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. First in 2011, and again this year, the JRC agreed to perform a thorough robustness and sensitivity analysis of the Global Innovation Index (GII).

A previous version of the GII model was submitted to the JRC in March 2012. The recommendations and flexibilities allowed by the JRC preliminary audit were taken into account in the final version of the GII model and are explained below as appropriate.

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

Composite indicators

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

This year the notion of weights as 'importance coefficients' was discarded to ensure a greater statistical coherence of the model, following the recommendations of the JRC.[1]

The GII includes four index measures:

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 Index is the ratio of the Output Sub-Index over the Input Sub-Index.

Economy rankings are provided for indicator, sub-pillar, pillar, and index scores.

The Innovation Efficiency Index serves to highlight those economies that have 'achieved more with less' and those that lag behind in terms of fulfilling 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 84 indicators, which fall within the following three categories:

1. quantitative/objective/hard data (62 indicators),
2. composite indicators/index data (16 indicators), and
3. survey/qualitative/subjective/soft data (6 indicators).

Hard data

Hard data series (62 indicators) are drawn from a variety of public and private sources such as United Nations agencies (the United Nations Educational, Scientific and Cultural Organization, the World Intellectual Property Organization), the World Bank, Thomson Reuters, and Standard & Poor's.

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 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 27 indicators, which can be broadly divided into five groups:

1. Indicators 4.1.3, 5.3.1, 6.2.3, and 6.3.1, which come in current US

dollars, were scaled by GDP in current US dollars.[3]

2. The count variables 3.3.3, 4.2.4, 5.2.4, 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 PPP terms, in current international dollars. 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 5.1.6, 7.2.2, 7.2.3, 7.3.1, 7.3.2, 7.3.3, and 7.3.4 were scaled by population (20–34 years old for 5.1.6, and 15–69 years old for the rest).

4. Variable 3.2.1, Electricity output in kWh per capita, was scaled by population to be consistent with 3.2.2, Electricity consumption in kWh per capita, which is scaled at the source by the International Energy Agency.

5. Sectoral indicators 5.3.2, 6.3.2, 7.2.1, 7.2.4, and 7.2.5 were scaled by the total corresponding to the particular statistic.[5]

Indices

Composite indicators come from a series of specialized agencies, such as the World Bank, the International Telecommunication Union (ITU), and the UN Public Administration Network (UNPAN). 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 were selected (15 in total).

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.

To give an example, GII sub-pillar 3.1 Information and communication technologies (ICT) 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, as 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, as they duplicate GII pillars 3 and 2, respectively. The e-Participation Index was developed separately by UNPAN in 2010.

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. Nonetheless, the six EOS questions included in 2011 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 141 economies, which were selected on the basis of the availability of data. Economies with a minimum indicator coverage of 54 indicators (63%) and with scores for at least two sub-pillars per pillar were retained. These criteria were determined jointly with the JRC in 2011. The last record available for each economy was considered, with a cut-off at year 2001. 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's 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. This year, 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 2 to Chapter 1).

Treatment of series with outliers

Potentially problematic indicators with outliers that could polarize results and unduly bias the rankings were treated with the rules listed below, following the recommendations of the JRC. This affected 35 hard data indicators.

First rule: Selection

The 35 problematic indicators were identified by a combination of skewness and kurtosis statistics:

- absolute value of skewness greater than 2, *and*
- kurtosis greater than 3.5.[6]

Second rule: Treatment

Series with one to four outliers (28 cases) were winsorised: 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.[7]

For series with five or more outliers (7 cases), skewness and/or kurtosis entered within the ranges specified above with transformation by natural logs.[8] Since only 'goods' were affected (i.e., indicators for which higher values indicate better outcomes, as opposed to 'bads'), the formula used was:

$$\ln\left[\frac{(\max - 1) \times (\text{country value} - \min)}{(\max - \min)} + 1\right]^{9}$$

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

Normalization

The 84 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.). In addition, for indices based on percent ranks, the percent ranks were recalculated for the sample of 141 economies.[10] The following formula was applied:

- Goods:

$$100 \times \frac{(\text{country value} - \min)}{(\max - \min)}$$

- Bads:

$$-100 \times \frac{(\text{country value} - \min)}{(\max - \min)} + 100$$

Notes

1 Paruolo et al. (2012) 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 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.). Only two sub-pillars are weighted 0.5: 7.2 Creative goods and services, and 7.3 Online creativity; while 22 indicators are weighted 0.5. Five indicators with Pearson correlation coefficients with their respective sub-pillar scores below 0.5 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": 3.2.4 Gross capital formation; 4.3.2 Market access for non-agricultural exports; 4.3.5 Intensity of local competition; 5.3.4 Foreign Direct Investment (FDI) net inflows; and 6.3.4 FDI net outflows. These criteria might need to be revised next year.

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

3 These indicators measure the gross loan portfolio of microfinance institutions; royalty and license fees' payments and receipts, and total computer software spending, respectively.

4 These count variables are mainly indicators that increase disproportionately with economic growth, and include: ISO 14001 environmental and ISO 9001 quality certificates issued; venture capital, joint venture, and strategic alliance deals; and resident patent, utility model, and trademark applications.

5 Creative exports of goods (services) are scaled by total exports of goods (services); high-tech exports minus re-exports (imports minus re-imports) by total exports minus re-exports (imports minus re-imports); and individual expenditure on recreation and culture by total individual consumption.

6 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 (141 economies).

7 This distributional issue affects the following variables: 1.2.3, 3.2.1, 4.2.2, 5.3.2, 5.3.4, 6.3.2, 6.3.4, 7.1.1, 7.1.2, 7.2.5 (1 outlier); 3.2.2, 4.3.3, 7.2.4 (2 outliers); 2.2.4, 3.3.3, 4.1.3, 4.2.3, 4.3.2, 4.3.4, 5.2.3, 5.3.1, 6.1.1, 6.1.3, 6.2.2, 7.2.2 (3 outliers); and 5.2.4, 6.3.1, 7.3.1 (4 outliers).

8 This distributional issue affects variables 2.2.3, 4.2.4, 5.1.6, 6.1.2, 6.2.4, 7.3.2 and 7.3.4.

9 The corresponding formula for 'bads' is:

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

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.

10 This concerns indicators 1.3.1, 1.3.2, 1.3.3, 4.1.1, and 4.2.1.

References

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

Paruolo, P., M. Saisana, and A. Saltelli. 2012. 'Ratings and Rankings: Voodoo or Science?' *Journal of the Royal Statistical Society* (in print). A draft version is available at http://arxiv.org/abs/1104.3009.