Ratings and rankings: voodoo or science?

Fifth Italian Congress of Econometrics and Empirical Economics (ICEEE-5th)
January 16-18, 2013, Genova, Italy

Paolo Paruolo (1), Michaela Saisana (2), Andrea Saltelli (2)

(1) University of Insubria, Varese, Italy
(2) European Commission, Joint Research Centre, Econometrics and Applied Statistics Unit
Outline

- Motivation & Objective of the study
- Overview of two international university rankings
  ARWU ranking & THES ranking
- Statistical coherence
- Conclusions
University rankings are used to judge about the performance of university systems
• These rankings are relevant to today’s discourse on Higher Education reform in the EU

• Also academics use ARWU

University rankings - yearly published

+ Very appealing for capturing a university’s multiple missions in a single number
+ Allow one to situate a given university in the worldwide context
- Can lead to misleading and/or simplistic policy conclusions

Question:
Can we say something about the quality of the university rankings and the reliability of the results?
### Overview - ARWU ranking

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Indicator</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of Education</td>
<td>Alumni of an institution winning Nobel Prizes and Fields Medals</td>
<td>10%</td>
</tr>
<tr>
<td>Quality of Faculty</td>
<td>Staff of an institution winning Nobel Prizes and Fields Medals</td>
<td>20%</td>
</tr>
<tr>
<td></td>
<td>Highly cited researchers in 21 broad subject categories</td>
<td>20%</td>
</tr>
<tr>
<td>Research Output</td>
<td>Articles published in Nature and Science</td>
<td>20%</td>
</tr>
<tr>
<td></td>
<td>Articles in Science Citation Index-expanded, Social Science Citation Index</td>
<td>20%</td>
</tr>
<tr>
<td>Academic performance</td>
<td>Academic performance with respect to the size of an institution</td>
<td>10%</td>
</tr>
</tbody>
</table>

### PROS and CONS

- 6 « objective » indicators
- Focus on research performance, overlooks other U. missions.
- Biased towards hard sciences intensive institutions
- Favours large institutions

### METHODOLOGY

- 6 indicators
- Best performing institution = 100; score of other institutions calculated as a percentage
- Weighting scheme chosen by rankers
- Linear aggregation of the 6 indicators
PROS and CONS

✓ Attempt to take into account teaching quality
✓ Two expert-based indicators: 50% of total (lack of transparency)
✓ Yearly changes in methodology
✓ Measures research quantity

METHODOLOGY

✓ 6 indicators
✓ Z-score calculated for each indicator; best performing institution = 100; other institutions are calculated as a percentage
✓ Weighting scheme: chosen by rankers
✓ Linear aggregation of the 6 indicators
1 – Same top10: Harvard, Cambridge, Princeton, Caltech, MIT and Columbia

2 - Greater variations in the middle to lower end of the rankings

3 - Europe is lagging behind: both ARWU (else SJTU) and THES rankings

4 – THES favours UK universities: all UK universities below the line (in red)
Question:
Can we say something about the quality of the university rankings and the reliability of the results?

Statistical coherence

We suggest to use as a measure of Importance of a variable in an overall index what is known as:

- Pearson’s correlation ratio
- First order effect
- Top marginal variance
- Main effect

\[ S_i = \frac{V_{X_i} \left( E_{X_{-i}} (Y|X_i) \right)}{V(Y)} \]

Question:
Can we say something about the quality of the university rankings and the reliability of the results?

Source: Paruolo, Saisana, Saltelli, 2013, J.Royal Stat. Society A
Pearson’s correlation ratio

\[ S_i = \frac{V_{X_i} \left( E_{X \sim_i} \left( Y \mid X_i \right) \right)}{V(Y)} \]

Features:
- it offers a precise definition of importance, that is ‘the expected reduction in variance of the CI that would be obtained if a variable could be fixed’;
- it can be used regardless of the degree of correlation between variables;
- it is model-free, in that it can be applied also in non-linear aggregations;
- it is not invasive, in that no changes are made to the CI or to the correlation structure of the indicators (unlike what we will see next on uncertainty analysis).
Using these points we can compute a statistics that tells us:
How much (on average) would the variance of the ARWU scores be reduced if I could fix the variable ‘Papers in Nature & Science’?

This measure $S_i$ shall be our ruler for ‘importance’; example: $S_i = 0.6 \rightarrow$ I could reduce the variation of the ARWU score by 60% by fixing ‘Papers in Nature & Science’.
Kernel regression

Fig. 1. 2009 HDI $y$ and life expectancy $x_1$: (a) CV criterion as a function of the smoothing parameter $h (h_{DPI}; h_{CV})$; (b) linearity test $p$-value as a function of $h (h_{DPI}; h_{CV})$; (c) main effects $S_i$ as a function of $h (h_{DPI}; h_{CV})$; (d) cross-plot of $y$ versus $x_1$ with a linear fit and local linear fits for $h_{DPI} = 0.0841$ (………) and $h_{CV} = 0.088$ (— — —)
One can hence compare the importance of an indicator as given by the nominal weight (assigned by developers) with the importance as measured by the first order effect \((Si)\) to test the index for coherence.
## Statistical coherence - ARWU

<table>
<thead>
<tr>
<th></th>
<th>( w_i )</th>
<th>( S_{i,\text{lin}} )</th>
<th>( S_{i,\text{CV}} )</th>
<th>( S_{i,\text{DPI}} )</th>
<th>( S_{i,\text{min}} )</th>
<th>( S_{i,\text{max}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2008 ARWU</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alumni winning Nobel Prize</td>
<td>0.10</td>
<td>0.64</td>
<td>0.65</td>
<td>0.67</td>
<td>0.65</td>
<td>0.76</td>
</tr>
<tr>
<td>Staff winning Nobel Prize</td>
<td>0.20</td>
<td>0.72</td>
<td>0.72</td>
<td>0.73</td>
<td>0.72</td>
<td>0.80</td>
</tr>
<tr>
<td>Highly cited researchers</td>
<td>0.20</td>
<td>0.81</td>
<td>0.85</td>
<td>0.87</td>
<td>0.85</td>
<td>0.90</td>
</tr>
<tr>
<td>Articles in <em>Nature and Science</em></td>
<td>0.20</td>
<td>0.87</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
<td>0.94</td>
</tr>
<tr>
<td>Articles in <em>Science and Social Sciences Citation Index</em></td>
<td>0.20</td>
<td>0.63</td>
<td>0.70</td>
<td>0.70</td>
<td>0.64</td>
<td>0.90</td>
</tr>
<tr>
<td>Academic performance (size adjusted)</td>
<td>0.10</td>
<td>0.71</td>
<td>0.76</td>
<td>0.75</td>
<td>0.72</td>
<td>0.88</td>
</tr>
</tbody>
</table>

1) Hypothesis of linearity is not rejected for two indicators for the ARWU (when evaluating the tests at hDPI and hCV)

2) Si’s are more similar to each other than the nominal weights, i.e. ranging between 0.14 and 0.19 (normalized Si’s to unit sum; CV estimates) when weights should either be 0.10 or 0.20.
1) Hypothesis of linearity not rejected for four indicators for the THES (when evaluating the tests at hDPI and hCV)

2) In THES, the combined importance of peer-review variables (recruiters and academia) appears larger than stipulated by developers, indirectly supporting the hypothesis of linguistic bias at times addressed to this measure. Further, the teacher/student ratio, a key variable aimed at capturing the teaching dimension, is much less important than it should be (normalized Si is 0.09, nominal weight is 0.20).
Conclusions

Two statistical tools developed by the JRC to test the quality of composite indicators

• Conceptual & Statistical coherence (non invasive)
• Uncertainty modeling and propagations (invasive)

More at:

http://composite-indicators.jrc.ec.europa.eu

(first Google hit on “composite indicators” over the last 10 years!)
References and Related Reading


