COIN Tool (beta version)

A quality assurance Excel-based tool for developers and users of composite indicators and scoreboards

Michaela Saisana
William Becker
Marcos Domíneuez-Torreiro
Daniel Vertesy

2017
This publication is a Technical report by the Joint Research Centre (JRC), the European Commission’s science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication.

Contact information
Michaela Saisana
Enrico Fermi 2749, 21027, Ispra (VA), Italy
michaela.saisana@jrc.ec.europa.eu

JRC Science Hub
https://ec.europa.eu/jrc

JRC105377

© European Union, 2017

The reuse of the document is authorised, provided the source is acknowledged and the original meaning or message of the texts are not distorted. The European Commission shall not be held liable for any consequences stemming from the reuse.

How to cite: Saisana M; Becker W; Dominguez-Torreiro M; Vertesy D. COIN Tool – beta version. A quality assurance Excel-based tool for developers and users of composite indicators and scoreboards. European Commission; 2017. JRC105377
# Table of contents

Acknowledgements ........................................................................................................ 2
Abstract ......................................................................................................................... 3
1 Introduction .................................................................................................................. 4
   1.1 Ten Step guide for constructing a composite indicator or gaining insights into the properties of a scoreboard ........................................................................................................ 4
   1.2 COIN Tool – How it is organized ........................................................................... 7
2 Computation of the composite indicator .................................................................... 8
   2.1 Database and conceptual framework (yellow tabs) .................................................. 8
      2.1.1 Organisation of the data .................................................................................. 8
   2.1.2 Conceptual Framework ....................................................................................... 9
   2.2 Treatment of outliers (green tabs) ......................................................................... 10
      2.2.1 Original dataset – detection of outliers ............................................................ 10
      2.2.2 Winsorization ................................................................................................. 11
      2.2.3 Box-cox transformations ................................................................................ 12
      2.2.4 Scatterplots ...................................................................................................... 13
   2.3 Weight adjustments and final ranking (purple tabs) .................................................. 14
      2.3.1 Outlier free dataset – descriptive statistics ....................................................... 14
      2.3.2 Indicator correlations and prospective weights .................................................. 14
      2.3.3 Adjustment of weights on the basis of upper-level correlations ...................... 16
      2.3.4 Ranking with adjusted weights ....................................................................... 17
      2.3.5 Heatmap of scores with adjusted weights ......................................................... 18
3 Scenaria (blue tabs) ..................................................................................................... 19
   3.1 Normalization ......................................................................................................... 19
      3.1.1 Min-max normalization .................................................................................... 19
      3.1.2 Z-score normalization ..................................................................................... 19
   3.2 Aggregation methods and rankings ........................................................................ 19
      3.2.1 Arithmetic and geometric averages ................................................................... 20
         3.2.1.1 Note on arithmetic v. geometric averages .................................................... 23
         3.2.1.2 Note on random weights .......................................................................... 23
         3.2.1.3 Note on trimmed means ......................................................................... 23
      3.2.2 Median and average rank ................................................................................ 24
      3.2.3 Borda rule ....................................................................................................... 25
      3.2.4 Copeland rule .................................................................................................. 26
         3.2.4.1 Outranking matrix .................................................................................... 26
         3.2.4.2 Copeland rule ......................................................................................... 27
   3.3 Scenaria .................................................................................................................. 28
4 Advanced features ....................................................................................................... 29
   4.1 Imputation of missing data .................................................................................... 29
   4.2 Shadow imputation ............................................................................................... 29
References ....................................................................................................................... 31
List of tables and figures ............................................................................................... 34
Acknowledgements

The authors are grateful to Daniela Benavente (consultant) who contributed extensively to enhance the quality and the functionalities in Excel of the COIN Tool.

Many participants at the JRC Annual Trainings on Composite Indicators and Scoreboards contributed to improve this tool further with their valuable comments and suggestions.
Abstract

The COIN Tool provides a practical Excel-based guide to the development of composite indicators and scoreboards, for policy-makers and researchers alike.

The COIN Tool aims to contribute to a better understanding of key methodological issues underpinning the development of composite indicators and to an improvement in the techniques currently used to build them. In particular, it contains a set of technical guidelines that can help constructors of composite indicators and scoreboards to improve the quality of their outputs. The COIN Tool is also helpful to users of composite indicators that wish to get a better understanding of the statistical properties of composite indicators and scoreboards.

The COIN Tool has been prepared by the Competence Centre on Composite Indicators and Scoreboards (COIN) at the European Commission’s Joint Research Centre. The COIN Tool implements many of the suggestions and recommendations provided in the 2008 OECD/JRC ‘Handbook on Constructing Composite Indicators: Methodology and User Guide’.

Further information on the topics treated in the COIN Tool and on other issues related to composite indicators and scoreboards can be found in the web page:


The COIN Tool starts from the premise that the developers of a composite indicator or scoreboard have already conducted a thorough literature review on the topic of interest, namely: definition(s) of the phenomenon, relevant studies, conceptual framework, methodological concerns.

The features included in the COIN Tool are the following:

- calculating descriptive statistics of the data,
- spotting and treating potentially problematic indicators that present highly skewed distributions,
- analysing the data correlation structure,
- estimating missing data,
- normalizing indicators (z-scores, min-max, ranks),
- aggregating indicators using (weighted) arithmetic averages, geometric averages, trimmed mean, median rank, summation of ranks, Borda rule, Copeland rule;
- conducting a simplified uncertainty analysis.

The COIN Tool in its current beta version is being tested by European Commission officials. The COIN Tool will be formally released in the fall of 2017.
1 Introduction

The use of composite indicators and scoreboards for designing and monitoring policies gained much interest in recent decades. Over 120 documents in the EU law online platform – EUR-Lex – include a reference to a composite indicator and over 1500 documents refer to a scoreboard of indicators. The first composite indicator from the Commission dates back to 1987. Today, the Commission services have developed more than 100 composite indicators and even more scoreboards. Examples are the Europe 2020 Index, the Regional Human Development Index and the Regional Poverty Index of DG REGIO, the European Innovation Union Scoreboard and the Small Business Act Principles of DG GROW, the Research Excellence Index and the Innovation Output Indicator of DG RTD, the Consumer Conditions Index and the Market Performance Index of DG JUST, the Digital Economy and Society Index of DG CNECT, the Banks’ contribution to EU Single Resolution Fund of DG FISMA, the Index for Risk Management of DG ECHO and the Cultural and Creative City Monitor of DG JRC.

In a nutshell, composite indicators are built by simplifying a policy concept into a summary figure by means of a conceptual framework and statistical analysis. Composite indicators are aggregations of observable indicators that aim to quantify underlying concepts that are not directly observable, such as competitiveness, freedom of the press or climate hazards. The resulting figures facilitate cross-country, -region, or -city comparisons and benchmarking. They help monitoring progress over time and evaluating ex-ante policy options based on multi-criteria analysis. Scoreboards of indicators have, to some extent, similar objectives to composite indicators, yet they do not consist of a mathematical aggregation.

Composite indicators are powerful practical tools that can help policy makers summarize complex and interdependent phenomena. They provide the big picture, are easy to interpret, easy to communicate, and attractive for the public. They are also drivers of behaviour and of change by forcing institutions and governments to question their standards. On the other hand, caution is needed to avoid situations where composite indicators may send misleading or partial policy messages because they are poorly constructed or misinterpreted.

1.1 Ten Step guide for constructing a composite indicator or gaining insights into the properties of a scoreboard

The table below presents a ‘decalogue’ for the construction of a composite indicator, or for assessing, inter alia, the statistical associations of the indicators in a scoreboard. The table which has been rearranged and extended from the information contained in the 2008 OECD/JRC Handbook. These steps have been put in practice in the JRC audits, conducted upon request of developers of multidimensional measures such as the INSEAD-WIPO-Cornell Global Innovation Index, UN Multidimensional Poverty Assessment Tool, the Composite Learning Index, the Environmental Performance Index, the Corruptions Perceptions Index, and the EU Competitiveness Index Index just to name a few.

This short ten-step guide stresses the importance of conducting an internal coherence assessment prior to the uncertainty and sensitivity analysis, so as to further refine and eventually correct the composite indicator structure. Expert opinion is needed in this phase in order to assess the results of the statistical analysis. Second, it stresses that there is a trade-off between multidimensionality and robustness in a composite indicator. One could have a very robust yet mono-dimensional index or a very volatile yet multi-dimensional one. This does not imply that the first index is better than the second one. In fact, this table suggests treating robustness analysis NOT as an attribute of a composite indicator but of the inference which the composite indicator has been called.
upon to support. Third, it highlights the iterative nature of the ten steps, which although presented consecutively in the OECD/JRC Handbook, the benefit to the developer is in the iterative nature of the steps.

Table 1. Ten Step Guide for Developing Composite Indicators and Scoreboards

<table>
<thead>
<tr>
<th>Step 1. Theoretical/Conceptual framework</th>
</tr>
</thead>
<tbody>
<tr>
<td>provides the basis for the selection and combination of variables into a meaningful composite indicator under a fitness-for-purpose principle (involvement of experts and stakeholders is important).</td>
</tr>
<tr>
<td>✓ Clear understanding and definition of the multidimensional phenomenon to be measured.</td>
</tr>
<tr>
<td>✓ Discuss the added-value of the composite indicator.</td>
</tr>
<tr>
<td>✓ Nested structure of the various sub-groups of the phenomenon (if relevant).</td>
</tr>
<tr>
<td>List of selection criteria for the underlying variables, e.g., input, output, process.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step 2. Data selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>should be based on the analytical soundness, measurability, country coverage, and relevance of the indicators to the phenomenon being measured and relationship to each other. The use of proxy variables should be considered when data are scarce (involvement of experts and stakeholders is important).</td>
</tr>
<tr>
<td>✓ Quality assessment of the available indicators.</td>
</tr>
<tr>
<td>✓ Discuss strengths and weaknesses of each selected indicator.</td>
</tr>
<tr>
<td>✓ Summary table on data characteristics, e.g., availability (across country, time), source, type (hard, soft or input, output, process), descriptive statistics (mean, median, skewness, kurtosis, min, max, variance, histogram).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step 3. Data treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>consists of imputing missing data, (eventually) treating outliers and/or making scale adjustments.</td>
</tr>
<tr>
<td>✓ Confidence interval for each imputed value that allows assessing the impact of imputation on the composite indicator results.</td>
</tr>
<tr>
<td>✓ Discuss and treat outliers, so as to avoid that they become unintended benchmarks (e.g., by applying Box-Cox transformations such square roots, logarithms, and other).</td>
</tr>
<tr>
<td>✓ Make scale adjustments, if necessary (e.g., taking logarithms of some indicators, so that differences at the lower levels matter more).</td>
</tr>
</tbody>
</table>

(back to step 2)

<table>
<thead>
<tr>
<th>Step 4. Multivariate analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>should be used to study the overall structure of the dataset, assess its suitability, and guide subsequent methodological choices (e.g., weighting, aggregation).</td>
</tr>
<tr>
<td>✓ Assess the statistical and conceptual coherence in the structure of the dataset (e.g., by principal component analysis and correlation analysis).</td>
</tr>
<tr>
<td>✓ Identify peer groups of countries based on the individual indicators and other auxiliary variables (e.g., by cluster analysis).</td>
</tr>
</tbody>
</table>

(back to Step 1)
Step 5. Normalisation

should be carried out to render the variables comparable.

- Make **directional adjustment**, so that higher values correspond to better performance in all indicators (or vice versa).
- Select a **suitable normalisation method** (e.g., min-max, z-scores, and distance to best performer) that respects the conceptual framework and the data properties.

Step 6. Weighting and aggregation

should be done along the lines of the theoretical/conceptual framework

- Discuss whether **compensability** among indicators should be allowed and up to which level of aggregation.
- Discuss whether **correlation** among indicators should be taken into account during the assignment of weights.
- Select a **suitable weighting and aggregation method** that respect the conceptual framework and the data properties. Popular weighting methods include equal weights, factor analysis derived weights, expert opinion, and data envelopment analysis. Popular aggregation methods include arithmetic average, geometric average, Borda, Copeland.

Internal coherence assessment (intermediate step). This step is briefly listed under step 9 in the Handbook but not thoroughly discussed. This assessment needs to be undertaken prior to the uncertainty and sensitivity analysis, so as to further refine the composite indicator structure (upon consultation with experts on the issue).

- Assess whether **dominance problems are present**, namely the composite indicator results are overly dominated by a small number of indicators and quantify the relative importance of the underlying components (e.g., by global sensitivity analysis, correlation ratios).
- Assess eventual "**noise**" added to the final composite indicator results by non-influential indicators.
- Assess the direction of impact of indicators and sub-dimensions, namely whether all components point to the same direction as the composite indicator (sign of correlation) and explain trade-offs.
- Assess whether certain indicators are statistically grouped under different dimensions than conceptualised and whether certain dimensions should be merged or split.
- Assess **eventual bias** introduced in the index (e.g., due to population size, population density)

(back to Step 1 and Step 2)

Step 7. Uncertainty and sensitivity analysis

should be undertaken to assess the robustness of the composite indicator scores/ranks to the underlying assumptions and to identify which assumptions are more crucial in determining the final classification. Important to note the trade-off between multidimensionality and robustness in a composite indicator, given that a mono-dimensional index is likely to be more robust than a multi-dimensional one. This does not imply that the first index is better than the second one. In fact, robustness analysis should NOT be treated as an attribute of the composite indicator but of the inference which the composite indicator has been called upon to support.

- Consider **different methodological paths** to build the index, and if available, **different conceptual frameworks**.
- **Identify the sources of uncertainty** underlying in the development of the composite indicator and provide the composite scores/ranks with confidence intervals.
- Explain why certain countries notably improve or deteriorate their relative position given the assumptions.
- Conduct **sensitivity analysis** to show what sources of uncertainty are more influential in determining the scores/ranks.
Step 8. Relation to other indicators
should be made to correlate the composite indicator (or its dimensions) with existing (simple or composite) indicators and to identify linkages through regressions.

- Correlate the composite indicator with relevant measurable phenomena and explain similarities or differences.
- Develop data-driven narratives on the results.
- Perform causality tests (if time series data are available).

Step 9. Decomposition into the underlying indicators
should be carried out to reveal drivers for good/bad performance.

- Profile country performance at the indicator level to reveal strengths and limitations.
- Perform causality tests (if time series data are available).

Step 10. Visualisation of the results
should receive proper attention given that it can influence (or help to enhance) interpretability.

- Identify suitable presentational tools for the targeted audience.
- Select the visualisation technique which communicates the most information without hiding vital information.
- Present the results in a clear, easy to grasp and accurate manner.


1.2 COIN Tool – How it is organized
The COIN Tool (beta version) is organised around three sections:

The first section “Computation of the composite indicator” guides the user through the different steps needed in order to:

- create the database and the conceptual framework (yellow tabs),
- how to go about treating the outliers (green tabs), and
- how to “statistically” adjust the weights in order to obtain coherence between an indicator’s importance and how it actually affects the ranking.

The second section “Scenaria” (blue tabs) guides the user through the normalisation and aggregation phases in constructing a composite indicator.

Finally, the third section “Advanced features” (gold tabs) currently includes insightful illustrations on the “no imputation” choice and how it is equivalent to a sort of “shadow imputation”.

Many more features and functionalities will be available in the COIN Tool when officially released in the fall of 2017.
2 Computation of the composite indicator

2.1 Database and conceptual framework (yellow tabs)

2.1.1 Organisation of the data

The dataset underpinning a scoreboard or a composite indicator should be copy-pasted in the tab “Database”. The user should:

- Organise the data in units (rows) x indicators (columns), grouping the indicators according to the conceptual framework.
- Indicate all dimensions pertaining to each indicator.
- Report relative weights assigned to each indicator and dimension, the COIN tool does not require weights to add up to one.
- Report the desired direction for each indicator (good = 1, bad = -1). Dimensions are all assumed to have positive direction (the higher the score, the better).
- Report the indicator and unit names.
- Report missing values as “n/a”.

![Figure 1. ‘Database’ tab](source: JRC, 2017)
The **COIN tool** supports the following structure:

- a maximum of 250 units (e.g. countries, universities, etc.), coded unit.001 to unit.250;
- a maximum of 99 indicators, coded ind.01 to ind.99;
- four dimension levels:
  - a maximum of 33 sub-pillars, coded sp.01 to sp.33;
  - a maximum of 11 pillars, coded p.01 to p.11;
  - a maximum of three sub-indices, coded si.1 to si.3; and
  - one final index, coded index.

**Notes**

1. For composite indicators with less than four dimension levels, the user should assign all dimensions to one supra-dimension (example, a framework with 10 sub-pillars, 3 pillars, no sub-index, one index, becomes sp.01 to sp.10, p.01 to p.03, si.1, index (si.1 and index results will be identical). Adding the “intermediary” level si.1 is crucial for all features to function.

2. Cells in light blue need to be filled in with the index data (or left blank).

3. Excel assigns a value of 0 to blank cells, it is therefore crucial to double check for blanks that could be taken as zero values in original data sources.

**2.1.2 Conceptual Framework**

The **COIN tool** “automatically” summarizes the information provided by the user in the tab “Database” into the tab “Framework” in the white cells.

**Figure 2. ‘Framework’ tab**

![Conceptual framework](#)

In this tab, the user should:

- Report the desired relative weights assigned to each dimension within its respective supra-dimension (cells in blue); the COIN tool does not require weights to add up to one (summing to one is done “automatically” within the COIN tool).
- Report the names of dimensions.

Notes (1) The direction of each dimension is assumed to be one (i.e. the higher the score, the better). If it is not the case, then the COIN tool will not function properly.

2.2 Treatment of outliers (green tabs)

2.2.1 Original dataset – detection of outliers

The COIN tool extracts the information provided in the tab “Database” and performs a series of computations and conditional formatting:

- The COIN tool detects zero values, missing data, and negative values.
- For each indicator, it calculates descriptive statistics: missing values, min, max, mean, standard deviation, skewness, kurtosis, median and first and third quartile.
- For each unit, it calculates the indicator coverage.

Indicators with potential outliers are detected by checking their third and fourth moments, i.e. absolute skewness > 2 AND kurtosis > 3.5 (the COIN tool includes an option to change these values). The COIN tool also detects potential outliers on the basis of the interquartile range, but this is for reference only.

Indicators with outliers should be treated either by winsorization or by transformation of the indicator.

Figure 3. ‘Original’ tab

Notes  (1) It is recommended to require at least 65 percent indicator coverage per unit and dimension (this requirement can be relaxed or stricter depending on the degree of correlation between indicators within a dimension).

(2) Excel assigns a value of 0 to blank cells, it is therefore crucial to double check for blanks that could be taken as zero values in original data sources.

2.2.2 Winsorization

This tab helps the user to treat indicators with skewness > 2 AND kurtosis > 3.5 AND less than 5 outliers by winsorization.

Winsorizing implies transforming the statistical series by limiting its extreme values (at the upper, lower or both ends) by assigning them the next best value. The method is usually used in the presence of few outlier values (roughly 5 percent of units).

- For problematic indicators detected in the tab “OD”, the COIN tool winsorizes 1 to 5 outlier values; the process stops at the level where absolute skewness and kurtosis enter into the required ranges.
- When winsorization is not effective in dealing with outliers, the COIN tool reports the indicator as being a candidate for Box-Cox transformation.

![Figure 4. 'Winsorization' tab](source: JRC, 2017.)
2.2.3 Box-cox transformations

This tab helps the user to treat indicators with skewness $> 2$ AND kurtosis $> 3.5$ AND six or more outliers by a Box-Cox transformation, which transforms the whole series of values in a non-linear way.

Figure 5. 'Box-Cox' tab


Formulas:

- new value = old value $^\lambda$ if $-5 < \lambda < 5$; and
- new value = $\ln$ (old value) if $\lambda = 0$ and old value $> 0$

Statistical packages check for the lambda value that provides the smallest standard deviation; but the Box-Cox power transformation is not a guarantee for normality, an analysis of skewness and kurtosis is still required.

The COIN tool includes three transformations based on Box-Cox:

Formulas:

- LN: In transformation such that new min = 0: new value = $\ln$ (old value - old min + 1)
- SQRT: square root such that new min = 0: new value = (old value - old min) $^0.5$
- LNMED: In transformation and normalization such that min = 0, max = 1, median = 0.5:
  new value = 0.5 $[\ln [1 + (\text{old value} - \text{min}) (\text{max} + \text{min} - 2 \text{ sample median}) / ((\text{sample median} - \text{min}) ^ 2)] / \ln [(\text{max} - \text{sample median}) / (\text{sample median} - \text{min})] * \text{direction} + 0.5 (1 - \text{direction})$
The third transformation, LNMED, is akin to the following two steps: first, a linear normalization to the (0, 1) range; and second a non-linear transformation aimed at bringing the median to 0.5. By bringing the median to 0.5, this normalization procedure generally solves for potential outliers. Formula in two steps:

- Linear min-max: \( Y = \frac{(\text{old value} - \text{min})}{(\text{max} - \text{min})} \times \text{direction} + 0.5 \times (1 - \text{direction}) \)
- Non-linear transformation: \( Z = \ln(1 + aY) / \ln(1 + a) \), where \( a \) is such that \( Z \) (sample median) = 0.5, so that \( a = (1 - 2 \times \text{sample median}) / \text{sample median}^2 \)

The COIN tool indicates which indicators still present problems, if any, for these an alternative transformation should be found outside of the COIN tool and copy-pasted in the corresponding column in the tab "Database".

### 2.2.4 Scatterplots

The tab "Scatterplots" includes a scroll down menu to visualize each indicator, as well as its winsorized and transformed versions. This tab helps to evidence the outliers.

**Figure 6. 'Scatterplots’ tab**

*Source: JRC, 2017.*
2.3 Weight adjustments and final ranking (purple tabs)

2.3.1 Outlier free dataset – descriptive statistics

The tab “OutlierFree” recovers the information from the green tabs and constructs a new dataset without outliers. This dataset is used for the adjustment made to the framework itself, i.e. adjustment of weights (including deletion of indicators, i.e. weights of 0). Descriptive statistics are computed again.

Tabs linking to this dataset are coloured in purple.

![Figure 7. 'OutlierFree’ tab](image)


2.3.2 Indicator correlations and prospective weights

The COIN Tool calculates correlations between indicators (Pearson coefficients r), taking into account the direction of effects:

- At this point all correlations are expected to be positive. Negative correlations imply either that the desired direction of the indicator is wrong; that there are trade-offs between indicators; that the sample is too small and not representative; or that there is random correlation (if the level of correlation is low). It is desirable not to have negative correlations within the same dimension. Note, however, that small samples might lead to spurious negative correlations.
- In composite indicators, weights must be understood as ‘scaling coefficients (as opposed to ‘importance coefficients’), with the aim of arriving at dimension scores that are balanced in their underlying components.
  - The user may decide to eliminate indicators that are randomly associated to any of the remaining indicators in the dimension (e.g. assign a weight of 0).
  - Highly collinear indicators (r > 0.92 roughly) within a given dimension need to be treated (either by eliminating one of the two, or counting them as a single
indicator, i.e. adjusting their relative weight); otherwise they will influence all principal component analysis and dominate the unit scores in the respective dimension.

**Figure 8. 'Correl' tab**

![Correl tab](image)


The COIN tool allows users to adjust relative weights (row: “prospective weights”) on the basis of this analysis. These prospective weights are reproduced in the tab “Correl rebalancing” (explanations below); the final determination of relative weights needs to be made in the context of the computation of the index.

**Note:** (1) Correlations of raw data adjusted for direction and outliers are the same as z-score and min-max correlations.
2.3.3 Adjustment of weights on the basis of upper-level correlations

The overall purpose of this tab is to help the user to arrive at a model that is balanced in its underlying components, i.e. with correlations of dimensions with its components that are of a similar range. Under somewhat strong assumptions, squared correlation coefficients give an indication of explained variance.

Figure 9. ‘Rebalancing’ tab

Composite indicator aggregates need to be computed for the purpose of adjusting weights:

- First, normalize each indicator taking into account the direction of indicators. Two options are available in the tab: min-max scores and z-scores (details in heading 4 Normalization). These computations are included in the hidden tabs “AggOldWeights” and “AggNewWeights”).
- Second, compute all aggregates. The COIN tool uses weighted arithmetic averages, widely used in constructing composite indicators (details in heading 5 Aggregates).
- Third, compute correlations of each indicator/dimension with its supra-dimension(s).

**Formula:** 
\[
\text{Correlation} = \text{correlation (ind.xx, dim.yy)}
\]
Weights are then adjusted as follows:

- Weighting down dimensions with HIGH correlations (example: weight of 0.5 instead of 1);
- Weighting up dimensions with LOW correlations (example: weight of 2 instead of 1);
- Assign weights of 0 for indicators with negative correlations or correlations close to 0.
- Weights do not need to add up to 1 (they are “internally” adjusted to a unity sum).

2.3.4 Ranking with adjusted weights

The tab “Ranking” presents the ranking and scores computed with adjusted weights from the outlier free dataset. There the ranking with initial weights is also reported, together with the difference in ranks between the two for each unit.

Figure 10. ‘Ranking’ tab

2.3.5 Heatmap of scores with adjusted weights

The tab “Heatmap” includes three examples of visual presentation of the final ranking and scores for index, sub-indices and pillars. These charts are using conditional formatting.

Figure 11. 'Heatmap' tab

3 Scenaria (blue tabs)

The tabs that follow are aimed at assessing the robustness and sensitivity of rankings to changes in modelling assumptions. Excel only allows for a limited number of assessments, advanced featured (Section 4) are presented for completeness, but other statistical packages should be used.

3.1 Normalization

3.1.1 Min-max normalization

Normalization is required to obtain indicator scores and compute composite indicator aggregates. To normalise indicators, the most commonly used is min-max normalization; at the indicator level, the direction of effects need to be taken into account.

The discussion of aggregates is left for the heading “Aggregates” below, however note that geometric averaging necessitates strictly positive values; this implies that normalized scores need to be strictly positive (for example set a minimum at 0.1).

Formulas:

- Normalization in the range \([0, 1]\): new value = (old value - min) / (max - min) * direction + 0.5 * (1 - direction)
- Normalization in the range \([\text{desired min}, \text{desired max}]\): new value = [ (old value - min) / (max - min) * direction + 0.5 * (1 - direction) ] * (desired max - desired min) + desired min

3.1.2 Z-score normalization

Z-score is another widely used normalization method; at the indicator level, the direction of effects need to be taken into account as well.

The discussion of aggregates is left for the heading “Aggregates” below, however note that geometric averaging necessitates strictly positive values; this implies that normalized scores need to be strictly positive (for example set a minimum at 0.1). Z-scores have mean 0 and standard deviation 1; to obtain strictly positive values the mean has to be increased (for example to 5 or even 10 as some outliers in the negative tail of the distribution might still get negative values).

Formulas:

- \([\text{mean 0, std 1}]\): z-score = (old value - indicator mean) / indicator std * direction
- \([\text{desired mean, desired std}]\): new value = (old value - indicator mean) / indicator std * direction * desired std + desired mean

3.2 Aggregation methods and rankings

Once the data has been normalized, to obtain scores and ranks the different indicators are aggregated into each supra-dimension (indicator scores into sub-pillar scores, sub-pillar scores into pillar scores, pillar scores into sub-index scores, and sub-pillar scores into the final index scores).

Several aggregation functions exist, the following are the formulas for the most commonly used (example for a total of \(M\) indicators):
3.2.1 Arithmetic and geometric averages

The “Minmax” and “Dataz” tabs compute weighted arithmetic for sub-pillar, pillar, sub-index and index scores (default scores). In addition, for index and sub-index scores, the tab computes arithmetic and geometric averages, for, in each case, new, equal and random weights.

- **Arithmetic mean (equal weights):** score = AVERAGE (normalised values)
- **Weighted arithmetic mean (unequal or random weights):** score = SUMPRODUCT (weights * normalised values)
- **Geometric mean (equal weights):** score = PRODUCT (normalised values) ^ (1 / M)
- **Weighted geometric mean (unequal or random weights):** score = EXP [SUMPRODUCT (weights, LN(normalised values))]

A ranking is then computed for each aggregate in the tab “Scenaria” (arithmetic mean rank, weighted geometric mean rank, etc.).
Figure 12. ‘Minmax’ tab

Figure 13. ‘Dataz’ tab

3.2.1.1 Note on arithmetic v. geometric averages

Arithmetic averages are fully compensatory, an important comparative advantage in few indicators can compensate comparative disadvantages in many indicators; geometric averages, in contrast, reward units with balanced profiles, and motivates them to improve in the dimensions in which their perform poorly, and not just in any dimension.

Note: Geometric means require pillar values above zero; a zero pillar value is highly improbable, but if computations were to break down, for Minmax the desired minimum should be set at 0.1, and for “Dataz” the mean should be set at minimum 5 (refer to heading 4 Normalization for details).

3.2.1.2 Note on random weights

It is advisable to assess the sensitivity of ranks to random weights. One can also use some other software and run a number of Monte Carlo simulations (e.g. 1’000) to obtain a confidence interval for ranks (e.g. range of 90% of ranks).

In Excel, weights can be randomly selected using a uniform distribution in a given range [desired min, desired max]:

Formula:

- Prior weight = RANDBETWEEN (desired min *100, desired max * 100) / 100
- Posterior weight = weight / sum (weights)

Note: This RANDBETWEEN Excel formula requires the desired min and max (Excel calls these the bottom and top values) to be greater than 1; thus the multiplication and division by 100 allows ranges with two decimals. For aggregation, the prior weights have to be re-scaled to unity sum; these posterior weights are obtained by dividing each weight by the sum of weights within the same dimension. By pressing F9, the weights are automatically changed and computations are automatically updated.

3.2.1.3 Note on trimmed means

For composite indicators with only one or two levels of aggregation, an alternative aggregation method is the computation of trimmed means for each unit (and the corresponding ranking); this method, however, is not advisable for dimensions with few components (e.g. less than 5 or 6):

- Trimmed mean, equal weights (the best and worst values are discarded): score = [SUM(normalised values) – LARGE (normalised values, 1) – SMALL (normalised values, 1)] / COUNT(normalised values – 2)
3.2.2 Median and average rank

In the 'Dataranks' tab, the computation of ranks on individual indicators from the original dataset helps in the interpretation of results when trying to argue why one unit is doing better than another within a given dimension:

- **Rank**: \( \text{rank} = \text{RANK} \ [\text{original value, range, } 0.5 \ast (1 - \text{direction})] \)
- **Median rank**: \( \text{median rank} = \text{MEDIAN} \) (ranks for the same unit across all indicators)
- **Average rank**: \( \text{average rank} = \text{AVERAGE} \) (ranks for the same unit across all indicators)

A ranking is then computed for each aggregate (median rank rank, average rank rank – no mistake in the double word “rank”), include in the tab "Scenaria".

**Figure 14. 'Dataranks' tab**


**Note**: Missing data distort results because for indicators with low unit-coverage, ranks will be lower (thus better).
3.2.3 Borda rule

In the ‘Borda’ tab, for $N_i$ units in indicator $i$, the top-ranked unit in that indicator gets $N_i - 1$ points; the second ranked unit gets $N_i - 2$ points and so on; the last ranked unit gets 0 points.

- **Borda points (unit/indicator):** $\text{Borda points} = N_i - \text{rank (rank computed in “Datarank”)}
- **Borda points (equal weights):** average Borda points $= \frac{\text{SUM (points)}}{\text{COUNT (points)}}$
- **Weighted Borda points:** weighted Borda points $= \frac{\text{SUMPRODUCT (weights scaled to unity sum * points)}}{\text{COUNT (points)}}$

A ranking is then computed for each aggregate (average and weighted Borda points), included in the tab “Scenaria”.

**Note:** Missing data distort results, because for indicators with low unit-coverage, Borda points will be lower (thus worse).

**Figure 15.** ‘Borda’ tab

3.2.4 Copeland rule

The Copeland rule requires the computation of the outranking matrix.

3.2.4.1 Outranking matrix

In the ‘Outranking Matrix’ tab, units are compared pairwise. For each comparison, all the weights corresponding to the indicators in which unit A has a better score than unit B are added up as evidence in favour of “A better than B” (abbreviated as AB). For N units, there are N*(N-1) comparisons to be made. The diagonal elements are set at 0 by definition. In practical terms, for each pairwise unit comparison the following formula is used:

Formulas

- With raw values: SUM across all indicators \( [(\text{weight for indicator } i) \times (1 + \text{direction of indicator } i) \times \text{SIGN(\text{raw value of unit A on indicator } i - \text{raw value of unit B on indicator } i))] / 2 \)
- With normalized values: SUM across all indicators \( [(\text{weight for indicator } i) \times \text{SIGN(\text{normalized value of unit A on indicator } i - \text{normalized value of unit B on indicator } i))} \]

Pairwise comparison values are entered in the so-called outranking matrix. Since the sum of weights is one, above/below diagonal entries add up to one.

Figure 16. ‘Outranking Matrix’ tab

3.2.4.2 Copeland rule

In the 'Copeland' tab, the outranking matrix is transformed as follows: all values greater than 0.5 are replaced with +1, all values lower than 0.5 with -1 and all ties (values of exactly 0.5) with 0. The diagonal elements are set at 0 by definition. The Copeland score for each unit is the sum of the values in a given row. A final ranking is then calculated.

Note: In general, some compensability/substitutability is desired at lower aggregation levels (sub-pillsars), aggregation methods listed in the previous section are thus appropriate. However, at higher aggregation levels (pillars, sub-indices, overall index), compensability is less desirable; the Copeland rule can then be used to aggregate dimensions.

Figure 17. ‘Copeland’ tab

3.3 Scenaria

In the ‘Scenaria’ tab, unit scores associated with composite indicators are generally not calculated under conditions of certainty. For each composite indicator, modelling choices are based on different criteria, such as expert opinion in the field (e.g. selection of indicators), common practice (e.g. min-max normalization), statistical analysis (e.g. treatment of outliers); simplicity (e.g. no imputation of missing data), etc.

The robustness of results to modelling choices can be assessed by computing rankings with a combination of Monte Carlo simulations (uncertainty analysis) and a multi-modelling approach (sensitivity analysis) involving, for instance, weights, the imputation of missing data, and the aggregation formula.

This tab simply gathers all the rankings calculated in the previous tabs, combining different normalisation and aggregation methods.

A median rank across all scenaria together with the rank interval (minimum and maximum rank) is also reported for each unit.

Figure 18. ‘Scenaria’ tab

4 Advanced features

4.1 Imputation of missing data

A composite indicator might be computed with no imputation of missing data; however, the imputation of missing data is highly recommended to undertake a statistical audit of the composite indicator, in particular to assess the robustness of results and their sensitivity of results to modelling choices. Usually the latest available data point within a specified period is used for the imputation of missing data. The period used should be relatively short, ideally less than 5 years.

4.2 Shadow imputation

The non-imputation of missing data is equivalent to assigning the sub-pillar score value to the particular indicator (or the pillar score if the sub-pillar score is not available either). In order to work with a complete dataset for the assessment of robustness of rankings, this tab performs a fake imputation of missing data by replacing missing values by the score of the unit on the respective sub-pillar or, if not available, in the respective pillar (these come from the scores computed with the adjusted weights). The values that differ from the values in the original dataset are detected in green (concerns missing data and outliers).

Figure 19. ‘MinmaxfakeImp’ tab

Figure 20. ‘DatazfakeImp’ tab

References


List of tables and figures

Table 1. Ten Step Guide for Developing Composite Indicators and Scoreboards......... 5

Figure 1. ‘Database’ tab ........................................................................................................ 8
Figure 2. ‘Framework’ tab ...................................................................................................... 9
Figure 3. ‘Original’ tab .........................................................................................................10
Figure 4. ‘Winsorization’ tab ..............................................................................................11
Figure 5. ‘Box-Cox’ tab ........................................................................................................12
Figure 6. ‘Scatterplots’ tab ..................................................................................................13
Figure 7. ‘OutlierFree’ tab ...................................................................................................14
Figure 8. ‘Correl’ tab..........................................................................................................15
Figure 9. ‘Rebalancing’ tab ..................................................................................................16
Figure 10. ‘Ranking’ tab......................................................................................................17
Figure 11. ‘Heatmap’ tab ....................................................................................................18
Figure 12. ‘Minmax’ tab ......................................................................................................21
Figure 13. ‘Dataz’ tab ..........................................................................................................22
Figure 14. ‘Dataranks’ tab ...................................................................................................24
Figure 15. ‘Borda’ tab.........................................................................................................25
Figure 16. ‘Outranking Matrix’ tab .....................................................................................26
Figure 17. ‘Copeland’ tab ....................................................................................................27
Figure 18. ‘Scenaria’ tab ......................................................................................................28
Figure 19. ‘MinmaxfakeImp’ tab .........................................................................................29
Figure 20. ‘DatazfakeImp’ tab............................................................................................30
Europe Direct is a service to help you find answers to your questions about the European Union.

Freephone number (*):

00 800 6 7 8 9 10 11

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).


HOW TO OBTAIN EU PUBLICATIONS

Free publications:

- one copy:
  - via EU Bookshop (http://bookshop.europa.eu);

- more than one copy or posters/maps:
  - from the European Union’s representations (http://ec.europa.eu/represent_en.htm);
  - from the delegations in non-EU countries (http://eeas.europa.eu/delegations/index_en.htm);
  - by contacting the Europe Direct service (http://europa.eu/europedirect/index_en.htm) or calling 00 800 6 7 8 9 10 11 (freephone number from anywhere in the EU) (*).

  (*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

Priced publications:

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre’s mission is to support EU policies with independent evidence throughout the whole policy cycle.