

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

Motivation & Objective

Classement de Shanghai : les universités françaises à la traîne

Seules 23 universités françaises figurent dans le classement de Shanghai des 500 meilleurs établissements mondiaux.

C'est un électrochoc qui secoue chaque année les universités tricolores en pleine torpeur estivale. Le classement de Shanghai, qui évalue les performances des meilleurs établissements d'enseignement supérieur mondiaux, vient d'être dévoilé. Ni chute ni progression spectaculaire : avec 23 universités dans le top 500 (22 l'an passé), et 3 dans le top 100 (4 l'an passé), la France se classe au septième rang des 37 pays, rétrocedant une place à la Suède. Des résultats décevants : en légère baisse par rapport à l'an passé, la France ne parvient toujours pas à rattraper son retard sur ses homologues britanniques et allemandes, dont une quarantaine d'universités sont classées.

Palmarès mondial des universités

Rang	Institution	Pays
1	Harvard	Etats-Unis
2	Stanford	Etats-Unis
3	Berkeley	Etats-Unis
4	Cambridge	Roy.-Uni
5	Massachusetts Inst. Tech. (MIT)	Etats-Unis
6	California Inst. Tech.	Etats-Unis
7	Columbia	Etats-Unis
8	Princeton	Etats-Unis
9	Chicago	Etats-Unis
10	Oxford	Roy.-Uni
42	Univ. Paris-VI	France
49	Univ. Paris-XI	France
73	Ecole normale sup. de Paris	France

Source : université de Jiao Tong.

Strasbourg-I ne peut qu'espérer que

Comment sauver l'université française

Des pistes pour remédier à la rétrogradation de nos établissements d'enseignement supérieur dans les classements internationaux

Le classement de Shanghai, qui fait désormais référence quels que soient ses défauts, vient de rétrograder encore une fois les universités françaises par rapport à leurs sœurs étrangères. Dans les cent premières, il n'en accepte plus que trois : Paris-VI (42^e), Paris-XI (49^e) et l'Ecole normale supérieure (73^e).

Jacques Blamont

Membre de l'Académie des sciences

0 ; revenus : 0. Ses ressources totales seraient donc sept fois inférieures à celles du Michigan. Les chiffres relatifs à d'autres universités américaines donnent un autre résultat. Attribuons les

les entreprises abonderaient les bourses. La gestion des droits d'inscription et des bourses ressortirait à l'autonomie universitaire.

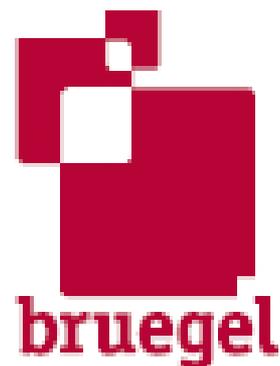
Une deuxième leçon à tirer de la pratique internationale est que l'université ne doit pas être asphyxiée par des gens qui n'y ont pas leur place. Une sélection à l'entrée s'impose, soit par une meilleure utilisation du baccalauréat (seuls seraient admis les titulaires d'une mention), soit

University rankings are used to judge about the performance of university systems

Motivation & Objective

- These rankings are relevant to today's discourse on Higher Education reform in the EU
- Also academics use ARWU

P. Aghion, M. Dewatripont, C. Hoxby, A. Sapir, A., “Higher aspirations: An agenda for reforming European universities” (Bruegel Blueprint Series N.5, 2008).



Motivation & Objective

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

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

METHODOLOGY

- ✓ 6 indicators
- ✓ Best performing institution =100; score of other institutions calculated as a percentage

PROS and CONS

- ✓ 6 « **objective** » indicators
- ✓ Focus on research performance, overlooks other aspects
- ✓ Biased towards hard sciences intensive institutions
- ✓ Favours large institutions

- ✓ Weighting scheme chosen by rankers
- ✓ Linear aggregation of the 6 indicators

Overview - THES ranking

Criteria	Indicator	Weight
Research Quality	Academic Opinion: Peer review, 6,354 academics	40%
	Citations per Faculty: Total citation/ Full Time Equivalent faculty	20%
Graduate Employability	Recruiter Review: Employers' opinion, 2,339 recruiters	10%
International Outlook	International Faculty: Percentage of international staff	5%
	International Students: Percentage of international students	5%
Teaching Quality	Student Faculty: Full Time Equivalent faculty/student ratio	20%

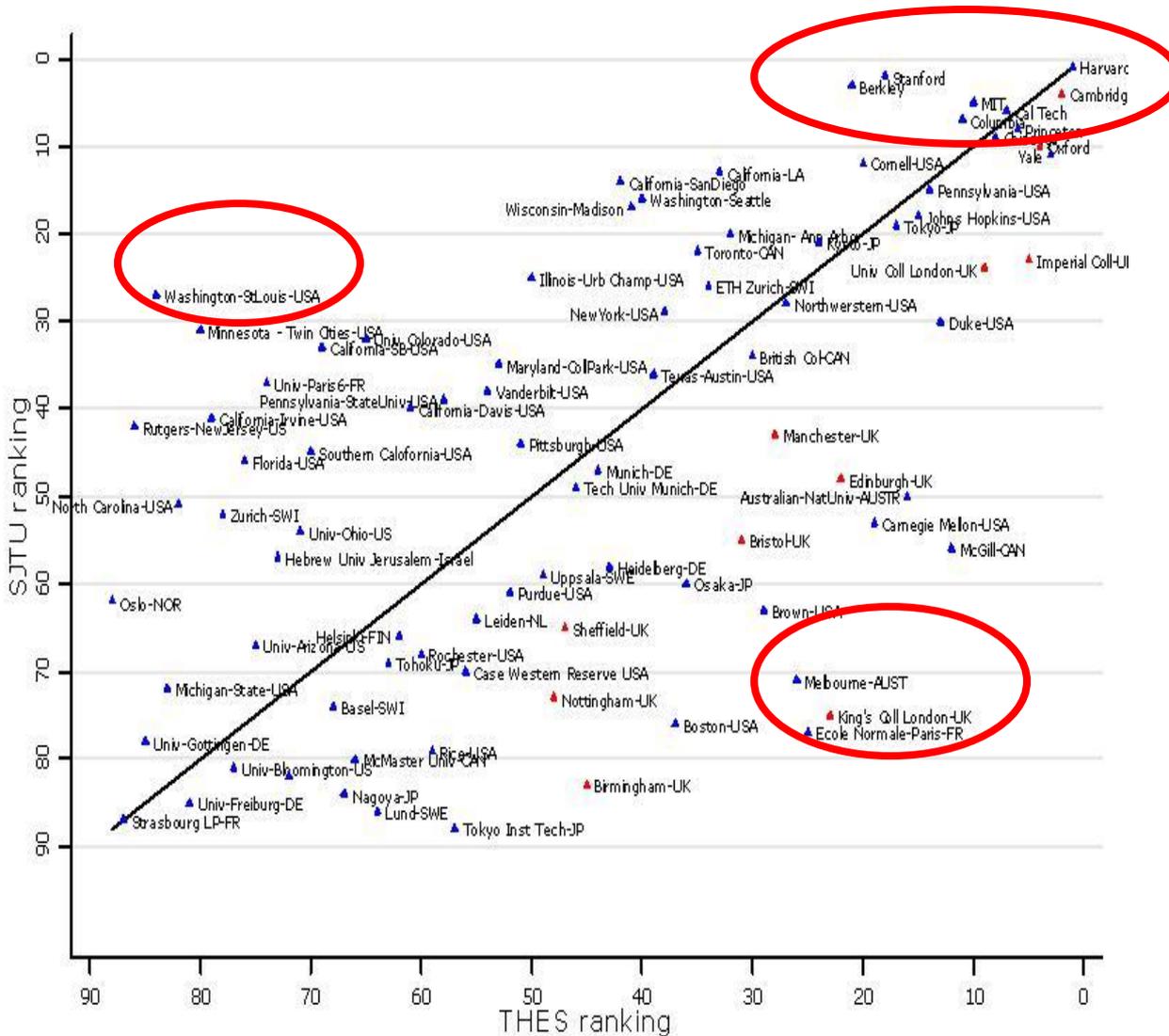
PROS and CC

- ✓ Attempt to take into account teaching quality
- ✓ Two expert-based indicators: 50% of total score (to ensure 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

Overview- Comparison (2007)



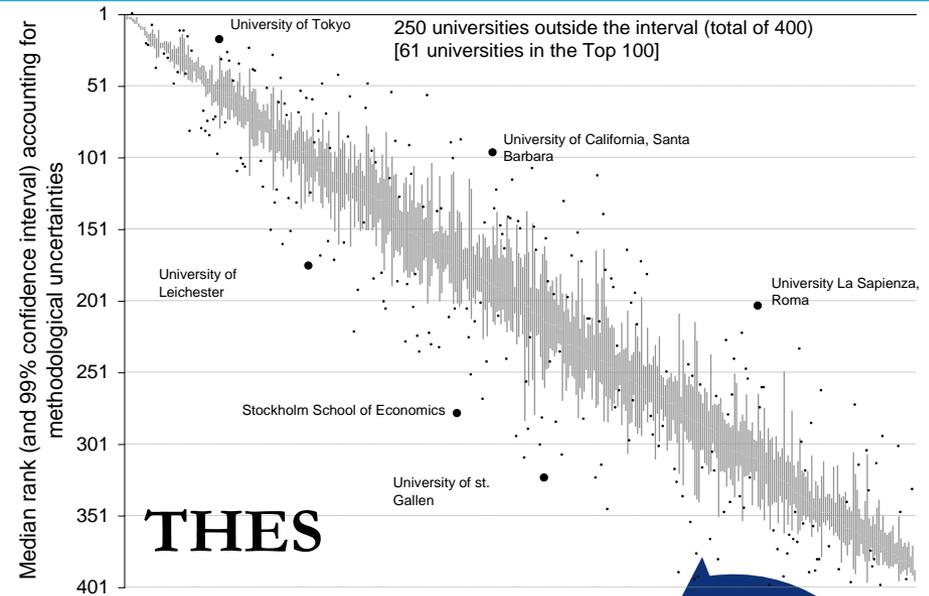
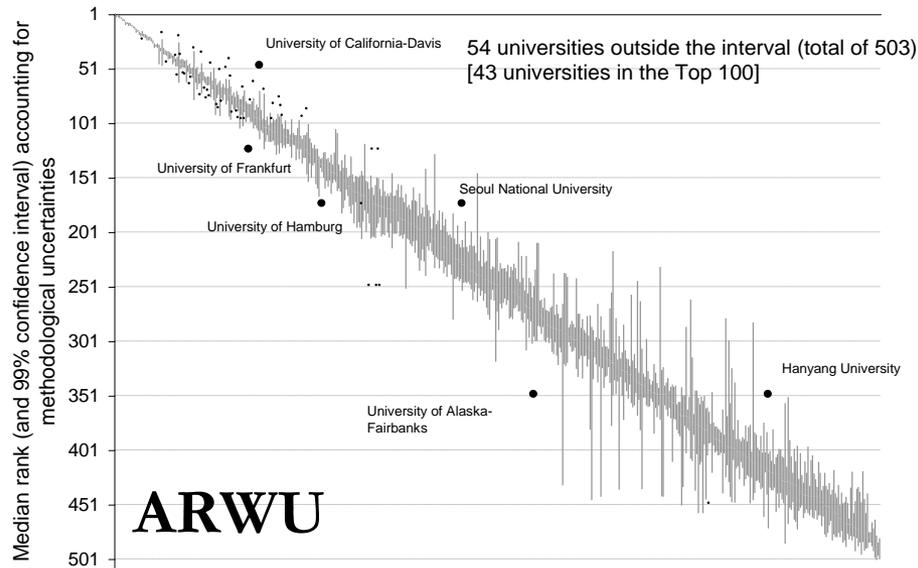
1 – Same top10: Harvard, Cambridge, Princeton, Caltech, MIT and Columbia

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)

Overview – Past results



Question:

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

Source: Saisana, D'Hombres, Saltelli, 2011, Research Policy 40, 165–177

Statistical coherence

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

$$S_i = \frac{V_{X_i} \left(E_{\mathbf{X}_{\sim i}} (Y | X_i) \right)}{V(Y)}$$

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

...

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

Statistical coherence

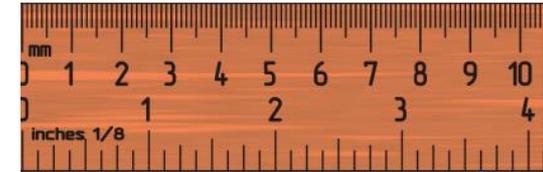
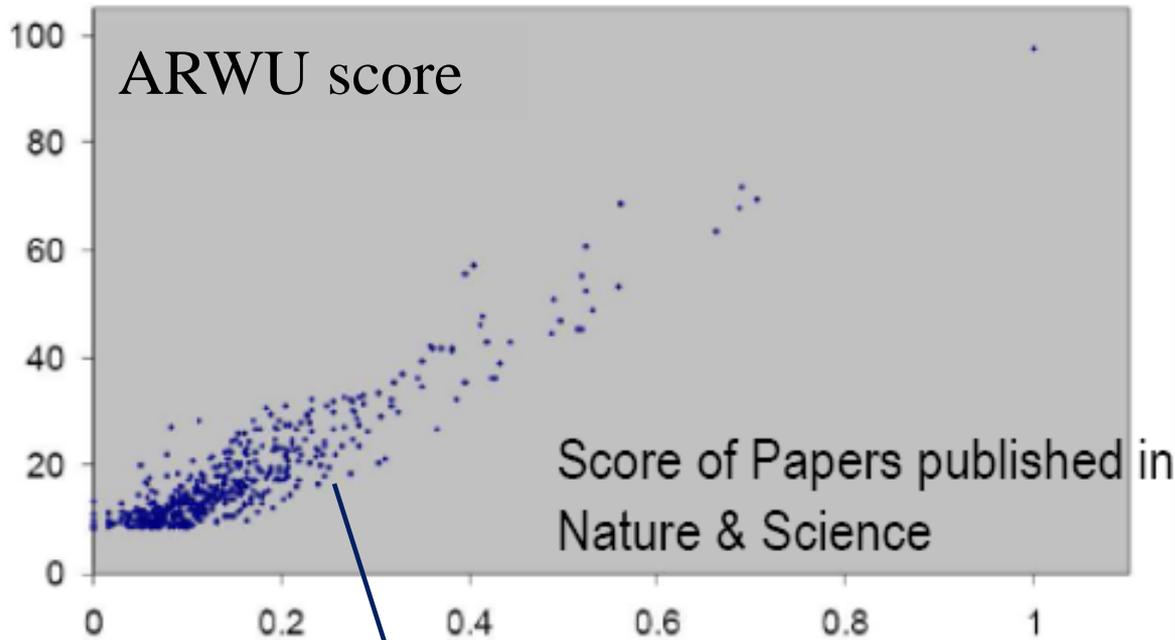
Pearson's correlation ratio

$$S_i = \frac{V_{X_i} \left(E_{\mathbf{X}_{\sim i}} (Y | X_i) \right)}{V(Y)}$$

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).

Statistical coherence



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'.

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'?

Kernel regression

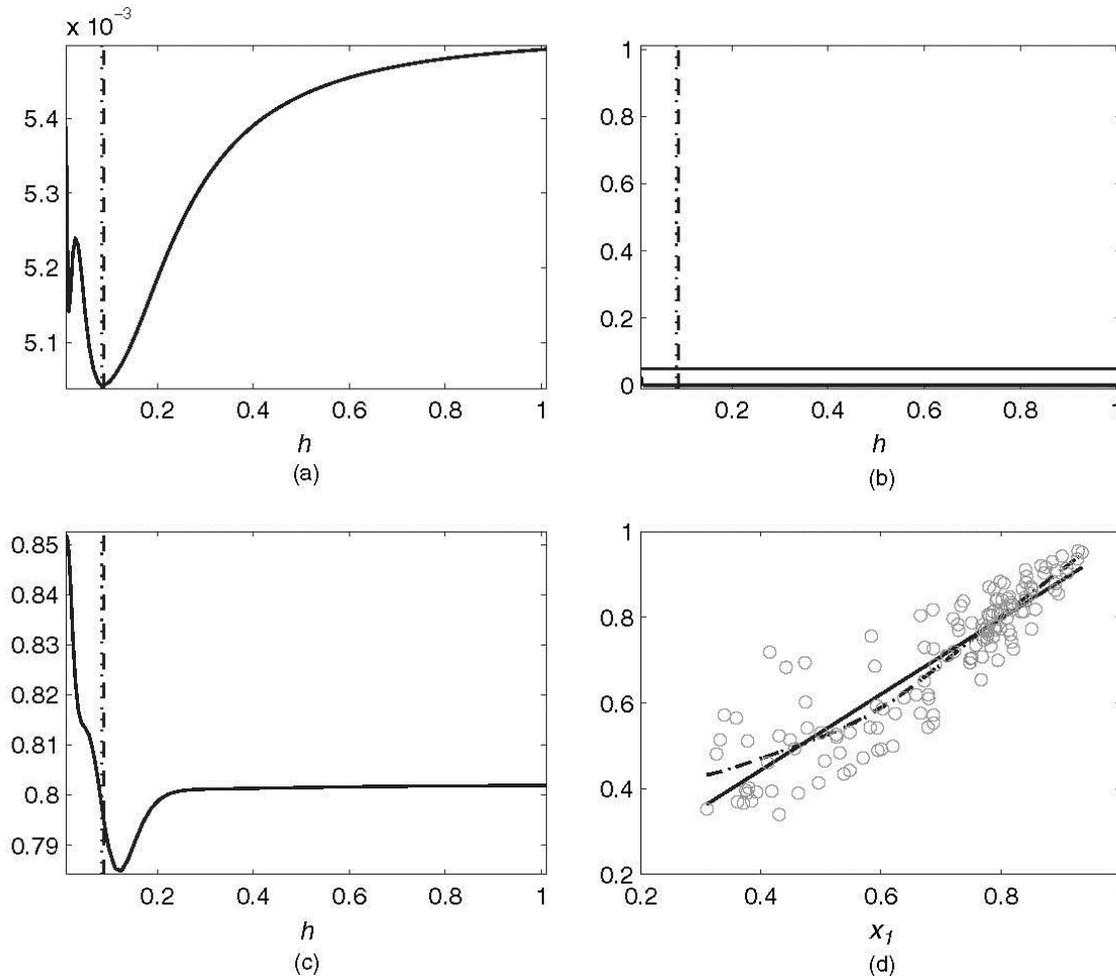


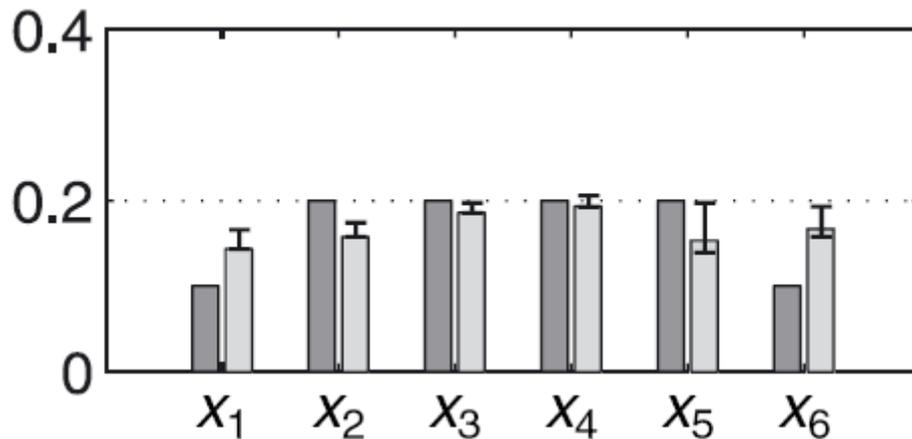
Fig. 1. 2009 HDI y and life expectancy x_1 : (a) CV criterion as a function of the smoothing parameter h (h_{DPI} ; h_{CV}); (b) linearity test p -value as a function of h (h_{DPI} ; h_{CV}); (c) main effects S_j as a function of h (h_{DPI} ; h_{CV}); (d) cross-plot of y versus x_1 with a linear fit and local linear fits for $h_{DPI} = 0.0841$ (.....) and $h_{CV} = 0.088$ (- · - ·)

Statistical coherence

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

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



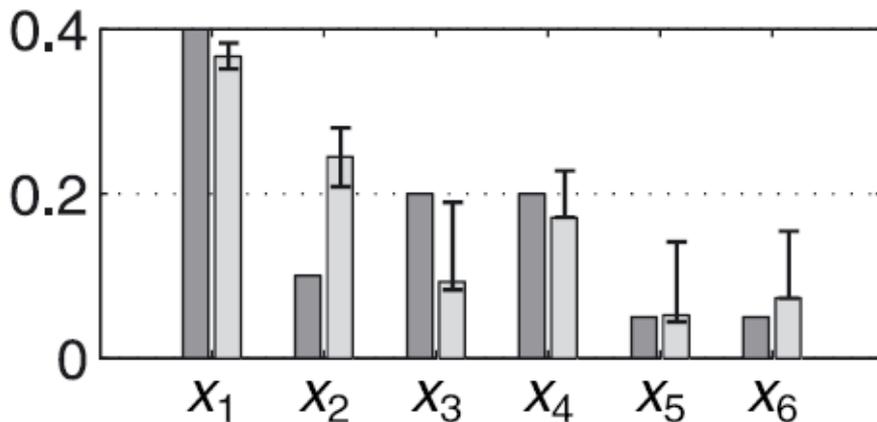
1) Hypothesis of linearity is not rejected for two indicators for the ARWU (when evaluating the tests at hDPI and hCV)

2) S_i 's are more similar to each other than the nominal weights, i.e. ranging between 0.14 and 0.19 (normalized S_i 's to unit sum; CV estimates) when weights should either be 0.10 or 0.20.

Statistical coherence - THES

	w_i	$S_{i,lin}$	$S_{i,CV}$	$S_{i,DPI}$	$S_{i,min}$	$S_{i,max}$
<i>2008 THES</i>						
Academic review	0.40	0.77	0.81	0.82	0.78	0.85
Recruiter review	0.10	0.45	0.54	0.54	0.46	0.62
Teacher/student ratio	0.20	0.19	0.21	0.20	0.18	0.42
Citations per faculty	0.20	0.38	0.38	0.41	0.38	0.50
International staff	0.05	0.10	0.12	0.12	0.10	0.31
International students	0.05	0.16	0.16	0.17	0.16	0.34

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 S_i 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

1. Paruolo P., Saisana M., Saltelli A., 2013, Ratings and Rankings: voodoo or science?. *J Royal Statistical Society A* **176(2)**.
2. Saisana M., Saltelli A., 2012, *JRC audit on the 2012 WJP Rule of Law Index*, In Agrast, M., Botero, J., Martinez, J., Ponce, A., & Pratt, C. *WJP Rule of Law Index® 2012*. Washington, D.C.: The World Justice Project.
3. Saisana M., Philippas D., 2012, *Sustainable Society Index (SSI): Taking societies' pulse along social, environmental and economic issues*, EUR 25578, Joint Research Centre, Publications Office of the European Union, Italy.
4. Saisana M., D'Hombres B., Saltelli A., 2011, Rickety Numbers: Volatility of university rankings and policy implications. *Research Policy* **40**, 165–177.
5. Saisana M., Saltelli A., Tarantola S., 2005, Uncertainty and sensitivity analysis techniques as tools for the analysis and validation of composite indicators. *J Royal Statistical Society A* **168(2)**, 307-323.
6. OECD/JRC, 2008, *Handbook on Constructing Composite Indicators. Methodology and user Guide*, OECD Publishing, ISBN 978-92-64-04345-9.