Appendix IV

Technical Notes

Technical Notes

Audit by the Joint Research Centre of the European Commission

The Joint Research Centre (JRC) of the European Commission has researched extensively on the complexity of composite indicators ranking economies' performances along policy lines. For the seventh consecutive year, the JRC has agreed to perform a thorough robustness and sensitivity analysis of the Global Innovation Index (GII) to look at some structural changes made to the list of indicators by the GII developing team (see Table 1 of Annex 2 to Chapter 1 for more details).

The recommendations from the JRC audit of the 2016 GII model were reviewed and incorporated into the 2017 GII model. Expanding on recommendations included in the GII 2016, this year an economy must have a minimum symmetric data coverage of at least 36 indicators in the Innovation Input Sub-Index (66%) and 18 indicators in the Innovation Output Sub-Index (66%), and it must have scores for at least two sub-pillars per pillar. The GII rules on data requirements will be continually strengthened in future years, incentivizing countries to further improve their data collection.

A final audit was performed in May 2017 on the 2017 GII model, the results of which are included in Annex 3 to Chapter 1.

Composite indicators

The GII relies on seven pillars. Each pillar is divided into three sub-pillars, and each sub-pillar is composed of two to five individual indicators. Each sub-pillar score is calculated as the weighted average of its individual indicators. Each pillar score is calculated as the weighted average of its sub-pillar scores.

The notion of weights as importance coefficients was, as in the previous three years, discarded to ensure a greater statistical coherence of the model, following the recommendations of the JRC.[1]

The GII includes three indices and one ratio:

1. The Innovation Input Sub-Index is the simple average of the first five pillar scores.
2. The Innovation Output Sub-Index is the simple average of the last two pillar scores.
3. The Global Innovation Index is the simple average of the Input and Output Sub-Indices.
4. The Innovation Efficiency Ratio is the ratio of the Output Sub-Index over the Input Sub-Index.

Country/economy rankings are provided for indicator, sub-pillar, pillar, and index scores.

The Innovation Efficiency Ratio serves to highlight those economies that have achieved more with less as well as those that lag behind in terms of achieving their innovation potential. In theory, assuming that innovation results go hand in hand with innovation enablers, efficiency ratios should evolve around the number one. This measure thus allows us to complement the GII by providing an insight that should be neutral to the development stages of economies.[2]

Individual indicators

The model includes 81 indicators, which fall into the following three categories:

1. quantitative/objective/hard data (57 indicators),
2. composite indicators/index data (19 indicators), and
3. survey/qualitative/subjective/soft data (5 indicators).

Hard data

Hard data series (57 indicators) are drawn from a variety of public and private sources such as United Nations agencies, including the United Nations Educational, Scientific and Cultural Organization (UNESCO), the United Nations Industrial Development Organization (UNIDO), the World Intellectual Property Organization (WIPO), the World Bank, the Joint Research Centre of the European Commission (JRC), PwC, Bureau van Dijk (BvD), Thomson Reuters, IHS Global Insight, and Google.

Indicators are often correlated with population, gross domestic product (GDP), or some other size-related factor; they require scaling by some relevant size indicator for economy comparisons to be valid. Most indicators are either scaled at the source or do not need to be scaled; for the rest, the scaling factor was chosen to represent a fair picture of economy differences. This affected 40 indicators, which can be broadly divided into four groups:

1. Indicators 2.1.1, 2.3.2, 3.2.3, 4.1.2, 4.1.3, 4.2.2, 5.1.3, 5.3.4, 6.2.3, and 6.3.4 were scaled by GDP in current US dollars.[3]
2. The count variables 3.3.3, 4.2.3, 5.2.4, 5.2.5, 6.1.1, 6.1.2, 6.1.3, 6.1.4, 6.2.4, 7.1.1, and 7.1.2 were scaled by GDP in purchasing power parity current international dollars (PPP$ GDP). This choice of denominator was dictated by a willingness to appropriately account for differences in development stages; in addition, scaling these variables by population would improperly bias results to the detriment of economies with large young or large ageing populations.[4]
3. Variables 3.2.1, 5.1.5, 6.2.2, 7.2.2, 7.2.3, 7.3.1, 7.3.2, 7.3.3, and 7.3.4 were scaled by population (total population for 3.2.1, population 25+ years old for 5.1.5, population 15–64 years old for 6.2.2, and population 15–69 years old for the rest).[5]
4. Sectoral indicators 5.3.1, 5.3.2, 5.3.3, 6.3.1, 6.3.2, 6.3.3, 7.2.1, and 7.2.5 were scaled by total trade; indicators 6.2.5 and 7.2.4 were scaled by the total unit corresponding to the particular statistic.[6]

Indices

Composite indicators come from a series of specialized agencies and academic institutions such as the World Bank, the International Telecommunication Union (ITU), the UN Public Administration Network (UNPAN), and Yale and Columbia Universities. Statisticians discourage the use of an 'index within an index' on two main grounds: the distorting effect of the use of different computing methodologies and the risk of duplicating variables. The normalization procedure partially solves for the former (more on this below). To avoid incurring the mistake of including a particular indicator more than once (directly and indirectly through a composite indicator), only indices with a narrow focus (19 in total) were selected.

Any remaining downside is outweighed by the gains in terms of model parsimony, acknowledgement of expert opinion, and focus on multi-dimensional phenomena that can hardly be captured by a single indicator.[7]

Survey data

Survey data are drawn from the World Economic Forum's Executive Opinion Survey (EOS). Survey questions are drafted to capture subjective perceptions on specific topics; five EOS questions were retained to capture phenomena strongly linked to innovative activities for which hard data either do not exist or have low economy coverage.

Country/economy coverage and missing data

This year's GII covers 127 economies, which were selected on the basis of the availability of data. Economies with a minimum indicator coverage of 36 indicators in the Innovation Input Sub-Index (66%) and 18 indicators in the Innovation Output Sub-Index (66%) were retained. This minimum data coverage threshold rule was adjusted—on the recommendation of the JRC—from the 60% minimum coverage for both Sub-Indices introduced in the GII 2016—to maintain and improve the significance of both the GII results and the country sample. In addition, all selected countries are required to have scores for at least two sub-pillars per pillar.

The last record available for each economy was considered, with a cut-off at year 2007, with four exceptions: indicators 2.2.2, 5.1.2, 6.2.5, and 7.2.4, for which time periods were extended to 2006.[8]

For the sake of transparency and replicability of results, no additional effort was made to fill missing values. Missing values are indicated with 'n/a' and are not considered in the sub-pillar score. However, the JRC audit assessed the robustness of the GII modelling choices (i.e., no imputation of missing data, fixed predefined weights, and arithmetic averages) by imputing missing data, applying random weights, and using geometric averages. Since 2012, on the basis of this assessment, a confidence interval is provided for each ranking in the GII as well as the Input and Output Sub-Indices (see Annex 3 to Chapter 1).

Treatment of series with outliers

Potentially problematic indicators with outliers that could polarize results and unduly bias the rankings were treated according to the rules listed below, following the recommendations of the JRC. This affected a total of 33 indicators; 31 out of the 57 hard data indicators and 2 out of the 19 composite indicators.

First rule: Selection

The identification of indicators as problematic used skewness or kurtosis. The problematic indicators had either:

- an absolute value of skewness greater than 2.25, *or*
- a kurtosis greater than 3.5.[9]

Second rule: Treatment

Series with one to five outliers (28 cases) were winsorized: The values distorting the indicator distribution were assigned the next highest value, up to the level where skewness and/or kurtosis entered within the ranges specified above.[10]

With one exception (see note 10) for series with five or more outliers (5 cases), skewness and/or kurtosis entered within the ranges specified above after multiplication by a given factor *f* and transformation by natural logs.[11] Since only 'goods' were affected (i.e., indicators for which higher values indicate better outcomes, as opposed to 'bads'), the formula used was:

$$\ln\left[\frac{(\max \times f - 1)\ (\text{economy value} - \min)}{\max - \min} + 1\right]^{12}$$

where 'min' and 'max' are the minimum and maximum indicator sample values.

Normalization

The 81 indicators were then normalized into the [0, 100] range, with higher scores representing better outcomes. Normalization was made according to the min-max method, where the min and max values were given by the minimum and maximum indicator sample values respectively, except for index and survey data, for which the original series' range of values was kept as min and max values (for example, [1, 7] for the World Economic Forum Executive Opinion Survey questions; [0, 100] for World Bank's World Governance Indicators; [0, 10] for ITU indices, etc.). The following formula was applied:

- **Goods:**

$$\frac{\text{economy value} - \min}{\max - \min} \times 100$$

- **Bads:**

$$\frac{\max - \text{economy value}}{\max - \min} \times 100$$

Notes

1 Paruolo et al. (2013) show that a theoretical inconsistency exists between the real theoretical meaning of weights and the meaning generally attributed to them by the standard practice in constructing composite indicators that use them as importance coefficients in combination with linear aggregation rules. The approach followed in the GII this year, in the last several years, is to assign weights of 0.5 or 1.0 to each component in a composite to ensure the highest correlations between them (i.e., indicator/sub-pillar, sub-pillar/pillar, etc.). Two sub-pillars (7.2 Creative goods and services, and 7.3 Online creativity) and 35 indicators (1.2.1, 1.2.2, 2.1.4, 2.1.5, 2.2.1, 2.2.3, 3.2.1, 3.2.2, 3.3.3, 4.2.2, 4.2.3, 4.3.1, 4.3.2, 5.1.3, 5.1.4, 5.1.5, 5.2.1, 5.2.4, 5.2.5, 5.3.1, 6.1.1, 6.1.2, 6.1.4, 6.1.5, 6.2.2, 6.2.3, 6.2.4, 6.2.5, 6.3.1, 6.3.2, 6.3.3, 7.1.2, 7.2.1, 7.2.2, and 7.2.3) are weighted 0.5; the rest have a weight of 1.0.

Seven indicators with Pearson correlation coefficients with their respective sub-pillar scores below 0.3 were kept in the model to ensure a conceptual coherence (as opposed to a statistical coherence) in the belief that some cyclical (as opposed to structural) dimension might be at the source of their behaviour as 'noise' (see also Annex 3 to Chapter 1): government expenditure on secondary education per pupil (2.1.2), graduates in science and engineering (2.2.2), gross capital formation (3.2.3), GERD financed by abroad (5.2.3), foreign direct investment net inflows (5.3.4), growth rate of GDP per person engaged (6.2.1), and printing and publishing output (7.2.4). Two out of the seven indicators listed above—2.1.2 and 7.2.4—are found to be non-influential in the GII framework for the first time this year. The remaining five indicators were found to be non-influential also in the GII 2016. On the other hand, two indicators that were found to be non-influential last year—GDP per unit of energy use (3.3.1) and Microfinance institutions' gross loan portfolio (4.1.3)—are instead found to be influential in this year's framework.

2 To account for differences in development, other composite indicators use weighting schemes differentiated by income level.

3 These indicators are expenditure on education (2.1.1); gross expenditure on R&D (GERD) (2.3.2); gross capital formation (3.2.3); domestic credit to private sector (4.1.2); microfinance institutions' gross loan portfolio (4.1.3); market capitalization (4.2.2); GERD performed by business enterprise (5.1.3); foreign direct investment net inflows (5.3.4); total computer software spending (6.2.3); and foreign direct investment net outflows (6.3.4).

4 These count variables are mainly indicators that increase disproportionately with economic growth. They include: ISO 14001 environmental certificates (3.3.3); venture capital deals (4.2.3); joint venture/strategic alliance deals (5.2.4); patent families filed in two or more offices (5.2.5); patent applications by origin (6.1.1); PCT applications by origin (6.1.2); utility model applications by origin (6.1.3); scientific and technical publications (6.1.4); ISO 9001 quality certificates (6.2.4); trademark application class count by origin (7.1.1); and industrial designs by origin (7.1.2).

5 These variables are electricity output (3.2.1); females employed with advanced degrees (5.1.5); new business density (6.2.2); national feature films produced (7.2.2); global entertainment and media market (7.2.3); generic (7.3.1) and country-code (7.3.2) top-level Internet domains; Wikipedia yearly edits (7.3.3); and video uploads on YouTube (7.3.4).

6 Intellectual property payments (5.3.1); high-tech imports less re-imports (5.3.2); ICT services imports (5.3.3); intellectual property receipts (6.3.1); high-tech exports less re-exports (6.3.2); ICT services exports (6.3.3); cultural and creative services exports (7.2.1); and creative goods exports (7.2.5) were scaled by total trade; high-tech and medium-high-tech output (6.2.5) and printing and publishing output (7.2.4) were scaled by total manufactures output.

7 For example, GII sub-pillar 3.1 Information and communication technologies (ICTs) is composed of four indices: ITU's ICT Access and Use sub-indices and UNPAN's Government Online Service and E-Participation indices. The first two are components of ITU's ICT Development Index together with an ICT skills sub-index that was not considered, because it duplicates GII pillar 2. Similarly, the Online Service Index is a component of UNPAN's E-Government Development Index together with two indices on Telecommunication Infrastructure and Human Capital that were not considered, because they duplicate GII pillars 3 and 2, respectively. The e-Participation Index was developed separately by UNPAN in 2010.

8 Indicator 2.2.2 (graduates in science and engineering): Because of the change in ISCED fields of classification and the transition to new questionnaires, when countries did not report detailed data, the UIS was not able to re-assign numbers into new field classifications. As a result, the UIS was not able to produce this indicator for these countries, so—per the recommendation of the UIS—the dataset from the GII 2016 was used. There was one economy affected: Hong Kong (China). Indicator 5.1.2 (firms offering formal training): The time period of data for Guinea includes 2006 for heightened coverage based on this economy's GII 2016 data availability. Indicator 6.2.5 (high-tech and medium-high-tech output) and indicator 7.2.4 (printing and publishing output): The time periods of data for Iceland, Madagascar, and Pakistan for these indicators include 2006 for heightened coverage based on these economies' GII 2016 data availability.

9 Based on Groeneveld and Meeden (1984), which sets the criteria of absolute skewness above 1 and kurtosis above 3.5. The skewness criterion was relaxed to account for the small sample at hand (127 economies).

10 This distributional issue affects the following variables: 1.2.3, 3.2.1, 3.3.3, 4.2.2, 5.3.2, 5.3.3, 6.1.5, 6.2.1, 6.2.2, 6.3.2, 7.1.1, 7.2.4, and 7.3.1 (1 outlier); 5.3.1, 7.2.1, 7.2.2, and 7.3.2 (2 outliers); 2.2.3, 4.1.3, 4.2.3, 5.2.5, 6.1.1, 6.1.2, 6.1.3, 6.2.5, and 6.3.3 (3 outliers); and 7.1.2 (4 outliers). The treatment criterion was relaxed this year to allow a single series (6.3.4) with 7 outliers—6 outliers given the next highest value and 1 given the next lowest value—to be winsorized instead of subjected to natural log transformation. This is because applying a log transformation at 1, 10, and 100 had the reverse effect, and instead of reducing skewness and kurtosis, it increased them.

11 This distributional issue affects variables 2.3.3, 4.3.3, 5.3.4, 6.3.1, and 7.2.5 (factor *f* of 1).

12 The corresponding formula for bads is:

$$\ln\left[\frac{(\max \times f - 1) \times (\max - \text{economy value})}{\max - \min} + 1\right]$$

These formulas achieve two things: converting all series into 'goods' and scaling the series to the range [1, max] so that natural logs are positive starting at 0.

References

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

Paruolo P., M. Saisana, and A. Saltelli. 2013. 'Ratings and Rankings: Voodoo or Science?' *Journal of the Royal Statistical Society* A 176(2), doi: 0964–1998/13/176000.