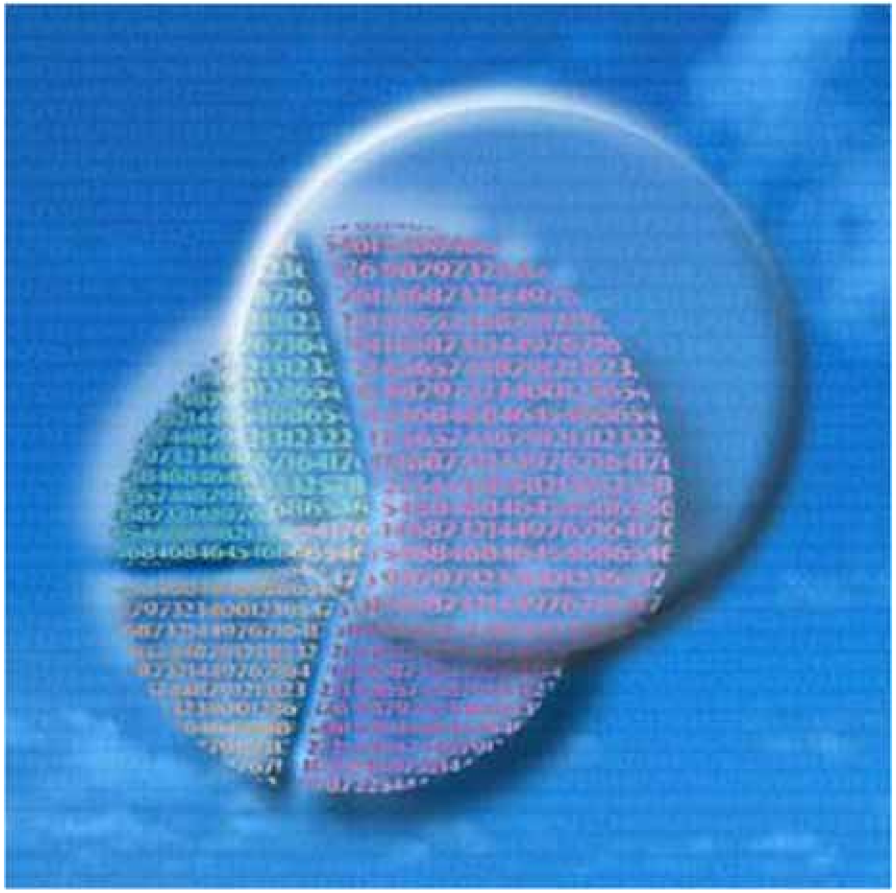
# Fallacies of rankings and ratings



SAMO 2013
University Nice Sophia Antipolis, Valrose Campus, Nice, France

Paolo Paruolo, Michaela Saisana, Andrea Saltelli, European Commission, Joint Research Centre

European Commission

andrea.saltelli@jrc.ec.europa.eu

# About indicators

# … and composite indicators

# Advocacy, analysis and quality

[…] composite indicators as an object populating a multidimensional space whose main axes are advocacy, analysis and quality […]

Saltelli, A., and Saisana, M., Advocacy, analysis and quality. The Bermuda triangle of Statistics, International Statistical Institute Conference, Hong Kong, August 2013, Statistics and Policy

# Advocacy, analysis and quality

These three dimensions (advocacy, analysis and quality) are not independent from one another.

[…]most developers adopt for transparency and simplicity linear aggregation procedures to build composite indicators which are fraught with considerable difficulties […]

In this case quality may suffer at the expenses of advocacy.

*ibidem*

European Commission

# THE ROLE OF COMPOSITE INDICATORS FOR MEASURING SOCIETAL PROGRESS

- ➔ **Ubiquitous; 5-fold increase in 6 y**
- ➔ **Statistics' best known face (to general public & media)**
- ➔ **Open the floor to plurality of norms and views**
- ➔ **Can provide analytic input to policy**

European Commission

# The Stiglitz-Sen-Fitoussi report

## Report by the Commission on the Measurement of Economic Performance and Social Progress

---

Professor Joseph E. Stiglitz, Chair, Columbia University

Professor Amartya Sen, Chair Adviser, Harvard University

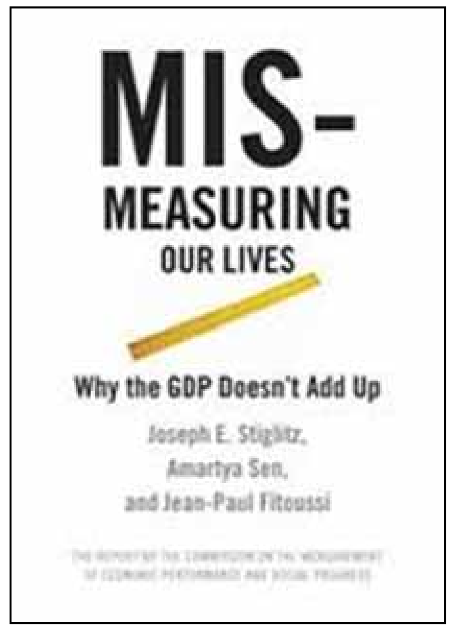
Professor Jean-Paul Fitoussi, Coordinator of the Commission, IEP

Europ
Commission

# More Statistical Indicators

**“the role of statistical indicators has increased over the last two decades”**

(Stiglitz report, 2009)




European Commission

# More Statistical Indicators

Why?

(i) more literacy,

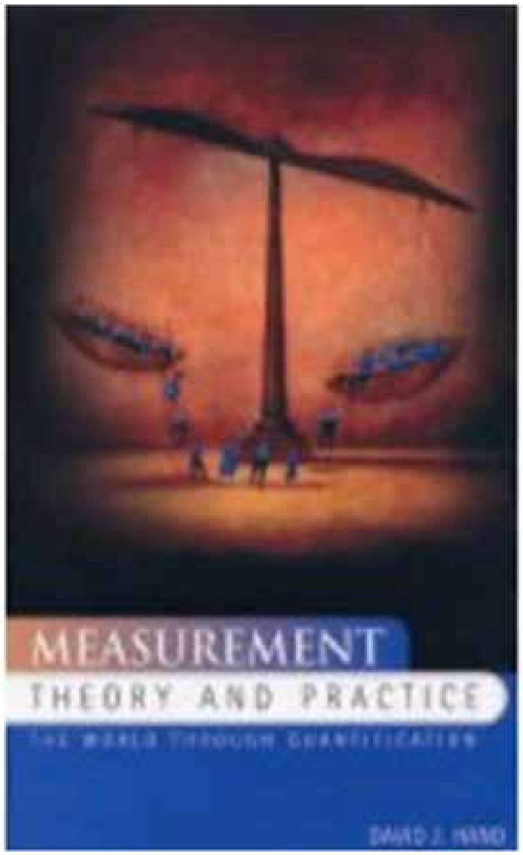
(ii) more complexity,

(iii) more information society

(Stiglitz report, 2009)

European Commission

# Caveats




"League tables [...] are an easy target for criticism.

[...] surgeon can refuse to operate on the difficult cases, schools can refuse to enter those pupils likely to do poor in examinations, health authorities can defer making appointments for some patients, so that the waiting lists look smaller, and so on."

European Commission

# Caveats

The Stiglitz report, on page 65, mentions: *[…] a general criticism that is frequently addressed at composite indicators, i.e. the arbitrary character of the procedures used to weight their various components.*

Adding: *[…] The problem is not that these weighting procedures are hidden, non-transparent or non-replicable – they are often very explicitly presented by the authors of the indices, and this is one of the strengths of this literature. The problem is rather that their normative implications are seldom made explicit or justified.*

European Commission

# Quality

European
Commission

# Quality of composite indicators

## Testing (composite) indicators: two approaches


Michaela Saisana, Andrea. Saltelli, and Stefano Tarantola, 2005, Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators.
*J. R. Statist. Soc. A* **168**(2), 307–323.

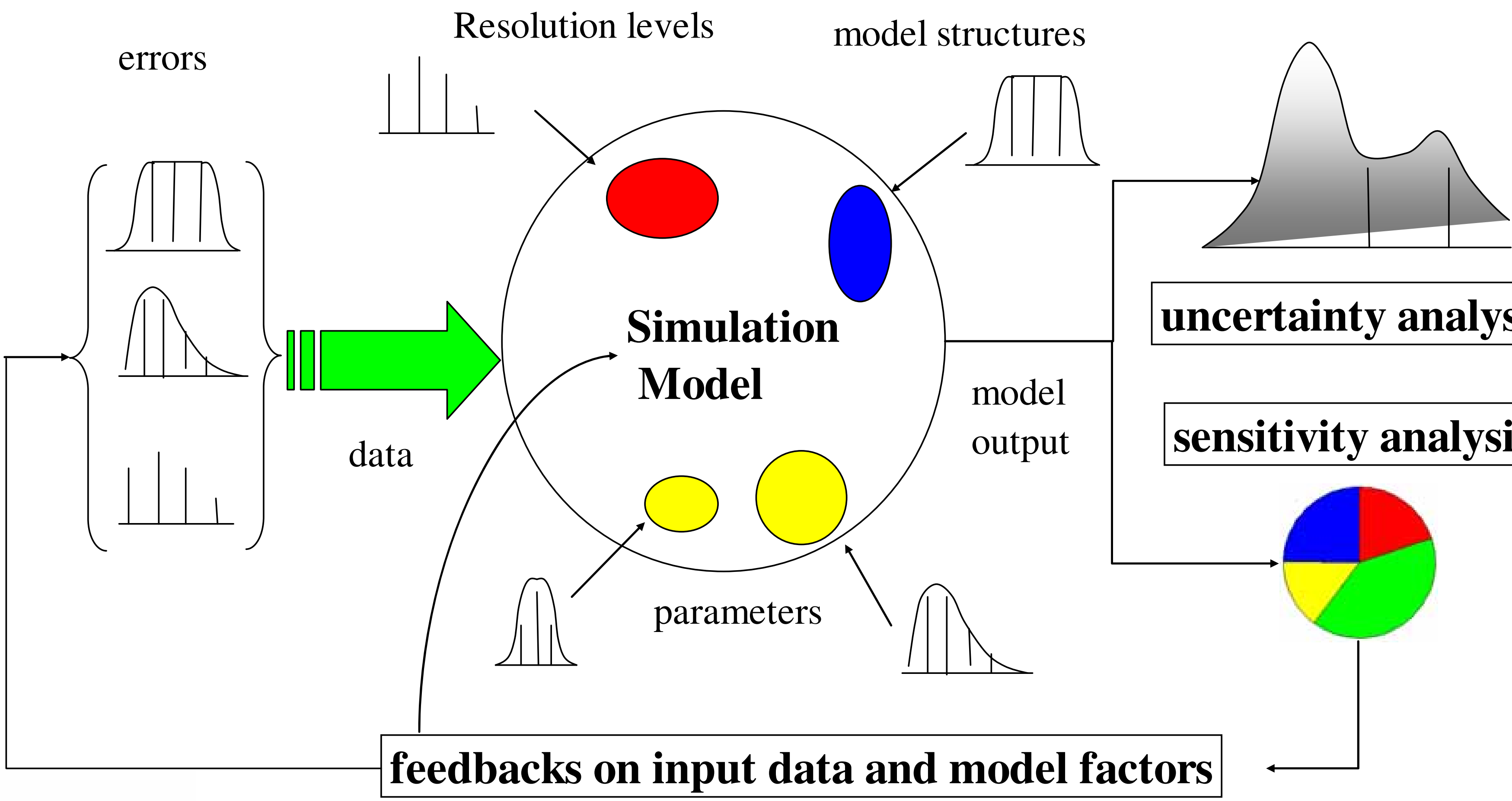
Paolo Paruolo, Michaela Saisana, Andrea Saltelli, 2013, Ratings and rankings: Voodoo or Science?,
*J. R. Statist. Soc. A*, **176** (2), 1-26

## Quality of composite indicators

# First: The invasive approach, University ranking example


Michaela Saisana, Béatrice d'Hombres, Andrea Saltelli, Rickety numbers: Volatility of university rankings and policy implications *Research Policy* (2011), **40**, 165-177

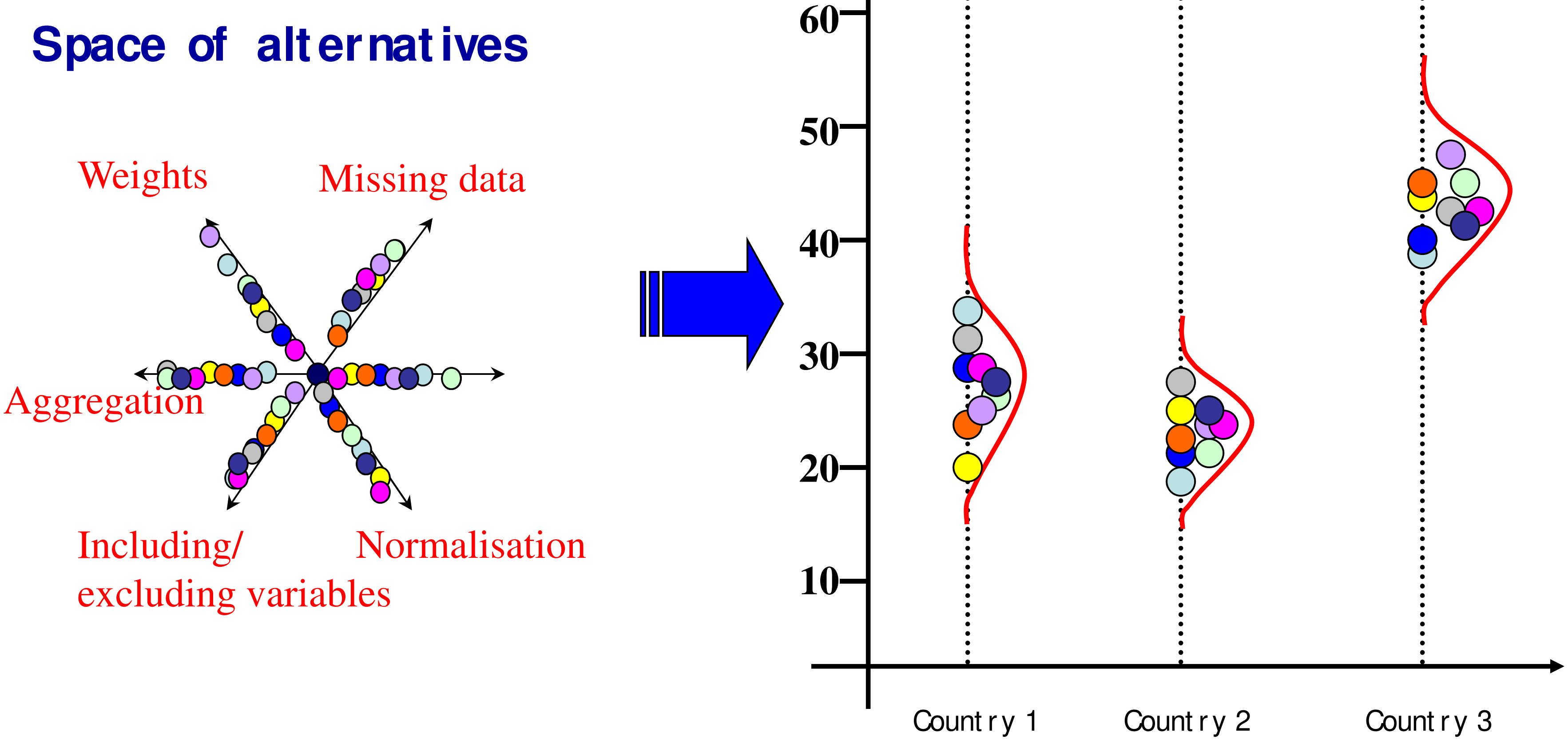
# (Invasive) Sensitivity Analysis



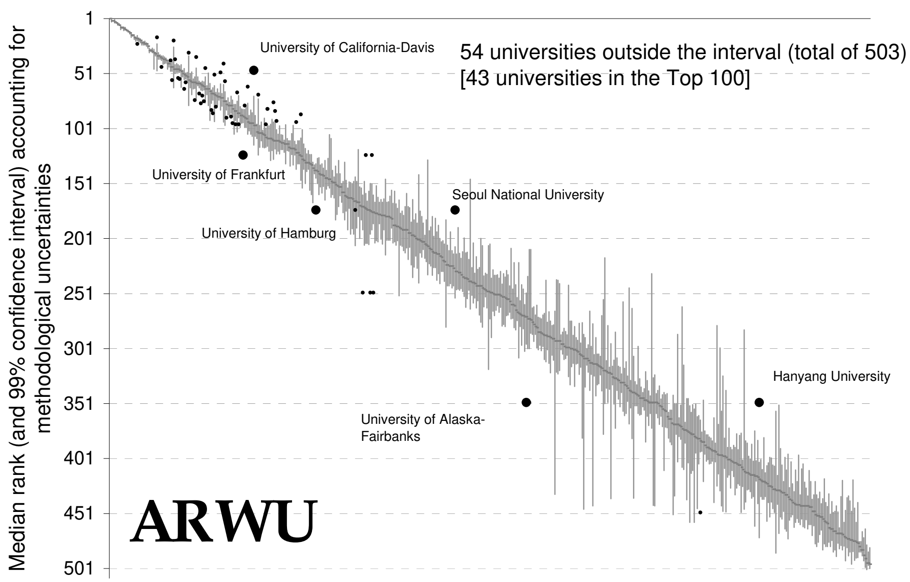
# Robustness analysis, of ARWU and THES

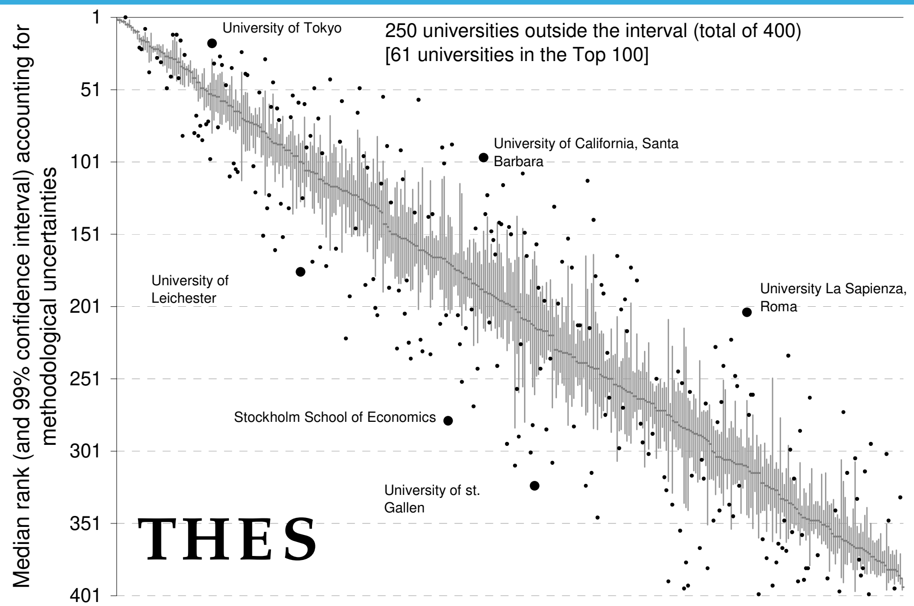
| Assumption | Alternatives |
|---|---|
| Number of indicators | ▪ all six indicators included or one-at-time excluded (6 options) |
| Weighting method | ▪ original set of weights,<br>▪ factor analysis,<br>▪ equal weighting,<br>▪ data envelopment analysis |
| Aggregation rule | ▪ additive,<br>▪ multiplicative,<br>▪ Borda multi-criterion |

# Sensitivity analysis



# Relative uncertainty of the two rankings




**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

European Commission

# ARWU: simulated ranks – Top20

*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

**Simulated rank range - SJTU 2008**

| | 1-5 | 6-10 | 11-15 | 16-20 | 21-25 | 26-30 | 31-35 | 36-40 | 41-45 | 46-50 | 51-55 | 56-60 | 61-65 | 66-70 | 71-75 | 76-80 | 81-85 | 86-90 | 91-95 | 96-100 | Original rank | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Harvard Univ | 100 | | | | | | | | | | | | | | | | | | | | 1 | USA |
| Stanford Univ | 89 | 11 | | | | | | | | | | | | | | | | | | | 2 | USA |
| Univ California - Berkeley | 97 | | | | | | | | | | | | | | | | | | | | 3 | USA |
| Univ Cambridge | 90 | 10 | | | | | | | | | | | | | | | | | | | 4 | UK |
| Massachusetts Inst Tech (MIT) | 74 | 26 | | | | | | | | | | | | | | | | | | | 5 | USA |
| California Inst Tech | 27 | 53 | 19 | | | | | | | | | | | | | | | | | | 6 | USA |
| Columbia Univ | 23 | 77 | | | | | | | | | | | | | | | | | | | 7 | USA |
| Princeton Univ | | 71 | 9 | 11 | 7 | | | | | | | | | | | | | | | | 8 | USA |
| Univ Chicago | | 51 | 34 | 13 | | | | | | | | | | | | | | | | | 9 | USA |
| Univ Oxford | | 99 | | | | | | | | | | | | | | | | | | | 10 | UK |
| Yale Univ | | 47 | 53 | | | | | | | | | | | | | | | | | | 11 | USA |
| Cornell Univ | | 27 | 73 | | | | | | | | | | | | | | | | | | 12 | USA |
| Univ California - Los Angeles | | 9 | 84 | 7 | | | | | | | | | | | | | | | | | 13 | USA |
| Univ California - San Diego | | | 41 | 46 | 9 | | | | | | | | | | | | | | | | 14 | USA |
| Univ Pennsylvania | | 6 | 71 | 23 | | | | | | | | | | | | | | | | | 15 | USA |
| Univ Washington - Seattle | | | 7 | 71 | 21 | | | | | | | | | | | | | | | | 16 | USA |
| Univ Wisconsin - Madison | | | 27 | 70 | | | | | | | | | | | | | | | | | 17 | USA |
| Univ California - San Francisco | | | 14 | 9 | 14 | | 11 | | 7 | 10 | | | | | | 6 | | | | 6 | 18 | USA |
| Tokyo Univ | | | 16 | 16 | 49 | 20 | | | | | | | | | | | | | | | 19 | Japan |
| Johns Hopkins Univ | | | 7 | 54 | 21 | 17 | | | | | | | | | | | | | | | 20 | USA |

- Harvard, Stanford, Berkley, Cambridge, MIT: top 5 in more than 75% of our simulations.
- Univ California SF: original rank 18th but could be ranked anywhere between the 6th and 100th position
- Impact of assumptions: much stronger for the middle ranked universities

# THES: simulated ranks – Top 20

*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

**Simulated rank range - THES 2008**

| | 1-5 | 6-10 | 11-15 | 16-20 | 21-25 | 26-30 | 31-35 | 36-40 | 41-45 | 46-50 | 51-55 | 56-60 | 61-65 | 66-70 | 71-75 | 76-80 | 81-85 | 86-90 | 91-95 | 96-100 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HARVARD University | 44 | 56 | | | | | | | | | | | | | | | | | | | 1 | USA |
| YALE University | 40 | 49 | 11 | | | | | | | | | | | | | | | | | | 2 | USA |
| University of CAMBRIDGE | 99 | | | | | | | | | | | | | | | | | | | | 3 | UK |
| University of OXFORD | 93 | 7 | | | | | | | | | | | | | | | | | | | 4 | UK |
| CALIFORNIA Institute of Technology | 46 | 50 | | | | | | | | | | | | | | | | | | | 5 | USA |
| IMPERIAL College London | 74 | 24 | | | | | | | | | | | | | | | | | | | 6 | UK |
| UCL (University College London) | 73 | 23 | | | | | | | | | | | | | | | | | | | 7 | UK |
| University of CHICAGO | | 80 | 19 | | | | | | | | | | | | | | | | | | 8 | USA |
| MASSACHUSETTS Institute of Technology | 14 | 13 | 17 | 16 | 11 | 11 | 7 | | | | | | | | | | | | | | 9 | USA |
| COLUMBIA University | 6 | 13 | 17 | 11 | 10 | 7 | 10 | 14 | | | | | | | | | | | | | 10 | USA |
| University of PENNSYLVANIA | | 37 | 56 | 6 | | | | | | | | | | | | | | | | | 11 | USA |
| PRINCETON University | 6 | 59 | 27 | 9 | | | | | | | | | | | | | | | | | 12 | USA |
| DUKE University | | | 27 | 11 | 9 | | 7 | | 10 | 6 | 9 | | 6 | | | | | | | | 13 | USA |
| JOHNS HOPKINS University | | | 20 | 10 | 9 | 9 | 7 | | 10 | | 6 | 6 | 7 | | | 6 | | | | | 13 | USA |
| CORNELL University | | 6 | 24 | 11 | | 7 | 6 | 7 | 9 | 9 | | 7 | | | | | | | | | 15 | USA |
| AUSTRALIAN National University | 10 | 30 | 29 | 31 | | | | | | | | | | | | | | | | | 16 | Australia |
| STANFORD University | | | 10 | 14 | 7 | 10 | 9 | 10 | 6 | 6 | | 7 | | | | | | | | | 17 | USA |
| University of MICHIGAN | | | 6 | 27 | 17 | 9 | 10 | 7 | 14 | 6 | | | | | | | | | | | 18 | USA |
| University of TOKYO | | | | 16 | 7 | 13 | 7 | | | | | | | | | | 6 | | 6 | | 19 | Japan |
| MCGILL University | | | 7 | 19 | 41 | 13 | 9 | 7 | | | | | | | | | | | | | 20 | Canada |

- 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 $10^{th}$-20th position, e.g., Duke Univ, John Hopkins Univ, Cornell Univ.

# Non invasive Sensitivity analysis

Second: The non-invasive approach

Comparing the weights as assigned by developers with 'effective weights' derived from sensitivity analysis.

And the linear aggregation paradox (weights are used as if they were importance coefficients while they are trade off coefficients)

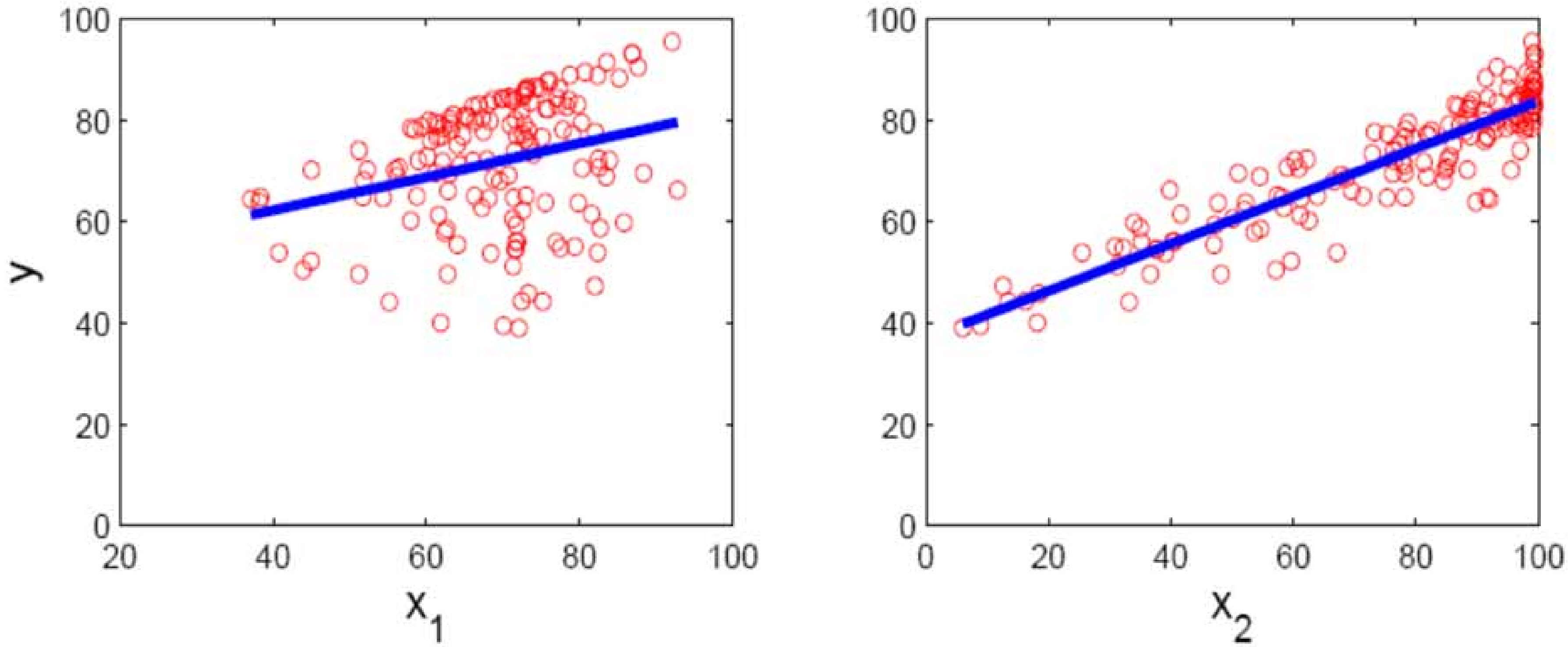
The linear aggregation paradox: weights are used as if they were importance coefficients while they are trade off coefficients

European Commission

# The linear aggregation paradox

An example. A dean wants to rank teachers based on 'hours of teaching' and 'number of publications', adding these two variables up she sees that teachers are practically ranked by publications.

European Commission




Dean's example: $y=x_1+x_2$.

Estimated $R_1^2 = 0.0759$, $R_2^2 = 0.826$,

$corr(x_1, x_2) = -0.151$, $V(x_1) = 116$, $V(x_2) = 614$, $V(y) = 162$.

$X_1$: hours of teaching $X_2$: number of publications

European Commission

# The linear aggregation paradox

To obviate this the dean substitutes the model

$y=1/2(x_1+x_2)$

with

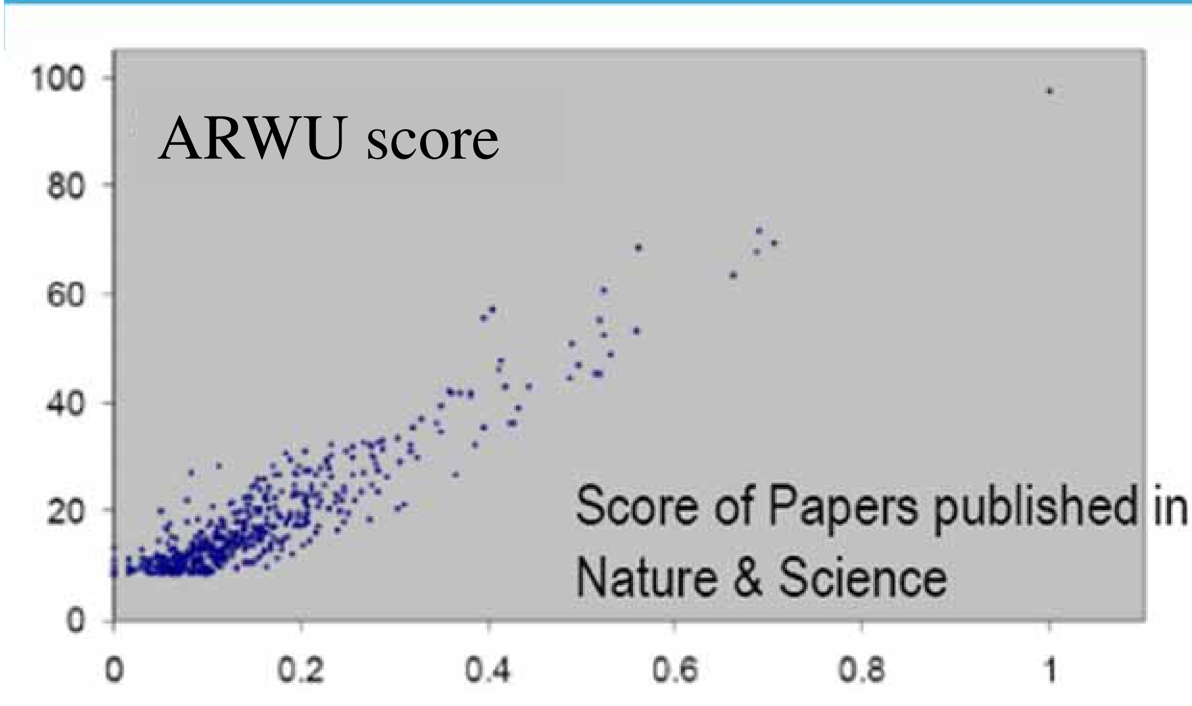
$y=0.7x_1+0.3x_2$

$X_1$: hours of teaching
$X_2$: number of publications

A professor comes by, looks at the last formula, and complains that publishing is disregarded in the department …

European Commission

# Statistical coherence

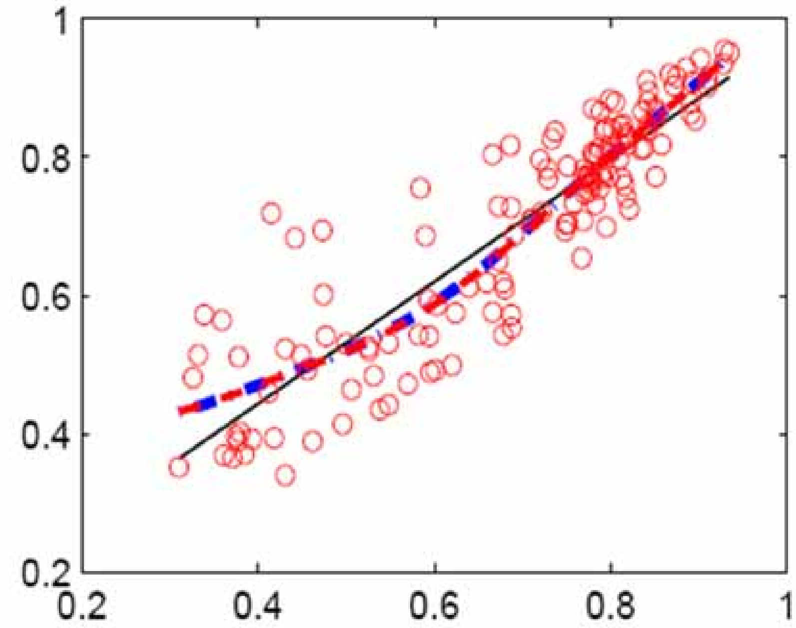



Using these points we can compute a statistics ($S_i$) that tells us:
How much (on average) would the variance of the ARWU scores be reduced if I could fix the variable 'Papers in Nature & Science'?

European Commission

index

Si [linear/ non linear] is the variance of the [linear/ non linear] interpolation curve




variable

European Commission

Pearson's correlation ratio

Smoothed curve

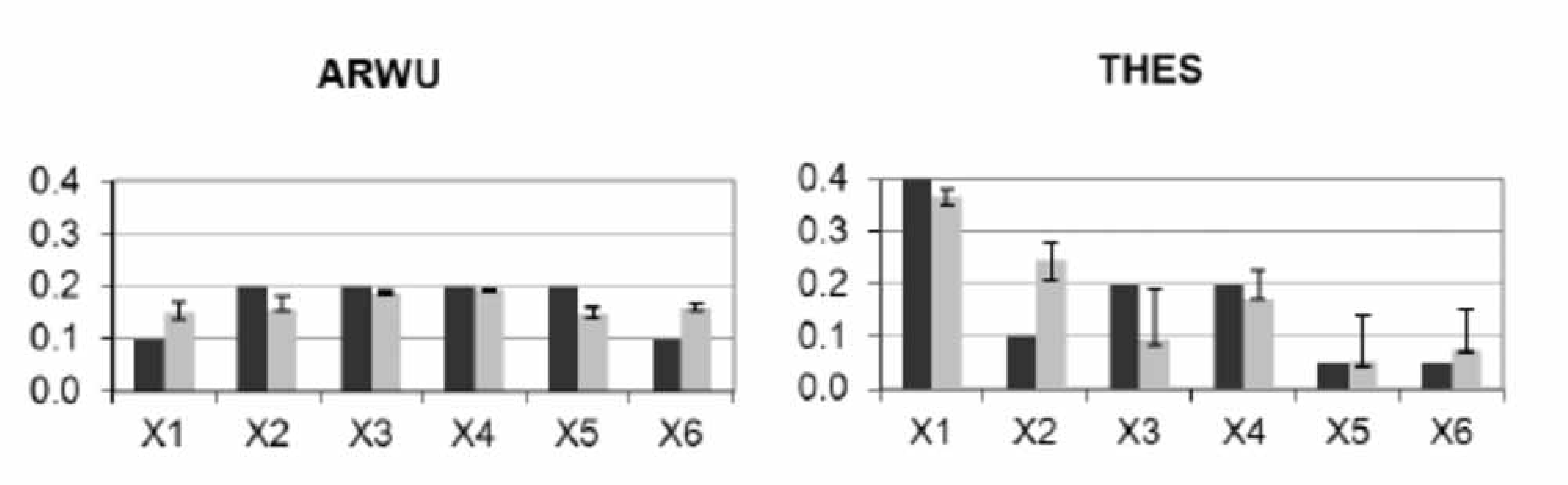
$$S_i \equiv \eta_i^2 := \frac{V_{x_i}\left(E_{\mathbf{x}_{\sim i}}\left(y \mid x_i\right)\right)}{V\left(y\right)}$$

First order sensitivity index

Unconditional variance

European Commission

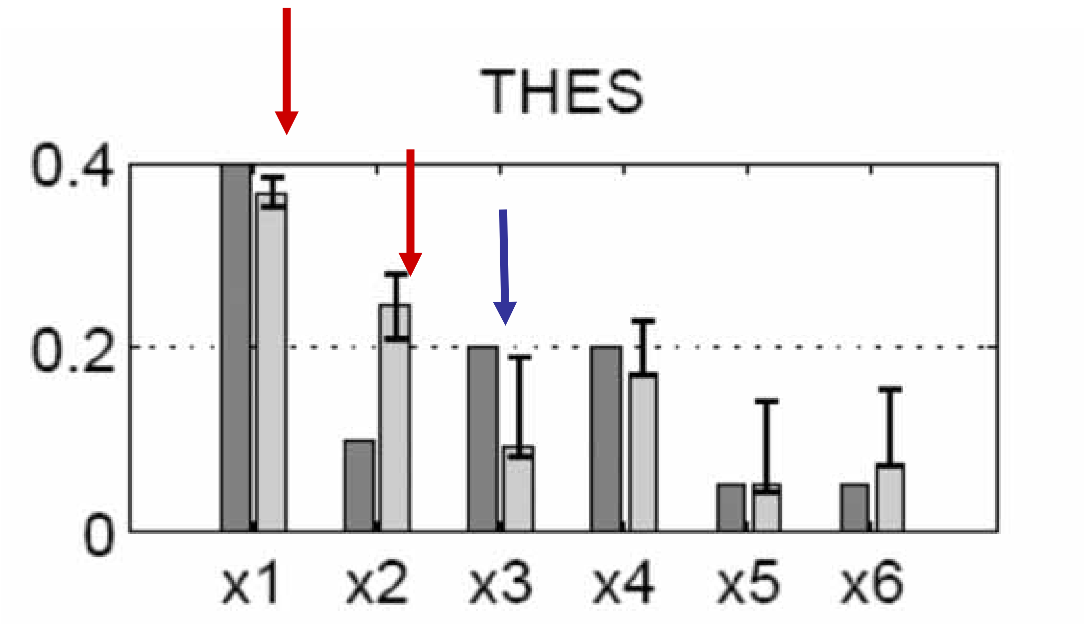
# University Rankings




Comparing the internal coherence of ARWU versus THES by testing the weights declared by developers with ‘effective’ importance measures.

# THES

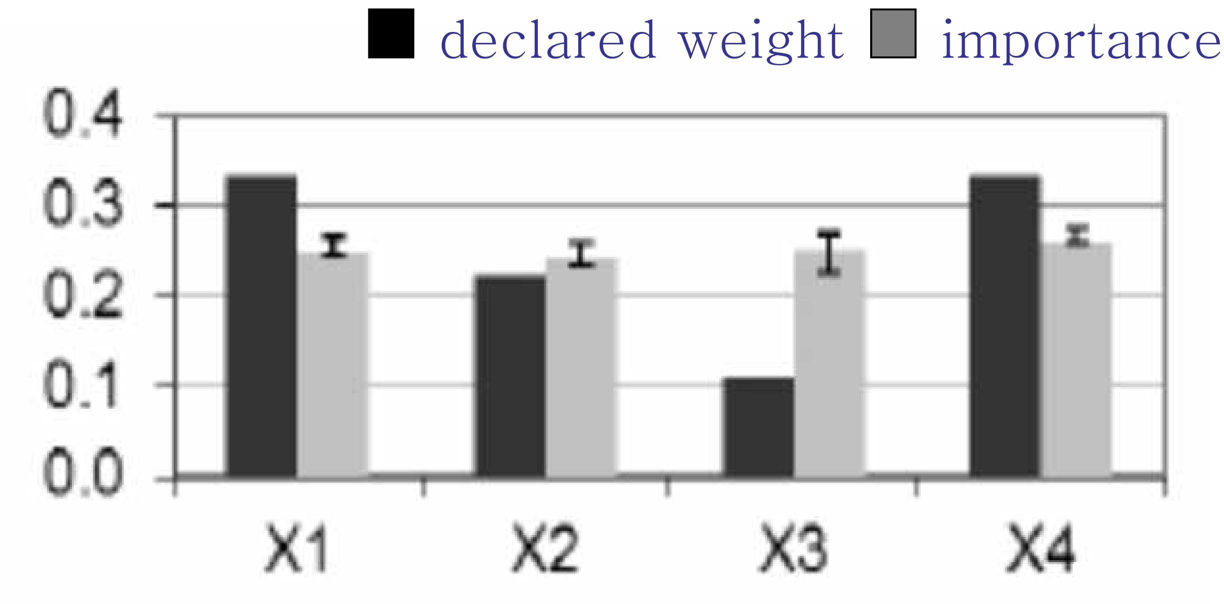
X1_Academic opinion: 6354 academics 40%
X2_Recruiters' opinion: 2339 recruiters 10%
X3_Full-time equivalent faculty/student ratio 20%
X4_Total citation/full time equivalent faculty 20%
X5_Percentage of full-time international staff 5%
X6_Percentage of full-time international students 5%




Issues with THES:
a) 'Opinion' variables' weight overall: >60% instead of 50

b) Faculty/student ratio: 10% instead of 20%

HDI 2009




Life expectancy, 33%

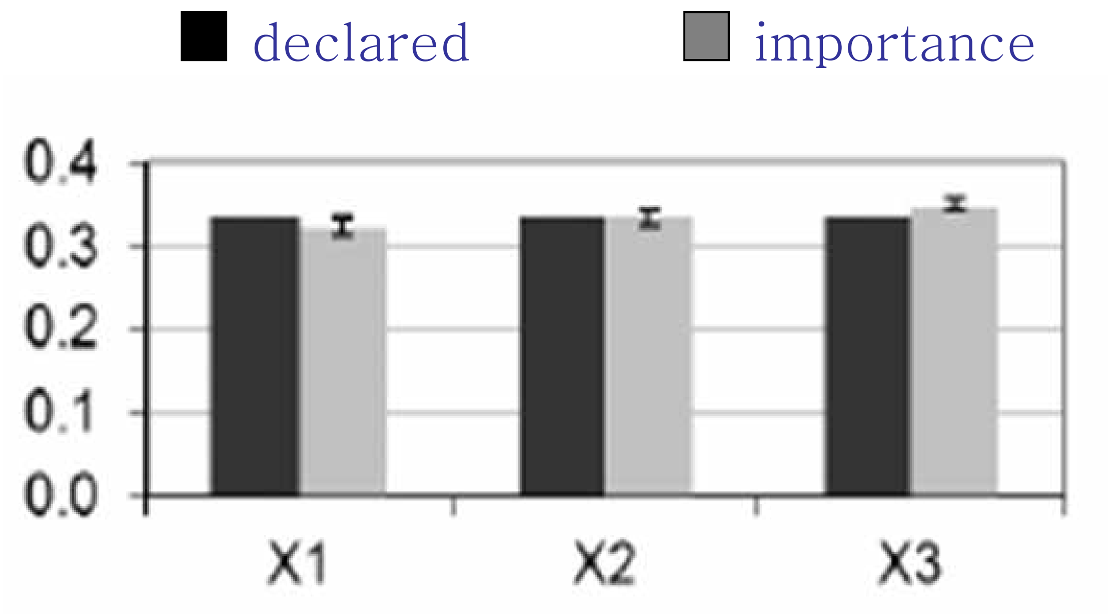
Adult literacy, 22%

Enrollment education, 11%

GDP per capita, 33%

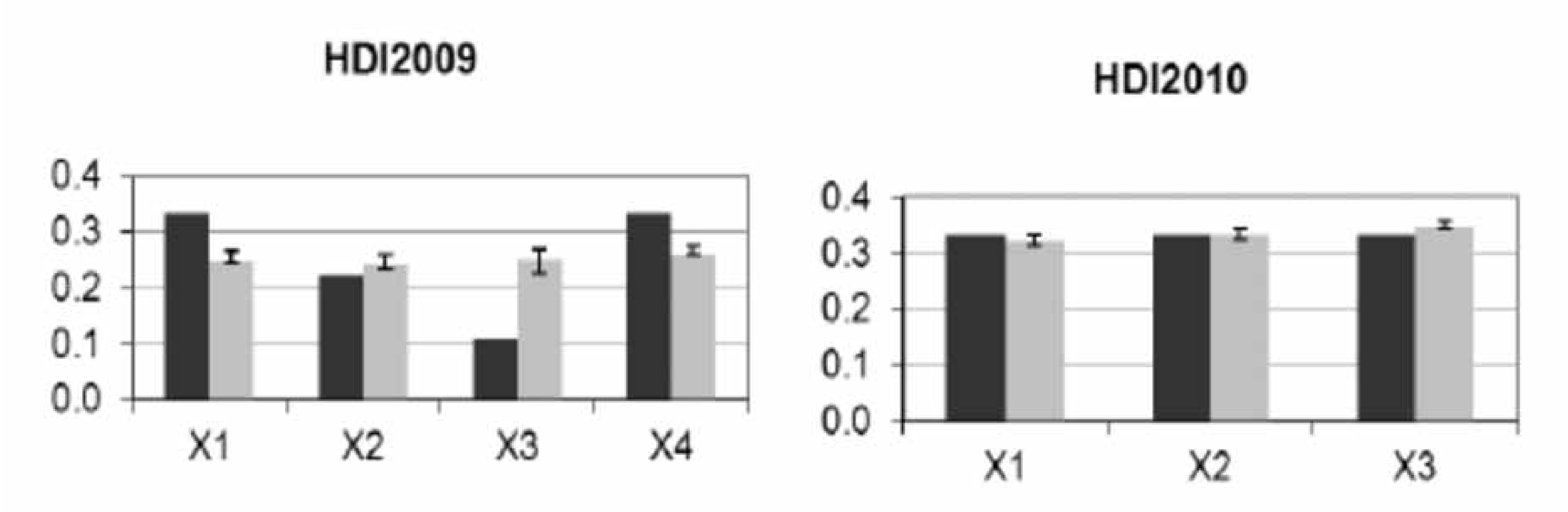
European Commission

HDI
2010




Life expectancy, 33%
Education, 33%
GNI per capita, 33%

declared weight    importance




HDI 2010 more coherent than HDI 2009

# The Sustainable Society Index (SSI-2008)

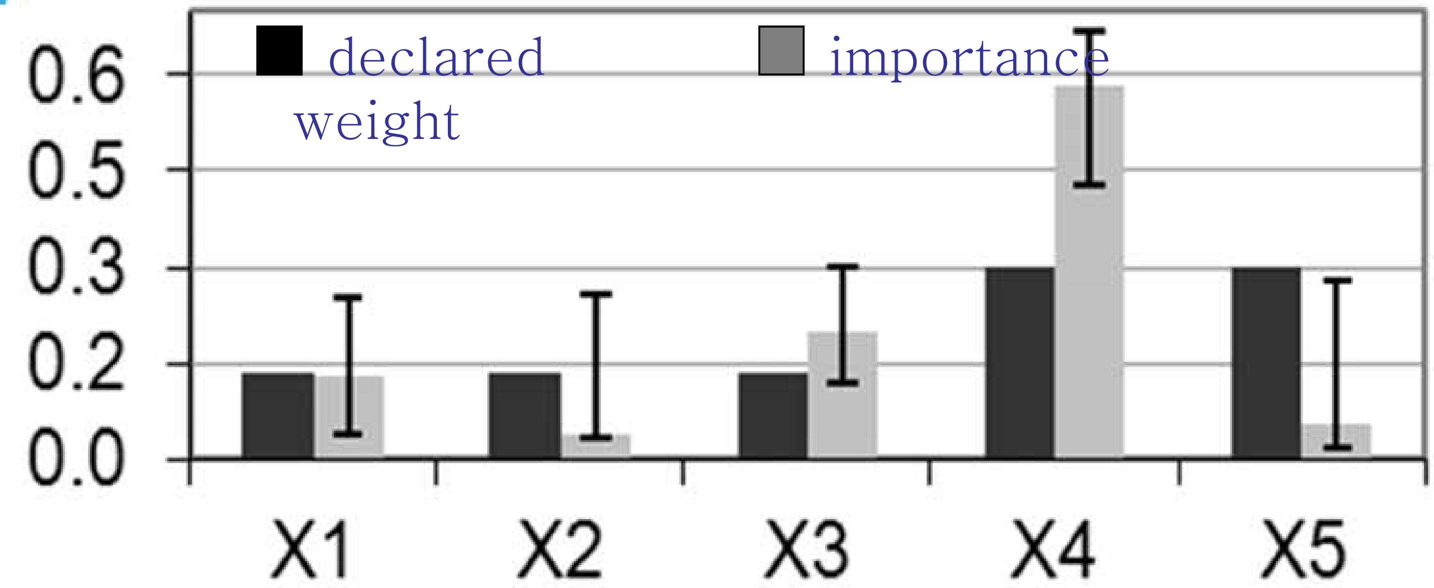
van de Kerk, G. and A. R. Manuel (2008). A comprehensive index for a sustainable society: The SSI, sustainable society index. Journal of Ecological Economics 66(2-3), 228-242.

See also

http://www.beyond-gdp.eu

European Commission

SSI 2008




Personal development, 0.13%

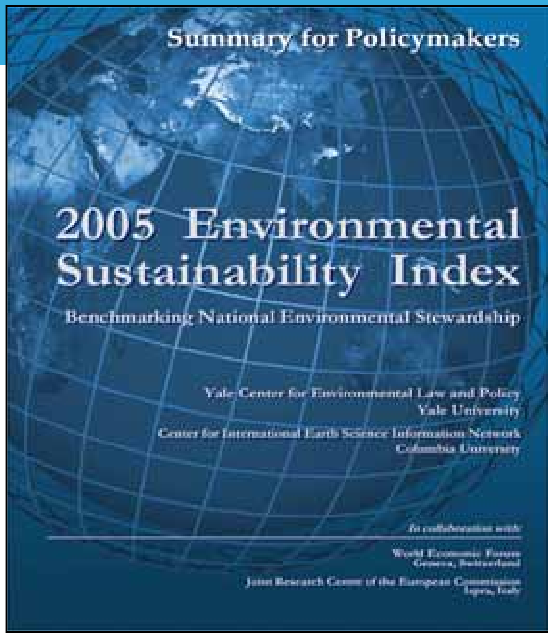
Healthy environment, 0.13%

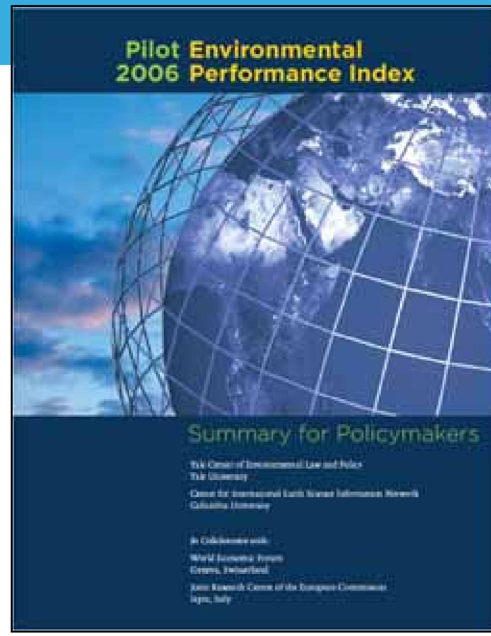
Well-balanced society, 0.13%

Sustainable use of resources, 30%

Sustainable World, 30%

European Commission





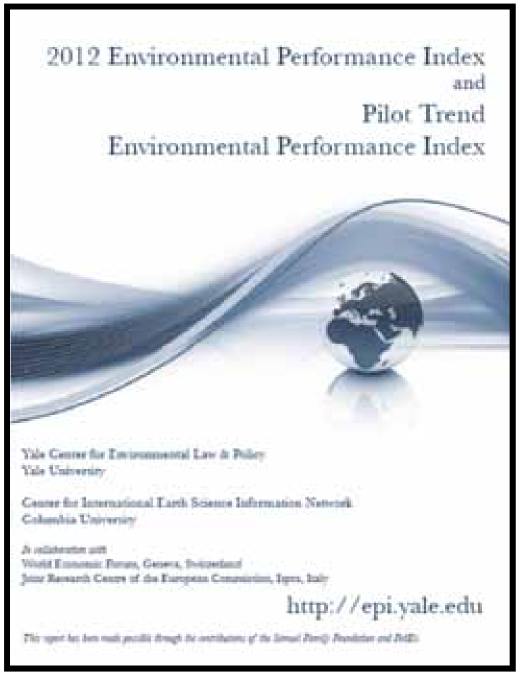


European Commission

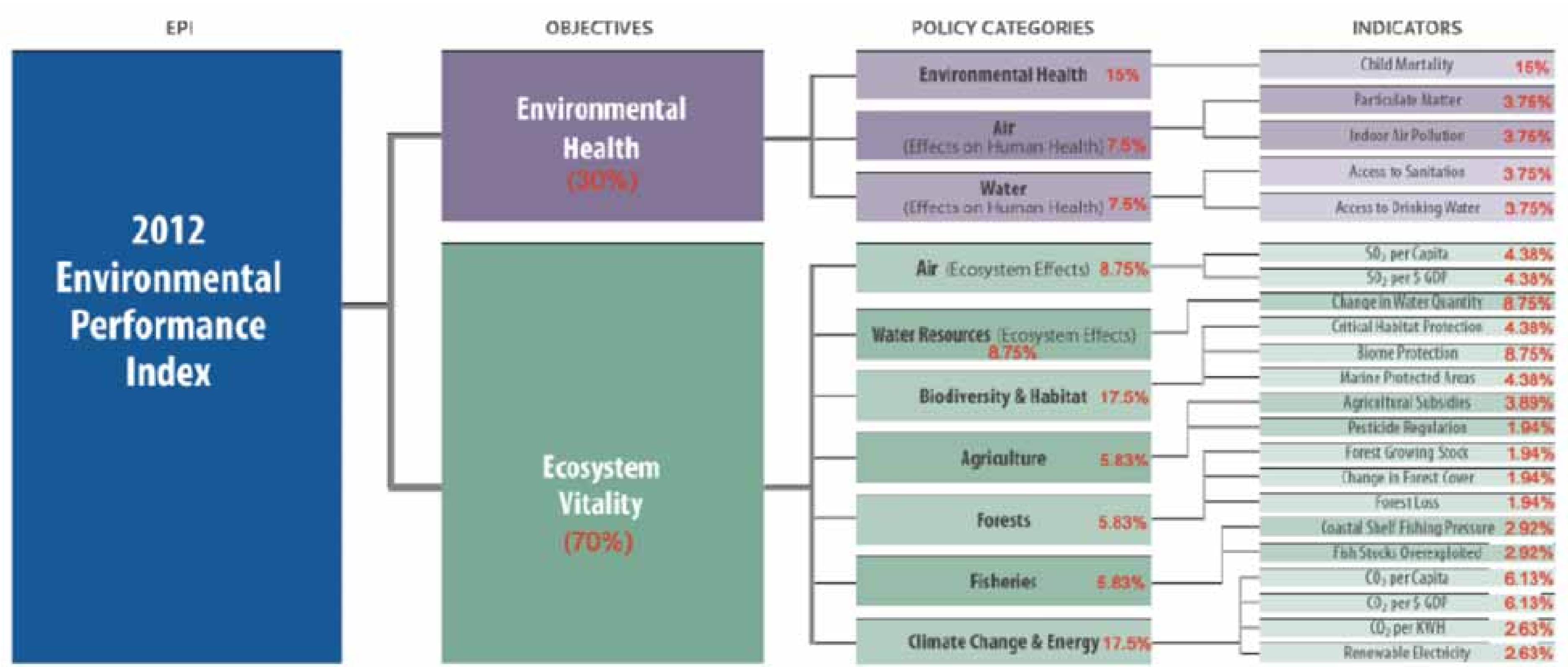
| EPI Rank | Country | Trend EPI Rank |
|---|---|---|
| 1 | Switzerland | 89 |
| 2 | Latvia | 1 |
| 3 | Norway | 84 |
| 4 | Luxembourg | 106 |
| 5 | Costa Rica | 113 |
| 6 | France | 19 |
| 7 | Austria | 71 |
| 8 | Italy | 12 |
| 9 | United Kingdom | 20 |
| 9 | Sweden | 63 |
| 11 | Germany | 56 |
| 12 | Slovakia | 7 |
| 13 | Iceland | 64 |
| 14 | New Zealand | 50 |
| 15 | Albania | 4 |
| 16 | Netherlands | 92 |
| 17 | Lithuania | 104 |
| 18 | Czech Republic | 25 |
| 19 | Finland | 54 |
| 20 | Croatia | 74 |

2012 Environmental Performance Index (EPI)

- Developed for 132 countries
- Based on 22 indicators grouped in
- Ten Policy Categories and Two Objectives



# EPI 2012 Framework




Weights for the two objectives in EPI 2012: 30-70

But in EPI 2010 they were 50-50

European Commission


## Appendix II. Preliminary Sensitivity Analysis

Michaela Saisana & Andrea Saltelli
European Commission – Joint Research Centre – IPSC, ITALY

The JRC analysis focused on:

1. Conceptual & statistical coherence in the EPI framework

2. Impact on EPI ranks of modeling assumptions (e.g. change of weights, aggregation formula)

3. Most sensitive (..to be read as least reliable) country ranks

European Commission

| EPI component | Importance measures for EPI | | Weights within EPI | Importance measures for the two EPI | | Weights within |
|---|---|---|---|---|---|---|
| | $S_i$ non linear[1] | $S_i$ linear[2] | | [illegible] | | |
| Environmental Health | 0.231 (0.057) | 0.329 | 30% | | | |
| Ecosystem Vitality | 0.489 (0.076) | 0.415 | 70% | | | |
| | | | | **Environmental Health** | | |
| Air Pollution (health) | 0.165 (0.092) | 0.267 | 8% | 0.455 (0.100) | 0.661 | 25% |
| Water & Sanitation (health) | 0.279 (0.122) | 0.289 | 8% | 0.925 (0.045) | 0.886 | 25% |
| Child Mortality | 0.415 (0.078) | 0.300 | 15% | 0.938 (0.022) | 0.918 | 50% |
| | | | 9% | **Ecosystem Vitality 2010** | | |
| Air pollution (ecosystem) | 0.108 (0.051) | 0.135 | 9% | [illegible] | [illegible] | [illegible] |
| Water (ecosystem) | 0.074 (0.059) | 0.166 | 18% | [illegible] | [illegible] | [illegible] |
| Biodiversity & Habitat | 0.438 (0.080) | 0.448 | 6% | [illegible] | [illegible] | [illegible] |
| Forestry | 0.121 (0.063) | 0.000 | 6% | [illegible] | [illegible] | [illegible] |
| Marine & Fisheries | 0.041 (0.032) | 0.015 | 6% | [illegible] | [illegible] | [illegible] |
| Agriculture | 0.166 (0.067) | 0.005 | 18% | [illegible] | [illegible] | [illegible] |
| Climate change | 0.116 (0.042) | 0.008 | 8% | 0.461 (0.081) | 0.446 | 25% |

**Table 1.** Importance measures for the EPI 2012 components *Source: European Commission Joint Research Centre*

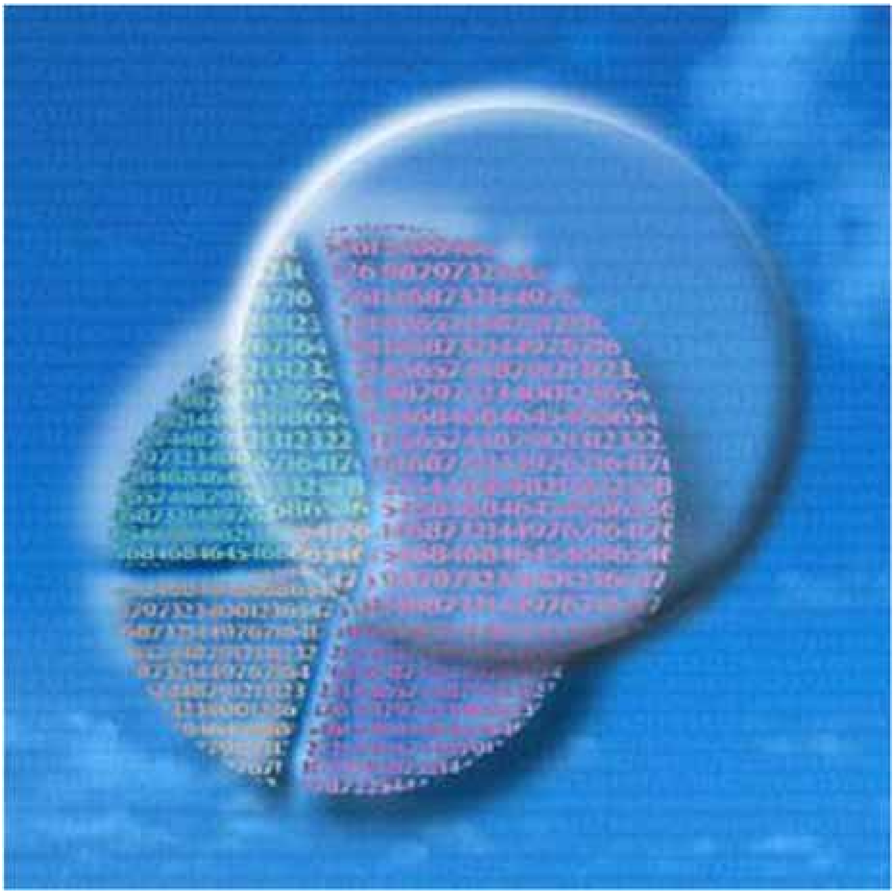
Notes: (1) Numbers represent the average kernel estimates of the Pearson correlation ratio ($\eta^2$) calculated by bootstrap (1000 samples). (2) Numbers represent the Pearson correlation coefficient (squared). (3) Bootstrap standard deviations for the correlation ratio are given in parenthesis. (4) Results are based on the data reported for 2010.

EPI relatively balanced in the two objectives

But
Forestry and Marine are "silent" indicators

European Commissi

# END



European
Commission