ANNEX 3

JOINT RESEARCH CENTRE STATISTICAL AUDIT OF THE 2018 GLOBAL INNOVATION INDEX

Michaela Saisana, Marcos Domínguez-Torreiro, and **Daniel Vértesy,** European Commission, Joint Research Centre (JRC), Ispra, Italy

Conceptual and practical challenges are inevitable when trying to understand and model the fundamentals of innovation at the national level worldwide. In its 11th edition, the 2018 Global Innovation Index (GII) considers these conceptual challenges in Chapter 1 and deals with practical challenges—related to data quality and methodological choices—by grouping country-level data over 126 countries and across 80 indicators into 21 sub-pillars, 7 pillars, 2 sub-indices, and, finally, an overall index. This annex offers detailed insights into the practical issues related to the construction of the GII, analysing in depth the statistical soundness of the calculations and assumptions made to arrive at the final index rankings. Statistical soundness should be regarded as a necessary but not sufficient condition for a sound GII, since the correlations underpinning the majority of the statistical analyses carried out herein 'need not necessarily represent the real influence of the individual indicators on the phenomenon being measured'.[1] Consequently, the development of the GII must be nurtured by a dynamic iterative dialogue between the principles of statistical and conceptual soundness or, to put it another way, between the theoretical understanding of innovation and the empirical observations of the data underlying the variables.

The European Commission's Competence Centre on Composite Indicators and Scoreboards at the Joint Research Centre (JRC) in Ispra has been invited for the eighth consecutive year to audit the GII. As in previous editions, the present JRC audit focuses on the statistical soundness of the multi-level structure of the index as well as on the impact of key modelling assumptions on the results.[2] The independent statistical assessment of the GII provided by the JRC guarantees the transparency and reliability of the index for both policy makers and other stakeholders, thus facilitating more accurate priority setting and policy formulation in this particular field.

As in past GII reports, the JRC analysis complements the country rankings with confidence intervals for the GII, the Innovation Input Sub-Index, and the Innovation Output Sub-Index in order to better appreciate the robustness of these ranks to the computation methodology. This year a discussion of the Innovation Efficiency Ratio and the caution that needs to be attached to it is added. Finally, the JRC analysis includes an assessment of the added value of the GII and a measure of the distance to the efficient frontier of innovation by using data envelopment analysis.

Figure 1.

Conceptual and statistical coherence in the GII 2018 framework

Step 4. Qualitative review
- Internal qualitative review (INSEAD, WIPO, Cornell University)
- External qualitative review (JRC, international experts)

Step 3. Statistical coherence
- Treatment of pairs of highly collinear variables as a single indicator
- Assessment of grouping indicators into sub-pillars, pillars, sub-indices, and the GII
- Use of weights as scaling coefficients to ensure statistical coherence
- Assessment of arithmetic average assumption
- Assessment of potential redundancy of information in the overall GII

Step 2. Data checks
- Check for data recency (78% of available data refer to 2016–2017)
- Availability requirements per country: coverage ≥ 66% for the Input and the Output Sub-Indices separately and at least two sub-pillars per pillar
- Check for reporting errors (interquartile range)
- Outlier identification and treatment (skewness and kurtosis)
- Direct contact with data providers

Step 1. Conceptual consistency
- Compatibility with existing literature on innovation and pillar definition
- Scaling factors per indicator to represent a fair picture of country differences (e.g., GDP, population)

Source: European Commission, Joint Research Centre, 2018.

Conceptual and statistical coherence in the GII framework

An earlier version of the GII model was assessed by the JRC in April–May 2018. Fine-tuning suggestions were taken into account in the final computation of the rankings in an iterative process with the JRC aimed at setting the foundation for a balanced index. The entire process followed four steps (see Figure 1).

Step 1: Conceptual consistency

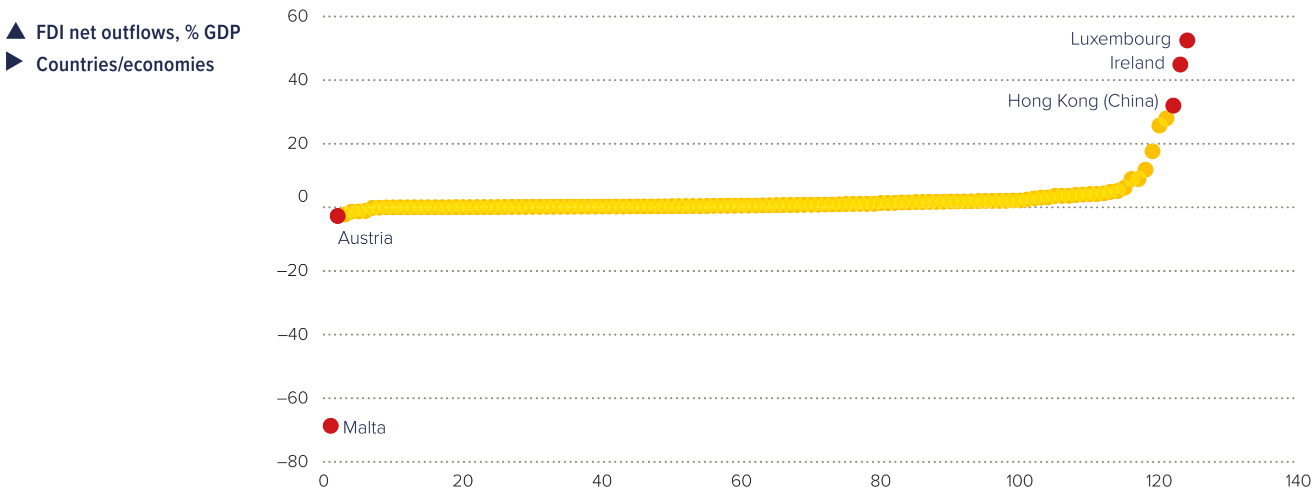
Eighty indicators were selected for their relevance to a specific innovation pillar on the basis of the literature review, expert opinion, country coverage, and timeliness. To represent a fair picture of country differences, indicators were scaled either at the source or by the GII team as appropriate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

Step 2: Data checks

The most recently released data within the period 2007–17 were used for each economy: 78% of the available data refer to 2016 or more recent years. In past editions, until 2015, countries were included if data availability was at least 60% across all variables in the GII framework. A more stringent criterion was adopted in 2016, following the JRC recommendation of past GII audits. That is, countries were included if data availability was at least 66% within each of the two sub-indices (i.e., 35 out of 53 variables within the Input Sub-Index and 18 out of the 27 variables in the Output Sub-Index) and at least two of the three sub-pillars in each pillar could be computed. This more stringent criterion for a country's inclusion in the GII was introduced in 2016 in order to ensure that country scores for the GII and for the two Input and Output Sub-Indices are not particularly sensitive to the missing values (as was the case for the Output Sub-Index scores of several countries in past editions). In practice, data availability for all countries included in the GII 2018 is very good: 80% of data are available for 87% (110 out of 126) of the countries in the sample. Potentially problematic indicators that could bias the overall results were identified on the

Figure 2.

Malta's outlier performance in FDI net outflows



Source: European Commission, Joint Research Centre, 2018.

Notes: Economies with the highest and lowest FDI outflow scores are highlighted. Skewness = –0.75; kurtosis = 28.16.

basis of two measures related to the shape of the distributions: skewness and kurtosis. In past editions, since 2011, values were treated if the indicators had absolute skewness greater than 2.0 and kurtosis greater than 3.5.[3] These criteria were decided jointly with the JRC back in 2011. In 2017, and after having analysed data in GII 2011–GII 2017, a less stringent criterion was adopted: an indicator was treated if the absolute skewness was greater than 2.25 and kurtosis greater than 3.5. These indicators were treated either by winsorization or by taking the natural logarithm (in case of more than five outliers; see Appendix IV Technical Notes in this report for details). In 2018, exceptional behaviour for the FDI net outflows (indicator 6.3.4) indicator was observed: Malta's outlier performance (see Figure 2) was not captured by the skewness and kurtosis criterion because of the symmetric behaviour of this indicator, whereby country values ranged between 68% and 52%. For this reason, and from this year on, it is recommended that the GII rule for the treatment of outliers be adjusted as follows:

- for indicators with absolute skewness greater than 2.25 and kurtosis greater than 3.5: use either winsorization or take the natural logarithm (in case of more than five outliers); and
- for indicators with absolute skewness less than 2.25 and kurtosis greater than 10.0: produce plots similar to the one presented in Figure 2 in order to identify potentially problematic values that need to be considered as outliers and treated accordingly.

Step 3: Statistical Coherence

Weights as scaling coefficients

Weights of 0.5 or 1.0 were jointly decided between the JRC and the GII team in 2012 to be scaling coefficients and not importance coefficients, with the aim of arriving at sub-pillar and pillar scores that were balanced in their underlying components (i.e., that indicators

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

	Sub-pillar	Institutions	Human capital and research	Infrastructure	Market sophistication	Business sophistication	Knowledge & technology outputs	Creative outputs
INNOVATION INPUT SUB-INDEX	1.1. Political environment	0.95	0.79	0.86	0.71	0.79	0.70	0.79
	1.2. Regulatory environment	0.92	0.71	0.72	0.62	0.74	0.66	0.72
	1.3. Business environment	0.85	0.67	0.70	0.62	0.66	0.64	0.63
	2.1. Education	0.57	0.77	0.55	0.38	0.52	0.50	0.52
	2.2. Tertiary education	0.63	0.81	0.67	0.50	0.51	0.53	0.56
	2.3. Research and development (R&D)	0.75	0.88	0.77	0.73	0.87	0.86	0.74
	3.1. Information and communication technologies (ICTs)	0.80	0.82	0.93	0.72	0.74	0.72	0.79
	3.2. General infrastructure	0.57	0.55	0.68	0.50	0.53	0.52	0.51
	3.3. Ecological sustainability	0.63	0.53	0.75	0.44	0.58	0.55	0.66
	4.1. Credit	0.63	0.53	0.55	0.86	0.57	0.50	0.58
	4.2. Investment	0.46	0.38	0.36	0.68	0.43	0.36	0.34
	4.3. Trade, competition, and market scale	0.52	0.65	0.72	0.70	0.62	0.63	0.61
	5.1. Knowledge workers	0.77	0.81	0.77	0.68	0.88	0.77	0.73
	5.2. Innovation linkages	0.58	0.50	0.53	0.52	0.77	0.60	0.64
	5.3. Knowledge absorption	0.64	0.64	0.63	0.56	0.84	0.79	0.64
INNOVATION OUTPUT SUB-INDEX	6.1. Knowledge creation	0.68	0.78	0.66	0.63	0.81	0.90	0.79
	6.2. Knowledge impact	0.54	0.61	0.62	0.47	0.62	0.79	0.62
	6.3. Knowledge diffusion	0.62	0.61	0.62	0.54	0.73	0.81	0.59
	7.1. Intangible assets	0.60	0.60	0.69	0.55	0.64	0.65	0.89
	7.2. Creative goods and services	0.70	0.65	0.72	0.63	0.68	0.70	0.83
	7.3. Online creativity	0.82	0.74	0.76	0.62	0.81	0.77	0.85

Source: European Commission, Joint Research Centre, 2018.

and sub-pillars can explain a similar amount of variance in their respective sub-pillars/pillars). Becker et al. (2017) and Paruolo et al. (2013) show that, in weighted arithmetic averages, the ratio of two nominal weights gives the rate of substitutability between two indicators, and hence can be used to reveal the relative importance of individual indicators. This importance can then be compared with ex-post measures of variables' importance, such as the non-linear Pearson correlation ratio. As a result of this analysis, 36 out of 80 indicators and two sub-pillars—7.2 Creative goods and services and 7.3 Online creativity—were assigned half weight while all other indicators and sub-pillars were assigned a weight of 1.0. In past GII editions, despite this weighting adjustment, a small number of indicators (seven in the GII 2017 edition) were found to be non-influential in the GII framework, implying that they could not explain at least 9% of countries' variation in the respective sub-pillar scores.[4] This year all 80 indicators are found to be sufficiently influential in the GII framework, which is worth highlighting as a very positive feature of this year's GII framework.

Principal components analysis and reliability item analysis

Principal component analysis (PCA) was used to assess 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 close to 60% (pillar 4: Market sophistication) up to 82% (pillar 1: Institutions) of the total variance in the three underlying sub-pillars. Furthermore, results confirm the expectation that the sub-pillars are more correlated with their own pillar than with any other pillar and that all correlation coefficients are close to or greater than 0.70 (see Table 1).

The five input pillars share a single statistical dimension that summarizes 82% of the total variance, and the five loadings (correlation

Table 2: Distribution of differences between pillar and GII rankings

Rank differences (positions)	Institutions	Human capital and research	Infrastructure	Market sophistication	Business sophistication	Knowledge and technology outputs	Creative outputs
More than 30	14.3%	11.9%	5.6%	21.4%	17.5%	8.7%	4.8%
20–29	11.9%	13.5%	17.5%	15.1%	11.9%	10.3%	9.5%
10–19	23.0%	26.2%	26.2%	27.8%	17.5%	25.4%	24.6%
10 or more*	**49.2%**	**51.6%**	**49.2%**	**64.3%**	**46.8%**	**44.4%**	**38.89%**
5–9	26.2%	23.0%	21.4%	16.7%	19.0%	27.8%	24.6%
Less than 5	21.4%	22.2%	22.2%	17.5%	31.0%	23.0%	31.7%
Same rank	3.2%	3.2%	7.1%	1.6%	3.2%	4.8%	4.8%
Total†	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100%**
Pearson correlation coefficient with the GII	0.89	0.89	0.90	0.79	0.92	0.93	0.93

Source: European Commission, Joint Research Centre, 2018.

* This column is the sum of the prior three rows.

† This column is the sum of all white rows.

coefficients) of these pillars are very similar to each other (0.84–0.92). This similarity suggests that the five pillars make roughly equal contributions to the variation of the Innovation Input Sub-Index scores, as envisaged by the developing team. The reliability of the Input Sub-Index, measured by the Cronbach alpha value, is very high at 0.94—well above the 0.70 threshold for a reliable aggregate.[5]

The two output pillars—Knowledge and technology outputs and Creative outputs—are strongly correlated with each other (0.81); they are also both strongly correlated with the Innovation Output Sub-index (0.95).

Finally, an important part of the analysis relates to clarifying the importance of the Input and Output Sub-Indices with respect to the variation of the GII scores. The GII is built as the simple arithmetic average of the five input sub-pillars and the two output sub-pillars, which implies that the input-related pillars have a weight of 5/7 versus a weight of 2/7 for the output-related pillars. Yet this does not imply that the input aspect is more important than the output aspect in determining the variation of the GII scores. In fact, the Pearson correlation coefficient of either the Input or the Output Sub-Index with the overall GII is 0.97 (and the two sub-indices have a correlation of 0.90), which suggests that the sub-indices are effectively placed on equal footing.

Overall, the tests so far show that the grouping of variables into sub-pillars, pillars, and an overall index is statistically coherent in the GII 2018 framework, and that the GII has a balanced structure at each aggregation level. Furthermore, this year all 80 indicators are found to be sufficiently influential in the GII framework—that is, each indicator explains at least 9% of countries' variation in the respective sub-pillar scores,[6] which is again worth highlighting as a very positive feature of this year's GII framework.

Added value of the GII

As already discussed, the Input and Output Sub-Indices correlate strongly with each other and with the overall GII. Furthermore, the five pillars in the Input Sub-Index have a very high statistical reliability. These results—the strong correlation between Input and Output Sub-Indices and the high statistical reliability of the five input pillars—may be interpreted by some as a sign of redundancy of information in the GII. The tests conducted by the JRC confirm that this is not the case. In fact, for more than 38% (up to 64%) of the 126 economies included in the GII 2018, the GII ranking and any of the seven pillar rankings differ by 10 positions or more (see Table 2). This is a desired outcome because it demonstrates the added value of the GII ranking, which helps to highlight other aspects of innovation that do not emerge directly by looking into the seven pillars separately. At the same time, this result points to the value of duly taking into account the GII pillars, sub-pillars, and individual indicators on their own merit. By doing so, country-specific strengths and bottlenecks on innovation can be identified and serve as an input for evidence-based policy making.

Step 4: Qualitative Review

Finally, the GII results—including overall country classifications and relative performances in terms of the Innovation Input or Output Sub-Indices—were evaluated to verify that the overall results are, to a great extent, consistent with current evidence, existing research, and prevailing theory. Notwithstanding these statistical tests and the positive outcomes on the statistical coherence of the GII structure, the GII model is and has to remain open for future improvements as better data, more comprehensive surveys and assessments, and new relevant research studies become available.

The impact of modelling assumptions on the GII results

An important part of the GII statistical audit is to check the effect of varying assumptions inside plausible ranges. Modelling assumptions with a direct impact on the GII scores and rankings relate to:

- setting up an underlying structure for the index based on a battery of pillars,
- choosing the individual variables to be used as indicators,
- deciding whether (and how) or not to impute missing data,
- deciding whether (and how) or not to treat outliers,
- selecting the normalization approach to be applied,
- choosing the weights to be assigned, and
- deciding on the aggregation rule to be implemented.

The rationale for these choices is manifold. For instance, expert opinion coupled with statistical analysis is behind the selection of the individual indicators, common practice and ease of interpretation suggests the use of a min-max normalization approach in the [0–100] range, the treatment of outliers is driven by statistical analysis, and simplicity and parsimony criteria seem to advocate for not imputing missing data. The unavoidable uncertainty stemming from the above-mentioned modelling choices is accounted for in the robustness assessment carried out by the JRC. More precisely, the methodology applied herein allows for the joint and simultaneous analysis of the impact of such choices on the aggregate scores, resulting in error estimates and confidence intervals calculated for the GII 2018 individual country rankings.

As suggested in the relevant literature on composite indicators,[7] the robustness assessment was based on Monte Carlo simulation and multi-modelling approaches, applied to 'error-free' data where potential outliers and eventual errors and typos have already been corrected in a preliminary stage. In particular, the three key modelling issues considered in the assessment of the GII were the treatment of missing data, the pillar weights, and the aggregation formula used at the pillar level.

Monte Carlo simulation comprised 1,000 runs of different sets of weights for the seven pillars in the GII. The weights were assigned to the pillars based on uniform continuous distributions centred in the reference values. The ranges of simulated weights were defined by taking into account both the need for a wide enough interval to allow for meaningful robustness checks and the need to respect the underlying principle of the GII that the Input and the Output Sub-Indices should be placed on an equal footing. As a result of these considerations, the limit values of uncertainty for the five input pillars are 10%–30%; the limit values for the two output pillars are 40%–60% (see Table 3).

The GII developing team, for transparency and replicability, has always opted not to estimate missing data. The 'no imputation' choice, which is common in similar contexts, might encourage economies not to report low data values. Yet this is not the case for the GII. After 11 editions of the GII, the index-developing team has not encountered any intentional no-reporting strategy. The consequence of the 'no imputation' choice in an arithmetic average is that it is equivalent to replacing an indicator's missing value for a given country with the respective sub-pillar score. Hence the available data (indicators) in the incomplete pillar may dominate, sometimes biasing the ranks up or down. To test the impact of the 'no imputation' choice, the JRC estimated missing data using the Expectation Maximization (EM) algorithm that was applied within each GII pillar.[8]

Regarding the aggregation formula, decision-theory practitioners challenge the use of simple arithmetic averages because of their fully compensatory nature, in which a comparative high advantage on a few indicators can compensate for a comparative disadvantage on many indicators.[9] To assess the impact of this compensability issue, the JRC relaxed the strong perfect substitutability assumption

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, 2018.

inherent in the arithmetic average and considered instead the geometric average, which is a partially compensatory approach that rewards economies with balanced profiles and motivates economies to improve in the GII pillars in which they perform poorly, and not just in *any* GII pillar.[10]

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

Uncertainty analysis results

The main results of the robustness analysis are shown with median ranks and 90% confidence intervals computed across the 4,000 Monte Carlo simulations for the GII and the two sub-indices (Figure 3 on page 78), and, for the first time this year, for the Efficiency Ratio (Figure 4 on page 79). The figures order economies from best to worst according to their reference rank (black line), the dot being the median rank over the simulations.

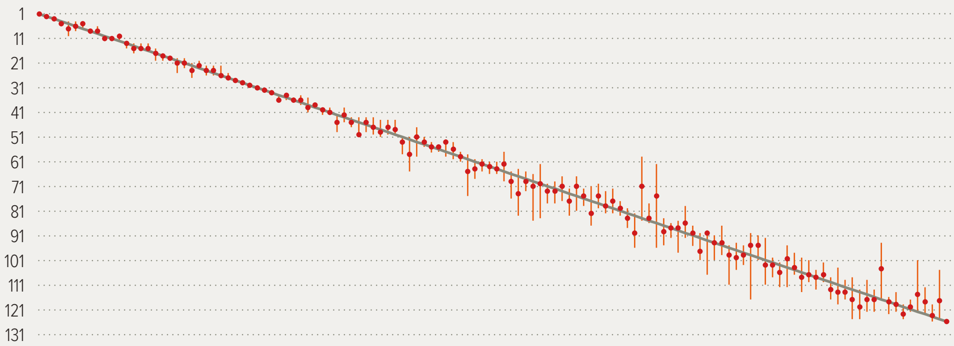
All published GII 2018 ranks lay within the simulated 90% confidence intervals, and for most economies these intervals are narrow enough for meaningful inferences to be drawn: there is a shift of fewer than 10 positions for 92 of the 126 economies. However, it is also true that only a small number of country ranks vary significantly with changes in weights and aggregation formula and because of the estimation of missing data. These six countries—Panama, The former Yugoslav Republic of Macedonia, Belarus, Rwanda, the Plurinational State of Bolivia, and Niger—have 90% confidence interval widths of more than 20 positions (up to 34 positions in the case of Belarus). Consequently, their GII ranks—between the 70th (Panama) and 122nd position (Niger) in the GII classification—should be interpreted cautiously and certainly not taken at face value. This is a remarkable improvement compared to GII versions until 2016, where more than 40 countries had confidence interval widths of more than 20 positions. This improvement in the confidence one can attach to the GII 2018 ranks is the direct result of the developers' choice since 2016 to adopt a more stringent criterion for an economy's inclusion, which requires at least 66% data availability within each of the two sub-indices. Some caution is also warranted in the Input Sub-Index for four economies—Bosnia and Herzegovina, Albania, Ukraine, and Panama—that have 90% confidence interval widths over 20 (up to 25 for Bosnia and Herzegovina). The Output Sub-Index is slightly more sensitive to the methodological choices: 14 countries—Panama, the United Republic of Tanzania, Oman, Paraguay, Mauritius, The former Yugoslav Republic of Macedonia, Ecuador, Zimbabwe, Namibia, Belarus, the Plurinational State of Bolivia, Guinea, Niger, and Togo—have 90% confidence interval widths over 20 (up

Figure 3.
Robustness analysis of the GII and Input and Output Sub-Indices

▲ GII 2018 ranks and interval of simulated ranks
▶ Countries/economies

● Median rank
— GII 2018 rank

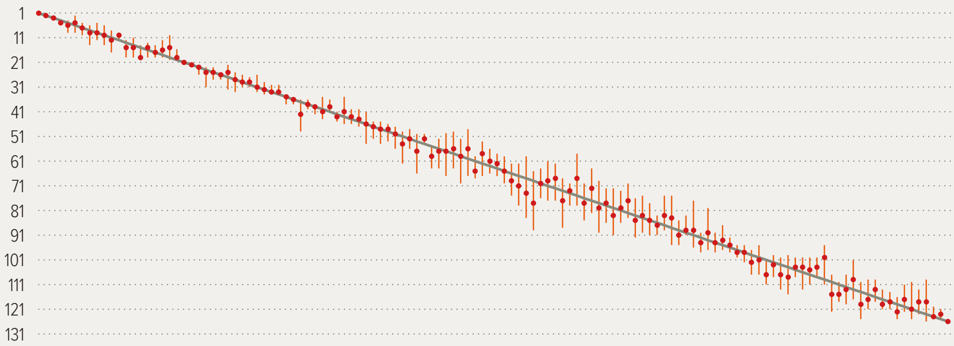
GII rank vs. median rank, 90% confidence intervals



▲ GII 2018 Input ranks and interval of simulated ranks
▶ Countries/economies

● Median rank
— GII 2018 Input rank

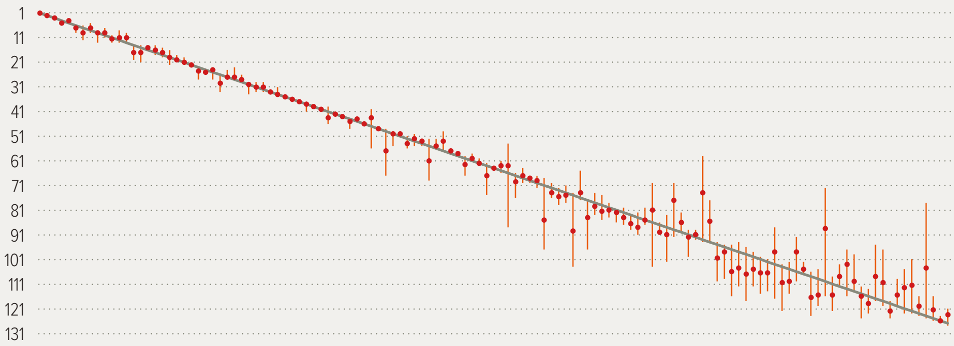
Input rank vs. median rank, 90% confidence intervals



▲ GII 2018 Output ranks and interval of simulated ranks
▶ Countries/economies

● Median rank
— GII 2018 Output rank

Output rank vs. median rank, 90% confidence intervals

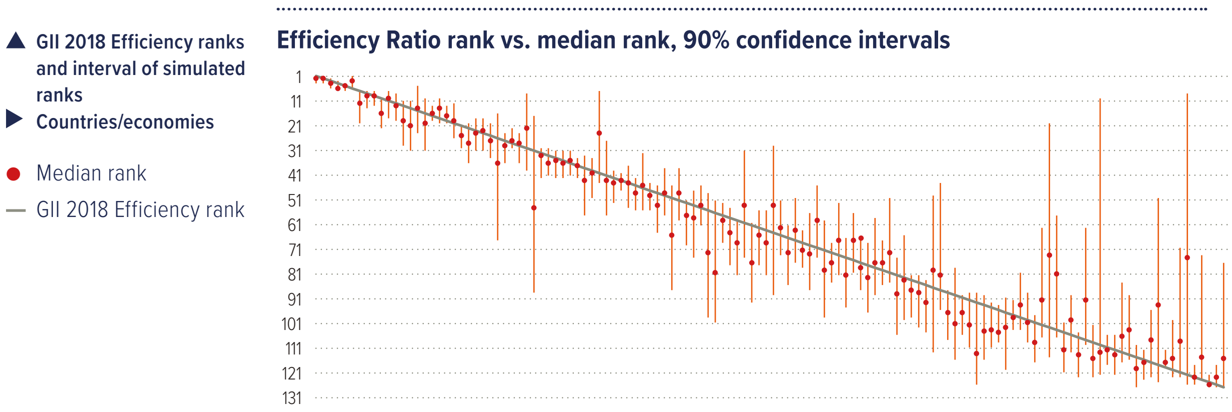


Source: European Commission, Joint Research Centre, 2018.

Notes: Median ranks and intervals are calculated over 4,000 simulated scenarios combining simulated weights, imputed versus missing values, and geometric versus arithmetic average at the pillar level. The Spearman rank correlation between the median rank and the GII 2018 rank is 0.996; between the median rank and the Innovation Input 2018 rank it is 0.997; and between the median rank and the Innovation Output 2018 rank it is 0.990.

Figure 4.

Robustness analysis of the Efficiency Ratio



Source: European Commission, Joint Research Centre, 2018.

Note: 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 within the Input and Output Sub-Indices. The Spearman rank correlation between the median rank and the Innovation Efficiency Ratio 2018 rank is 0.969

to 47 for Togo). This sensitivity is mostly the consequence of the estimation of missing data and the fact that there are only two pillars: this means that changes to the imputation method, weights, or aggregation formula have a more notable impact on the country ranks in the Innovation Output Sub-Index.

Although a few economy ranks, in the GII 2018 overall or in the two sub-indices, appear to be sensitive to the methodological choices, the published rankings for the vast majority can be considered representative of the plurality of scenarios simulated herein. Taking the median rank as the yardstick for an economy's expected rank in the realm of the GII's unavoidable methodological uncertainties, 75% of the economies are found to shift fewer than three positions with respect to the median rank in the GII, or in the Input and Output Sub-Indices.

For full transparency and information, Table 4 reports the GII 2018 Index and Input and Output Sub-Indices economy ranks together with the simulated 90% confidence intervals in order to better appreciate the robustness of the results to the choice of weights, of the aggregation formula, and the impact of estimating missing data (where applicable).

Emphasizing the identification of and relation between innovation input and output indicators seems irresistible from a policy perspective since doing so may possibly shed light on the effectiveness of innovation systems and policies. Yet this statistical audit shows that Innovation Efficiency Ratios, calculated as ratios of indices, have to be approached with care. Upon the request of the GII developing team, this year's JRC audit addresses the following question: How much confidence can one attach to the GII innovation efficiency scores and ranks for the countries worldwide? The Innovation Efficiency Ratio is calculated as the ratio of the Innovation Output Sub-Index score over the Innovation Input Sub-Index score. It shows how much innovation output a given country is getting for its inputs. Figure 4 shows the median ranks and 90% confidence intervals computed across the 4,000 Monte Carlo simulations for the Innovation Efficiency Ratio.

All published GII 2018 Innovation Efficiency ranks lay within the simulated 90% confidence intervals, but for most economies these intervals are too wide for meaningful inferences to be drawn: there is a shift of more than 20 positions for 60 of the 126 economies. Hence, while propagating the uncertainty in the two GII sub-indices to their sum—the GII—has a modest impact on the GII ranks (merely six countries shift more than 20 positions), this same

Table 4: GII 2018 and Input/Output Sub-Indices: Ranks and 90% confidence intervals

	GII 2018		Input Sub-Index		Output Sub-Index	
Country/Economy	Rank	Interval	Rank	Interval	Rank	Interval
Switzerland	1	[1, 1]	2	[2, 3]	1	[1, 1]
Netherlands	2	[2, 3]	9	[5, 12]	2	[2, 2]
Sweden	3	[2, 3]	3	[2, 4]	3	[3, 3]
United Kingdom	4	[4, 6]	4	[4, 6]	6	[6, 9]
Singapore	5	[4, 10]	1	[1, 1]	15	[14, 21]
United States of America	6	[4, 8]	6	[2, 9]	7	[6, 12]
Finland	7	[4, 7]	5	[4, 9]	8	[5, 9]
Denmark	8	[7, 9]	7	[5, 10]	13	[9, 13]
Germany	9	[6, 9]	17	[14, 19]	5	[4, 5]
Ireland	10	[10, 12]	18	[12, 19]	9	[8, 13]
Israel	11	[10, 12]	19	[10, 20]	11	[10, 13]
Korea, Republic of	12	[9, 12]	14	[11, 19]	12	[8, 13]
Japan	13	[13, 15]	12	[9, 12]	18	[15, 19]
Hong Kong (China)	14	[13, 17]	8	[6, 14]	21	[19, 22]
Luxembourg	15	[13, 16]	25	[23, 28]	4	[4, 6]
France	16	[13, 16]	16	[13, 19]	16	[14, 16]
China	17	[15, 20]	27	[22, 32]	10	[7, 11]
Canada	18	[17, 20]	10	[6, 14]	26	[26, 33]
Norway	19	[18, 19]	13	[12, 19]	24	[24, 26]
Australia	20	[19, 25]	11	[8, 17]	31	[29, 33]
Austria	21	[19, 23]	20	[16, 20]	28	[23, 28]
New Zealand	22	[21, 27]	15	[14, 20]	30	[29, 34]
Iceland	23	[20, 23]	22	[22, 23]	19	[16, 22]
Estonia	24	[22, 26]	26	[25, 28]	17	[14, 18]
Belgium	25	[22, 26]	21	[21, 21]	23	[23, 28]
Malta	26	[22, 27]	28	[25, 33]	14	[14, 20]
Czech Republic	27	[25, 28]	30	[27, 31]	20	[18, 21]
Spain	28	[27, 28]	23	[22, 26]	27	[24, 28]
Cyprus	29	[29, 30]	33	[30, 34]	22	[21, 22]
Slovenia	30	[29, 31]	31	[26, 33]	29	[26, 29]
Italy	31	[30, 31]	29	[26, 31]	32	[29, 33]
Portugal	32	[32, 33]	32	[29, 34]	33	[32, 34]
Hungary	33	[32, 33]	41	[36, 41]	25	[23, 28]
Latvia	34	[34, 37]	35	[35, 38]	38	[37, 41]
Malaysia	35	[34, 36]	34	[30, 34]	39	[38, 39]
Slovakia	36	[35, 37]	39	[38, 42]	36	[35, 36]
Bulgaria	37	[34, 38]	44	[40, 46]	34	[31, 34]
United Arab Emirates	38	[35, 41]	24	[23, 31]	54	[52, 55]
Poland	39	[38, 39]	38	[36, 40]	40	[40, 41]
Lithuania	40	[40, 42]	36	[35, 38]	44	[44, 48]
Croatia	41	[39, 41]	42	[41, 45]	42	[41, 42]
Greece	42	[42, 49]	40	[35, 44]	52	[51, 56]
Ukraine	43	[39, 45]	75	[58, 79]	35	[35, 36]
Thailand	44	[43, 47]	52	[48, 56]	45	[43, 45]
Viet Nam	45	[43, 51]	65	[59, 70]	41	[39, 46]
Russian Federation	46	[43, 49]	43	[35, 46]	56	[52, 56]
Chile	47	[43, 50]	45	[40, 47]	53	[50, 55]
Moldova, Republic of	48	[44, 51]	79	[72, 86]	37	[37, 38]
Romania	49	[44, 50]	49	[46, 53]	48	[47, 48]
Turkey	50	[44, 50]	62	[53, 67]	43	[43, 44]
Qatar	51	[51, 58]	47	[45, 52]	60	[59, 67]
Montenegro	52	[51, 65]	51	[49, 62]	55	[52, 69]
Mongolia	53	[47, 59]	66	[62, 75]	47	[40, 56]
Costa Rica	54	[51, 55]	64	[58, 67]	51	[49, 51]
Serbia	55	[53, 57]	56	[52, 64]	58	[56, 58]
Mexico	56	[54, 57]	54	[50, 54]	61	[58, 61]
India	57	[52, 59]	63	[56, 66]	57	[49, 57]
South Africa	58	[53, 60]	48	[45, 54]	65	[61, 66]
Georgia	59	[57, 61]	53	[50, 66]	62	[60, 63]
Kuwait	60	[58, 75]	81	[73, 86]	49	[48, 67]
Saudi Arabia	61	[60, 68]	46	[41, 54]	78	[74, 83]
Uruguay	62	[60, 65]	67	[62, 79]	59	[57, 59]
Colombia	63	[61, 66]	50	[47, 56]	72	[70, 76]

Country/Economy	GII 2018 Rank	GII 2018 Interval	Input Sub-Index Rank	Input Sub-Index Interval	Output Sub-Index Rank	Output Sub-Index Interval
Brazil	64	[61, 66]	58	[49, 64]	70	[67, 72]
Iran, Islamic Republic of	65	[57, 69]	93	[80, 97]	46	[45, 47]
Tunisia	66	[65, 76]	77	[64, 82]	63	[62, 75]
Brunei Darussalam	67	[64, 83]	37	[36, 49]	112	[103, 112]
Armenia	68	[65, 73]	94	[90, 98]	50	[49, 55]
Oman	69	[66, 85]	57	[50, 67]	75	[74, 104]
Panama	70	[62, 84]	78	[69, 90]	66	[54, 88]
Peru	71	[70, 78]	59	[52, 70]	83	[83, 89]
Bahrain	72	[69, 78]	70	[64, 76]	74	[71, 78]
Philippines	73	[67, 77]	82	[70, 84]	68	[64, 70]
Kazakhstan	74	[72, 83]	55	[55, 64]	91	[89, 100]
Mauritius	75	[67, 81]	61	[59, 68]	89	[70, 92]
Morocco	76	[72, 79]	84	[75, 90]	69	[67, 70]
Bosnia and Herzegovina	77	[71, 87]	68	[59, 84]	82	[80, 87]
Kenya	78	[70, 80]	91	[77, 96]	64	[63, 65]
Jordan	79	[73, 82]	88	[75, 95]	67	[66, 76]
Argentina	80	[72, 82]	72	[62, 77]	81	[81, 86]
Jamaica	81	[77, 83]	83	[76, 92]	76	[65, 77]
Azerbaijan	82	[80, 88]	76	[70, 85]	87	[86, 91]
Albania	83	[82, 96]	69	[65, 89]	95	[94, 110]
The former Yugoslav Republic of Macedonia	84	[59, 85]	71	[61, 77]	93	[59, 94]
Indonesia	85	[78, 86]	90	[83, 91]	73	[72, 79]
Belarus	86	[62, 96]	60	[48, 67]	110	[72, 116]
Dominican Republic	87	[84, 95]	92	[90, 98]	77	[77, 97]
Sri Lanka	88	[86, 92]	95	[87, 97]	80	[78, 84]
Paraguay	89	[85, 98]	89	[85, 95]	86	[70, 104]
Lebanon	90	[79, 92]	87	[75, 89]	94	[77, 94]
Botswana	91	[87, 95]	74	[70, 79]	107	[102, 107]
Tanzania, United Republic of	92	[90, 101]	106	[100, 113]	71	[68, 97]
Namibia	93	[89, 107]	80	[72, 91]	103	[88, 117]
Kyrgyzstan	94	[91, 101]	85	[78, 91]	101	[100, 115]
Egypt	95	[87, 99]	105	[100, 108]	79	[75, 85]
Trinidad and Tobago	96	[95, 111]	86	[83, 91]	104	[103, 122]
Ecuador	97	[94, 105]	96	[92, 98]	97	[95, 116]
Cambodia	98	[95, 103]	103	[100, 113]	84	[82, 91]
Rwanda	99	[90, 117]	73	[68, 88]	120	[109, 120]
Senegal	100	[91, 101]	102	[99, 108]	90	[82, 90]
Tajikistan	101	[92, 111]	104	[99, 115]	88	[83, 102]
Guatemala	102	[100, 108]	107	[100, 111]	96	[95, 109]
Uganda	103	[102, 112]	98	[95, 102]	111	[108, 122]
El Salvador	104	[95, 112]	97	[95, 100]	113	[97, 116]
Honduras	105	[98, 107]	99	[97, 107]	106	[92, 110]
Madagascar	106	[100, 114]	119	[116, 125]	85	[80, 87]
Ghana	107	[101, 110]	108	[100, 108]	102	[101, 114]
Nepal	108	[105, 113]	101	[101, 111]	114	[99, 114]
Pakistan	109	[102, 110]	120	[111, 122]	92	[89, 93]
Algeria	110	[108, 117]	100	[95, 107]	116	[113, 123]
Cameroon	111	[104, 119]	115	[109, 121]	98	[94, 112]
Mali	112	[109, 117]	118	[114, 121]	100	[98, 112]
Zimbabwe	113	[108, 125]	121	[110, 125]	99	[96, 118]
Malawi	114	[113, 125]	111	[110, 118]	108	[106, 124]
Mozambique	115	[109, 122]	112	[107, 119]	109	[105, 120]
Bangladesh	116	[113, 122]	114	[110, 125]	105	[102, 115]
Bolivia, Plurinational State of	117	[94, 117]	109	[95, 111]	117	[95, 118]
Nigeria	118	[116, 123]	116	[109, 118]	115	[112, 125]
Guinea	119	[114, 122]	124	[120, 125]	118	[97, 120]
Zambia	120	[119, 125]	123	[109, 126]	119	[118, 125]
Benin	121	[117, 122]	110	[107, 122]	123	[116, 124]
Niger	122	[101, 122]	113	[101, 117]	122	[101, 123]
Côte d'Ivoire	123	[112, 123]	122	[113, 123]	127	[121, 128]
Burkina Faso	124	[119, 126]	117	[113, 121]	121	[105, 123]
Togo	125	[105, 125]	125	[121, 125]	125	[116, 126]
Yemen	126	[125, 126]	126	[125, 126]	124	[78, 125]

Source: European Commission, Joint Research Centre, 2018.

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

			Number of countries that *improve*		Number of countries that *deteriorate*	
Index or Sub-Index	Uncertainty tested (pillar level only)	Spearman rank correlation	by 20 or more positions	between 10 and 19 positions	by 20 or more positions	between 10 and 19 positions
GII	Geometric vs. arithmetic average	0.994	0	0	0	4
	EM imputation vs. no imputation of missing data	0.989	2[1]	4	0	3
	Geometric average and EM imputation vs. arithmetic average and missing values	0.984	4[2]	2	0	7
Input Sub-Index	Geometric vs. arithmetic average	0.996	0	0	0	1
	EM imputation vs. no imputation of missing data	0.994	0	2	0	2
	Geometric average and EM imputation vs. arithmetic average and missing values	0.992	0	4	0	2
Output Sub-Index	Geometric vs. arithmetic average	0.997	0	0	0	3
	EM imputation vs. no imputation of missing data	0.962	4[2]	12[3]	2[4]	15[5]
	Geometric average and EM imputation vs. arithmetic average and missing values	0.961	4[2]	9	2[4]	14

Source: European Commission, Joint Research Centre, 2018.

Notes:

1 The former Yugoslav Republic of Macedonia, the Plurinational State of Bolivia.

2 The former Yugoslav Republic of Macedonia, Belarus, the Plurinational State of Bolivia, Togo.

3 Panama, Mauritius, Paraguay, Lebanon, Namibia, Rwanda, El Salvador, Honduras, Nepal, Guinea, Niger, Côte d'Ivoire.

4 Oman, the United Republic of Tanzania.

5 Kuwait, Tunisia, Albania, Dominican Republic, Kyrgyzstan, Trinidad and Tobago, Ecuador, Tajikistan, Uganda, Ghana, Cameroon, Zimbabwe, Malawi, Mozambique, Bangladesh.

uncertainty propagation to their ratio has a very high impact on the country ranks. This is not a challenge specific to the GII framework per se but a statistical property that comes with ratios of composite indicators. Hence developers and users of indices alike need to take efficiency ratios of this nature with great caution. The JRC recommendation to the GII team would be to draw policy inference from the Input-Output performance in way similar to the way they plot GII scores against the economies' level of economic development and to comment on those pairs/groups of economies that have similar Innovation Input levels but very different Innovation Output levels. Economies that are at the same Output level but have very different Input levels should be treated the same way. Additional plots of the Innovation Efficiency Ratios against either the GII scores or economies' GDP per capita levels would offer additional insights in this respect.

Sensitivity analysis results

Complementary to the uncertainty analysis, sensitivity analysis has been used to identify which of the modelling assumptions have the highest impact on certain country ranks.

Table 5 summarizes the impact of changes of the EM imputation method and/or the geometric aggregation formula, with fixed weights at their reference values (as in the original GII). Similar to last year's results, this year neither the GII nor the Input or Output Sub-Index are found to be heavily influenced by the imputation of missing data, or the aggregation formula. Depending on the combination of the choices made, only six countries—The former Yugoslav Republic of Macedonia, Belarus, the Plurinational State of Bolivia, Togo, Oman, and the United Republic of Tanzania—shift rank by 20 positions or more.

All in all, the published GII 2018 ranks are reliable and for the vast majority of countries the simulated 90% confidence intervals are narrow enough for meaningful inferences to be drawn. Nevertheless, the readers of the GII 2018 report should consider country ranks in the GII 2018 and in the Input and Output Sub-Indices not only at face value but also within the 90% confidence intervals in order to better appreciate to what degree a country's rank depends on the modelling choices. Since 2016, following the JRC recommendation in past GII audits, the developers' choice to apply the 66% indicator coverage threshold separately to the Input and Output Sub-Indices in the GII 2018 has led to a net increase in the reliability of country ranks for the GII and the two sub-indices. Furthermore, the adoption in 2017 of less stringent criterion for the skewness and kurtosis (greater than 2.25 in absolute value and greater than 3.5, respectively) has not introduced any bias in the estimates.

Efficiency frontier in the GII by Data Envelopment Analysis

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

Several innovation-related policy issues at the national level entail an intricate balance between global priorities and country-specific strategies. Comparing the multi-dimensional performance on innovation by subjecting countries to a fixed and common set of weights may prevent acceptance of an innovation index on the grounds that a given weighting scheme might not be fair to a particular country. An appealing feature of the Data Envelopment Analysis (DEA) literature applied in real decision-making settings is to determine endogenous weights that maximize the overall score of each decision-making unit given a set of other observations.

In this section, the assumption of fixed pillar weights common to all countries is relaxed once more; this time country-specific weights that maximize a country's score are determined endogenously by DEA.[11] In theory, each country is free to decide on the relative contribution of each pillar to its score, so as to achieve the best possible score in a computation that reflects its innovation strategy. In practice, the DEA method assigns a higher (lower) contribution to those pillars in which a country is relatively strong (weak). Reasonable constraints on the weights are applied to preclude the possibility of a country achieving a perfect score by assigning a zero weight to weak pillars: for each country, the share of each pillar score (i.e., the pillar score multiplied by the DEA weight over the total score) has upper and lower bounds of 5% and 20% respectively. The DEA score is then measured as the weighted average of all seven pillar scores, where the weights are the country-specific DEA weights, compared to the best performance among all other countries with those same weights. The DEA score can be interpreted as a measure of the 'distance to the efficient frontier'.

Table 6 on page 84 presents the pie shares and DEA scores for the top 25 countries in the GII 2018, next to the GII 2018 ranks and efficiency ratio ranks. All pie shares are in accordance with the starting point of granting leeway to each country when assigning shares, while not violating the (relative) upper and lower bounds. The pie shares are quite diverse, reflecting the different national innovation strategies. These pie shares can also be seen to reflect countries' comparative advantage in certain GII pillars vis-à-vis all other countries and all pillars. For example, Switzerland and Singapore are the only two economies this year that obtain a perfect DEA score of 1.00. In the case of Switzerland this is achieved by assigning 18% to 19% of its DEA score to a mix of input and output pillars, namely Human capital and research, Business sophistication, Knowledge and technology outputs, and Creative outputs. Instead, merely 6% to 10% of Switzerland's DEA score comes from three input pillars, namely Institutions, Infrastructure, and Market sophistication. Using a different mix, Singapore would assign 14% to 20% of its DEA score of 1.00 to all five input pillars—Institutions, Human capital and research, Infrastructure, Market sophistication, and Business sophistication—while merely 5% to 6% of its DEA score comes from the two output pillars capturing Knowledge and technology outputs and Market sophistication. Switzerland and Singapore are closely followed by Sweden, the Netherlands, the United Kingdom, Finland, the United States of America, and Denmark, which score between 0.94 (Denmark) and 0.98 (Sweden) in terms of efficiency. Figure 5 on page 85 shows how close the DEA scores and the GII 2018 scores are for all 126 economies (Pearson correlation of 0.993). Note that, by construction, the version of DEA used herein is closer to the GII than to the Efficiency Ratio calculated as the Output Sub-Index score divided by the Input Sub-Index score (with a Pearson correlation of 0.680).

The Efficiency Ratio and the DEA score embed very different concepts of efficiency, leading to completely different results and insights. A high score in the Innovation Efficiency Ratio is obtained by scoring higher on the Output Sub-Index than on the Input Sub-Index, irrespective of the actual scores in these two sub-indices. In contrast, a high score in the DEA approach can be obtained by having comparative advantages on several GII pillars (irrespective of these being input or output pillars). The DEA scores are therefore closer to the GII scores than to the Innovation Efficiency Ratio.

Table 6: Pie shares (absolute terms) and efficiency scores for the top 25 economies in the GII 2018

	INPUT PILLARS					OUTPUT PILLARS						
Country/Economy	Institutions	Human capital and research	Infrastructure	Market sophistication	Business sophistication	Knowledge and technology outputs	Creative outputs	Efficient frontier rank (DEA)	GII rank	Difference from GII rank	Efficiency Ratio rank	Difference from GII rank
Switzerland	0.08	0.18	0.10	0.06	0.19	0.19	0.19	1	1	0	1	0
Netherlands	0.20	0.10	0.20	0.05	0.20	0.05	0.20	4	2	–2	4	–2
Sweden	0.20	0.20	0.20	0.05	0.20	0.05	0.10	3	3	0	10	–7
United Kingdom	0.20	0.20	0.20	0.20	0.05	0.05	0.10	4	4	0	21	–17
Singapore	0.18	0.20	0.14	0.18	0.19	0.06	0.05	1	5	4	63	–58
United States of America	0.20	0.05	0.20	0.20	0.20	0.05	0.10	7	6	–1	22	–16
Finland	0.20	0.20	0.20	0.06	0.20	0.05	0.09	6	7	1	24	–17
Denmark	0.20	0.20	0.20	0.20	0.05	0.05	0.10	7	8	1	29	–21
Germany	0.20	0.20	0.20	0.10	0.05	0.05	0.20	10	9	–1	9	0
Ireland	0.20	0.20	0.20	0.05	0.20	0.05	0.10	13	10	–3	13	–3
Israel	0.05	0.20	0.20	0.20	0.20	0.05	0.10	13	11	–2	14	–3
Korea, Republic of	0.20	0.20	0.20	0.20	0.05	0.05	0.10	10	12	2	20	–8
Japan	0.20	0.09	0.20	0.20	0.20	0.05	0.06	10	13	3	44	–31
Hong Kong (China)	0.20	0.05	0.20	0.20	0.20	0.05	0.10	9	14	5	54	–40
Luxembourg	0.20	0.05	0.20	0.10	0.20	0.05	0.20	20	15	–5	2	13
France	0.20	0.20	0.20	0.20	0.05	0.05	0.10	16	16	0	32	–16
China	0.05	0.05	0.20	0.20	0.20	0.10	0.20	24	17	–7	3	14
Canada	0.20	0.20	0.20	0.20	0.08	0.05	0.07	16	18	2	61	–43
Norway	0.20	0.20	0.20	0.20	0.07	0.05	0.08	18	19	1	52	–33
Australia	0.20	0.20	0.20	0.20	0.05	0.05	0.10	13	20	7	76	–56
Austria	0.20	0.20	0.20	0.08	0.20	0.05	0.07	20	21	1	53	–32
New Zealand	0.20	0.20	0.20	0.20	0.05	0.05	0.10	19	22	3	59	–37
Iceland	0.20	0.05	0.20	0.10	0.20	0.05	0.20	22	23	1	23	0
Estonia	0.20	0.05	0.20	0.20	0.10	0.05	0.20	24	24	0	12	12
Belgium	0.20	0.20	0.20	0.07	0.20	0.05	0.08	22	25	3	38	–13

Source: European Commission, Joint Research Centre, 2018.

Notes: Pie shares are in absolute terms, bounded by 0.05 and 0.20 for all seven pillars. In the GII 2018, however, the five input pillars each have a fixed weight of 0.10; the two output pillars each have a fixed weight of 0.25.

Conclusions

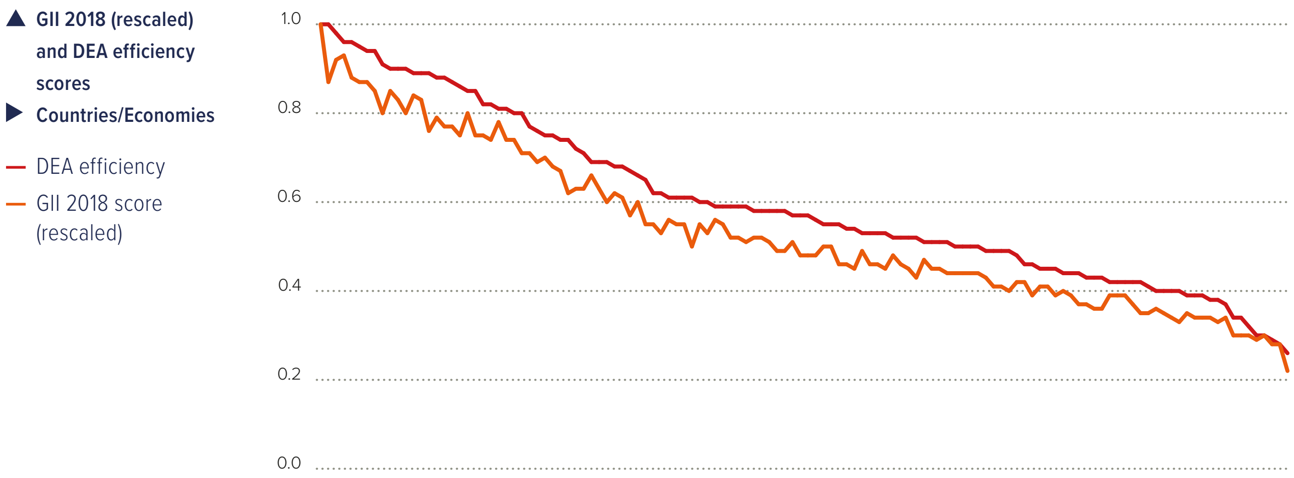
The JRC analysis suggests that the conceptualized multi-level structure of the GII 2018—with its 80 indicators, 21 sub-pillars, 7 pillars, 2 sub-indices, up to an overall index—is statistically sound and balanced: that is, each sub-pillar makes a similar contribution to the variation of its respective pillar. This year, the refinements made by the developing team have helped to enhance the already strong statistical coherence in the GII framework, where for all 80 indicators their capacity to distinguish countries' performance is maintained at the sub-pillar level or higher.

The no-imputation choice for not treating missing values, common in relevant contexts and justified on grounds of transparency and replicability, can at times have an undesirable impact on some country scores, with the additional negative side-effect that it may encourage countries not to report low data values. The adoption, since 2016, by the GII team of a more stringent data coverage threshold (at least 66% for the input- and output-related indicators, separately) has notably improved the confidence in the country ranks for the GII and the two sub-indices.

Additionally, the choice of the GII team, which was made in 2012, to use weights as scaling

Figure 5.

GII 2018 scores and DEA 'distance to the efficient frontier' scores



Source: European Commission, Joint Research Centre, 2018.

Note: For comparison purposes, we have rescaled the GII scores by dividing them with the best performer in the overall GII 2018.

coefficients during the development of the index constitutes a significant departure from the traditional, yet erroneous, vision of weights as a reflection of indicators' importance in a weighted average. It is hoped that such a consideration will be made also by other developers of composite indicators to avoid situations where bias sneaks in when least expected.

The strong correlations between the GII components are proven not to be a sign of redundancy of information in the GII. For more than 38.9% (up to 64.3%) of the 126 economies included in the GII 2018, the GII ranking and the rankings of any of the seven pillars differ by 10 positions or more. This demonstrates the added value of the GII ranking, which helps to highlight other components of innovation that do not emerge directly from looking into the seven pillars separately. At the same time, this finding points to the value of duly taking into account the GII pillars, sub-pillars, and individual indicators on their own merits. By doing so, country-specific strengths and bottlenecks in innovation can be identified and serve as an input for evidence-based policy making.

All published GII 2018 ranks lie within the simulated 90% confidence intervals that take into account the unavoidable uncertainties in the estimation of missing data, the weights (fixed vs. simulated), and the aggregation formula (arithmetic vs. geometric average) at the pillar level. For the vast majority of countries these intervals are narrow enough for meaningful inferences to be drawn: the intervals comprise fewer than 10 positions for 73% (92 out of 126) of the economies. Some caution is needed mainly for six countries—Panama, The former Yugoslav Republic of Macedonia, Belarus, Rwanda, the Plurinational State of Bolivia, and Niger—with ranks that are highly sensitive to the methodological choices. The Input and the Output Sub-Indices have the same modest degree of sensitivity to the methodological choices related to the imputation method, weights, or aggregation formula. Country ranks, either in the GII 2018 or in the two sub-indices, can be considered representative of the many possible scenarios: 75% of the countries shift fewer than three positions with respect to the median rank in the GII or either of the Input and Output Sub-Indices.

All things considered, the present JRC audit findings confirm that the GII 2018 meets international quality standards for statistical soundness, which indicates that the GII index is a reliable benchmarking tool for innovation practices at the country level around the world.

Finally, the ‘distance to the efficient frontier’ measure calculated with Data Envelopment Analysis could complement the Innovation Efficiency Ratio as a measure of efficiency, even if it is conceptually closer to the GII score than to the efficiency ratio. A word of caution on taking Innovation Efficiency Ratios alone as a yardstick for the identification of and relation between innovation input and output indicators has been added in this year’s GII audit. In fact, the same amount of uncertainty in the Input and Output Sub-Indices propagated to their sum—that is, to the GII or to their ratio—is found to result in notably different impact on country ranks: six countries shifting more than 20 positions in the case of the GII compared to 60 of the 126 economies shifting more than 20 positions in the case of the Innovation Efficiency Ratio. Not being a challenge specific to the GII framework but a statistical property that comes with ratios of composite indicators, developers and users of indices alike need to be very careful when considering efficiency ratios of this nature. The JRC recommendation to the GII team would be to gain policy insights from plots of Input against Output performance, and from plots of the Innovation Efficiency Ratios against either the GII scores or economies’ GDP per capita levels.

The GII should not be seen as the ultimate and definitive ranking of countries with respect to innovation. On the contrary, the GII best represents an ongoing attempt by the Cornell University, the business school INSEAD, and the World Intellectual Property Organization to find metrics and approaches that better capture the richness of innovation, continuously adapting the GII framework to reflect the improved availability of statistics and the theoretical advances in the field. In any case, the GII should be regarded as a sound attempt to pave the way for better and more informed innovation policies worldwide.

Notes

1 OECD/EC JRC, 2008, p. 26.

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

3 Groeneveld and Meeden (1984) set the criteria for absolute skewness above 1 and kurtosis above 3.5. The skewness criterion was relaxed in the GII case after having conducted ad-hoc tests in the GII 2008-2018 timeseries.

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

5 Nunnally, 1978.

6 See note 4.

7 Saisana et al., 2005; Saisana et al., 2011; Vértesy 2016; Vértesy and Deiss, 2016.

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: (1) The expectation E-step: Given a set of parameter estimates, such as a mean vector and covariance matrix for a multivariate normal distribution, the E-step calculates the conditional expectation of the complete-data log likelihood given the observed data and the parameter estimates. (2) The maximization M-step: Given a complete-data log likelihood, the M-step finds the parameter estimates to maximize the complete-data log likelihood from the E-step. The two steps are iterated until the iterations converge.

9 Munda, 2008.

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

11 A question that arises from the GII approach is whether there is a way to benchmark countries’ multi-dimensional performance on innovation without imposing a fixed and common set of weights that may not be fair to a particular country. The original question in the DEA literature was how to measure each unit’s relative efficiency in production compared to a sample of peers, given observations on input and output quantities and, often, no reliable information on prices (Charnes and Cooper, 1985). A notable difference between the original DEA question and the one applied here is that no differentiation between inputs and outputs is made (Cherchye et al., 2008; Melyn and Moesen, 1991). To estimate DEA-based distance to the efficient frontier scores, we consider the $m = 7$ pillars in the GII 2018 for $n = 126$ countries, with y_{ij} the value of pillar j in country i. The objective is to combine the pillar scores per country into a single number, calculated as the weighted average of the m pillars, where w_i represents the weight of the i-th pillar. In absence of reliable information about the true weights, the weights that maximize the DEA-based scores are endogenously determined. This gives the following linear programming problem for each country j:

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

subject to

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

where

$j = 1, \ldots, 7,$
$i = 1, \ldots, 126$

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

References and related reading

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

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

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

Cherchye, L., W. Moesen, N. Rogge, T. Van Puyenbroeck, M. Saisana, A. Saltelli, R. Liska, and S. Tarantola. 2008. 'Creating Composite Indicators with DEA and Robustness Analysis: The Case of the Technology Achievement Index'. *Journal of Operational Research Society* 59: 239–51.

Groeneveld, R. A. and G. Meeden. 1984. 'Measuring Skewness and Kurtosis'. *The Statistician* 33: 391–99.

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

Melyn, W. and W. Moesen. 1991. 'Towards a Synthetic Indicator of Macroeconomic Performance: Unequal Weighting when Limited Information Is Available'. *Public Economics Research Paper* 17. Leuven: Centre for Economic Studies.

Munda, G. 2008. *Social Multi-Criteria Evaluation for a Sustainable Economy.* Berlin Heidelberg: Springer-Verlag.

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

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

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

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

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

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

Schneider, T. 2001. 'Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. *Journal of Climate* 14: 853–71.

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

Vértesy, 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,19–21 September 2016, Ghent, Belgium.