Joint Research Centre (JRC) statistical audit of the 2023 Global Innovation Index

Begoña Cabeza Martínez, Jaime Lagüera González, Ana Rita Neves, Panagiotis Ravanos, Michaela Saisana, Oscar Smallenbroek and Carlos Tacao Moura, European Commission, JRC, Ispra, Italy

WIPO

Conceptual and practical challenges are inevitable when trying to understand and model the fundamentals of innovation at the national level worldwide. Now in its 16th edition, the Global Innovation Index (GII) 2023, 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 80 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 informed 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 13th 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 allow a better appreciation of the robustness of these rankings to the choice of 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.

Box 1 Conceptual and statistical coherence in the GII 2023 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 (88 percent of available data refer to 2020 or a later year)
- 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 the GII Advisory Board members)
- a one-off qualitative audit (by the WIPO Internal Oversight Section)[2]
- external qualitative review (by JRC-COIN and international experts)

Source: European Commission, Joint Research Centre, 2023.

Conceptual and statistical coherence within the GII framework

The GII model was assessed by the JRC-COIN in June 2023. Suggestions for fine-tuning certain aspects 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 Box 1.

Step 1: Conceptual consistency

A total of 80 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. On the advice of JRC-COIN, the GII developers normalized nine more indicators to a 0–100 range in the 2023 edition, so that all indicators have the same range, which facilitates their individual contributions to the overall index score.

The 2023 edition of the GII includes some changes to the indicators.

- Printing and other media (indicator 7.2.4 in last year's edition) has been dropped from the Creative goods and services sub-pillar (7.2).
- In sub-pillar 6.2 Knowledge impact, the New businesses indicator has been replaced by a new indicator, Unicorn valuation (6.2.2).
- Indicator 6.3.5 ISO 9001 (previously ISO 9001 quality certificates) moved from sub-pillar 6.2 Knowledge impact to sub-pillar 6.3 Knowledge diffusion.
- The computation methodology of the following indicators changed: 1.3.2 Entrepreneurship policies and culture, 4.1.1 Finance for startups and scaleups and 7.3.3 GitHub commits. The first two indicators are computed using a five-year moving average in the period 2015–2022 (2017–2021 in the GII 2022), while the third indicator now includes both commit pushes received and sent (the indicator comprised only commit pushes received in the GII 2022).
- Finally, sub-pillar 1.1 Political environment has been renamed Institutional environment.

Step 2: Data checks

The data used for each economy were those most recently released within the period 2013 to 2023: 88 percent of the available data refer to 2020 or a later year. 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 based on previous 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 17 out of the 26 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 2023 is very good: 80 percent of data is available for 85 percent of the economies covered (equivalent to 112 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. In 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.[3] 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, 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 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.[4]

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 used as 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, two indicators have a weight of 0.5 – 1.2.1 Regulatory quality and 1.2.2 Rule of law – both in the input sub-pillar 1.2 Regulatory environment. Two sub-pillars are also given a weight of 0.5 – 7.2 Creative goods and services and 7.3 Online creativity.

Despite this weighting adjustment, four indicators (3.2.3 Gross capital formation, 4.1.3 Loans from microfinance institutions, 5.3.4 FDI net inflows and 6.2.1 Labor productivity growth) 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.[5] These four 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 or Labor productivity growth, which calls either for better formulation of these indicators or for better metrics for those concepts. That said, the other 76 indicators out of the total of 80 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 63 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 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 (0.86–0.94). 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 (0.96 and 0.97).

Finally, the two sub-indices are equally important in the overall GII. The GII is built as a simple arithmetic average of the Input Sub-Index and the Output Sub-Index. In fact, the Pearson correlation coefficients of the two sub-indices with the GII (0.96 and 0.97, respectively), and the correlation between themselves (0.91), suggests that they are effectively placed on an equal footing.

Concluding remarks

Overall, tests so far show that the grouping of variables into sub-pillars, pillars and an overall index is statistically coherent within the GII 2023 framework and that the GII has a balanced structure at each aggregation level. Furthermore, this year, all but four of the 80 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.[6] The only recommended possible refinement to the GII framework relates to four indicators – 3.2.3 Gross capital formation, 4.1.3 Loans from microfinance institutions, 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. Contrary to expectations, an economy's innovation level has proved to be almost entirely independent of the Gross capital formation, Loans from microfinance institutions, FDI net inflows and Labor productivity growth within the country concerned.

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

		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 Institutional environment	0.93	0.77	0.84	0.66	0.80	0.71	0.72
	1.2 Regulatory environment	0.86	0.70	0.73	0.56	0.70	0.62	0.63
	1.3 Business environment	0.83	0.43	0.43	0.40	0.52	0.39	0.35
	2.1 Education	0.52	0.78	0.66	0.56	0.61	0.58	0.61
	2.2 Tertiary education	0.57	0.81	0.73	0.61	0.60	0.58	0.62
	2.3 Research and development (R&D)	0.69	0.89	0.76	0.78	0.91	0.90	0.83
	3.1 Information and communication technologies (ICTs)	0.68	0.83	0.93	0.75	0.76	0.75	0.81
	3.2 General infrastructure	0.70	0.74	0.80	0.64	0.71	0.67	0.62
	3.3 Ecological sustainability	0.56	0.61	0.81	0.54	0.66	0.66	0.67
	4.1 Credit	0.58	0.70	0.65	0.83	0.65	0.64	0.65
	4.2 Investment	0.59	0.58	0.52	0.79	0.63	0.60	0.62
	4.3 Trade, diversification, and market scale	0.40	0.68	0.70	0.76	0.63	0.71	0.68

Table 1 Continued

	5.1 Knowledge workers	0.64	0.86	0.81	0.72	0.93	0.83	0.81
	5.2 Innovation linkages	0.80	0.79	0.74	0.72	0.91	0.83	0.76
	5.3 Knowledge absorption	0.65	0.73	0.73	0.68	0.89	0.79	0.78
Innovation Output Sub-Index	6.1 Knowledge creation	0.61	0.85	0.72	0.73	0.87	0.92	0.84
	6.2 Knowledge impact	0.61	0.70	0.68	0.74	0.76	0.87	0.72
	6.3 Knowledge diffusion	0.55	0.74	0.79	0.67	0.78	0.90	0.75
	7.1 Intangible assets	0.45	0.70	0.66	0.65	0.68	0.72	0.91
	7.2 Creative goods and services	0.65	0.75	0.79	0.72	0.82	0.81	0.82
	7.3 Online creativity	0.73	0.81	0.79	0.70	0.85	0.80	0.84

Source: European Commission, Joint Research Centre, 2023.

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 that this is not the case. In fact, for more than 34 percent (up to 70 percent) of the 132 economies included in the GII 2023, 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 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 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	27.3	7.6	5.3	12.1	6.8	6.1	3.8
20–29	21.2	14.4	9.8	13.6	13.6	6.8	7.6
10–19	22.0	23.5	30.3	32.6	22.7	22.0	35.6
10 or more*	70.5	45.5	45.5	58.3	43.2	34.8	47.0
5–9	18.2	32.6	28.8	18.9	24.2	30.3	24.2
Less than 5	11.4	18.2	22.0	21.2	28.0	31.1	23.5
Same rank	0.0	3.8	3.8	1.5	4.5	3.8	5.3
Total**	100	100	100	100	100	100	100
Spearman rank correlation coefficient with the GII	0.74	0.91	0.93	0.86	0.91	0.94	0.94

Source: European Commission, Joint Research Centre, 2023.
Notes: * This row is the sum of the previous three rows. ** This row is the sum of all non-orange.

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 within 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 2023 individual economy rankings.

As suggested by the relevant literature on composite indicators,[7] the robustness assessment was based on Monte Carlo simulation and multi-modeling approaches, applied to “error-free” data where potential outliers, 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 4,000 runs of different sets of weights for the seven GII pillars. Weights were assigned to the pillars based on random perturbations 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, 2023.
Note: The Amelia II package was used to create 100 expectation–maximization (EM) imputed data sets. The median values of these data were used for the uncertainty analysis.

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 16 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).[8]

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.[9]

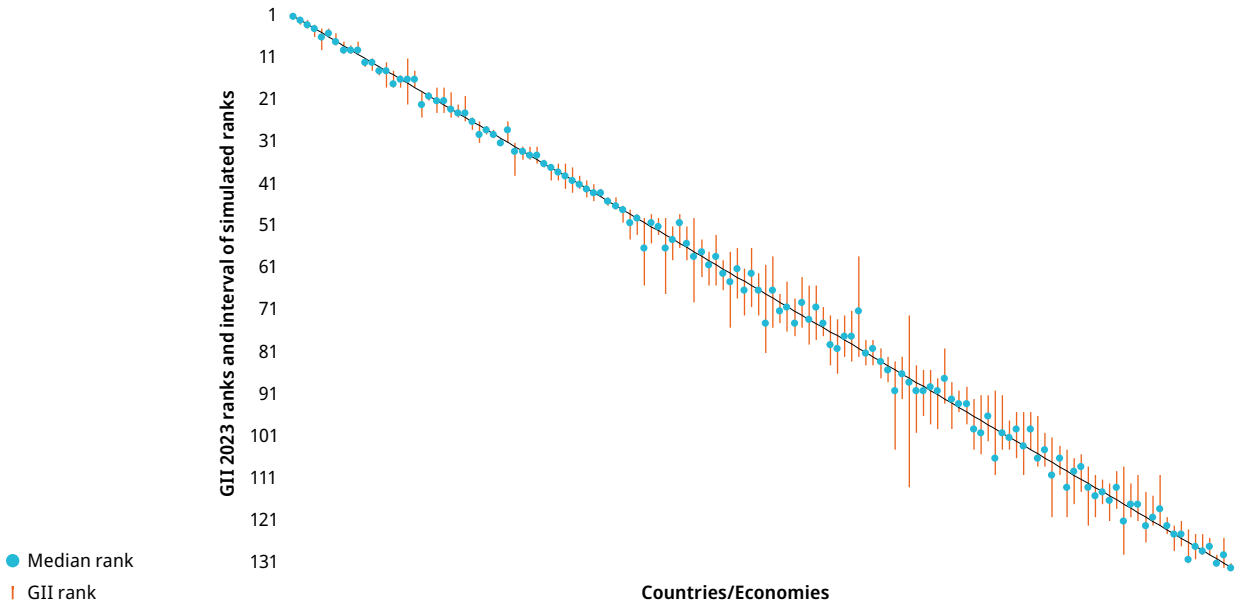
Four models were tested based on the combination of no imputation versus EM imputation and arithmetic versus geometric average. A random combination of these choices plus a random set of perturbed weights were used in a total of 4,000 simulations for the GII and each of the two sub-indices (see Table 3 for a summary of the uncertainties considered).

Uncertainty analysis results

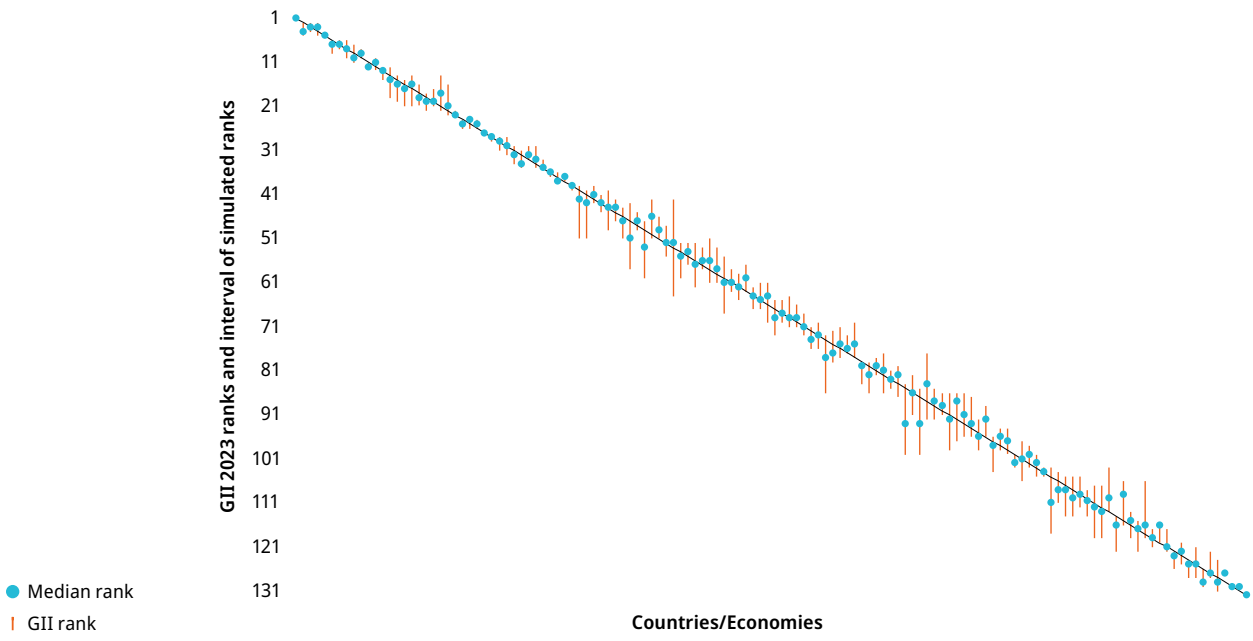
The main results of the robustness analysis are shown in Figure 1, with median ranks and 90 percent confidence intervals computed across the 4,000 Monte Carlo simulations for the GII and the two sub-indices. Economies are in ascending order (best to worst performing) according to their reference rank (black line), with the dot representing the median rank over the simulations.

Figure 1 Robustness analysis of the GII, Input and Output Sub-Indices

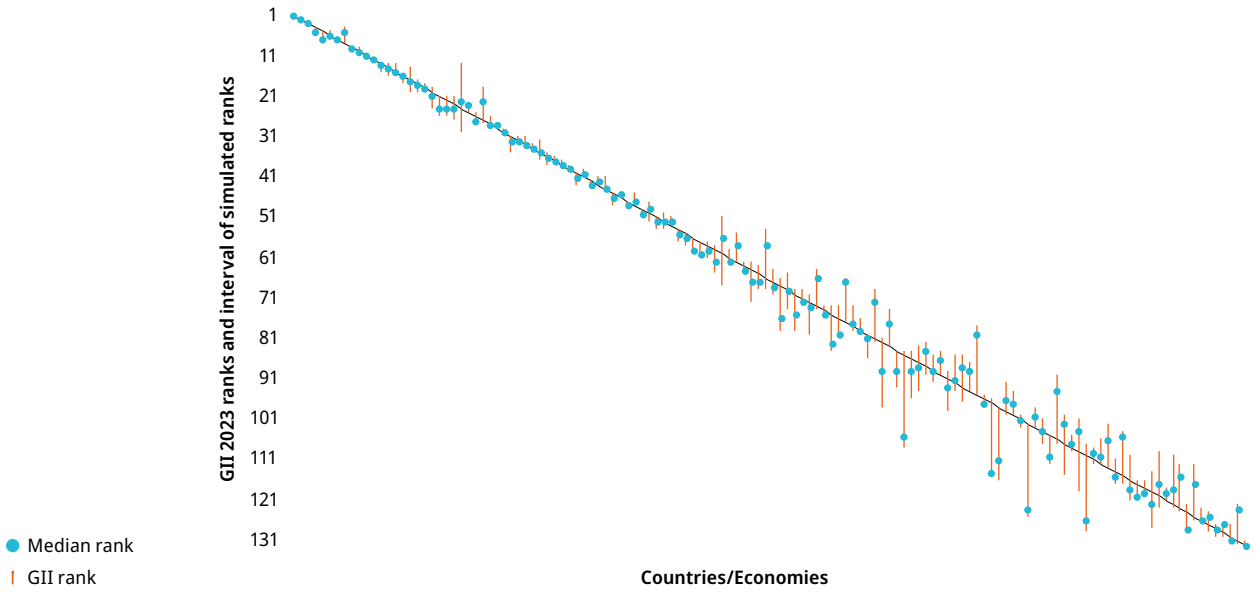
(a) GII rank vs. median rank, 90 percent confidence intervals



(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, 2023.
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 2023 rank is 0.995; between the median rank and the Innovation Input 2023 rank is 0.999; and between the median rank and the Innovation Output 2023 rank is 0.991.

All published GII 2023 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 89 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. Five economies – Zimbabwe, Botswana, Bahrain, Belarus and Brunei Darussalam – have 90 percent confidence interval widths of more than 20 positions (21, 24, 21, 24 and 41 positions, respectively). Consequently, their rankings (117th, 85th, 67th, 80th and 87th) 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 2023 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 one economy – Brunei Darussalam – which has a 90 percent confidence interval width of more than 20 positions (22). A similar degree of caution is needed in the Output Sub-Index for three economies – Botswana, Côte d'Ivoire and Ghana – which have 90 percent confidence interval widths of more than 20 positions (up to 24 for Ghana). 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 (three compared to 13 in the GII 2019 edition), which is a noteworthy improvement.

Although the rankings for a few economies, in the GII 2023 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, 80 percent of the economies are found to shift fewer than three positions with respect to the median rank in the GII and the Input Sub-Index; however, the percentage for the Output Sub-Index is lower, at 62 percent.

In order to offer full transparency and complete information, Table 4 reports the GII 2023 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[10] 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 2023 and Input/Output Sub-Indices: rankings and 90 percent confidence intervals

	GII 2023		Input Sub-Index		Output Sub-Index	
	Rank	Interval	Rank	Interval	Rank	Interval
Switzerland	1	[1, 1]	3	[2, 4]	1	[1, 1]
Sweden	2	[2, 3]	4	[2, 5]	3	[3, 3]
United States	3	[2, 4]	2	[2, 5]	4	[4, 6]
United Kingdom	4	[3, 6]	6	[6, 9]	2	[2, 2]
Singapore	5	[4, 9]	1	[1, 1]	12	[12, 13]
Finland	6	[4, 6]	5	[4, 5]	9	[9, 10]
Netherlands (Kingdom of the)	7	[5, 8]	10	[8, 10]	5	[5, 8]
Germany	8	[7, 10]	13	[13, 15]	6	[5, 6]
Denmark	9	[8, 10]	7	[6, 8]	10	[9, 10]
Republic of Korea	10	[7, 10]	12	[10, 13]	7	[7, 8]
France	11	[11, 13]	17	[14, 21]	11	[11, 11]
China	12	[11, 14]	25	[24, 26]	8	[4, 8]
Japan	13	[13, 15]	11	[11, 12]	14	[13, 16]
Israel	14	[12, 18]	21	[14, 22]	13	[13, 15]
Canada	15	[14, 18]	9	[7, 11]	20	[19, 24]
Estonia	16	[15, 18]	14	[12, 19]	16	[16, 18]
Hong Kong, China	17	[11, 22]	8	[6, 10]	24	[13, 30]
Austria	18	[14, 18]	18	[16, 21]	15	[13, 16]
Norway	19	[19, 25]	15	[14, 20]	28	[26, 29]
Iceland	20	[19, 21]	20	[17, 21]	25	[23, 25]
Luxembourg	21	[18, 24]	22	[16, 23]	23	[21, 27]
Ireland	22	[18, 24]	26	[24, 26]	18	[17, 20]
Belgium	23	[19, 25]	23	[22, 23]	22	[21, 26]
Australia	24	[22, 25]	16	[15, 21]	30	[29, 30]
Malta	25	[20, 26]	27	[27, 27]	17	[14, 20]
Italy	26	[25, 28]	35	[33, 35]	19	[18, 20]
New Zealand	27	[26, 31]	24	[24, 26]	31	[31, 35]
Cyprus	28	[27, 29]	33	[30, 33]	21	[21, 26]

Table 4 Continued

Spain	29	[28, 30]	28	[28, 29]	26	[25, 27]
Portugal	30	[30, 31]	31	[30, 34]	29	[28, 29]
Czech Republic	31	[26, 31]	34	[30, 35]	27	[19, 28]
United Arab Emirates	32	[31, 39]	19	[18, 22]	54	[54, 57]
Slovenia	33	[32, 35]	29	[28, 31]	38	[37, 39]
Lithuania	34	[32, 35]	32	[31, 35]	37	[36, 37]
Hungary	35	[32, 36]	36	[36, 37]	33	[31, 34]
Malaysia	36	[35, 37]	30	[28, 32]	46	[45, 46]
Latvia	37	[37, 40]	38	[37, 38]	39	[38, 40]
Bulgaria	38	[36, 40]	45	[42, 47]	34	[33, 35]
Türkiye	39	[36, 42]	52	[48, 55]	32	[31, 33]
India	40	[37, 43]	46	[44, 51]	35	[32, 37]
Poland	41	[39, 42]	50	[42, 51]	36	[35, 38]
Greece	42	[40, 44]	42	[39, 43]	41	[39, 41]
Thailand	43	[41, 45]	44	[40, 49]	43	[41, 43]
Croatia	44	[42, 44]	43	[41, 45]	44	[41, 44]
Slovakia	45	[44, 46]	51	[46, 51]	45	[45, 48]
Viet Nam	46	[44, 47]	57	[53, 58]	40	[40, 43]
Romania	47	[46, 50]	55	[52, 57]	47	[47, 49]
Saudi Arabia	48	[47, 54]	37	[36, 38]	67	[64, 70]
Brazil	49	[48, 53]	59	[53, 61]	49	[49, 50]
Qatar	50	[49, 65]	39	[39, 40]	70	[69, 79]
Russian Federation	51	[48, 55]	58	[51, 61]	53	[51, 53]
Chile	52	[49, 53]	48	[45, 49]	56	[56, 60]
Serbia	53	[49, 67]	41	[40, 51]	64	[62, 72]
North Macedonia	54	[51, 59]	49	[47, 60]	58	[57, 61]
Ukraine	55	[48, 56]	78	[70, 78]	42	[42, 44]
Philippines	56	[51, 59]	69	[64, 71]	52	[50, 54]
Mauritius	57	[49, 69]	40	[39, 51]	72	[70, 80]
Mexico	58	[54, 63]	77	[73, 77]	51	[51, 54]
South Africa	59	[57, 65]	71	[68, 73]	57	[57, 61]
Republic of Moldova	60	[53, 65]	81	[78, 82]	50	[47, 52]
Indonesia	61	[59, 66]	64	[62, 67]	63	[62, 65]
Iran (Islamic Republic of)	62	[57, 75]	87	[85, 100]	48	[45, 48]
Uruguay	63	[56, 68]	56	[52, 62]	73	[64, 74]
Kuwait	64	[61, 72]	67	[65, 73]	65	[63, 69]
Georgia	65	[56, 70]	54	[52, 60]	77	[66, 77]
Colombia	66	[62, 72]	63	[57, 63]	71	[69, 73]
Bahrain	67	[60, 81]	47	[43, 58]	86	[84, 96]
Mongolia	68	[58, 75]	79	[79, 84]	60	[51, 68]
Oman	69	[67, 74]	65	[61, 67]	78	[73, 79]
Morocco	70	[64, 76]	90	[86, 91]	55	[55, 58]
Jordan	71	[68, 77]	70	[66, 71]	76	[73, 81]
Armenia	72	[63, 75]	83	[81, 85]	62	[55, 62]
Argentina	73	[65, 79]	84	[80, 87]	59	[58, 65]

Table 4 Continued

Costa Rica	74	[65, 78]	66	[61, 70]	81	[69, 82]
Montenegro	75	[70, 77]	62	[59, 65]	83	[74, 83]
Peru	76	[72, 84]	60	[55, 68]	84	[84, 93]
Bosnia and Herzegovina	77	[73, 86]	75	[72, 79]	80	[80, 86]
Jamaica	78	[72, 82]	82	[77, 86]	69	[65, 74]
Tunisia	79	[71, 83]	96	[89, 96]	61	[59, 63]
Belarus	80	[58, 82]	88	[77, 92]	66	[54, 69]
Kazakhstan	81	[78, 84]	68	[65, 70]	87	[83, 94]
Uzbekistan	82	[78, 84]	72	[71, 76]	88	[82, 90]
Albania	83	[80, 87]	73	[70, 76]	94	[87, 94]
Panama	84	[82, 88]	93	[86, 96]	75	[73, 84]
Botswana	85	[83, 104]	61	[58, 63]	110	[107, 129]
Egypt	86	[82, 92]	99	[94, 100]	74	[73, 76]
Brunei Darussalam	87	[72, 113]	53	[42, 64]	125	[112, 126]
Pakistan	88	[84, 100]	113	[103, 113]	68	[66, 79]
Azerbaijan	89	[85, 96]	76	[71, 78]	104	[101, 107]
Sri Lanka	90	[85, 98]	103	[100, 105]	79	[76, 80]
Cabo Verde	91	[87, 99]	74	[73, 86]	106	[90, 107]
Lebanon	92	[80, 93]	86	[82, 91]	95	[78, 95]
Senegal	93	[88, 99]	95	[92, 99]	93	[85, 97]
Dominican Republic	94	[90, 95]	89	[85, 92]	96	[95, 98]
El Salvador	95	[89, 98]	102	[98, 103]	90	[84, 90]
Namibia	96	[92, 104]	80	[79, 86]	111	[108, 112]
Bolivia (Plurinational State of)	97	[91, 105]	91	[86, 99]	101	[100, 103]
Paraguay	98	[91, 102]	101	[97, 106]	92	[85, 94]
Ghana	99	[90, 110]	107	[105, 114]	85	[84, 108]
Kenya	100	[91, 104]	104	[103, 105]	91	[89, 99]
Cambodia	101	[97, 104]	97	[96, 104]	100	[94, 100]
Trinidad and Tobago	102	[95, 106]	92	[86, 97]	108	[105, 109]
Rwanda	103	[95, 110]	85	[84, 100]	113	[102, 113]
Ecuador	104	[95, 104]	98	[94, 99]	99	[92, 100]
Bangladesh	105	[96, 108]	114	[114, 122]	89	[85, 92]
Kyrgyzstan	106	[100, 108]	94	[87, 96]	112	[106, 112]
Madagascar	107	[101, 120]	125	[121, 128]	82	[81, 98]
Nepal	108	[103, 110]	106	[104, 111]	103	[98, 103]
Nigeria	109	[104, 120]	116	[113, 119]	98	[98, 116]
Lao People's Democratic Republic	110	[106, 117]	100	[100, 103]	120	[109, 123]
Tajikistan	111	[105, 114]	109	[105, 112]	107	[100, 115]
Côte d'Ivoire	112	[108, 122]	112	[107, 119]	102	[102, 125]
United Republic of Tanzania	113	[110, 120]	105	[103, 118]	123	[112, 124]
Togo	114	[111, 117]	120	[116, 120]	105	[105, 112]
Nicaragua	115	[112, 121]	110	[108, 114]	118	[116, 120]
Honduras	116	[109, 118]	115	[106, 116]	114	[111, 117]
Zimbabwe	117	[108, 129]	127	[122, 128]	97	[96, 115]
Zambia	118	[112, 120]	111	[107, 119]	122	[110, 123]

Table 4 Continued

Algeria	119	[110, 121]	118	[106, 119]	116	[110, 121]
Benin	120	[114, 126]	108	[105, 114]	128	[127, 130]
Uganda	121	[115, 122]	117	[115, 122]	121	[118, 121]
Guatemala	122	[110, 122]	121	[117, 122]	115	[104, 117]
Cameroon	123	[120, 124]	123	[120, 125]	117	[116, 121]
Burkina Faso	124	[122, 128]	119	[117, 121]	127	[124, 129]
Ethiopia	125	[121, 127]	130	[130, 131]	109	[101, 119]
Mozambique	126	[123, 131]	128	[124, 131]	124	[122, 129]
Mauritania	127	[124, 130]	122	[122, 126]	129	[127, 130]
Guinea	128	[124, 129]	131	[126, 132]	119	[114, 128]
Mali	129	[125, 129]	129	[124, 129]	126	[123, 126]
Burundi	130	[129, 131]	126	[126, 130]	130	[127, 131]
Niger	131	[125, 132]	124	[124, 128]	131	[122, 132]
Angola	132	[131, 132]	132	[131, 132]	132	[131, 132]

Source: European Commission, Joint Research Centre, 2023.
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 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, three economies, Ghana, Côte d'Ivoire and Brunei Darussalam – shift rank by more than 20 positions.

All in all, the published GII 2023 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 2023 report should consider an economy's ranking in the GII 2023 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 2023 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	0.995	0	0	1*	1
	EM imputation vs. no imputation of missing data	0.995	0	1	0	2
	Geometric average and EM imputation vs. arithmetic average and missing values	0.993	0	3	0	4
Input Sub-Index	Geometric vs. arithmetic average	0.997	0	0	0	2
	EM imputation vs. no imputation of missing data	0.996	0	2	0	1
	Geometric average and EM imputation vs. arithmetic average and missing values	0.995	0	1	0	4
Outpu Sub-Index	Geometric vs. arithmetic average	0.999	0	0	0	1
	EM imputation vs. no imputation of missing data	0.988	0	9	2**	4
	Geometric average and EM imputation vs. arithmetic average and missing values	0.986	0	9	2***	5

Source: European Commission, Joint Research Centre, 2023.
Notes: EM is the expectation–maximization algorithm. * Brunei Darussalam (down from 87th to 112th place in the GII). ** Ghana (down from 85th to 107th position in the Output Sub-Index) and Côte d'Ivoire (down from 102nd to 123rd in the Output Sub-Index). *** Ghana (down from 85th to 108th in the Output Sub-Index) and Côte d'Ivoire (down from 102nd to 125th 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.[11] In theory, each economy is free to decide on the relative contribution of each innovation pillar to its score, so as to achieve the best possible score in a computation that reflects its innovation strategy. In practice, the DEA method assigns a higher (lower) contribution to those pillars in which an economy is relatively strong (weak). Reasonable constraints 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 lower and upper 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 2023 alongside their respective GII 2023 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, Sweden, the United States and Singapore are the only economies to obtain a perfect DEA score of 1.00. In the case of Switzerland, this is achieved by assigning 20 percent of its DEA score to the two output pillars, namely Knowledge and technology outputs and Creative outputs, while between 5 and 20 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 two input pillars – Human capital and research and Business sophistication – and to one output pillar – Knowledge and technology outputs – while between 5 and 17 percent of its DEA score comes from the output pillar capturing Creative outputs and from the input pillars measuring Institutions, Infrastructure and Market sophistication. Switzerland, Sweden, the United States Singapore and are closely followed by Finland (0.98) and the United Kingdom (0.97) in terms of efficiency. Figure 2 shows how close the DEA scores and the GII 2023 scores are for all 132 economies (Pearson correlation of 0.992).[12]

Table 6 Pie shares (absolute terms) and efficiency scores for the top 25 GII 2023 economies

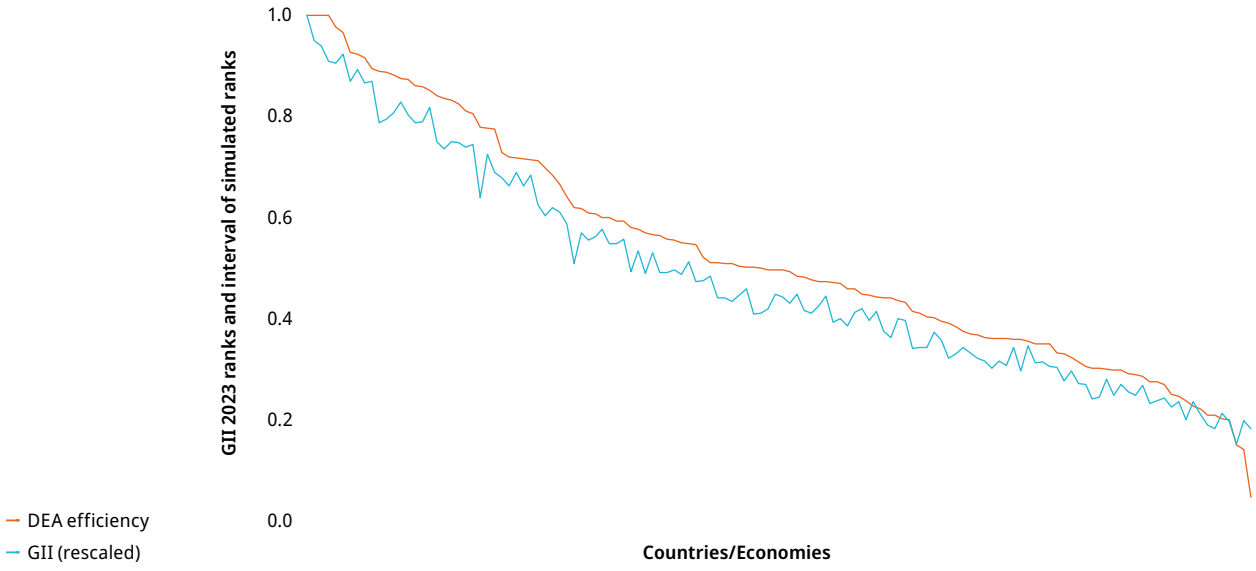
	Input pillars					Output pillars					
	Institutions	Human capital and research	Infrastructure	Market sophistication	Business sophistication	Knowledge and technology outputs	Creative outputs	Efficient frontier score (DEA)	Efficient frontier rank (DEA)	GII rank	Difference from GII rank
Switzerland	0.05	0.20	0.05	0.20	0.10	0.20	0.20	1.00	1	1	0
Sweden	0.05	0.20	0.17	0.05	0.20	0.20	0.13	1.00	1	2	1
United States	0.05	0.20	0.06	0.20	0.20	0.20	0.09	1.00	1	3	2
Singapore	0.20	0.20	0.05	0.16	0.20	0.14	0.05	1.00	1	5	4
Finland	0.19	0.20	0.20	0.05	0.18	0.13	0.05	0.98	5	6	1
United Kingdom	0.05	0.20	0.20	0.20	0.10	0.20	0.05	0.97	6	4	-2
Denmark	0.20	0.20	0.20	0.05	0.20	0.05	0.10	0.93	7	9	2
Netherlands (Kingdom of the)	0.20	0.15	0.20	0.05	0.19	0.16	0.05	0.92	8	7	-1
Republic of Korea	0.05	0.20	0.20	0.05	0.20	0.10	0.20	0.92	9	10	1
Germany	0.05	0.20	0.20	0.20	0.17	0.05	0.13	0.89	10	8	-2
Hong Kong, China	0.20	0.20	0.20	0.20	0.05	0.05	0.10	0.89	11	17	6
Canada	0.20	0.20	0.20	0.20	0.07	0.05	0.08	0.89	12	15	3
Japan	0.10	0.20	0.20	0.20	0.16	0.09	0.05	0.88	13	13	0
France	0.05	0.20	0.20	0.20	0.16	0.05	0.14	0.88	14	11	-3
Israel	0.05	0.20	0.12	0.18	0.20	0.20	0.05	0.87	15	14	-1
Austria	0.20	0.20	0.20	0.05	0.20	0.05	0.10	0.86	16	18	2
Estonia	0.20	0.06	0.20	0.20	0.20	0.05	0.09	0.86	17	16	-1
China	0.05	0.20	0.20	0.19	0.11	0.20	0.05	0.85	18	12	-6
Norway	0.20	0.20	0.20	0.07	0.20	0.05	0.08	0.84	19	19	0
Australia	0.20	0.20	0.20	0.20	0.07	0.05	0.08	0.84	20	24	4
Iceland	0.20	0.20	0.20	0.06	0.20	0.05	0.09	0.83	21	20	-1
Luxembourg	0.20	0.10	0.20	0.05	0.20	0.05	0.20	0.83	22	21	-1
Belgium	0.18	0.20	0.20	0.05	0.20	0.12	0.05	0.81	23	23	0

Table 6 Continued

Ireland	0.20	0.16	0.20	0.05	0.20	0.14	0.05	0.81	24	22	-2
Malta	0.20	0.05	0.20	0.10	0.20	0.05	0.20	0.78	26	25	-1

Source: European Commission, Joint Research Centre, 2023.
Notes: Pie shares are in absolute terms, bounded by 0.05 and 0.20 for all seven innovation pillars. In the GII 2023 ranking, however, each of the five input pillars has a fixed weight of 0.10 while each of the two output pillars has a fixed weight of 0.25. Darker 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 the DEA ranking. For countries with an efficient frontier score (DEA) equal to 1, there usually exist multiple alternative sets of pillar weights resulting in the same score (i.e., 1). The pillar shares depicted in this table for the first four countries (Switzerland, Sweden, United States and Singapore) were derived based on one of these alternative sets of weights. Different sets of pillar weights for these countries may arise from the use of different solver software, all of which, however, correspond to a DEA efficient frontier score of 1.

Figure 2 GII 2023 scores and DEA "distance to the efficiency frontier" scores



Source: European Commission, Joint Research Centre, 2023.
Notes: For comparison purposes, the GII scores were rescaled by dividing them by the result of the best performer in the overall GII 2023 (Switzerland).

Conclusion

The JRC-COIN analysis suggests that the conceptualized multilevel structure of the GII 2023 – with its 80 indicators, 21 sub-pillars, seven pillars and two 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 to enhance an already strong statistical coherence within the GII framework, in which the capacity of the 80 indicators to distinguish between economies' performances is maintained at the sub-pillar level or lower in all but four 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 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 not to be a sign of redundancy of information within the GII. For more than 34 percent (up to 70 percent) of the 132 economies included in the GII 2023, 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 2023 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 majority of economies, such intervals are narrow enough for meaningful inferences to be drawn: the intervals comprise 10 or fewer positions for 67 percent (89 out of 132) of the economies. Some caution is needed, mainly for five countries – Bahrain, Belarus, Botswana, 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 2023 or in the two sub-indices, can be considered to be representative of the many possible scenarios: 80 percent of the economies shift fewer than three positions with respect to the median rank within the GII, 87 percent within the Input and 62 percent within the Output Sub-Indices.

All things considered, the present JRC-COIN audit findings confirm that the GII 2023 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 16 years of constant refinement, to pave the way for better and more informed innovation policies worldwide.

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(2), 239–251.

Groeneveld, R.A. and G. Meeden (1984). Measuring skewness and kurtosis. *The Statistician*, 33(4), 391–399.

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: 167–185.

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), 609–634.

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), 307–323.

Saisana, M., B. D'Hombres and A. Saltelli (2011). Rickety numbers: Volatility of university rankings and policy implications. *Research Policy*, 40(1), 165–177.

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, NY, Fontainebleau and Geneva: Cornell University, INSEAD and 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(5), 853–871.

Vertesy, D. (2016). 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.

Endnotes

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 Available at: https://www.wipo.int/export/sites/www/about-wipo/en/oversight/docs/iaod/audit/audit-gii-exec-summary.pdf, IOD Ref: IA 2022-03, April 14, 2023.

3 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.

4 There is one indicator that fits these criteria in the 2023 GII normalized data, namely FDI net inflows, which has a skewness of -1.81 and kurtosis of 20. To aid transparency, the GII developers could elaborate on their (non) treatment of this indicator. As mentioned in Step 3: Statistical coherence, this is also one of the four indicators that appears to be non-influential in the GII framework, which may be partly attributable to the high level of skewness of the indicator.

5 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.

6 See note 3.

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

8 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.

9 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.

10 Saisana *et al*., 2018.

11 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 2023 for $n = 132$ economies, with yij 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:

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

subject to

$wij \geq 0$, 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).

12 For four countries – Benin, Burkina Faso, Mauritania and Niger – the DEA score is lower than the (rescaled) GII score because the restrictions appended in the DEA model to restrict the contribution of each of the seven pillars to no less than 5 percent and no more than 20 percent are less favorable for those countries compared to the GII weights.