficient with the GII	0.76	0.92	0.93	0.86	0.91	0.95	0.93

Source: European Commission, Joint Research Centre, 2022.
Notes: * This column is the sum of the previous three rows. ** This column is the sum of all non-pink rows.

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 in order to verify that the overall results are, to a great extent, consistent with current evidence, existing research and prevailing theory. Notwithstanding such statistical tests and the positive outcomes regarding the statistical coherence of the GII structure, the GII model is, and has to remain, open to 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 GII scores and rankings relate to:
- setting up an underlying structure for the index based on pillars;
- choosing the individual variables to be used as indicators;
- deciding whether (and how) to impute missing data;
- deciding whether (and how) 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 2022 individual economy rankings.

As suggested by 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 at 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 GII pillars. Weights were assigned to the pillars based on uniform continuous distributions centered on 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 and 30 percent, whereas the limit values for the two output pillars are between 40 and 60 percent (Table 3).

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, 2022.

For transparency and replicability purposes, the GII team has always opted to not estimate missing data. The "no imputation" choice, which is common in similar contexts, might encourage economies to not report low data values. However, this is not the case for the GII. After 15 editions, the GII team has not encountered any strategy of deliberate non-reporting of 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, which was applied within each GII pillar and then compared to the no-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 (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]

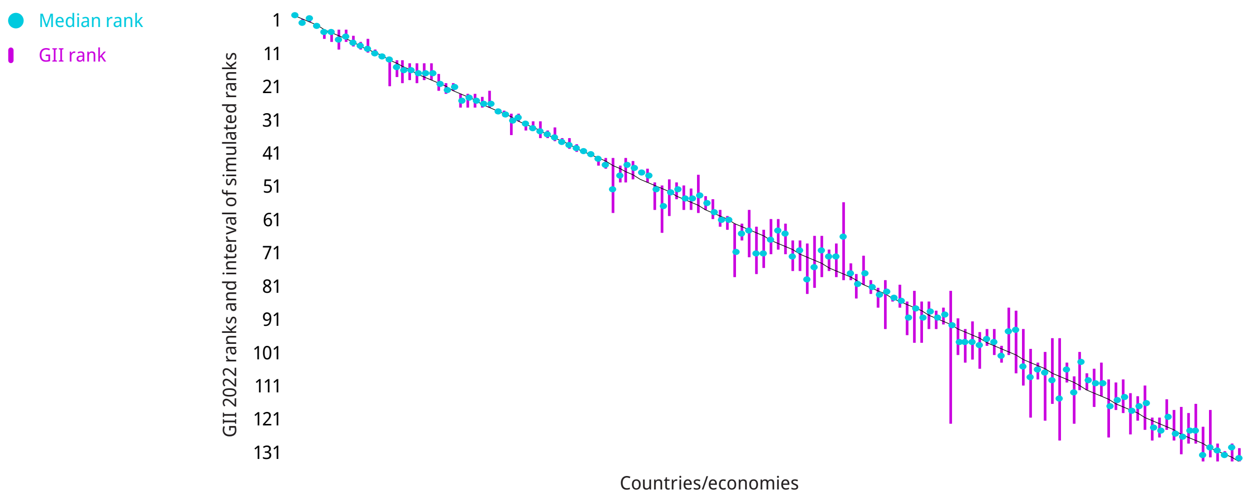
Four models were tested, based on the combination of no imputation versus EM imputation and arithmetic versus geometric average, using 1,000 simulations per model (random weights versus fixed weights), for a total of 4,000 simulations for the GII and each of the two sub-indices (see Table 3 for a summary of the uncertainties considered).

Uncertainty analysis results

The main results of the robustness analysis are shown in Figure 2, with median ranks and 90 percent confidence intervals computed across the 4,000 Monte Carlo simulations for the GII and the two sub-indices. Economies are in ascending order (best to worst performing) according to their reference rank (black line), with the dot representing the median rank over the simulations.

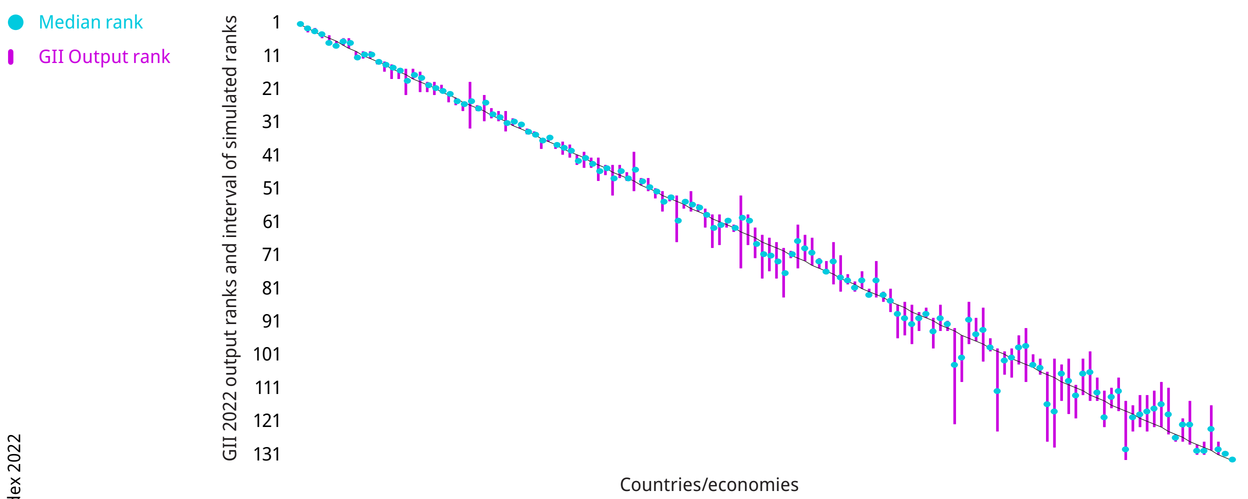
Figure 2 Robustness analysis of the GII, Input and Output Sub-Indices

(a) GII rank vs. median rank, 90 percent confidence intervals



(b) Input rank vs. median rank, 90 percent confidence intervals

(c) Output rank vs. median rank, 90 percent confidence intervals



Source: European Commission, Joint Research Centre, 2022.

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 2022 rank is 0.998; between the median rank and Innovation Input 2022 rank is 0.998; and between the median rank and the Innovation Output 2022 rank is 0.996.

All published GII 2022 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 10 or fewer positions for 97 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 when imputing missing data. Three economies – Belarus, Brunei Darussalam and Zimbabwe – have 90 percent confidence interval widths over 20 positions (23, 39 and 30 positions, respectively). Consequently, their GII rankings – at the 77th (Belarus), 92nd (Brunei Darussalam) and 107th position (Zimbabwe) in the GII classification – should be interpreted cautiously and certainly not taken at face value. However, 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 2022 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 five economies – Belarus, the Islamic Republic of Iran, Brunei Darussalam, Uganda and the United Republic of Tanzania – that have 90 percent confidence interval widths of more than 20 positions (up to 27 for Belarus and the Islamic Republic of Iran). A similar degree of caution is also needed in the Output Sub-Index for five economies – Zimbabwe, Nigeria, the United Republic of Tanzania, Belarus and Côte d'Ivoire – that have 90 percent confidence interval widths of more than 20 positions (up to 29 for Zimbabwe). 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 (five compared to 13 in the GII 2019 edition), which is a noteworthy improvement.

Although the rankings for a few economies, in the GII 2022 overall or in the two sub-indices, appear to be sensitive to methodological choices, the published rankings for the vast majority can be considered as 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 2022 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).

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[8] 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 to instead 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 a similar Innovation Input level but very different Innovation Output level, and vice versa.

Table 4 GII 2022 and Input/Output Sub-Indices: rankings and 90 percent confidence intervals

	GII 2022		Input Sub-Index		Output Sub-Index	
	Rank	Interval	Rank	Interval	Rank	Interval
Switzerland	1	[1, 1]	3	[2, 4]	1	[1, 1]
United States	2	[2, 3]	2	[2, 4]	5	[4, 7]
Sweden	3	[2, 3]	4	[2, 5]	2	[2, 3]
United Kingdom	4	[4, 4]	7	[5, 9]	3	[2, 3]
Netherlands	5	[5, 8]	10	[7, 13]	6	[6, 8]
Republic of Korea	6	[5, 9]	16	[10, 18]	4	[4, 5]
Singapore	7	[5, 11]	1	[1, 1]	14	[13, 17]
Germany	8	[5, 9]	12	[11, 16]	7	[5, 7]

Table 4 Continued

	GII 2022		Input Sub-Index		Output Sub-Index	
	Rank	Interval	Rank	Interval	Rank	Interval
Finland	9	[7, 10]	6	[5, 7]	9	[9, 11]
Denmark	10	[9, 11]	8	[7, 11]	10	[9, 11]
China	11	[8, 12]	21	[17, 24]	8	[5, 8]
France	12	[11, 12]	13	[12, 16]	11	[9, 11]
Japan	13	[13, 13]	11	[8, 14]	12	[12, 13]
Hong Kong, China	14	[14, 22]	5	[4, 8]	25	[18, 32]
Canada	15	[14, 19]	9	[8, 12]	23	[23, 25]
Israel	16	[14, 21]	22	[14, 25]	16	[14, 22]
Austria	17	[15, 20]	17	[14, 21]	21	[19, 21]
Estonia	18	[15, 21]	15	[10, 21]	22	[22, 24]
Luxembourg	19	[15, 20]	20	[17, 23]	18	[15, 21]
Iceland	20	[15, 20]	24	[23, 26]	17	[14, 17]
Malta	21	[18, 23]	27	[27, 28]	13	[12, 15]
Norway	22	[21, 24]	14	[11, 19]	29	[27, 30]
Ireland	23	[21, 23]	25	[22, 26]	19	[18, 21]
New Zealand	24	[24, 28]	23	[20, 26]	28	[26, 29]
Australia	25	[24, 28]	19	[15, 20]	32	[31, 32]
Belgium	26	[24, 28]	26	[23, 26]	24	[24, 27]
Cyprus	27	[25, 28]	29	[28, 30]	20	[18, 22]
Italy	28	[23, 28]	31	[30, 34]	15	[14, 17]
Spain	29	[29, 30]	28	[27, 29]	26	[25, 27]
Czech Republic	30	[29, 31]	33	[31, 36]	27	[22, 30]
United Arab Emirates	31	[30, 36]	18	[16, 22]	52	[51, 57]
Portugal	32	[30, 32]	32	[31, 34]	31	[29, 31]
Slovenia	33	[33, 35]	30	[29, 32]	35	[35, 38]
Hungary	34	[32, 35]	36	[34, 38]	34	[33, 34]
Bulgaria	35	[32, 37]	47	[42, 51]	30	[27, 33]
Malaysia	36	[35, 37]	35	[32, 36]	37	[37, 38]
Turkiye	37	[34, 38]	49	[43, 56]	33	[32, 34]
Poland	38	[37, 39]	41	[37, 43]	36	[35, 36]
Lithuania	39	[37, 40]	34	[31, 36]	47	[45, 48]
India	40	[39, 41]	42	[38, 46]	39	[37, 41]
Latvia	41	[40, 41]	39	[37, 43]	42	[41, 44]
Croatia	42	[42, 42]	45	[42, 49]	40	[40, 43]
Thailand	43	[43, 45]	48	[42, 54]	44	[43, 46]
Greece	44	[43, 46]	44	[41, 49]	49	[47, 49]
Mauritius	45	[43, 59]	40	[38, 55]	54	[52, 66]
Slovakia	46	[45, 50]	54	[49, 56]	45	[43, 52]
Russian Federation	47	[43, 50]	46	[39, 52]	50	[47, 51]
Viet Nam	48	[44, 49]	59	[54, 62]	41	[39, 44]
Romania	49	[45, 50]	56	[51, 60]	43	[41, 48]
Chile	50	[46, 50]	43	[41, 46]	57	[55, 57]
Saudi Arabia	51	[50, 58]	37	[35, 39]	65	[62, 71]
Qatar	52	[51, 65]	38	[37, 47]	67	[65, 75]
Iran (Islamic Republic of)	53	[49, 60]	73	[64, 91]	38	[36, 40]
Brazil	54	[50, 55]	58	[49, 63]	53	[52, 54]
Serbia	55	[51, 58]	55	[47, 59]	58	[56, 62]
Republic of Moldova	56	[52, 58]	78	[72, 82]	46	[43, 47]
Ukraine	57	[48, 59]	75	[64, 80]	48	[39, 51]
Mexico	58	[54, 58]	70	[59, 72]	55	[54, 56]
Philippines	59	[55, 61]	76	[68, 80]	51	[50, 53]
Montenegro	60	[58, 63]	51	[49, 59]	72	[64, 72]
South Africa	61	[60, 64]	69	[62, 72]	61	[59, 62]
Kuwait	62	[62, 78]	66	[63, 77]	66	[64, 77]
Colombia	63	[62, 67]	63	[54, 65]	70	[69, 71]
Uruguay	64	[58, 72]	57	[49, 62]	76	[66, 79]
Peru	65	[63, 77]	52	[47, 64]	81	[80, 83]
North Macedonia	66	[64, 75]	60	[54, 69]	77	[70, 81]
Morocco	67	[61, 71]	87	[83, 89]	56	[51, 57]
Costa Rica	68	[61, 70]	67	[62, 72]	71	[61, 74]
Argentina	69	[62, 71]	77	[67, 78]	62	[61, 63]
Bosnia and Herzegovina	70	[67, 76]	64	[58, 75]	75	[72, 76]
Mongolia	71	[67, 76]	81	[77, 88]	64	[58, 67]
Bahrain	72	[68, 83]	50	[44, 64]	86	[84, 94]

Table 4 Continued

	GII 2022		Input Sub-Index		Output Sub-Index	
	Rank	Interval	Rank	Interval	Rank	Interval
Tunisia	73	[66, 81]	89	[80, 91]	59	[58, 68]
Georgia	74	[66, 78]	61	[56, 69]	82	[72, 83]
Indonesia	75	[70, 76]	72	[65, 79]	74	[71, 74]
Jamaica	76	[68, 78]	88	[81, 91]	60	[58, 67]
Belarus	77	[56, 79]	86	[63, 90]	63	[52, 74]
Jordan	78	[74, 79]	71	[65, 78]	78	[76, 79]
Oman	79	[77, 84]	62	[54, 65]	87	[85, 97]
Armenia	80	[72, 80]	82	[78, 87]	73	[65, 73]
Panama	81	[79, 83]	83	[79, 90]	80	[75, 80]
Uzbekistan	82	[81, 87]	68	[66, 76]	91	[85, 93]
Kazakhstan	83	[79, 93]	65	[60, 71]	97	[86, 102]
Albania	84	[84, 85]	80	[74, 82]	89	[86, 89]
Sri Lanka	85	[80, 87]	102	[92, 106]	68	[66, 77]
Botswana	86	[85, 95]	74	[69, 84]	94	[94, 108]
Pakistan	87	[82, 97]	111	[98, 112]	69	[68, 83]
Kenya	88	[85, 97]	103	[100, 108]	79	[78, 81]
Egypt	89	[85, 93]	97	[94, 99]	83	[81, 84]
Dominican Republic	90	[88, 93]	90	[86, 95]	92	[90, 93]
Paraguay	91	[87, 92]	94	[91, 98]	84	[80, 87]
Brunei Darussalam	92	[82, 121]	53	[46, 70]	129	[115, 129]
Azerbaijan	93	[90, 101]	79	[75, 83]	110	[109, 119]
Kyrgyzstan	94	[93, 103]	85	[79, 93]	108	[103, 114]
Ghana	95	[91, 102]	105	[101, 109]	88	[87, 93]
Namibia	96	[94, 105]	84	[81, 90]	113	[107, 114]
Cambodia	97	[93, 98]	92	[88, 98]	102	[94, 103]
Ecuador	98	[93, 101]	96	[92, 99]	98	[95, 99]
Senegal	99	[98, 103]	93	[91, 100]	105	[101, 106]
El Salvador	100	[87, 101]	101	[92, 102]	95	[84, 97]
Trinidad and Tobago	101	[88, 106]	95	[86, 103]	103	[92, 108]
Bangladesh	102	[93, 110]	112	[109, 125]	90	[89, 98]
United Republic of Tanzania	103	[99, 119]	100	[97, 118]	99	[98, 123]
Tajikistan	104	[103, 108]	104	[99, 113]	101	[98, 107]
Rwanda	105	[100, 120]	91	[87, 103]	123	[110, 124]
Madagascar	106	[96, 115]	125	[120, 128]	85	[85, 95]
Zimbabwe	107	[96, 126]	120	[110, 127]	93	[92, 121]
Nicaragua	108	[103, 109]	99	[94, 111]	112	[99, 114]
Côte d'Ivoire	109	[107, 121]	109	[101, 117]	106	[105, 126]
Guatemala	110	[100, 111]	117	[111, 120]	96	[89, 96]
Nepal	111	[106, 111]	106	[101, 114]	111	[101, 112]
Lao People's Democratic Republic	112	[105, 116]	98	[96, 108]	122	[108, 122]
Honduras	113	[103, 113]	108	[97, 109]	116	[107, 117]
Nigeria	114	[108, 125]	113	[105, 119]	107	[101, 128]
Algeria	115	[109, 117]	110	[100, 115]	118	[115, 123]
Myanmar	116	[108, 118]	122	[116, 126]	104	[100, 104]
Ethiopia	117	[112, 124]	126	[123, 128]	100	[99, 106]
Zambia	118	[113, 120]	118	[111, 121]	115	[110, 116]
Uganda	119	[110, 123]	116	[103, 126]	120	[112, 123]
Burkina Faso	120	[119, 126]	114	[112, 118]	124	[124, 126]
Cameroon	121	[119, 125]	124	[117, 131]	114	[111, 122]
Togo	122	[114, 123]	115	[111, 119]	125	[119, 126]
Mozambique	123	[117, 126]	123	[118, 128]	119	[112, 122]
Benin	124	[116, 130]	107	[102, 116]	131	[127, 131]
Niger	125	[118, 127]	119	[116, 126]	126	[114, 127]
Mali	126	[115, 127]	128	[120, 128]	121	[111, 122]
Angola	127	[122, 132]	129	[128, 132]	117	[114, 132]
Yemen	128	[117, 131]	132	[125, 132]	109	[101, 118]
Mauritania	129	[127, 132]	121	[117, 126]	132	[131, 132]
Burundi	130	[129, 131]	127	[124, 132]	130	[126, 130]
Iraq	131	[127, 132]	130	[113, 131]	127	[127, 130]
Guinea	132	[128, 132]	131	[129, 132]	128	[126, 130]

Source: European Commission, Joint Research Centre, 2022.

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.

Sensitivity analysis results

Complementary to the uncertainty analysis, sensitivity analysis has been used to identify which of the modeling assumptions have the greatest impact on certain country rankings. 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 in 2022 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 three economies – United Republic of Tanzania, Zimbabwe and Nigeria – shift rank by more than 20 positions.

All in all, the published GII 2022 rankings 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 2022 report should consider an economy's ranking in the GII 2022 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 modeling choices.

These confidence intervals also have to be taken into account when comparing economy rank changes from one year to the next at the GII or Innovation Sub-Index level in order to avoid drawing erroneous conclusions about an economy's 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 2022 has led to a net increase in the reliability of economy rankings for both the GII and the two sub-indices. Furthermore, the adoption in 2017 of less stringent criteria for 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 the most sensitive rankings

			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	0.992	0	3	1 *	2
	EM imputation vs. no imputation of missing data	0.993	0	2	0	2
	Geometric average and EM imputation vs. arithmetic average and missing values	0.991	0	4	0	8
Input Sub-Index	Geometric vs. arithmetic average	0.997	0	0	0	2
	EM imputation vs. no imputation of missing data	0.995	0	2	0	1
	Geometric average and EM imputation vs. arithmetic average and missing values	0.993	0	3	0	4
Output Sub-Index	Geometric vs. arithmetic average	0.998	0	0	0	0
	EM imputation vs. no imputation of missing data	0.986	0	7	2 **	7
	Geometric average and EM imputation vs. arithmetic average and missing values	0.985	0	11	3 ***	6

Source: European Commission, Joint Research Centre, 2022.

Notes: EM, expectation–maximization (EM) algorithm. * Brunei Darussalam (down from 92nd to 121st in the GII). ** United Republic of Tanzania (down from 99th to 122nd in the Output Sub-Index), Zimbabwe (down from 93rd to 119th in the Output Sub-Index). *** United Republic of Tanzania (down from 99th to 121st in the Output Sub-Index), Zimbabwe (down from 93rd to 117th in the Output Sub-Index) and Nigeria (down from 107th to 128th 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 be unfair 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 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 be unfair 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.[9] 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 weights. The DEA score can be interpreted as a measure of the "distance to the efficiency frontier."

Table 6 presents pie shares and DEA scores for the top 25 economies in the GII 2022 alongside their respective GII 2022 rankings. 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, the United States and Sweden are the only economies to obtain a perfect DEA score of 1.00. In the case of Switzerland, this is achieved by assigning 19 percent of its DEA score to the two output pillars, namely Knowledge and technology outputs and Creative outputs, while 9 percent to 16 percent of Switzerland's DEA score comes from the five remaining pillars. Using a different approach, Sweden has assigned 20 percent of its DEA score to three input pillars – Human capital and research, Infrastructure and Business sophistication – and to one output pillar – Knowledge and technology outputs – while just 5 to 10 percent of its DEA score comes from the output pillar capturing Creative outputs, and from the input pillars measuring Institutions and Market sophistication. Switzerland, the United States and Sweden are closely followed by Singapore (0.99) and the United Kingdom (0.97) in terms of efficiency. Figure 3 shows how close the DEA scores and the GII 2022 scores are for all 132 economies (Pearson correlation of 0.995).

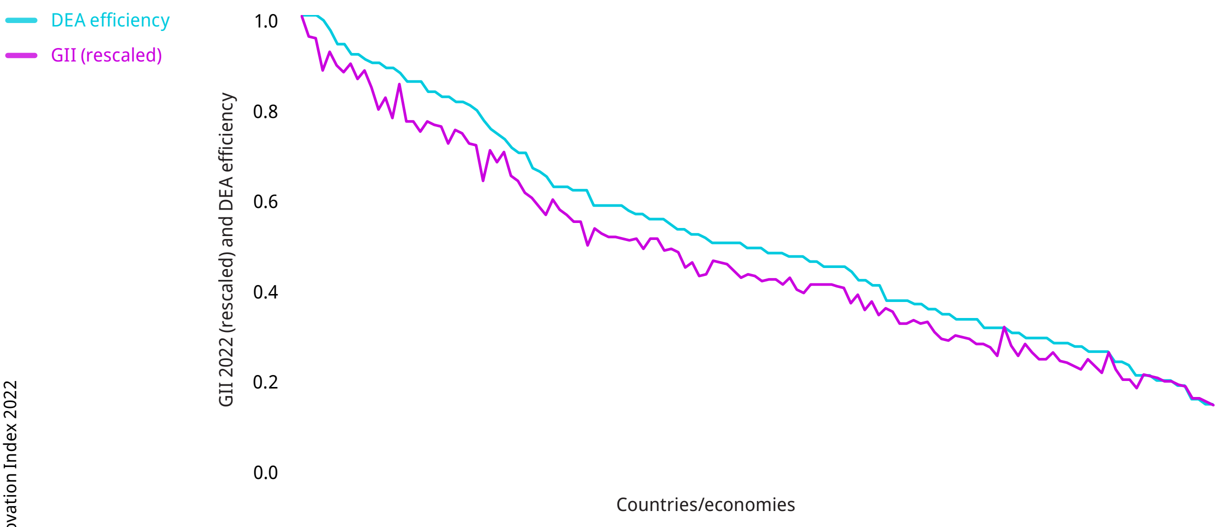
Table 6 Pie shares (absolute terms) and efficiency scores for the top 25 GII 2022 economies

	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	0.09	0.10	**0.15**	**0.13**	**0.16**	**0.19**	**0.19**	1.00	1	1	0
United States	0.06	**0.19**	0.07	**0.20**	**0.20**	**0.19**	0.09	1.00	1	2	1
Sweden	0.05	**0.20**	**0.20**	0.10	**0.20**	**0.20**	0.05	1.00	1	3	2
Singapore	**0.20**	**0.20**	**0.15**	**0.15**	**0.20**	0.05	0.05	0.99	4	7	3
United Kingdom	0.05	**0.20**	**0.20**	**0.20**	0.10	0.05	**0.20**	0.97	5	4	–1
Republic of Korea	0.05	**0.20**	**0.20**	0.05	**0.20**	0.10	**0.20**	0.94	6	6	0
Finland	**0.20**	**0.20**	**0.20**	0.05	**0.20**	0.10	0.05	0.94	6	9	3
Netherlands	**0.20**	**0.20**	**0.20**	0.05	**0.20**	0.05	0.10	0.92	8	5	–3
Denmark	**0.20**	**0.20**	**0.20**	0.10	**0.20**	0.05	0.05	0.92	8	10	2
Germany	0.11	**0.20**	**0.20**	**0.19**	0.05	0.05	**0.20**	0.91	10	8	–2
France	0.08	**0.20**	**0.20**	**0.20**	0.07	0.05	**0.20**	0.90	11	12	1
Hong Kong, China	**0.20**	0.10	**0.20**	**0.20**	0.05	0.05	**0.20**	0.90	11	14	3
Japan	**0.20**	**0.14**	**0.20**	**0.15**	**0.20**	0.05	0.05	0.89	13	13	0
Canada	**0.20**	**0.20**	**0.20**	**0.20**	0.10	0.05	0.05	0.89	13	15	2
China	0.05	0.05	**0.20**	**0.17**	**0.20**	**0.13**	**0.20**	0.88	15	11	–4
Austria	**0.20**	**0.20**	**0.20**	0.10	**0.20**	0.05	0.05	0.86	16	17	1
Estonia	**0.20**	0.05	**0.20**	**0.20**	**0.19**	0.05	0.11	0.86	16	18	2
Norway	**0.20**	**0.20**	**0.20**	0.10	**0.20**	0.05	0.05	0.86	16	22	6
Israel	0.05	0.12	**0.20**	**0.18**	**0.20**	**0.20**	0.05	0.84	19	16	–3
Luxembourg	**0.20**	0.05	**0.20**	0.10	**0.20**	0.05	**0.20**	0.84	19	19	0
Iceland	**0.20**	0.10	**0.20**	0.05	**0.20**	0.05	**0.20**	0.83	21	20	–1
Australia	**0.20**	**0.20**	**0.20**	**0.15**	**0.15**	0.05	0.05	0.83	21	25	4
Malta	**0.20**	0.05	**0.20**	0.10	**0.20**	0.05	**0.20**	0.82	23	21	–2
Ireland	**0.20**	**0.20**	**0.20**	0.05	**0.20**	0.10	0.05	0.82	23	23	0
New Zealand	**0.20**	**0.20**	**0.20**	0.12	**0.18**	0.05	0.05	0.81	25	24	–1

Source: European Commission, Joint Research Centre, 2022.

Notes: Pie shares are in absolute terms, bounded by 0.05 and 0.20 for all seven innovation pillars. In the GII 2022, however, each of the five input pillars has a fixed weight of 0.10 while each of the two output pillars each has a fixed weight of 0.25. Darker and bolder colors represent a higher contribution by 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 according to GII 2022 ranking.

Figure 3 GII 2022 scores and DEA "distance to the efficiency frontier" scores



Source: European Commission, Joint Research Centre, 2022.

Notes: For comparison purposes, the GII scores were rescaled by dividing them by the result of the best performer in the overall GII 2022 (Switzerland).

Conclusion

The JRC-COIN analysis suggests that the conceptualized multilevel structure of the GII 2022 – 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 over the years have helped enhance an already strong statistical coherence within the GII framework, in which the capacity of the 81 indicators to distinguish between economies' performances is maintained at the sub-pillar level or higher in all but five cases.

The decision to not 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 to not 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 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.

Strong correlations between the GII components are proven to not be a sign of redundancy of information within the GII. For more than 34 percent (up to 67 percent) of the 132 economies included in the GII 2022, 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 not immediately apparent from a separate analysis of each pillar. At the same time, this finding points to there being value in duly considering the 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 2022 rankings lie within the simulated 90 percent confidence intervals that take into consideration the unavoidable uncertainties inherent in an 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, such intervals are narrow enough for meaningful inferences to be drawn: the intervals comprise fewer than 10 positions for 73 percent (97 out of 132) of economies. Some caution is needed, mainly for three countries – Belarus, Brunei Darussalam and Zimbabwe – whose GII rankings 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 2022 or in the two sub-indices, can be considered to be representative of the many possible scenarios: 75 up to 83 percent of economies shift fewer than three positions with respect to the median rank within the GII or either of the Input or Output Sub-Indices.

All things considered, the present JRC-COIN audit findings confirm that the GII 2022 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 be unfair 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 15 years of constant refinement, 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–GII 2018* series range.

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 Saisana *et al*., 2018.

9 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 2022 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:

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

subject to

$w_{ij} \geq 0$, where, j=1,...,7, i = 1,...,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 and 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.

Saisana, M., M. Dominguez Torreiro and D. Vertesy (2018). Joint Research Centre statistical audit of the 2018 Global Innovation Index. In Cornell University, INSEAD and WIPO, *Global Innovation Index 2018: Energizing the World with Innovation*. Ithaca, Fontainebleau and Geneva: Cornell University, INSEAD and the World Intellectual Property Organization.

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 September 2016.

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