Quality Assurance and Robustness

The European Commission’s science and knowledge service
Joint Research Centre

William Becker
william.becker@ec.europa.eu
rankings are here to stay, and it is therefore worth the time and effort to get them right
Rankings range from irresponsible musings by self-appointed experts and money-making schemes by commercial organizations to, at their best, serious efforts by academic or research organizations (Aitbach, 2015).

Notwithstanding recent attempts to establish good practice in composite indicator construction (OECD, 2008), “there is no recipe for building composite indicators that is at the same time universally applicable and sufficiently detailed” (Cherchye et al., 2007).

Booysen (2002, p.131) summarises the debate on composite indicators by noting that “not one single element of the methodology of composite indexing is above criticism”.

Andrews et al. (2004)] argue that “many indices rarely have adequate scientific foundations to support precise rankings: [...] typical practice is to acknowledge uncertainty in the text of the report and then to present a table with unambiguous rankings”
Tools for “serious efforts”

Quality Assurance

Robustness
Tools for “serious efforts”

Quality Assurance

Robustness
Tools for “serious efforts”

Quality Assurance

Ensuring statistical coherence (see previous lectures)
How do indicators contribute to the composite?

Robustness

How sensitive is the composite indicator to its assumptions?
• Transparency
• Exploration of uncertainty
• Anticipate criticism
We use Monte Carlo methods and sensitivity analysis

We use Monte Carlo methods and sensitivity analysis
Definition of the university is broad

A university – as the name suggests – tends to encompass a broad range of purposes and dimensions, focus and missions. It is difficult to condense into a compact measure.

But measuring and benchmarking excellence is still in demand:

- Governance
- Accountability
- Transparency

The growing mobility of students and researchers has also created a market for these measures among the prospective students and their families.
UK universities tumble in world rankings amid Brexit concerns

Uncertainty over research funding and immigration rules blamed for decline, as Cambridge slips out of top three for first time

Top 200 universities in the world - the table

The surveys that informed the rankings were carried out before the UK voted in June to leave the EU
An extreme impact of global rankings

**What** - 2005 THES created a major controversy in Malaysia: country’s top two universities slipping by almost 100 places compared to 2004.

**Why** - change in the ranking methodology (not well known fact and of limited comfort)

**Impact** - Royal commission of inquiry to investigate the matter. A few weeks later, the Vice-Chancellor of the University of Malaysia stepped down.
Composite indicators move money
(policy-makers are listening)

<table>
<thead>
<tr>
<th>Composite indicator</th>
<th>Budget programme</th>
<th>2018 Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>International Logistics Performance Index</td>
<td>Customs 2020</td>
<td>80.2 Million</td>
</tr>
<tr>
<td>Inform Index for Risk Management</td>
<td>Humanitarian aid</td>
<td>1.1 Billion</td>
</tr>
<tr>
<td>Corruption index, Press freedom Index, Ease of Doing Business Index + others.</td>
<td>Instrument for Pre-accession Assistance</td>
<td>1.7 Billion</td>
</tr>
<tr>
<td>Worldwide Governance Indicator</td>
<td>Development Cooperation Instrument</td>
<td>3.0 Billion</td>
</tr>
</tbody>
</table>
Part 1: Quality assurance—Weighting
“a general criticism that is frequently addressed at composite indicators, i.e. the arbitrary character of the procedures used to weight their various components”

- the Stiglitz Report, 2010
\[ y = \sum_{i=1}^{d} w_i x_{ji}, \quad j = 1, 2, \ldots, n \]
\[ y_j = \sum_{i=1}^{d} w_i \cdot c_{ji}, \quad j = 1, 2, \cdots, n \]
What do we want to achieve by weighting?

- Equal weighting: indicators should be equally important/influential
- Principle component analysis
What do we want to achieve by weighting?

- Equal weighting: indicators should be equally important/influential
- Principle component analysis: let the data speak! (clue: data doesn’t speak on its own)
- Data envelopment analysis: give the benefit of the doubt – let units maximise their scores

For methods with only one set of weights, there are at least two differing objectives:

- Balance indicators (or make indicators agree with pre-specified importance)
- Maximise information transfer/compression (PCA)
Weights are typically assigned to reflect importance, but...

weights do not equal measures
Weights are typically assigned to reflect importance, but...

weights do not equal measures
A composite indicator to rank teachers...
A composite indicator to rank teachers...
A composite indicator to rank teachers...

\[ y = \frac{1}{3} (x_1 + x_2 + x_3) \]

All have been standardised to have unit variance, but \( x_2 \) and \( x_3 \) have a correlation of 0.7. After calculating the values of the composite indicator, \( R^2 \) is used to check influence:

\[ R_i^2 := \text{corr}^2(y, x_i) \]

\[ R_1^2 = 0.227 \]
\[ R_2^2 = 0.657 \]
\[ R_3^2 = 0.657 \]

Increase the weight of \( x_1 \)? Teachers will complain that the index unfairly favours publications!

Correlations are very common in composite indicators.
How can we measure importance?
Measures of dependence

Fundamentally, we are interested in how the random variable $y$ depends on the random variable $x$.

1. Covariance:

\[ \text{cov}(y, x_i) := E[(y - \mu_y)(x_i - \mu_i)] \]

\[ \mu_i = E(x_i) \]
Measures of dependence

Fundamentally, we are interested in how the random variable $y$ depends on the random variable $x$.

1. Covariance: $cov(y, x_i) := E[(y - \mu_y)(x_i - \mu_i)]$
Measures of dependence

Fundamentally, we are interested in how the random variable $y$ depends on the random variable $x$.

1. Covariance:

$$\text{cov}(y, x_i) := E[(y - \mu_y)(x_i - \mu_i)]$$
Measures of dependence

Fundamentally, we are interested in how the random variable \( y \) depends on the random variable \( x \).

1. Covariance:

\[ \text{cov}(y, x_i) := \mathbb{E}[(y - \mu_y)(x_i - \mu_i)] \]
Measures of dependence

Fundamentally, we are interested in how the random variable $y$ depends on the random variable $x$.

1. Covariance:

$$\text{cov}(y, x_i) := E[(y - \mu_y)(x_i - \mu_i)]$$
Measures of dependence

Fundamentally, we are interested in how the random variable $y$ depends on the random variable $x$.

1. Covariance: $\text{cov}(y, x_i) := E[(y - \mu_y)(x_i - \mu_i)]$
Measures of dependence

Fundamentally, we are interested in how the random variable $y$ depends on the random variable $x$.

1. Covariance: $\text{cov}(y, x_i) := \text{E}[(y - \mu_y)(x_i - \mu_i)]$

(2+) + (2-) = 0
Measures of dependence

Fundamentally, we are interested in how the random variable $y$ depends on the random variable $x$.

1. Covariance:
   \[ \text{cov}(y, x_i) := E[(y - \mu_y)(x_i - \mu_x)] \]

2. Pearson correlation coefficient:
   \[ R_i := \text{corr}(y, x_i) := \frac{\text{cov}(y, x_i)}{\sigma_y \sigma_x} \]
   - Standardises covariance so that $R \in [-1, 1]$: 1 or -1 is perfect correlation, 0 is no correlation. Allows comparability.

3. Coefficient of determination: $R_i^2$
   - $R_i^2$ is a measure of linear dependence
   - $R_i^2 \in [0, 1]$: higher values indicate stronger dependence.

More generally, $R_i^2$ can be defined as \( \frac{\text{variance explained by regression}}{\text{total variance}} \).
Measuring the importance of indicators

Correlation coefficient

$$R_i := \text{corr}(y, x_i) := \frac{\text{cov}(y, x_i)}{\sigma_y \sigma_i}.$$  

$$R_i^2 := \text{corr}^2(y, x_i),$$

Ok but only measures linear dependence. Not always the case.

Correlation ratio*

$$S_i \equiv \eta_i^2 := \frac{V_{x_i} \left( \mathbb{E}_{x \sim i} (y \mid x_i) \right)}{V(y)},$$

Also known as “main effect index”, “first-order sensitivity index”, “nonlinear R^2”...

- Allows for nonlinear dependence
- Easily estimated by regression

*First conceived by Karl Pearson in 1905
Nonlinearity in the main effect

\[ S_i \equiv \eta_i^2 := \frac{V_{x_i}(E_{x_i}(y | x_i))}{V(y)} \]
Local linear regression

Weighted polynomial regression.

\[ \hat{m}(x_i) = \frac{\sum_{j=1}^{n} w(x_{ji} - x_i; h)y_j}{\sum_{j=1}^{n} w(x_{ji} - x_i; h)} \]

Can we tune the weights?

yes

(Optimisation approach)
Balancing the indicators

Minimisation algorithm → Trial weights → Nonlinear regression → $S_i$ → Difference → Target $S_i$
Balancing the indicators

Minimisation algorithm → Trial weights → Nonlinear regression → $S_i$ → Difference → Target $S_i$
Balancing the indicators

Minimisation algorithm → Trial weights → Nonlinear regression → Objective function → Target $S_i$ → Difference → $S_i$
Balancing the indicators

Minimisation algorithm → Trial weights → Nonlinear regression

Target $S_i$ → Difference

Objective function

$$\sum_{i=1}^{d} (\tilde{S}_i^* - \tilde{S}_i(w))^2$$
What about information?

Fundamentally, we want \( y \) to contain as much information from the \( x_i \) as possible. One way to measure this is through *mutual information* (a measure from information theory). MI requires knowledge of joint and marginal distributions. But has a link with (linear) correlation:

\[
I(y, x_i) = -\frac{1}{2}\ln(1 - R_i^2)
\]

With \( S_i \) we can generalise this to nonlinear dependence.

Implication: we can use \( S_i \) to measure the *mutual information* between each \( x_i \) and \( y \).

We can also use this to try to tune the weights to maximise the information.
Maximising information

- Maximisation algorithm
- Trial weights
- Nonlinear regression
- Sum of $S_i$

Objective function
Actually this is almost the same as PCA

PCA weighting rests on the idea that there is one or more “latent phenomena” which is driving the indicators (i.e. much in the spirit of composite indicators). [The ideal case is that we find just one]

PCA finds the linear combination of the indicators which explains the most variance of the indicators.
Actually this is almost the same as PCA

PCA weighting rests on the idea that there is one or more “latent phenomena” which is driving the indicators (i.e. much in the spirit of composite indicators). [The ideal case is that we find just one]

PCA finds the linear combination of the indicators which explains the most variance of the indicators.
Information and number of indicators

What happens when we add more indicators?

\[
\lim_{n \to \infty} I(x_i, y) = -\frac{1}{2} \ln(1 - \rho)
\]

\[
\lim_{n \to \infty} R^2 = \rho
\]

The average MI tends to a limit which is defined by the average correlation between the indicators.

[in linear/Gaussian case… should roughly hold in practice]

Implications:

- Large frameworks can still work from an information perspective
- Size of framework should be (in part) determined from an information PoV
1. Teaching
   • Reputation survey
   • Staff-student ratio
   • Doctorates
   • Institutional income

2. Research
   • Reputation survey
   • Research income
   • Research productivity

3. Citations

4. International outlook
   • Percentage international students and staff
   • Institutional collaboration

5. Industry Income

https://www.mentimeter.com/app
<table>
<thead>
<tr>
<th>Academic Reputation</th>
<th>Employer Reputation</th>
<th>Faculty/Student Ratio</th>
<th>Citations per Faculty</th>
<th>International Faculty</th>
<th>International Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>40%</td>
<td>10%</td>
<td>20%</td>
<td>20%</td>
<td>5%</td>
<td>5%</td>
</tr>
<tr>
<td>Opinion-based survey</td>
<td>Opinion-based survey</td>
<td>Teacher/student ratio</td>
<td>Citations divided by faculty</td>
<td>Proportion of international faculty members</td>
<td>Proportion of international students</td>
</tr>
<tr>
<td>Academic Reputation</td>
<td>Employer Reputation</td>
<td>Faculty/Student Ratio</td>
<td>Citations per Faculty</td>
<td>International Faculty</td>
<td>International Students</td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------------</td>
<td>-----------------------</td>
<td>-----------------------</td>
<td>-----------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>40%</td>
<td>10%</td>
<td>20%</td>
<td>20%</td>
<td>5%</td>
<td>5%</td>
</tr>
<tr>
<td>Opinion-based survey</td>
<td>Opinion-based survey</td>
<td>Teacher/student ratio</td>
<td>Citations divided by faculty</td>
<td>Proportion of international faculty members</td>
<td>Proportion of international students</td>
</tr>
</tbody>
</table>
QS input data
All of the largest rank differences labelled here are such that the Times rank is higher than the QS rank.
QS Regression

x1  Academic Reputation
x2  Employer Reputation
x3  Faculty Student
x4  Citations per Faculty
x5  International Faculty
x6  International Students
The diagram illustrates the QS Target Importance compared to Si for various categories:

- **Academic Reputation**
- **Employer Reputation**
- **Faculty Student Citations per Faculty**
- **Citations per Faculty**
- **International Faculty**
- **International Students**

The x-axis represents the categories, while the y-axis shows the importance ranging from 0 to 0.5. The bars in blue represent the Target, and the bars in red represent the Original.
Times Regression

x1  Citations
x2  Industry Income
x3  International Outlook
x4  Research
x5  Teaching
Times target importance compared to Si

- Citations
- Industry Income
- International Outlook
- Research
- Teaching

Target
Original
<table>
<thead>
<tr>
<th></th>
<th>Original</th>
<th>Optimised</th>
<th>Opt (+)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic Reputation</td>
<td>0.4</td>
<td>0.60</td>
<td>0.32</td>
</tr>
<tr>
<td>Employer Reputation</td>
<td>0.1</td>
<td>0.37</td>
<td>0.00</td>
</tr>
<tr>
<td>Faculty Student</td>
<td>0.2</td>
<td>0.32</td>
<td>0.30</td>
</tr>
<tr>
<td>Citations per Faculty</td>
<td>0.2</td>
<td>0.28</td>
<td>0.31</td>
</tr>
<tr>
<td>International Faculty</td>
<td></td>
<td>0.05</td>
<td>0.07</td>
</tr>
<tr>
<td>International Students</td>
<td></td>
<td>0.05</td>
<td>0.00</td>
</tr>
</tbody>
</table>
The graph illustrates the times target importance compared to Si for different categories. The categories included are Citations, Industry Income, International Outlook, Research, and Teaching.

The table below shows the original and optimised importance values for each category:

<table>
<thead>
<tr>
<th>Category</th>
<th>Original</th>
<th>Optimised</th>
<th>Optimised (pos weights)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citations</td>
<td>0.30</td>
<td>0.46</td>
<td>0.44</td>
</tr>
<tr>
<td>Industry Income</td>
<td>0.03</td>
<td>-0.08</td>
<td>0.11</td>
</tr>
<tr>
<td>International Outlook</td>
<td>0.08</td>
<td>0.19</td>
<td>0.16</td>
</tr>
<tr>
<td>Research</td>
<td>0.30</td>
<td>0.03</td>
<td>0.00</td>
</tr>
<tr>
<td>Teaching</td>
<td>0.30</td>
<td>0.39</td>
<td>0.29</td>
</tr>
</tbody>
</table>
CIAO Toolbox

Composite Indicator Analysis and Optimisation

- Optimisation of weights (balancing + [soon] infomax)
- Different types of aggregation
- Nonlinear dependence modelling

(very beta at the moment...)
Can we tune the weights?

yes

(Optimisation approach)
Can we tune the weights?

Should

hmm
New Rank | Original Rank | University
---|---|---
1 | 5 | California Institute of Technology (Caltech)
2 | 1 | Massachusetts Institute of Technology (MIT)
3 | 17 | Johns Hopkins University
4 | 4 | University of Cambridge
5 | 66 | University of Illinois at Urbana-Champaign
6 | 6 | University of Oxford
7 | 2 | Stanford University
8 | 7 | UCL (University College London)
9 | 3 | Harvard University
10 | 24 | Duke University
<table>
<thead>
<tr>
<th>New Rank</th>
<th>Original Rank</th>
<th>University</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>Massachusetts Institute of Technology</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>University of Oxford</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>California Institute of Technology</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>Imperial College London</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>Stanford University</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>University of Cambridge</td>
</tr>
<tr>
<td>7</td>
<td>17</td>
<td>Johns Hopkins University</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>Princeton University</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>ETH Zurich</td>
</tr>
<tr>
<td>10</td>
<td>18</td>
<td>Duke University</td>
</tr>
</tbody>
</table>
A final thought
A final thought
A final thought
The Datasaurus Dozen

https://CRAN.R-project.org/package=datasauRus
Summary

• Understand objective of weighting (esp. balancing vs. information objectives)
• Estimate $S_i$ of indicators using nonlinear regression (allows for correlation)
• Optimise weights to agree with target “importance” using numerical algorithm
• CIAO toolbox if you want to really explore

All this can be found in recent paper:

Takeaway

Weights don’t equal importance in composite indicators. Composite indicators often involve a tradeoff between statistics and communication.

Open Questions

Does importance = $S_i$? When we assign “importances” to variables, are we implicitly taking some correlation into account?
Part 2: Robustness and uncertainty analysis
Which steps in the construction of a composite indicator are uncertain?

- Step 10. Presentation & dissemination
- Step 9. Association with other variables
- Step 8. Back to the indicators
- Step 7. Robustness & sensitivity
- Step 6. Weighting & aggregation
- Step 5. Normalisation of data
- Step 4. Multivariate analysis
- Step 3. Data treatment (missing, outliers)
- Step 2. Selection of indicators
- Step 1. Developing the framework

ALL OF THEM
Which steps in the construction of a composite indicator are uncertain?

- Step 10. Presentation & dissemination
- Step 9. Association with other variables
- Step 8. Back to the indicators
- Step 7. Robustness & sensitivity
- Step 6. Weighting & aggregation
- Step 5. Normalisation of data
- Step 4. Multivariate analysis
- Step 3. Data treatment (missing, outliers)
- Step 2. Selection of indicators
- Step 1. Developing the framework

ALL OF THEM
Which steps in the construction of a composite indicator are uncertain?

Step 10. Presentation & dissemination
Step 9. Association with other variables
Step 8. Back to the indicators
Step 7. Robustness & sensitivity
Step 6. Weighting & aggregation
Step 5. Normalisation of data
Step 4. Multivariate analysis
Step 3. Data treatment (missing, outliers)
Step 2. Selection of indicators
Step 1. Developing the framework

Even here: uncertain pdfs, limited exploration of assumptions...
What weights? How to aggregate?
How? Min/max, standardise, rank...
Which methods? Assume normality?
How to impute? What is an outlier?
The right ones? Irrelevant ones? Missing indicators?
What is the definition of “excellence”/“innovation”/etc...?

But, the same is true for any model.
Uncertainty and Sensitivity Analysis

We need:

• To quantify the uncertainty in our assumptions (assign probabilities to alternative assumptions)
• To propagate this through our composite indicator (Monte Carlo)
• To quantify/visualise uncertainty in the scores of our composite indicator (confidence intervals, pdfs, scatter plots, sensitivity analysis tools)

Uncertainty analysis
How uncertain is the output (CI scores) given the uncertainty in the input (assumptions made in its construction)?

Sensitivity analysis
How much uncertainty is caused by each assumption?
Model averaging: whenever a choice in the composite setting-up may not be strongly supported or if you may not trust one single model, we’ll recommend you to use more models.
it is essential to vary assumptions simultaneously to fully explore the “assumption space”.

Only child 1:

\[ \text{chaos} = f_1(C_1) \]

Child 1 plus child 2:

\[ \text{chaos} = f_1(C_1) + f_2(C_2) + f_{12}(C_1, C_2) \neq f_1(C_1) + f_2(C_2) \]

Only child 2:

\[ \text{chaos} = f_2(C_2) \]

Testing the behaviour of one child at a time would not at all explore the full space of chaos.
Uncertainty analysis of composite indicators requires some programming—currently using Matlab but moving to R.

Testable assumptions include:
- Normalisation method (A₁)
- Weighting method (A₂)
- Perturbations of weights (A₃)
- Set of indicators included (A₄)
- Data imputation method (…)
- Structure of composite (…)
- ++ anything you can program! (…)

Sample this function many times at random $A_i$ values, and record the output each time.
What do uncertainty and sensitivity analyses tell you?

→ NOT to verify whether the two global university rankings are **legitimate models** to measure university performance

→ To test whether the rankings and/or their associated inferences are **robust or volatile with respect to changes in selected methodological assumptions** within a plausible and legitimate range.

Uncertainty analysis

- Normalisation
  - Min/max
  - Rank
- Aggregation
  - Arithmetic
  - Geometric
- Weights
  - +/-25% of nominal weight value
Activate simultaneously different sources of uncertainty that cover a wide spectrum of methodological assumptions

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Alternatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of indicators</td>
<td>• all six indicators included or one-at-time excluded (6 options)</td>
</tr>
<tr>
<td>Weighting method</td>
<td>• original set of weights, factor analysis, equal weighting, data envelopment analysis</td>
</tr>
<tr>
<td>Aggregation rule</td>
<td>• additive, multiplicative, Borda multi-criterion</td>
</tr>
</tbody>
</table>

70 scenarios

Estimate the FREQUENCY of the university ranks obtained in the different simulations
**Impact of assumptions:** much stronger for the middle ranked universities.
- Impact of uncertainties on the university ranks is even more apparent.
- M.I.T.: ranked 9th, but confirmed only in 13% of simulations (plausible range [4, 35])
- Very high volatility also for universities ranked 10th-20th position, e.g., Duke Univ, John Hopkins Univ, Cornell Univ.

<table>
<thead>
<tr>
<th>University</th>
<th>Rank</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>HARVARD University</td>
<td>1</td>
<td>USA</td>
</tr>
<tr>
<td>YALE University</td>
<td>2</td>
<td>USA</td>
</tr>
<tr>
<td>University of CAMBRIDGE</td>
<td>3</td>
<td>UK</td>
</tr>
<tr>
<td>University of OXFORD</td>
<td>4</td>
<td>UK</td>
</tr>
<tr>
<td>CALIFORNIA Institute of Technology</td>
<td>5</td>
<td>USA</td>
</tr>
<tr>
<td>IMPERIAL College London</td>
<td>6</td>
<td>UK</td>
</tr>
<tr>
<td>UCL (University College London)</td>
<td>7</td>
<td>UK</td>
</tr>
<tr>
<td>University of CHICAGO</td>
<td>8</td>
<td>USA</td>
</tr>
<tr>
<td>MASSACHUSETTS Institute of Technology</td>
<td>9</td>
<td>USA</td>
</tr>
<tr>
<td>COLUMBIA University</td>
<td>10</td>
<td>USA</td>
</tr>
<tr>
<td>University of PENNSYLVANIA</td>
<td>11</td>
<td>USA</td>
</tr>
<tr>
<td>PRINCETON University</td>
<td>12</td>
<td>USA</td>
</tr>
<tr>
<td>DUKE University</td>
<td>13</td>
<td>USA</td>
</tr>
<tr>
<td>JOHNS HOPKINS University</td>
<td>14</td>
<td>USA</td>
</tr>
<tr>
<td>CORNELL University</td>
<td>15</td>
<td>USA</td>
</tr>
<tr>
<td>AUSTRALIAN National University</td>
<td>16</td>
<td>Australia</td>
</tr>
<tr>
<td>STANFORD University</td>
<td>17</td>
<td>USA</td>
</tr>
<tr>
<td>University of MICHIGAN</td>
<td>18</td>
<td>USA</td>
</tr>
<tr>
<td>University of TOKYO</td>
<td>19</td>
<td>Japan</td>
</tr>
<tr>
<td>MCGILL University</td>
<td>20</td>
<td>Canada</td>
</tr>
</tbody>
</table>

Simulated rank range - THES 2008

Legend:
- Frequency lower 15%
- Frequency between 15 and 30%
- Frequency between 30 and 50%
- Frequency greater than 50%

Note: Frequencies lower than 4% are not shown
Uncertainty analysis – ARWU results

54 universities outside the interval (total of 503)
[43 universities in the Top 100]
Uncertainty analysis – THE results

250 universities outside the interval (total of 400) [61 universities in the Top 100]
The Global Talent Competitiveness Index 2017 (GTCI)
Talent and Technology

Technology & Work: Disruption & Creation

The advance of technology is disrupting the world of work.

The new nature of work

High connectedness: Collaboration and co-creation
Work life blend
The job for life no longer exists: Multi-career is the norm

New jobs require new skills & attitudes

Beyond automation

Flexibility is key. 30% of US & European workers are free agents.

The GTCI ranks countries by their ability to grow, attract and retain talent.
### Uncertainty analysis for the GTCI 2017: Weights, missing data, aggregation, and normalisation

<table>
<thead>
<tr>
<th>I. Uncertainty in the treatment of missing values</th>
<th>REFERENCE</th>
<th>ALTERNATIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>No estimation of missing data</td>
<td>Expectation Maximisation (EM)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II. Uncertainty in the aggregation formula at pillar level</th>
<th>REFERENCE</th>
<th>ALTERNATIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arithmetic average</td>
<td>Geometric average</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III. Uncertainty in the method of normalisation</th>
<th>REFERENCE</th>
<th>ALTERNATIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partial normalisation</td>
<td>Full normalisation</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IV. Uncertainty in the weights</th>
<th>GTCI sub-index</th>
<th>Pillar</th>
<th>Reference value for the weight (within the sub-index)</th>
<th>Distribution assigned for robustness analysis (within the sub-index)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Input</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Enable</td>
<td></td>
<td>0.25</td>
<td>$U[0.15,0.35]$</td>
</tr>
<tr>
<td></td>
<td>Attract</td>
<td></td>
<td>0.25</td>
<td>$U[0.15,0.35]$</td>
</tr>
<tr>
<td></td>
<td>Grow</td>
<td></td>
<td>0.25</td>
<td>$U[0.15,0.35]$</td>
</tr>
<tr>
<td></td>
<td>Retain</td>
<td></td>
<td>0.25</td>
<td>$U[0.15,0.35]$</td>
</tr>
<tr>
<td></td>
<td>Output</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vocational and Technical Skills</td>
<td></td>
<td>0.50</td>
<td>$U[0.40,0.60]$</td>
</tr>
<tr>
<td></td>
<td>Global Knowledge Skills</td>
<td></td>
<td>0.50</td>
<td>$U[0.40,0.60]$</td>
</tr>
</tbody>
</table>
Sensitivity Analysis

Which assumptions are driving the uncertainty in the rankings?

We can also go further and delve into the sensitivity analysis literature (variance-based SA, moment-independent measures, etc).
A brief glimpse of SA...

If we want to go a step further we can calculate formal sensitivity measures telling us, e.g:

“Of the assumptions tested, the uncertainty in the choice of normalisation method causes 40% of the uncertainty in rankings”

Let $A_1, A_2, A_3,...$ be the assumptions used in the composite indicator construction, which result in a set of scores:

\[
\text{scores} = \text{CI}(A_1, A_2, A_3, A_4,...)
\]

The uncertainty in the scores, as a result of uncertainty in the $A_i$, can be captured by $\text{var(scores)}$. Sensitivity analysis can decompose this variance into portions attributable to each assumption.

\[
\text{var(scores)} = \sum_{i=1}^{k} V_i + \sum_{i}^{k} \sum_{j<i} V_{i,j} + \cdots + V_{1,2,...,k}
\]

\[
V_i = \text{var}_{A_i}[E_{A_{\sim i}}(\text{scores}|A_i)]
\]

\[
V_{i,j} = \text{var}_{A_iA_j}[E_{A_{\sim i,j}}(\text{scores}|A_i, A_j)] - \text{var}_{A_i}[E_{A_{\sim i}}(\text{scores}|A_i)] - \text{var}_{A_j}[E_{A_{\sim j}}(\text{scores}|A_j)]
\]

Requires a specific design.

Very informative but perhaps hard to communicate...

Uncertainty and validation

We cannot explore the full uncertainty of a composite indicator: we can only explore some of the assumptions.

“Given the assumptions that were tested, the outcome of the composite indicator is shown to vary in the following way...” [lower bound on uncertainty]

In general it is not possible to validate a composite indicator (build compelling evidence that it is an effective model).

(but, the same problems apply to some extent to any model)
From irresponsible musings to serious efforts

Rankings range from irresponsible musings by self-appointed experts and money-making schemes by commercial organizations to, at their best, serious efforts by academic or research organizations. (Aitbach, 2015)

Composite indicators

• Are very widely used
• Fill a demand for which there is no other alternative

Therefore we should use available tools to **increase robustness and credibility**:

1. **Transparency**—detailed description of methodology, data sources, assumptions
2. **Statistical soundness**—analysis of correlations, data structure, effects of weights, etc.
3. **Uncertainty and sensitivity analysis**—check effect of alternative but plausible assumptions. Honestly acknowledge uncertainty.
References


Stay in touch

EU Science Hub: ec.europa.eu/jrc

Twitter: @EU_ScienceHub

Facebook: EU Science Hub - Joint Research Centre

LinkedIn: Joint Research Centre

YouTube: EU Science Hub