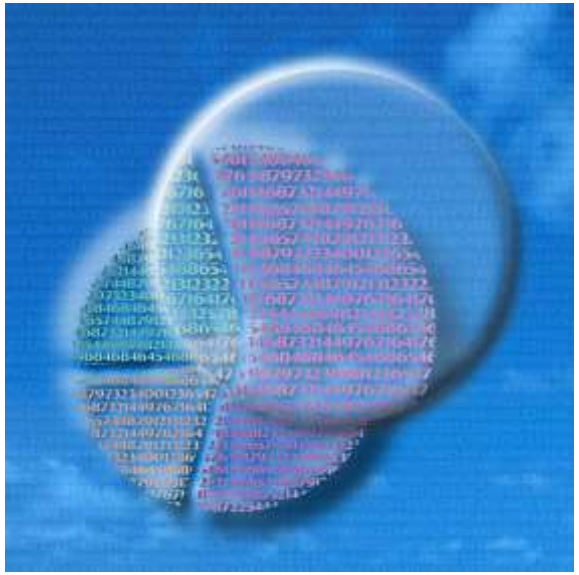


The indiscreet charm of composite indicators

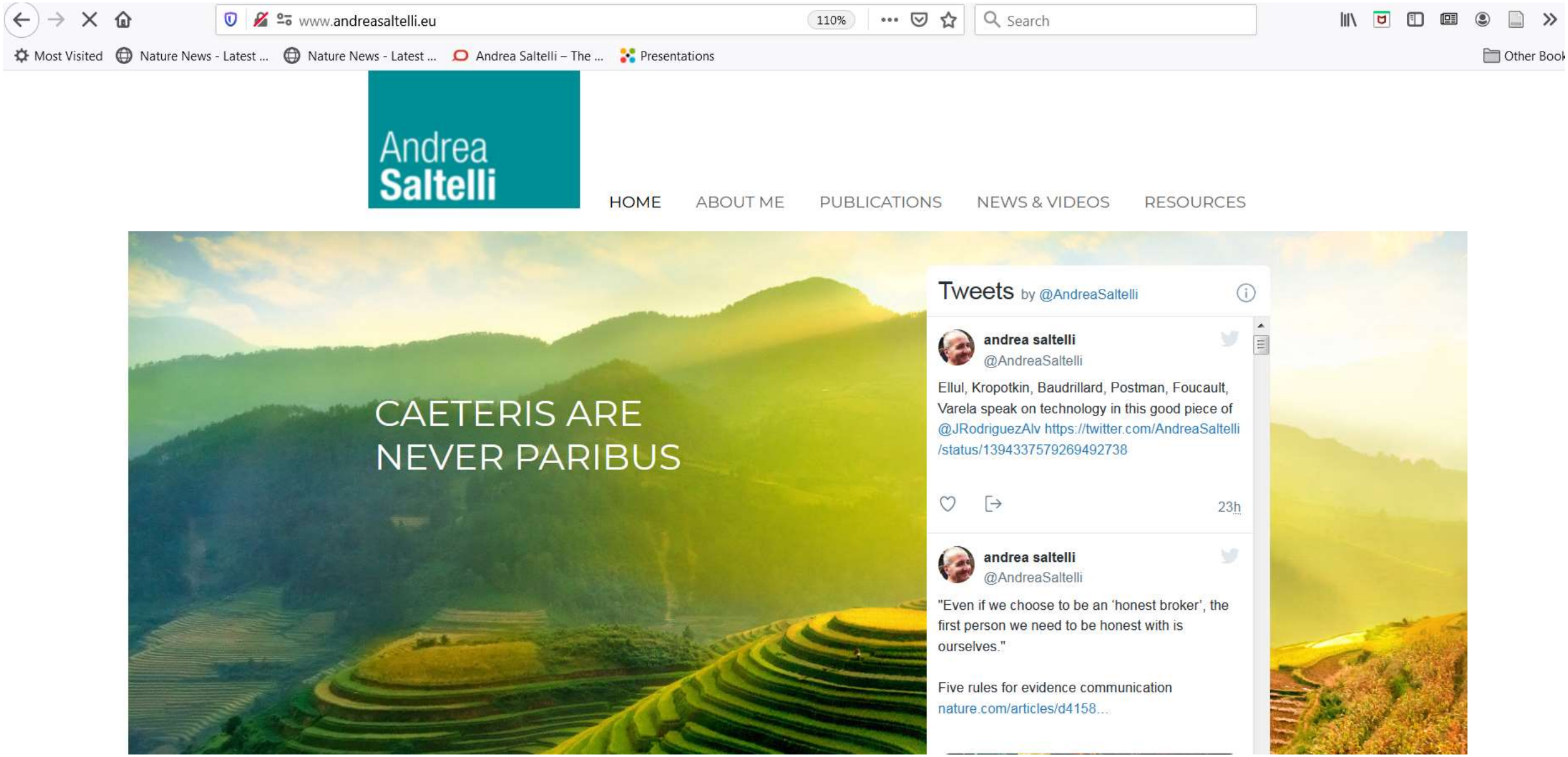
andrea.saltelli@gmail.com

Open Evidence Research, Universitat Oberta de Catalunya (UOC), Barcelona

University of Bergamo
Dipartimento di Scienze Economiche

May 19, 2021

www.andreasaltelli.eu



Composite indicators: What are they?

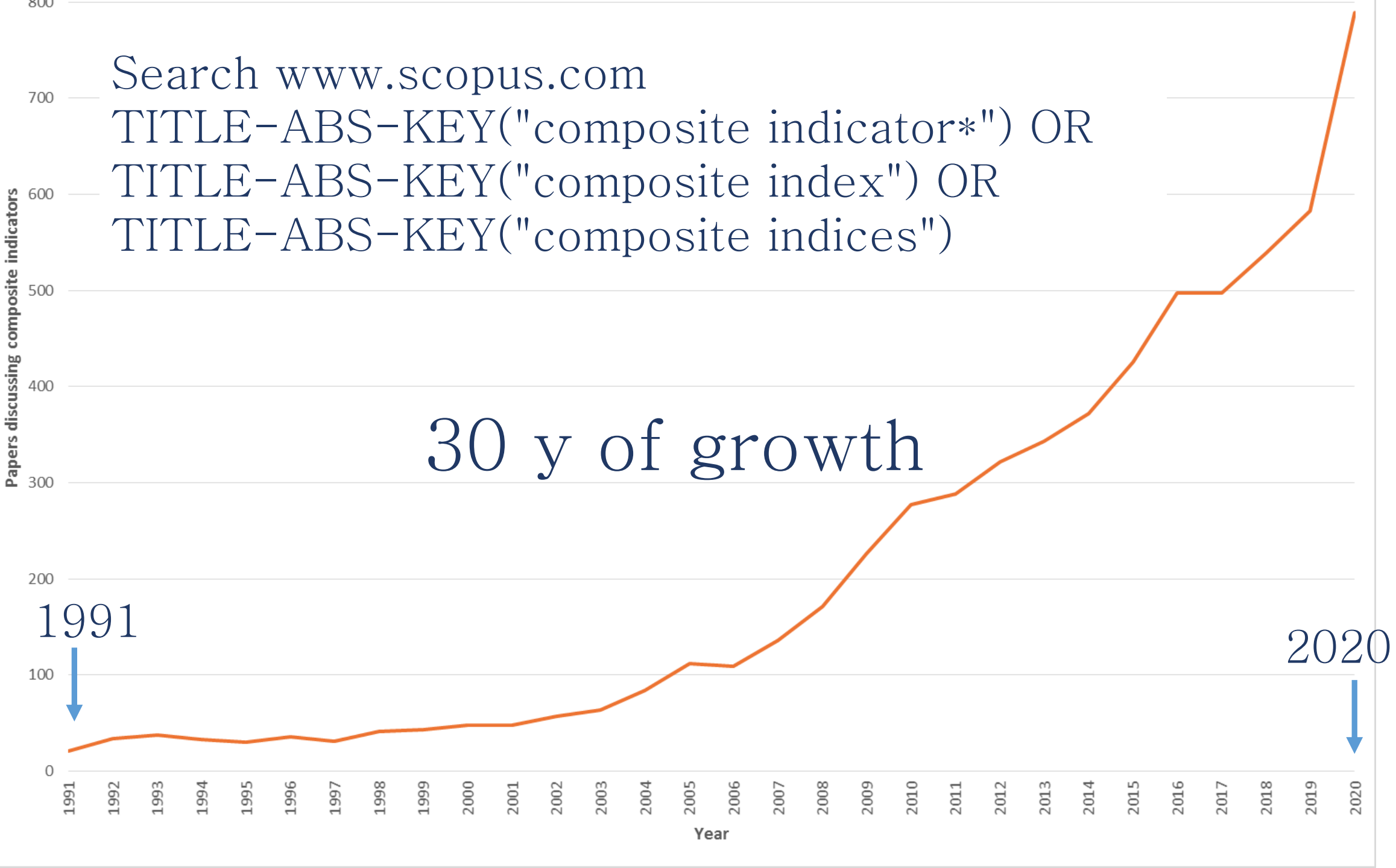
Eight factors further disaggregated into 44 sub-factors

1.1 Government powers are effectively limited by the legislature
1.2 Government powers are effectively limited by the judiciary
1.3 Government powers are effectively limited by independent auditing and review
1.4 Government officials are sanctioned for misconduct
1.5 Government powers are subject to non-governmental checks
1.6 Transition of power is subject to the law

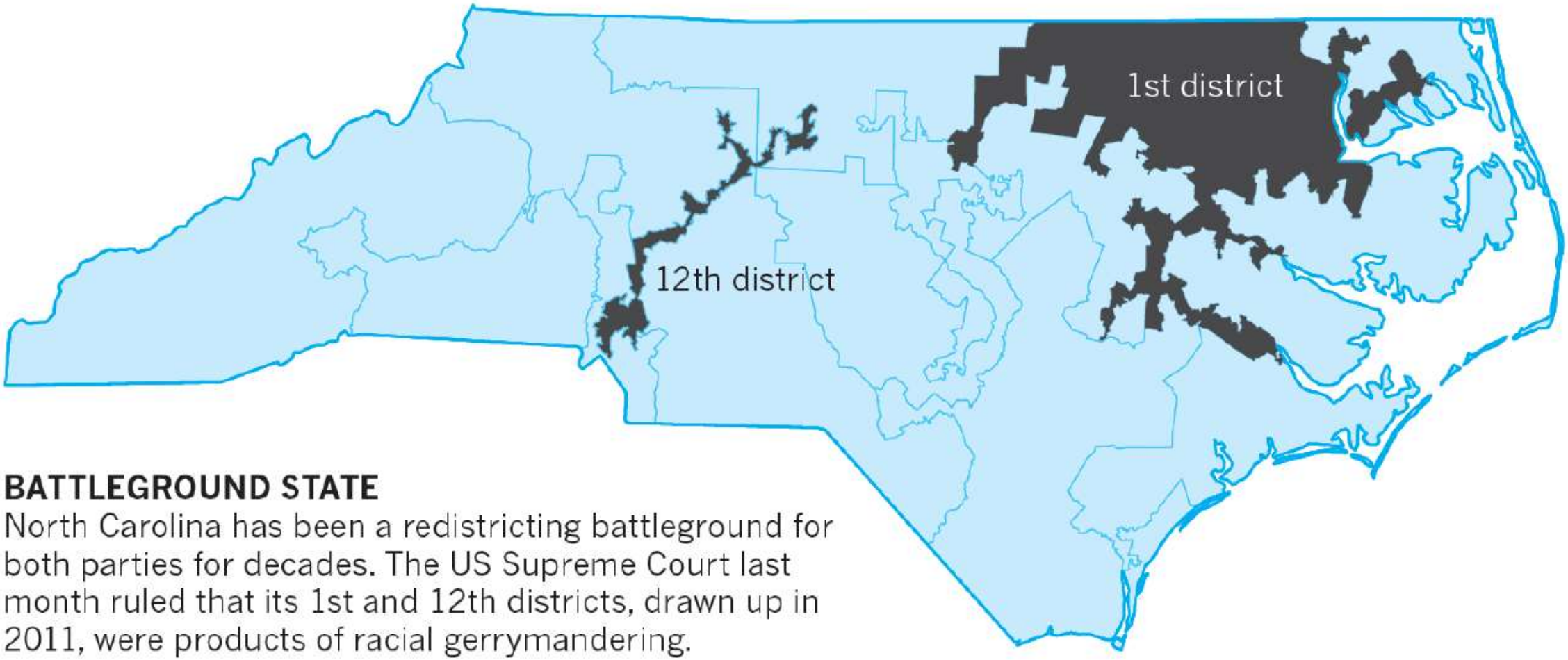
One of the eight factors with its 6 sub factors …

https://worldjusticeproject.org/sites/default/files/documents/WJP-ROLI-2019-Single%20Page%20View-Reduced_0.pdf

Ubiquity of composite indicators

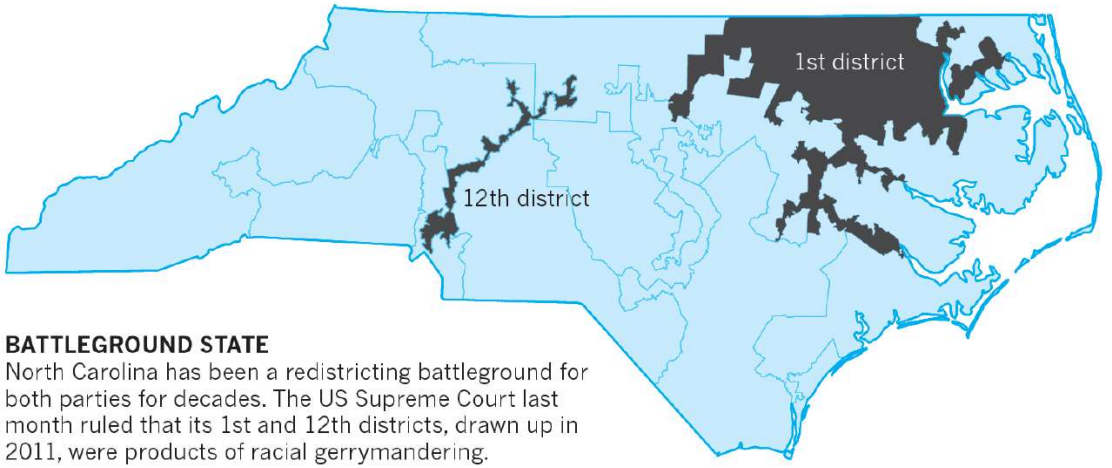

Search www.scopus.com
TITLE-ABS-KEY("composite indicator*") OR
TITLE-ABS-KEY("composite index") OR
TITLE-ABS-KEY("composite indices")
30 y of growth
1991
2020
Papers discussing composite indicators
Year
0
100
200
300
400
500
600
700
800
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020

Making the case for gerrymandering?



BATTLEGROUND STATE
North Carolina has been a redistricting battleground for both parties for decades. The US Supreme Court last month ruled that its 1st and 12th districts, drawn up in 2011, were products of racial gerrymandering.

Nature June 2017 article on the mathematics of ‘nailing’ gerrymandering



BATTLEGROUND STATE
North Carolina has been a redistricting battleground for both parties for decades. The US Supreme Court last month ruled that its 1st and 12th districts, drawn up in 2011, were products of racial gerrymandering.

“[US] ranked 55th of 158 nations — last among Western democracies — in a 2017 index of voting fairness (Electoral Integrity Project)”

Carrie Arnold, 2017, The mathematicians who want to save democracy, 200, NATURE, VOL 546, 8 JUNE 2017.

Quality of composite indicators

Specific elements of quality for composite indicators

OECD JRC

RELEVANCE

In the context of composite indicators, relevance has to be evaluated considering the degree to which it meet current and potential needs of the users

[…] ensure that the right range of domains is covered in a balanced way

ACCURACY

The credibility of data products refers to confidence that users place in … the image of the data producer, i.e., the brand image …

[crucial] that the data are perceived to be produced professionally and that practices are transparent

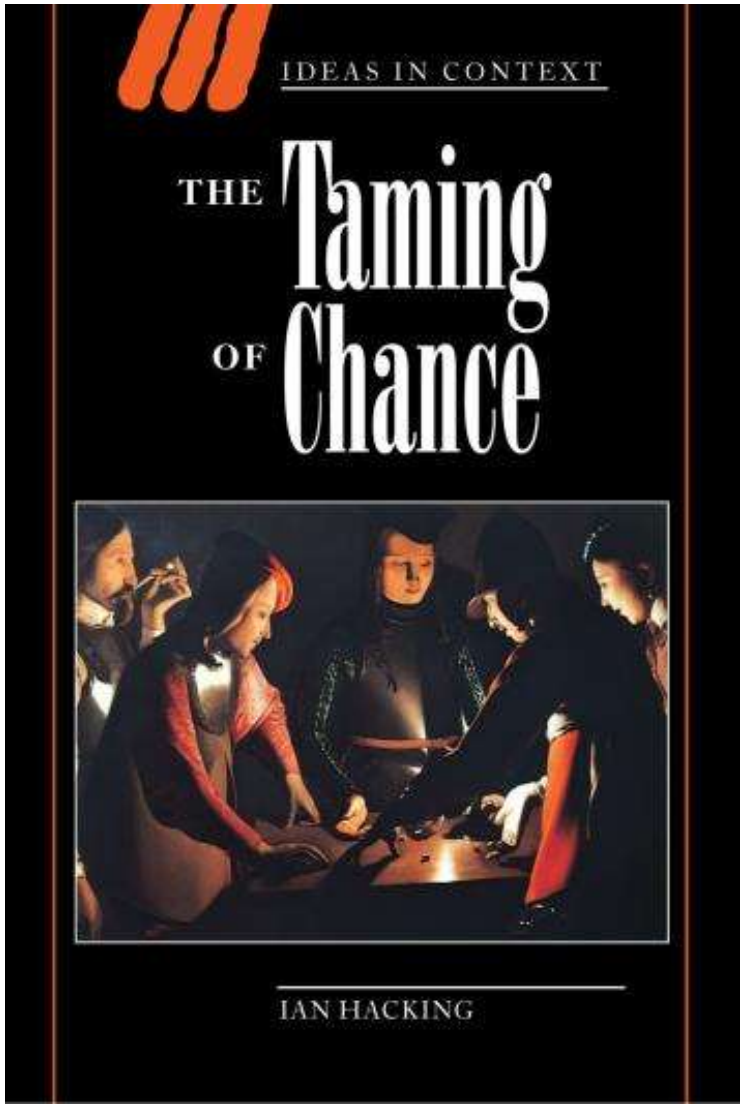
(for example, data are not manipulated, nor their release timed in response to political pressure)

COHERENCE

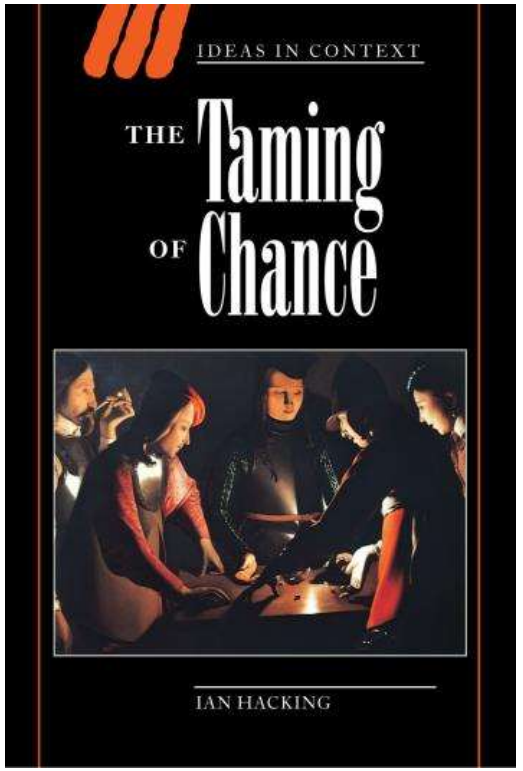
… ensure coherence over time and across countries … Coherence across countries implies that from country to country the data are based on common concepts, definitions, classifications and methodology, or that any differences can be justified

History

The first scoreboard?



Ian Hacking, 1990, The taming of chance, Cambridge University Press.

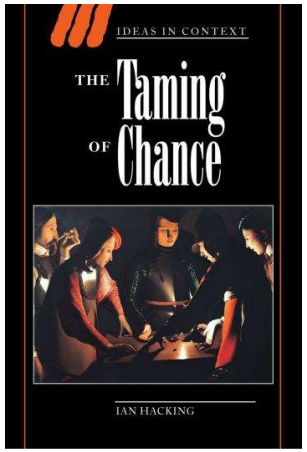


Statistics ⟷ nation state ⟷ Modernity

Leibnitz, ‘philosophical godfather of Prussian official statistics’.

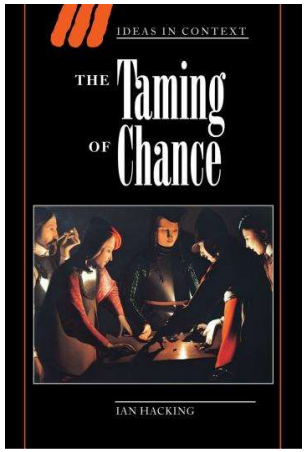
Gottfried Wilhelm Leibniz (1646-1716)

His proposal to the Prince Frederik of Prussia, 1700



56 categories to ‘measure the power of a state’, the first scoreboard;

- number of marriageable girls,
- able bodied capable to carry arms,
- diseases,
- child mortality,
- …
- number of Jews



Gottfried Wilhelm Leibniz (1646-1716)

Is a theory for composite indicators possible?

European Commission

Elements for a comprehensive assessment of public indicators

Paul-Marie Boulanger

2014

Editor: Andrea Saltelli

Report EUR 26921 EN

Joint Research Centre

Paul-Marie Boulanger

Paul-Marie Boulanger, 2014, Elements for a comprehensive assessment of public indicators, Report EUR 26921 EN. http://publications.jrc.ec.europa.eu/repository/bitstream/JRC92162/lbna26921enn.pdf

CI as boundary objects, between analysis and advocacy, as:

- instruments of democratization of expertise;
- instruments of social discovery
- semiotic objects

Paul-Marie Boulanger, 2014, Elements for a comprehensive assessment of public indicators, Report EUR 26921 EN.
http://publications.jrc.ec.europa.eu/repository/bitstream/JRC92162/lbna26921enn.pdf

A triadic conception of the sign as structure connecting three elements:

- the sign properly said (S);
- an object (O) and
- an “interpretant”(I)

Charles Sanders Peirce,
the father of semiotics
1839-1914

“This monkey possess a sophisticated repertory of vocal signs for signaling the presence of a predator”

African vervet monkey
(Cercopithecus aethiops)

It can distinguish

- a terrestrial stalking one such as a leopard,
- an aerial raptor such as an eagle or
- a ground predator such as a snake

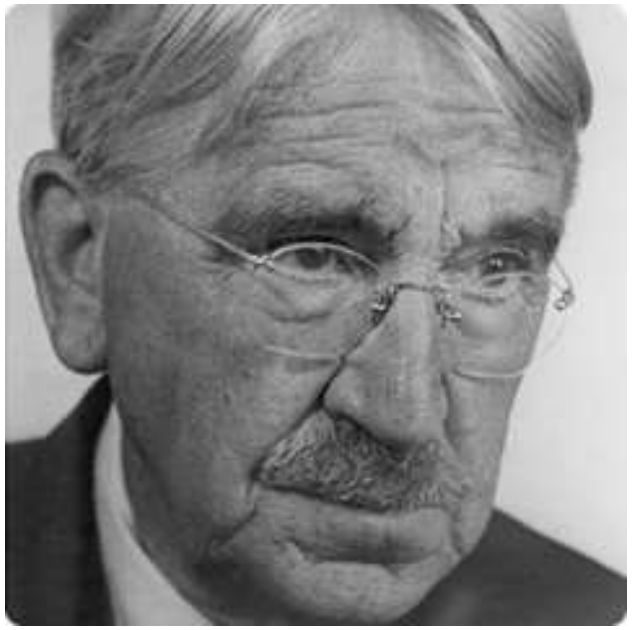
African vervet monkey (Cercopithecus aethiops)

Sign ⟷ Cry

Object ⟷ Predator

Interpretant ⟷ Behaviour

Composite indicators as instrumental to the creation of a new public, through a process of social discovery (J. Dewey)



John Dewey
1859-1952

Dewey, J., 1938. The Public and its Problems, Read Book Ltd. Edition, 2013.

Why are ‘social discoveries’ needed?

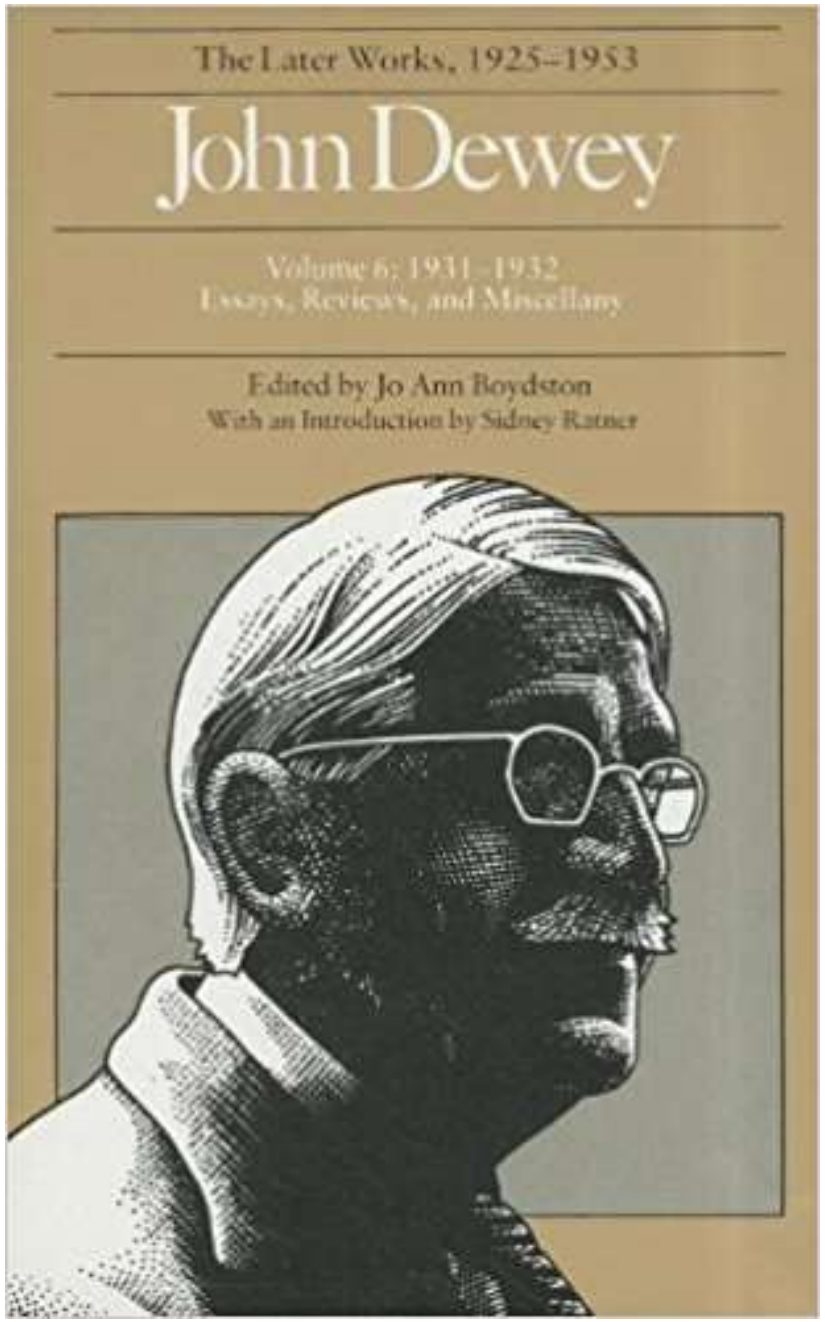
Because there are ‘publics’ affected by transaction taking place somewhere else.

“[…] machine age has so enormously expanded, multiplied, intensified and complicated the scope of the indirect consequences […] that the resultant public cannot identify and distinguish itself”

Dewey, J., 1938. The Public and its Problems, Read Book Ltd. Edition, 2013.

Social facts – unlike physical facts, are only meaningful in a context of desired ends

From J. Dewey ‘Social Science and Social Control’ in John Dewey: The Later Works, 1925-1953: 1931-1932, Vol. 6-ExLibrary,



Building a composite indicator can be seen as a process of social discovery for which a model of extended participation comes natural.
Frames and indicators are co-produced in the process which must be designed as to have a meaningful ‘interpretant’, or ‘end-in-sight’

Paul-Marie Boulanger, 2014, Elements for a comprehensive assessment of public indicators, Report EUR 26921 EN. http://publications.jrc.ec.europa.eu/repository/bitstream/JRC92162/lbna26921enn.pdf

Critique of composite indicators: the Fitoussi-Stiglitz-Sen report

"The role [of statistical indicators] has increased significantly over the last two decades.

This reflects improvements in the level of **education** in the population, increases in the **complexity** of modern economies and the widespread use of **information technology**"

Jean-Paul Fitoussi,
Amartya Sen, Joseph Stiglitz

CMEPSP (2009). Commission on the Measurement of Economic Performance and Social Progress, URL: http://ec.europa.eu/eurostat/documents/118025/118123/Fitoussi+ Commission+ report, last accessed June 2017.

“a general criticism … frequently addressed at composite indicators, i.e. the arbitrary character of the procedures used to weight their various components […]

Jean-Paul Fitoussi,
Amartya Sen, Joseph Stiglitz

[…] an aggregation procedure always means putting relative values on the items that are introduced in the index”

CMEPSP (2009). Commission on the Measurement of Economic Performance and Social Progress, URL: http://ec.europa.eu/eurostat/documents/118025/118123/Fitoussi+ Commission+ report, last accessed June 2017.

"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.

Jean-Paul Fitoussi,
Amartya Sen, Joseph Stiglitz

The problem is rather that their normative implications are seldom made explicit or justified"

CMEPSP (2009). Commission on the Measurement of Economic Performance and Social Progress, URL: http://ec.europa.eu/eurostat/documents/118025/118123/Fitoussi+ Commission+ report, last accessed June 2017.

Critique of composite indicators: Ravallion

There are types two indices: those built on economic theory / monetary aggregates / shadow prices and all others (=mashup indices)

Martin Ravallion

+ existing measures of e.g. development or poverty (Human Development Index, HDI, the Multidimensional Poverty Index, MPI) are bad at coping with tradeoffs

Martin Ravallion, 2010, Mashup indices of development, Policy Research Working Paper 5432 , The World Bank Development Research Group, http://documents.worldbank.org/curated/en/454791468329342000/pdf/WPS5432.pdf

Martin Ravallion

“To illustrate the distinction, consider two stylized examples of composite indices, both formed from the data on household assets and consumer durables found in the Demographic and Health Surveys (DHS). For index A the variables and their weights are set by the analyst, who has some concept of ―economic welfare‖ in mind, and thinks this is related to certain variables in the DHS, which are aggregated based on the analyst‘s judgments. For index B, the variables and weights are instead based on a regression model calibrated to another survey data set for which a comprehensive measure of consumption (though still containing measurement errors) could be derived. The model is calibrated to common variables in the expenditure survey and the DHS, and the regression model is used to predict wealth in the DHS. A is a mashup index, B is not.”

(M. Ravallion)

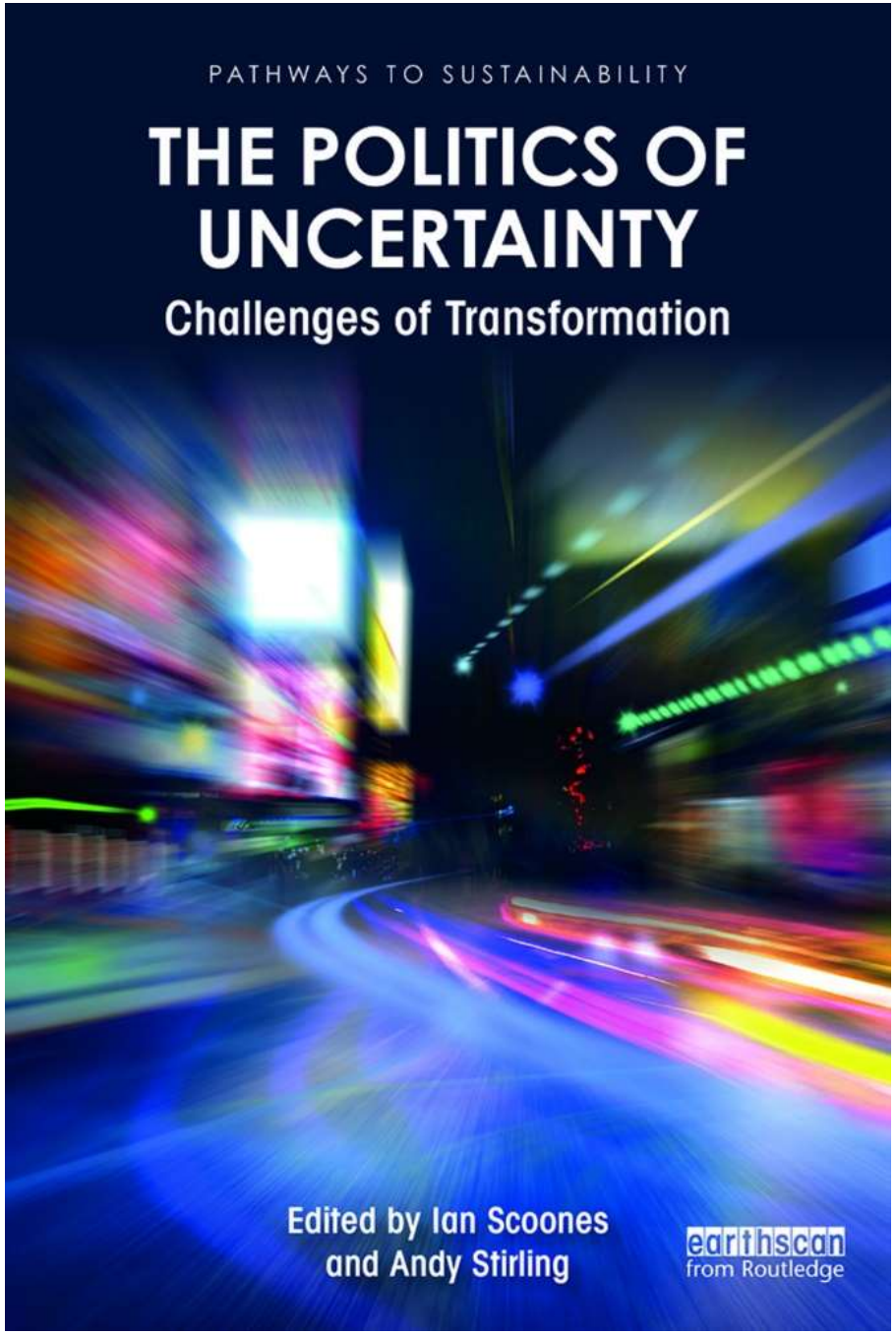
A recent critique of reductionism

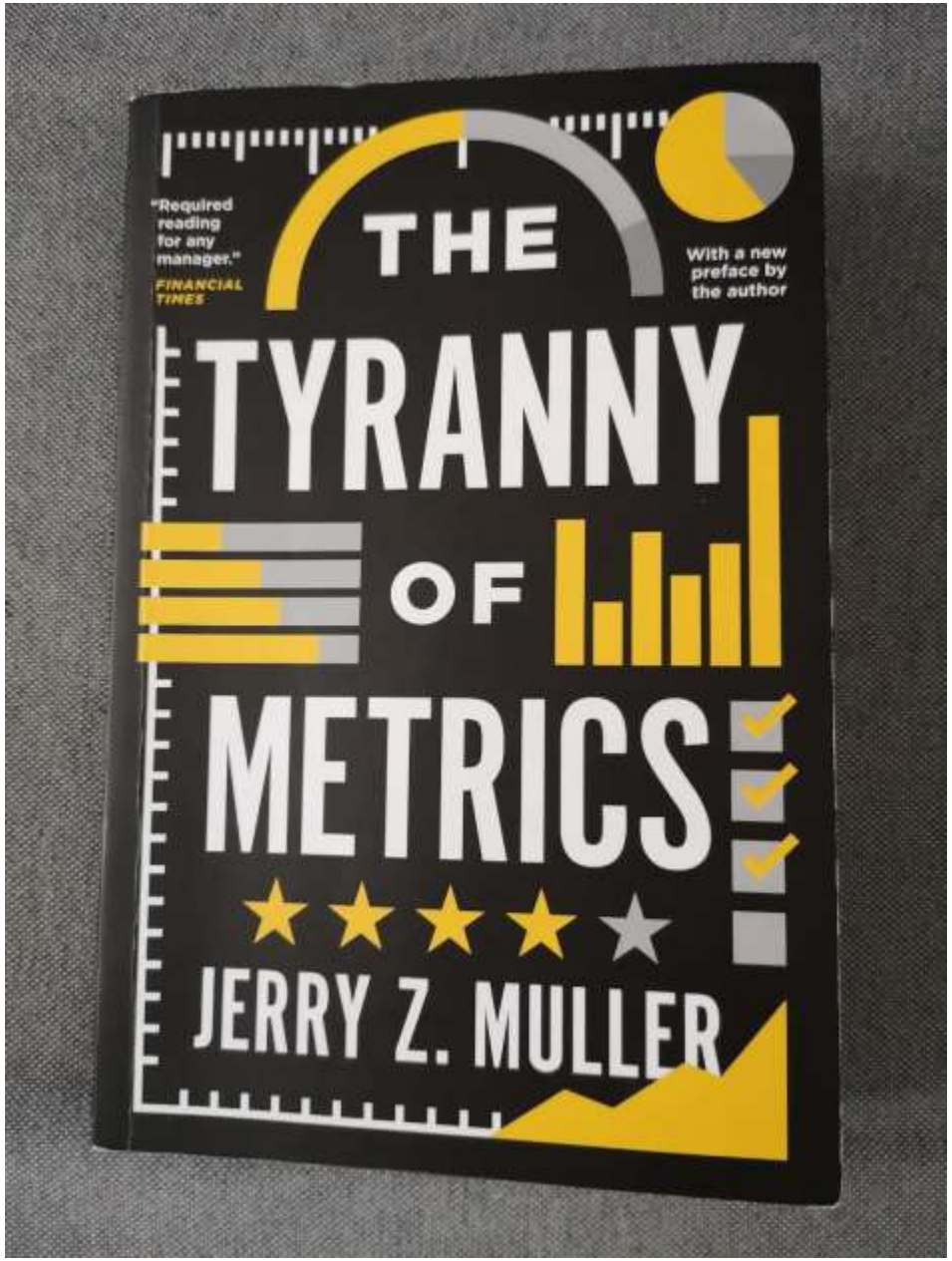
4

THE UNRAVELLING OF TECHNOCRATIC ORTHODOXY?

Contemporary knowledge politics in technology regulation

Patrick van Zwanenberg

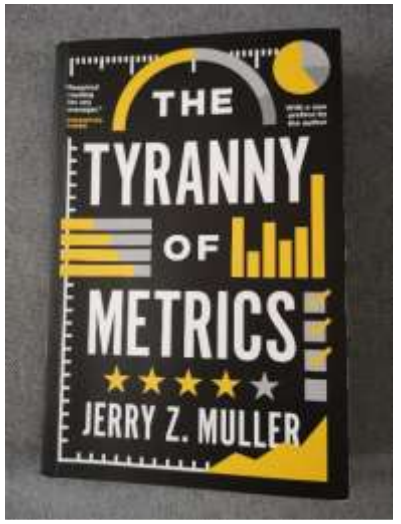




Can composite indicators do harm?

J. Z. Muller, The tyranny of metrics. Princeton University Press , 2018.

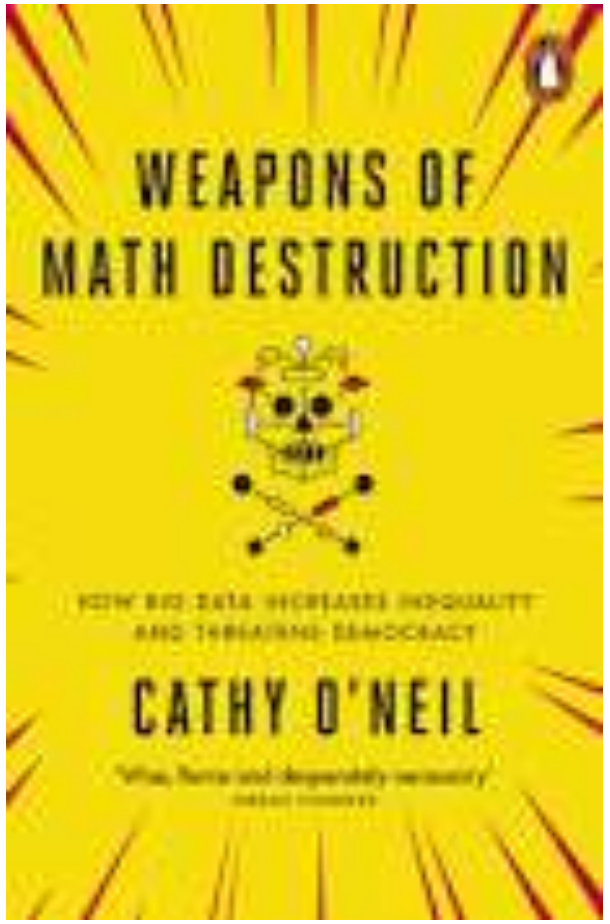
Unintended consequences: a litany

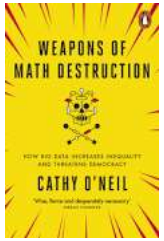


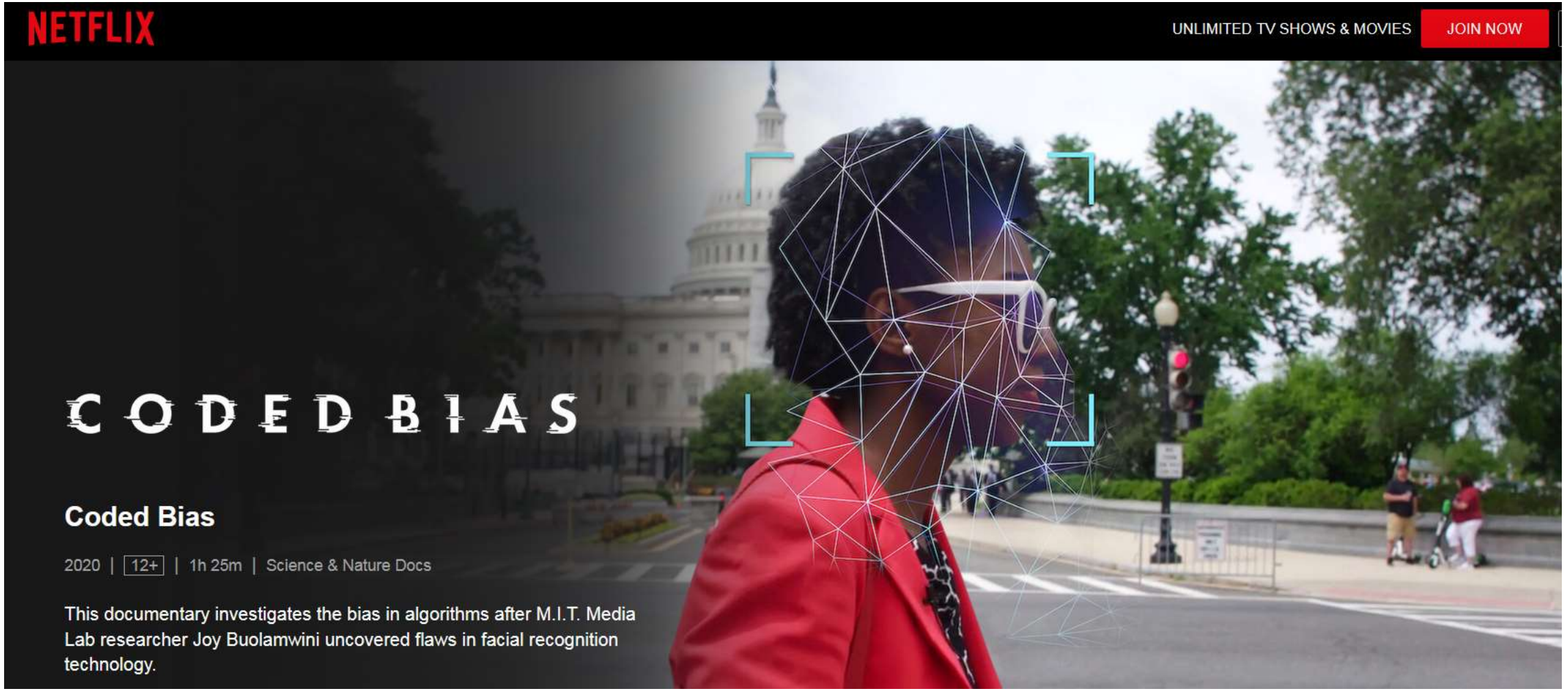
- Goal displacement
- Short termism
- Diminishing utility
- Rule cascade
- Discouraging risk taking
- Discouraging innovation
- Rewarding luck
- Discouraging cooperation and common purpose
- Degrading work
- Time waste
- Loss of productivity

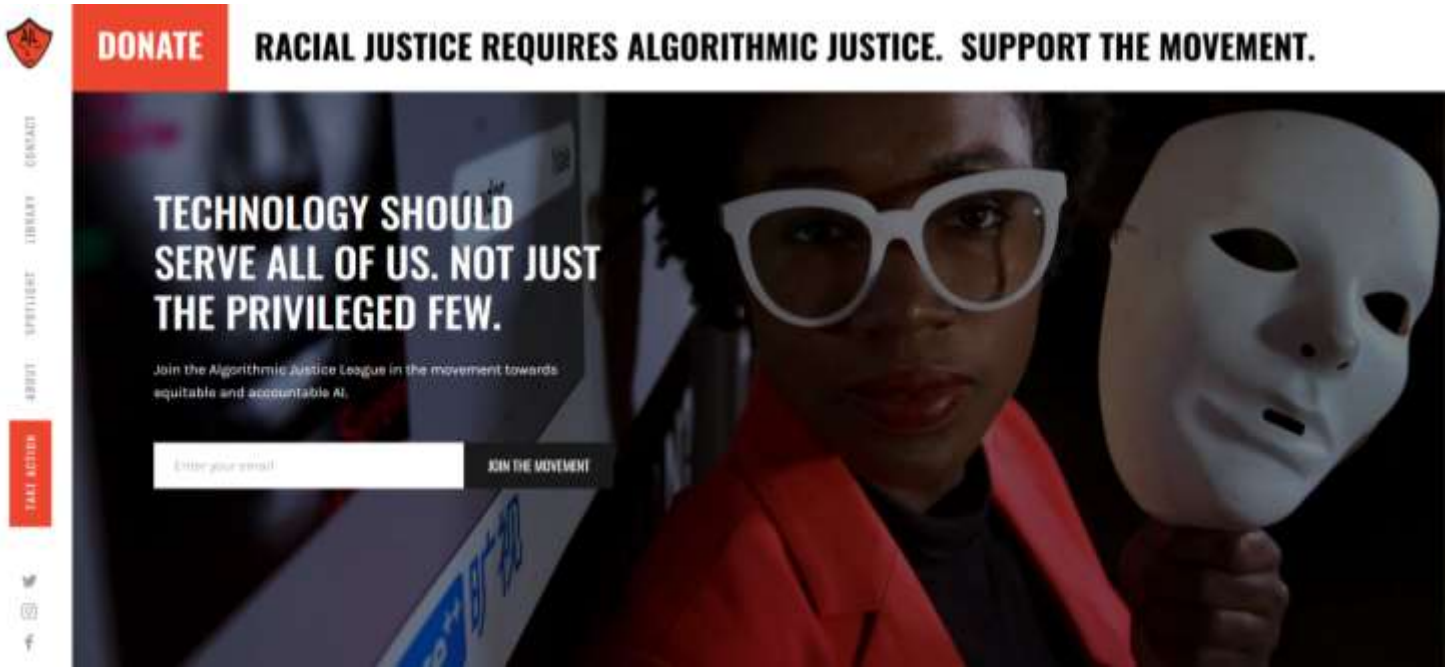
Weapons of math destruction: opaque, do harm, do scale

Cathy O'Neil




WEAPONS OF
MATH DESTRUCTION
CATHY O'NEIL


NETFLIX
UNLIMITED TV SHOWS & MOVIES
JOIN NOW
CODED BIAS
Coded Bias
2020 | 12+ | 1h 25m | Science & Nature Docs
This documentary investigates the bias in algorithms after M.I.T. Media Lab researcher Joy Buolamwini uncovered flaws in facial recognition technology.



Algorithmic Justice League

The New York Times
Bloomberg Business
Forbes
TIME
FORTUNE
TED
WIRED
The Telegraph

Since composite indicators are here to stay, how can we make them defensible?

… or how can we deconstruct them?

Tools for evidence appraisal such sensitivity analysis and sensitivity auditing can be useful to gauge (and possibly deconstruct or reinforce) these measures

Sensitivity analysis

PDF

Info

Series A
Statistics in Society

Explore this journal >

Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators

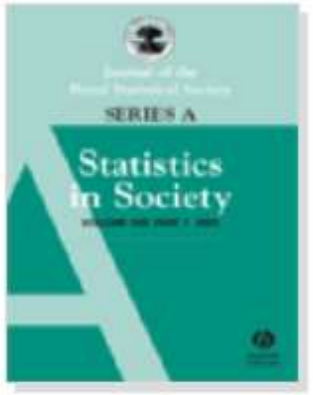
M. Saisana, A. Saltelli, S. Tarantola

First published: 3 March 2005 Full publication history

DOI: 10.1111/j.1467-985X.2005.00350.x View/save citation

Cited by (CrossRef): 181 articles Check for updates

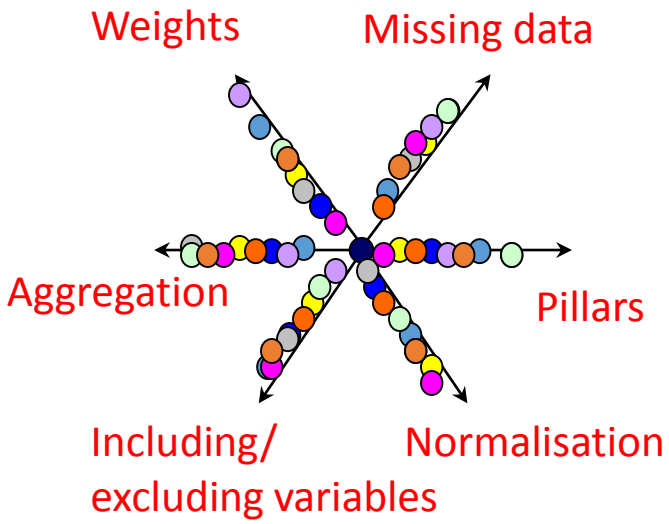
Citation tools

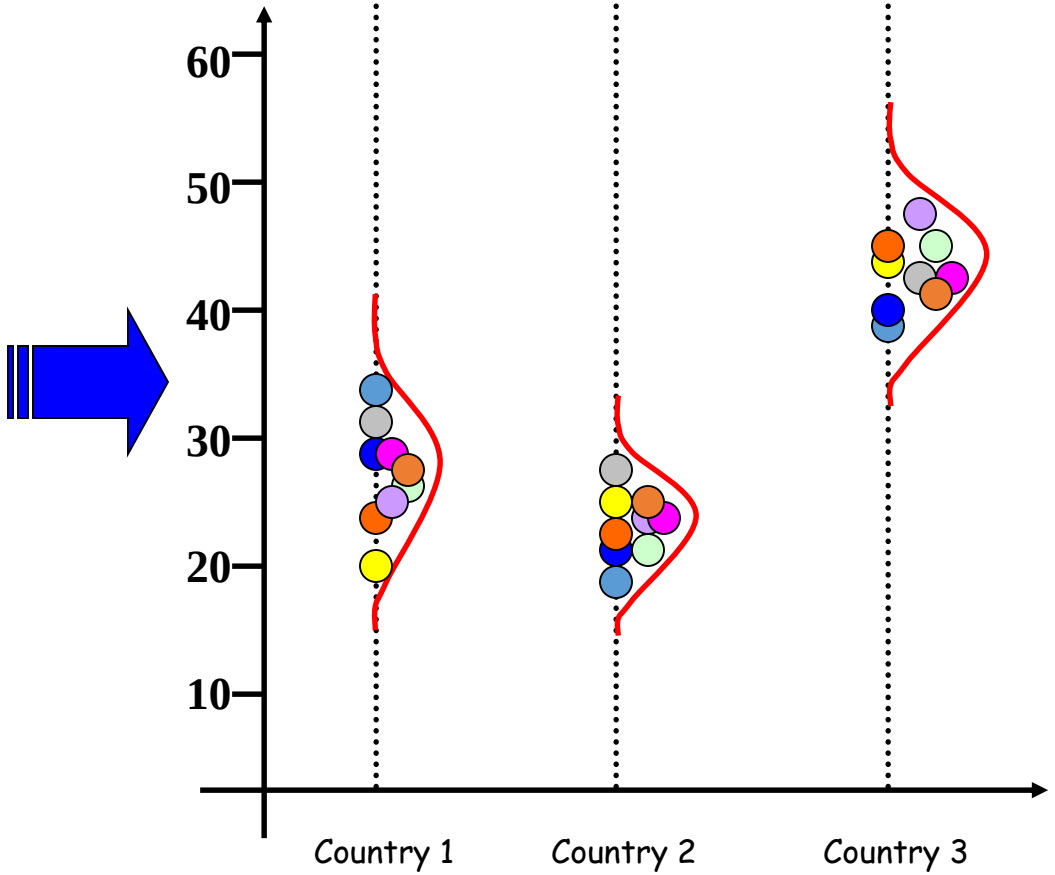


View issue TOC
Volume 168, Issue 2
March 2005
Pages 307–323

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

Space of alternatives





Sensitivity analysis to compare volatility of ranking

Research Policy 40 (2011) 165–177

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

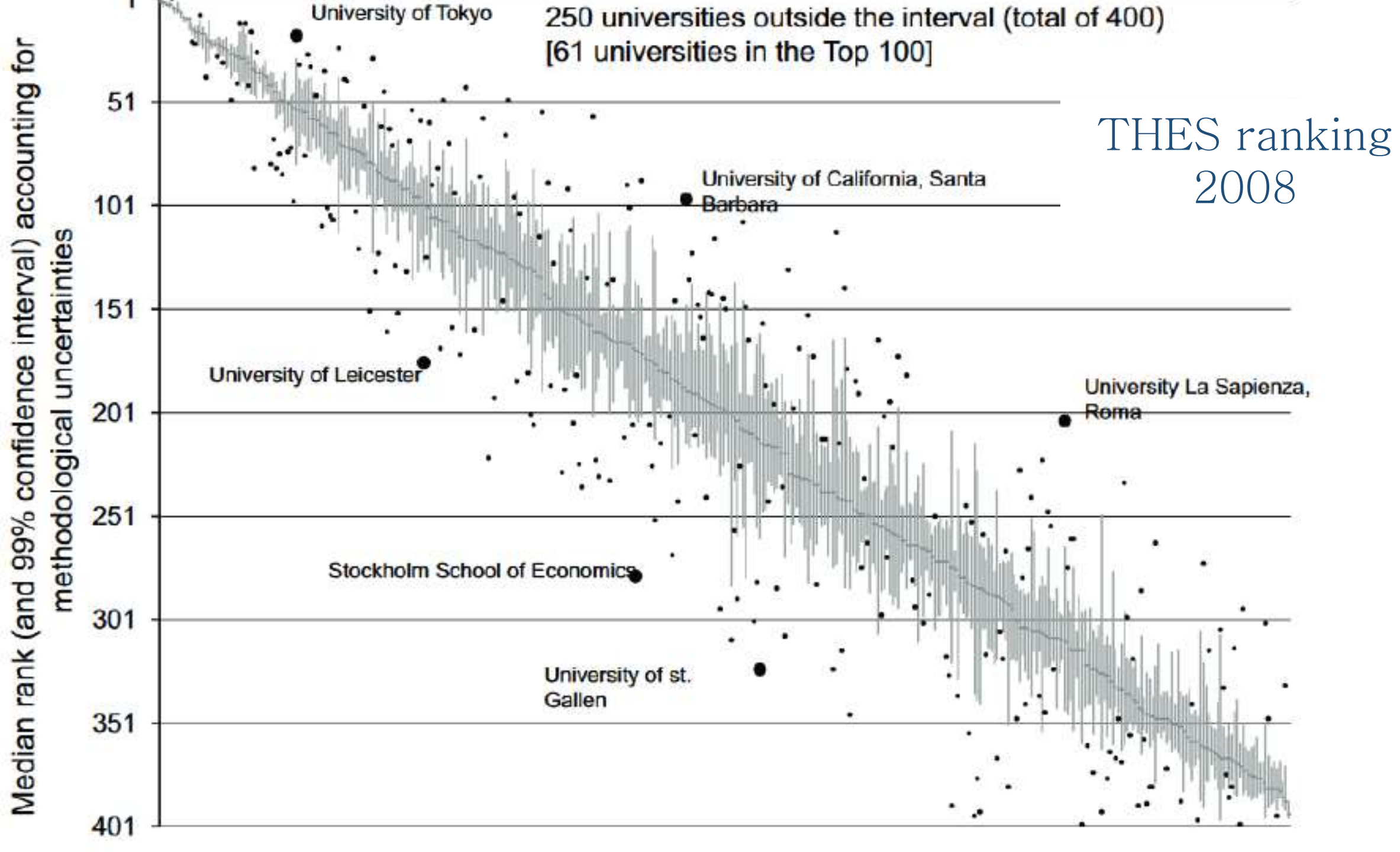
Rickety numbers: Volatility of university rankings and policy implications

Michaela Saisana*, Béatrice d'Hombres, Andrea Saltelli

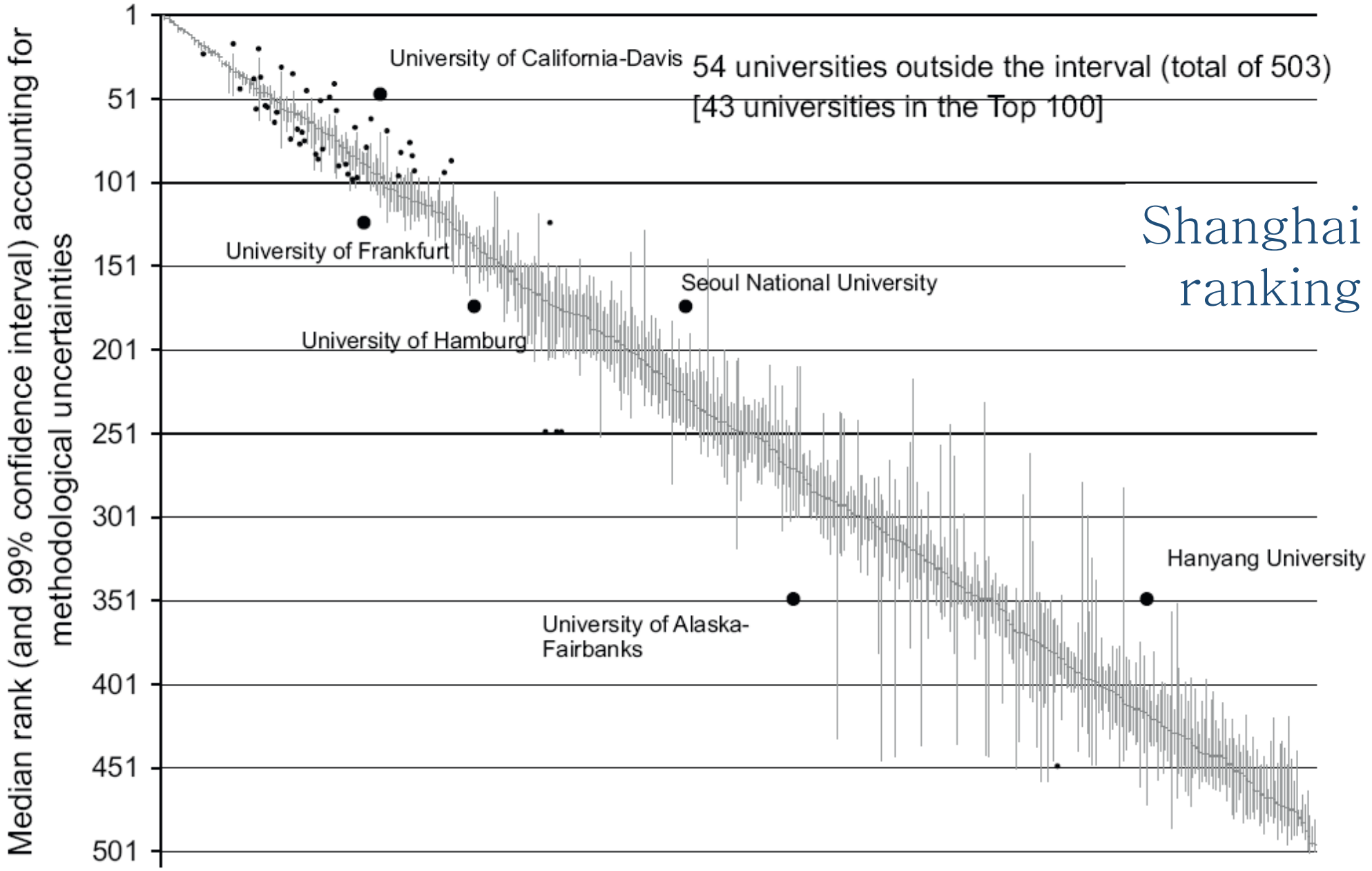
Econometrics and Applied Statistics, Joint Research Centre, European Commission, Enrico Fermi 2749, 21027 Ispra, Italy

Sensitivity analysis to compare volatility of ranking

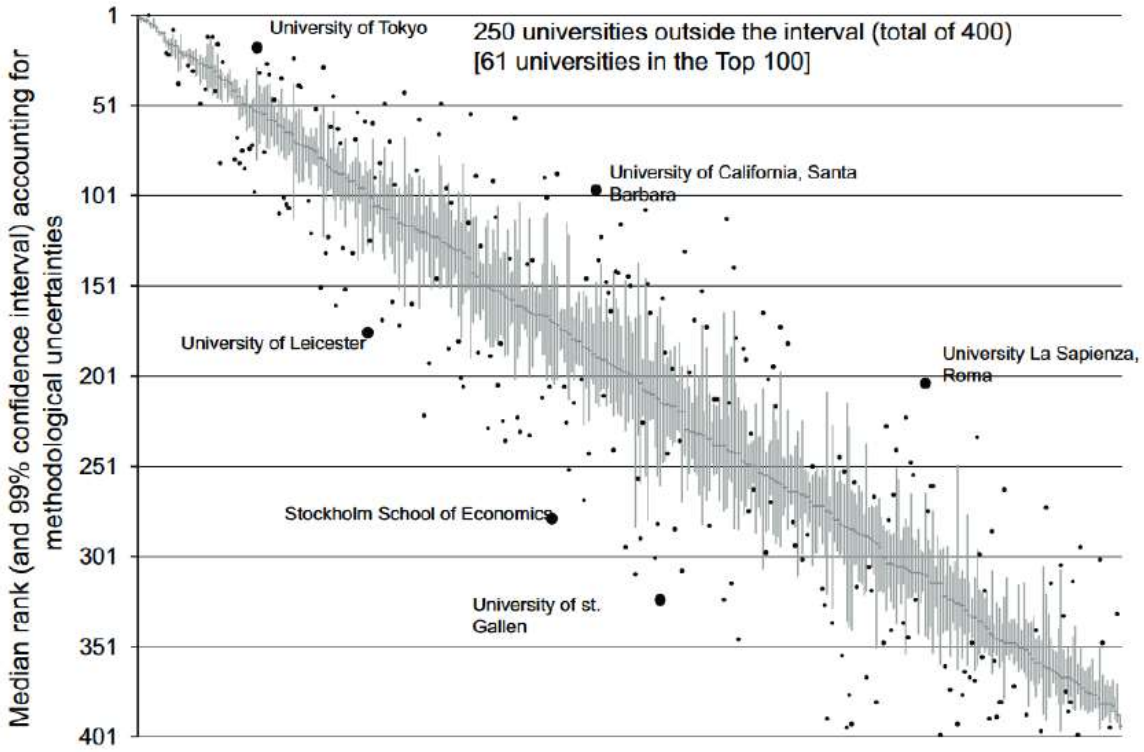
An ‘invasive’ analysis as the developers’ choices are questioned/varied



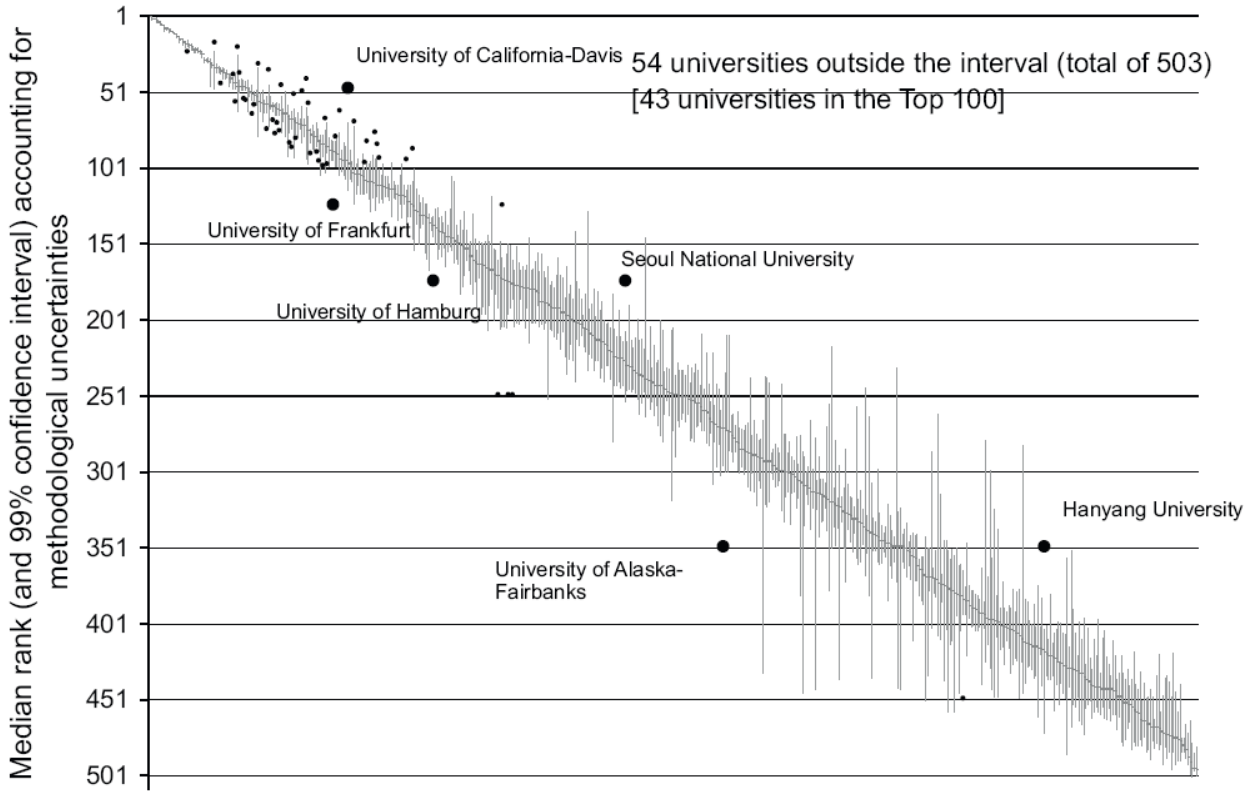
THES ranking 2008



Shanghai ARWU ranking 2008



THES ranking 2008



Shanghai ARWU ranking 2008

Incidentally: these university rankings have also damaged the educational systems

« processus de Bologne (en 1999) + stratégie de Lisbonne (en 2000), → passage d'une logique de service public à une logique de marché, concurrentielle et gestionnaire »

Le classement de Shanghai. Histoire, analyse et critique

Fabien Eloire

DANS L'HOMME & LA SOCIÉTÉ **2010/4 (n° 178)**, PAGES 17 À 38

One can test whether assigned weights correspond to real importance

Journal of the Royal Statistical Society
SERIES A Statistics in Society

J. R. Statist. Soc. A (2013)
176, *Part* 3, *pp.* 609–634

Ratings and rankings: voodoo or science?

Paolo Paruolo

University of Insubria, Varese, Italy

and Michaela Saisana and Andrea Saltelli

European Commission, Ispra, Italy

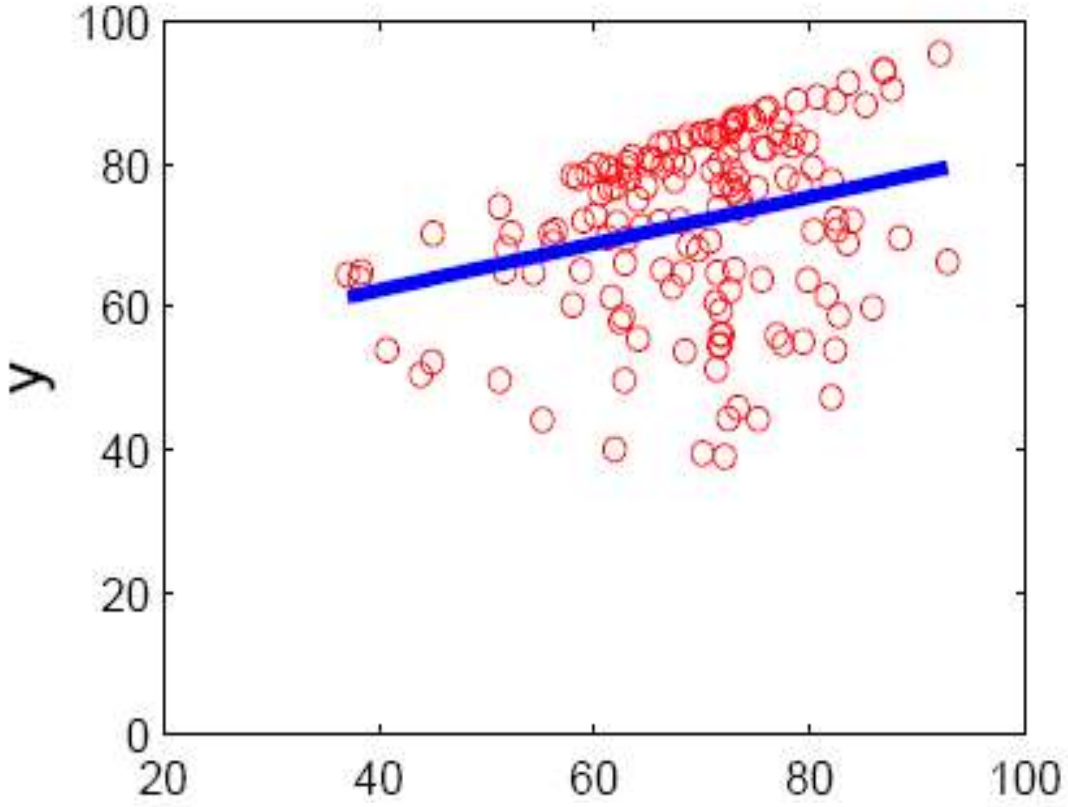
The linear aggregation paradox: **weights are used as if they were importance coefficients** while they are trade off coefficients

An example. A dean wants to rank teachers based on ‘hours of teaching’ and ‘number of publications …

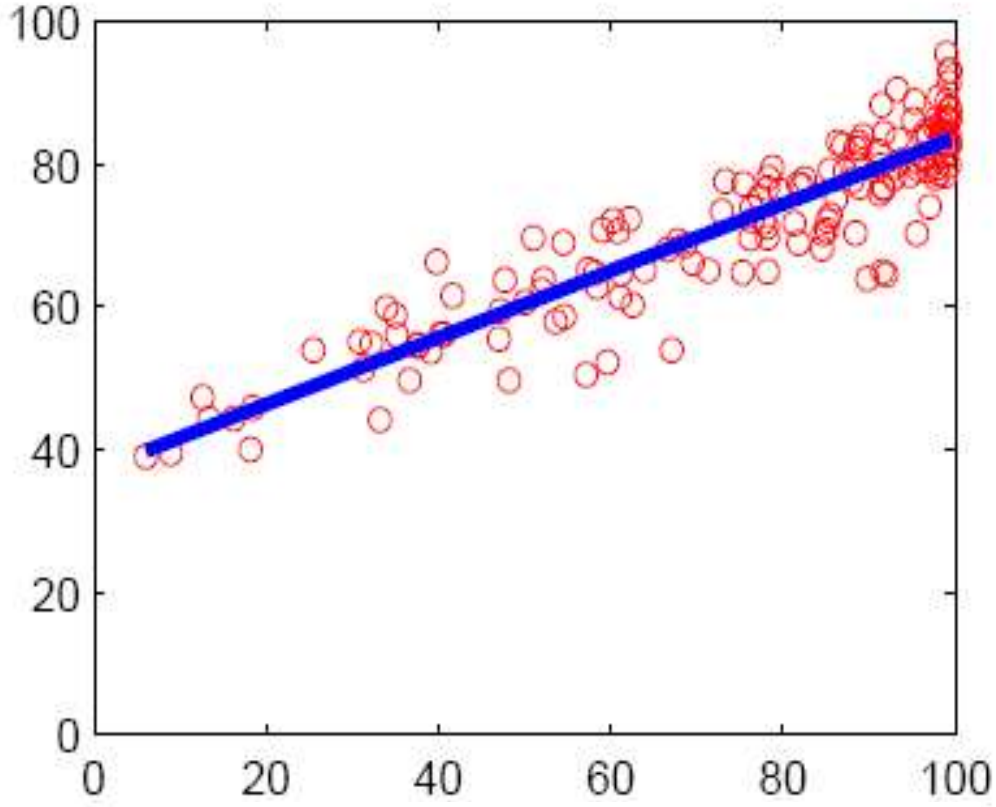
$$Y=0.5X_1 + 0.5X_2$$

X_1: hours of teaching
X_2: number of publications

… adding these two variables up she sees that teachers are practically ranked by publications alone



hours of teaching



number of publications

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

Estimated $R_{ht}^2 = 0.0759$, $R_{np}^2 = 0.826$

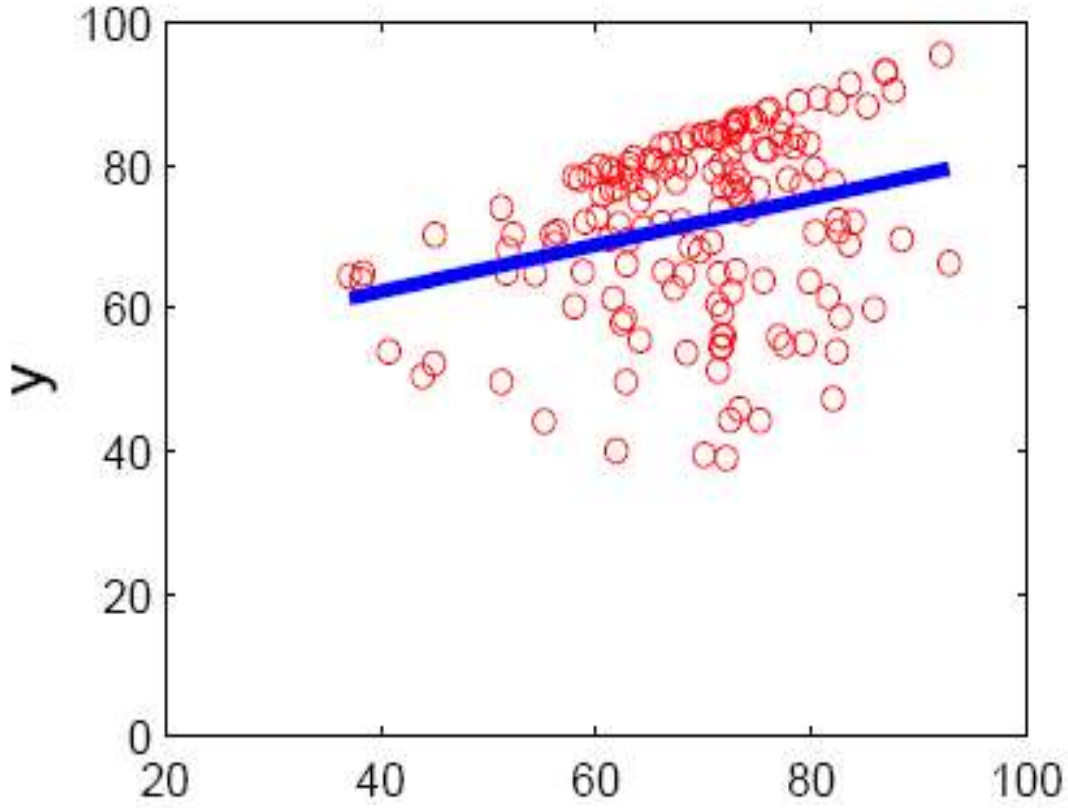
To obviate this the dean substitutes the model

$$y=0.5x_1+0.5x_2$$

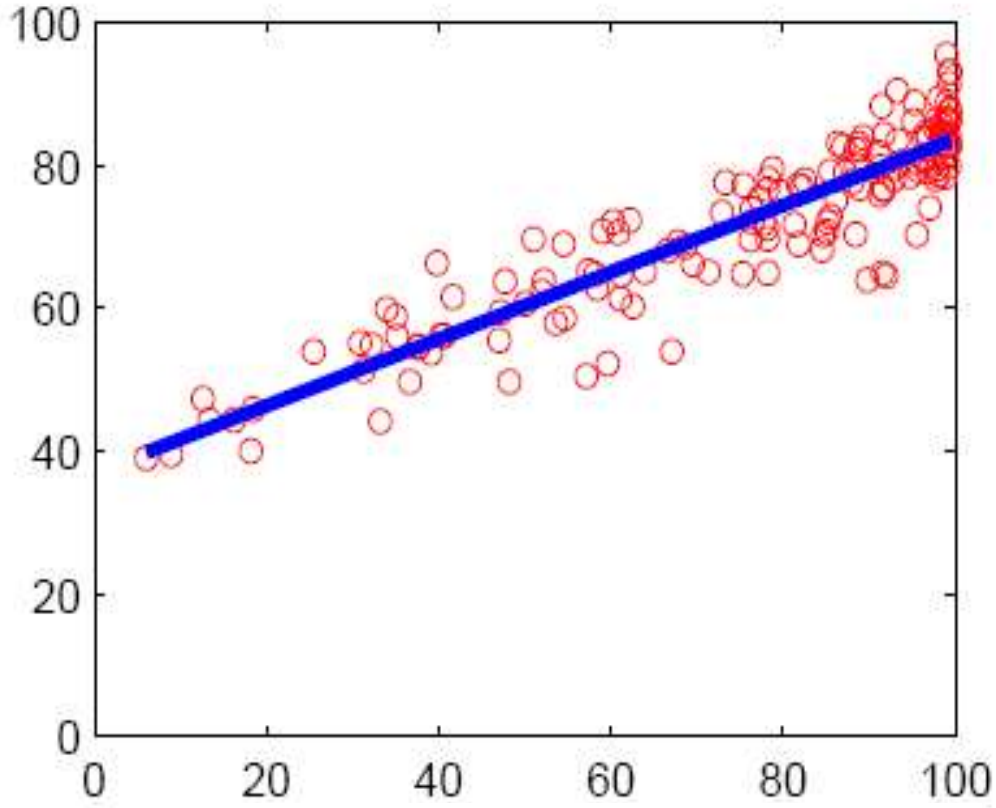
with

$$y=0.7x_1+0.3x_2$$

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

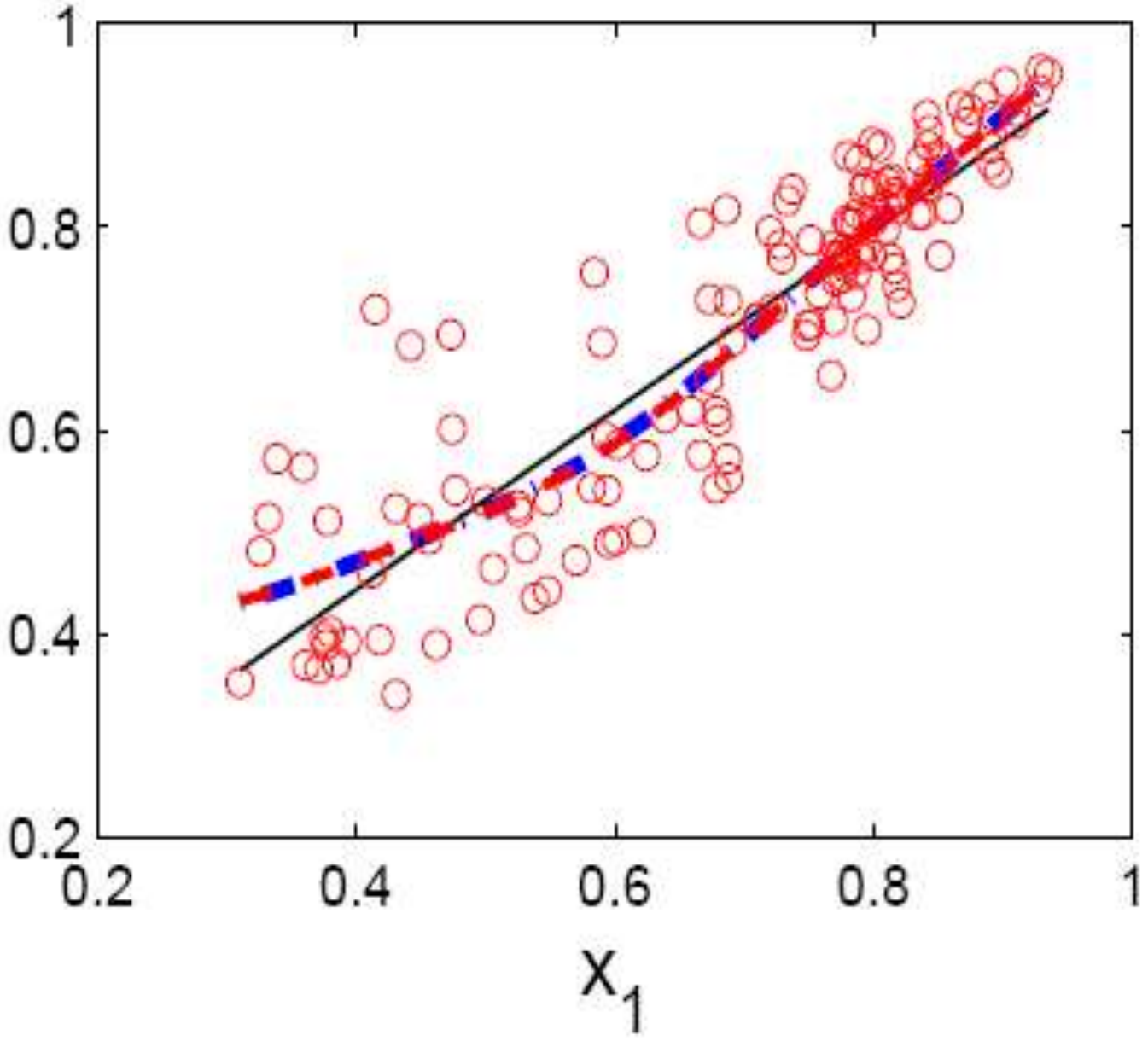


hours of teaching



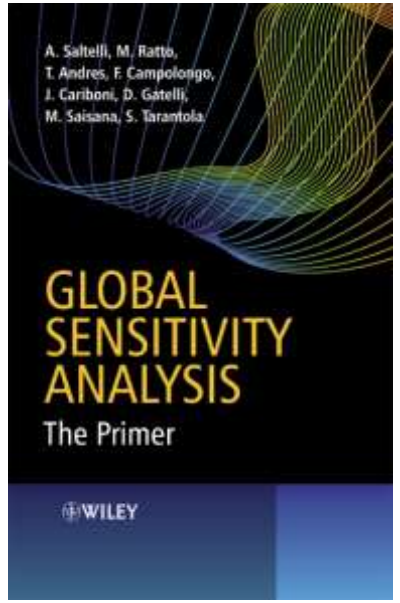
number of publications

Using a result from sensitivity analysis the scatterplots → numbers reflecting the importance of a variable



The straight line corresponds to R^2

The variance of the moving average is a sensitivity measure



Pearson's correlation ratio

Smoothed curve

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

First order sensitivity index

Unconditional variance

S_i is the expected fractional variance reduction that would be achieved on average if x_i could be fixed

Why (an exercise) ?

$$S_i = \frac{V(E(Y|x_i))}{V(Y)}$$

$$V\big(E(Y|x_i)\big) + E\big(V(Y|x_i)\big) = V(Y)$$

Normalization paradox: **weights are assigned as to add up to one.** This is questionable.

Given a simple CI $Y = w_1 x_1 + w_2 x_2$

If both x_1 and x_2 are standardized the importance of x_1 is $S_1 = \frac{w_1^2}{w_1^2 + w_2^2}$

and $S_1 + S_2 = 1$

Thus the relative importance of x_1, x_2

is not $\frac{w_1}{w_2}$ but $\frac{w_1^2}{w_2^2}$ …

… and the absolute importance are not $\frac{w_1}{w_1+w_2}$ and $\frac{w_2}{w_1+w_2}$

but $\frac{w_1^2}{w_1^2+w_2^2}$ and $\frac{w_2^2}{w_1^2+w_2^2}$

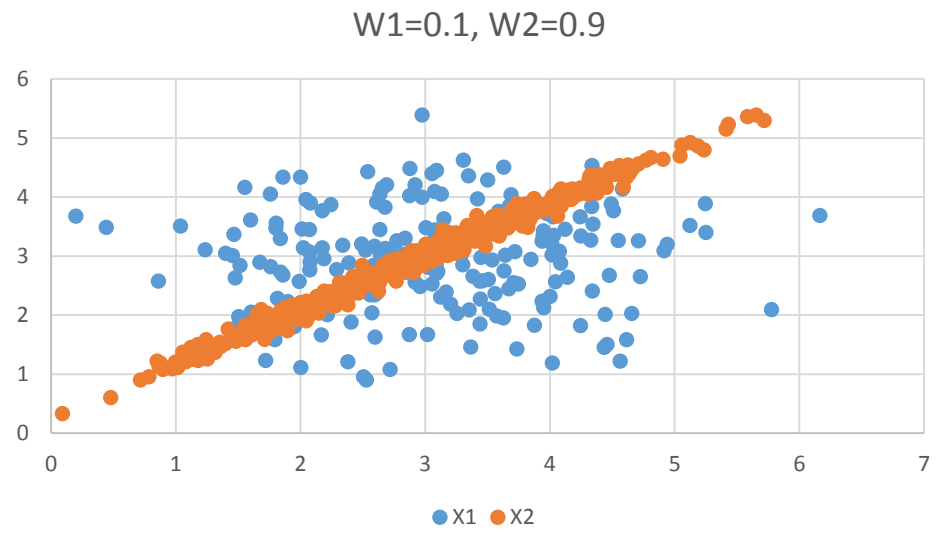
Implications?

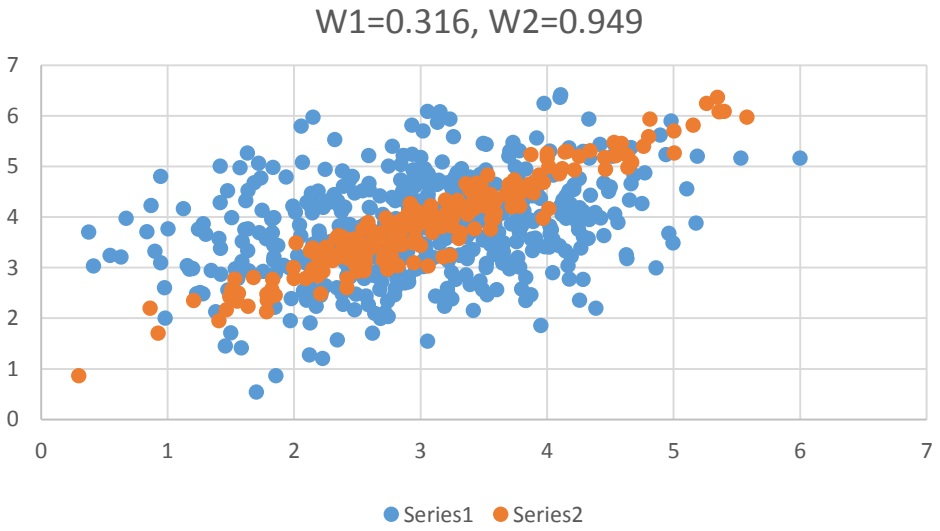
	x_1	x_2
Presumed importance	10%	90%
Real importance	1.2%	98.8%

$$\frac{w_1}{w_1+w_2}, \frac{w_2}{w_1+w_2}$$

$$\frac{w_1^2}{w_1^2+w_2^2}, \frac{w_2^2}{w_1^2+w_2^2}$$

This holds if we use our definition of importance (what expected fraction of the variance of Y would be reduced on average if x_1 could be fixed) but you can verify this empirically using scatterplots





$$w_1 = 0.1,\ w_2 = 0.9$$

$$w_1 = 0.32,\ w_2 = 0.95$$

Comparing assigned weights versus measured importance for the 2009 and 2010 versions of the Human Development index

Journal of the Royal Statistical Society
SERIES A Statistics in Society

J. R. Statist. Soc. A (2013)
176, *Part* 3, *pp.* 609–634

Ratings and rankings: voodoo or science?

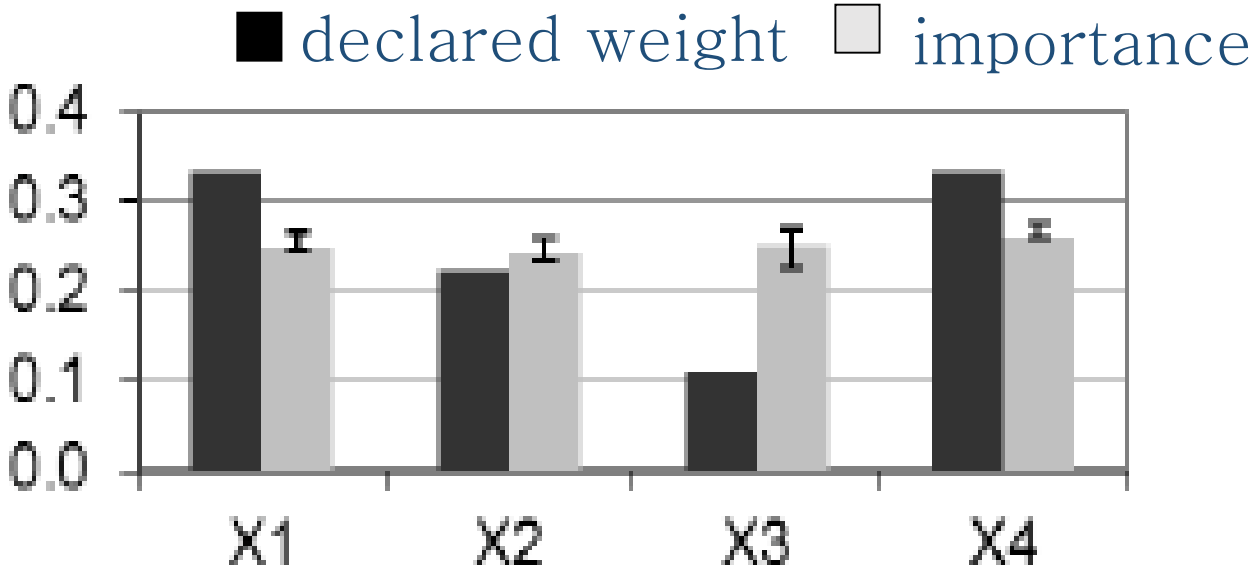
Paolo Paruolo

University of Insubria, Varese, Italy

and Michaela Saisana and Andrea Saltelli

European Commission, Ispra, Italy

HDI
2009



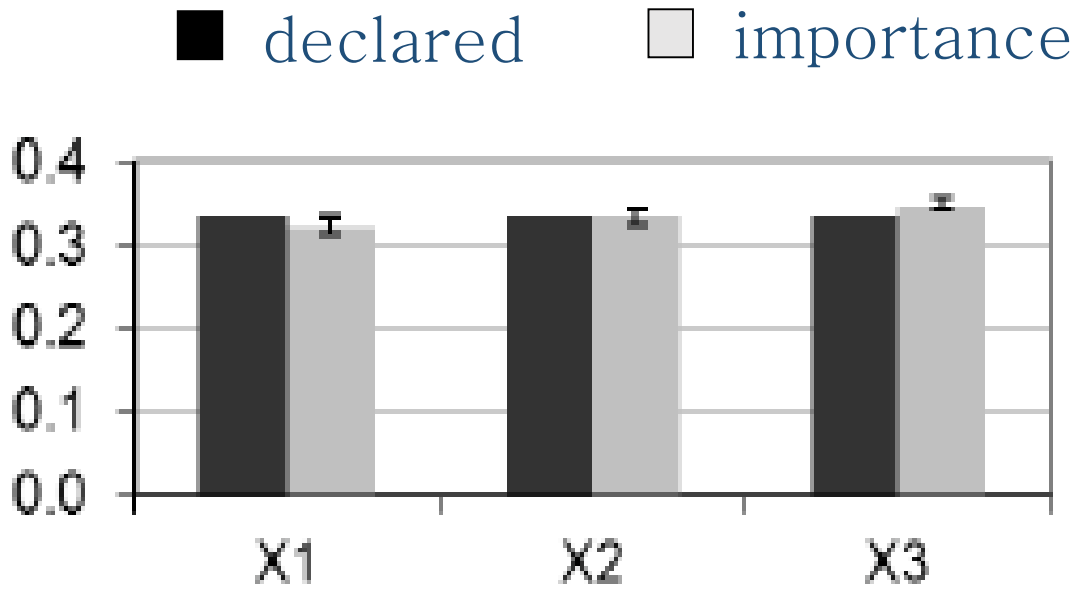
Life expectancy, 33%

Adult literacy, 22%

Enrollment education, 11%

GDP per capita, 33%

HDI
2010

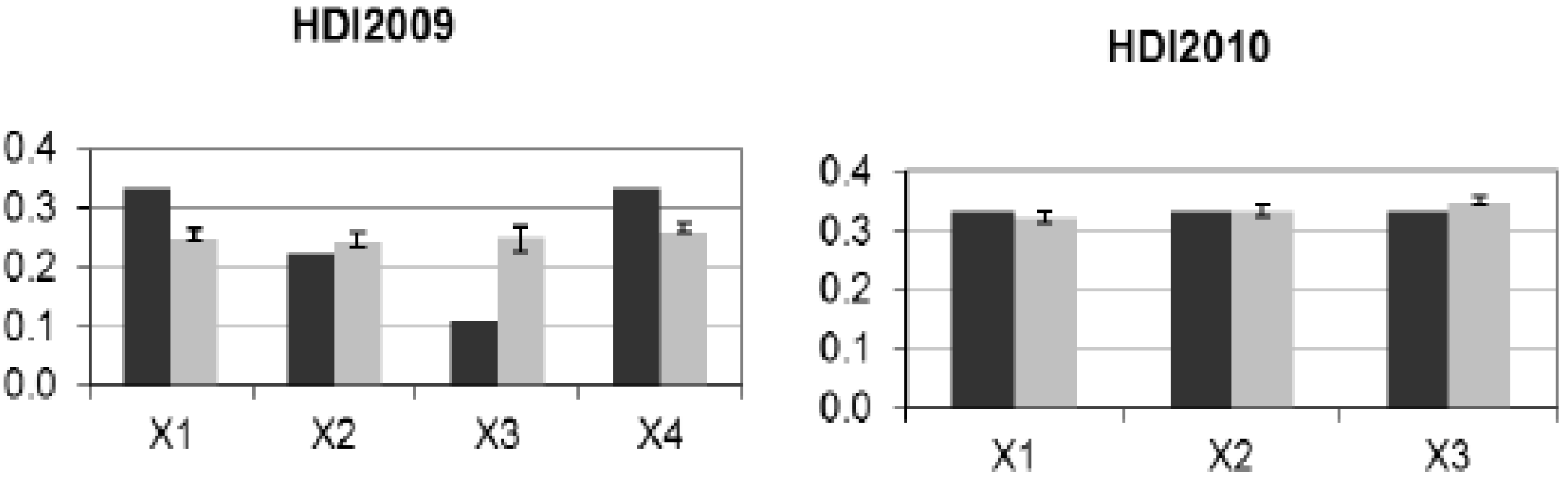


Life expectancy, 33%

Education, 33%

GNI per capita, 33%

declared weight importance



HDI 2010 more coherent than HDI 2009

Sensitivity auditing

COMMENT · 24 JUNE 2020

Five ways to ensure that models serve society: a manifesto

Pandemic politics highlight how predictions need to be transparent and humble to invite insight, not blame.

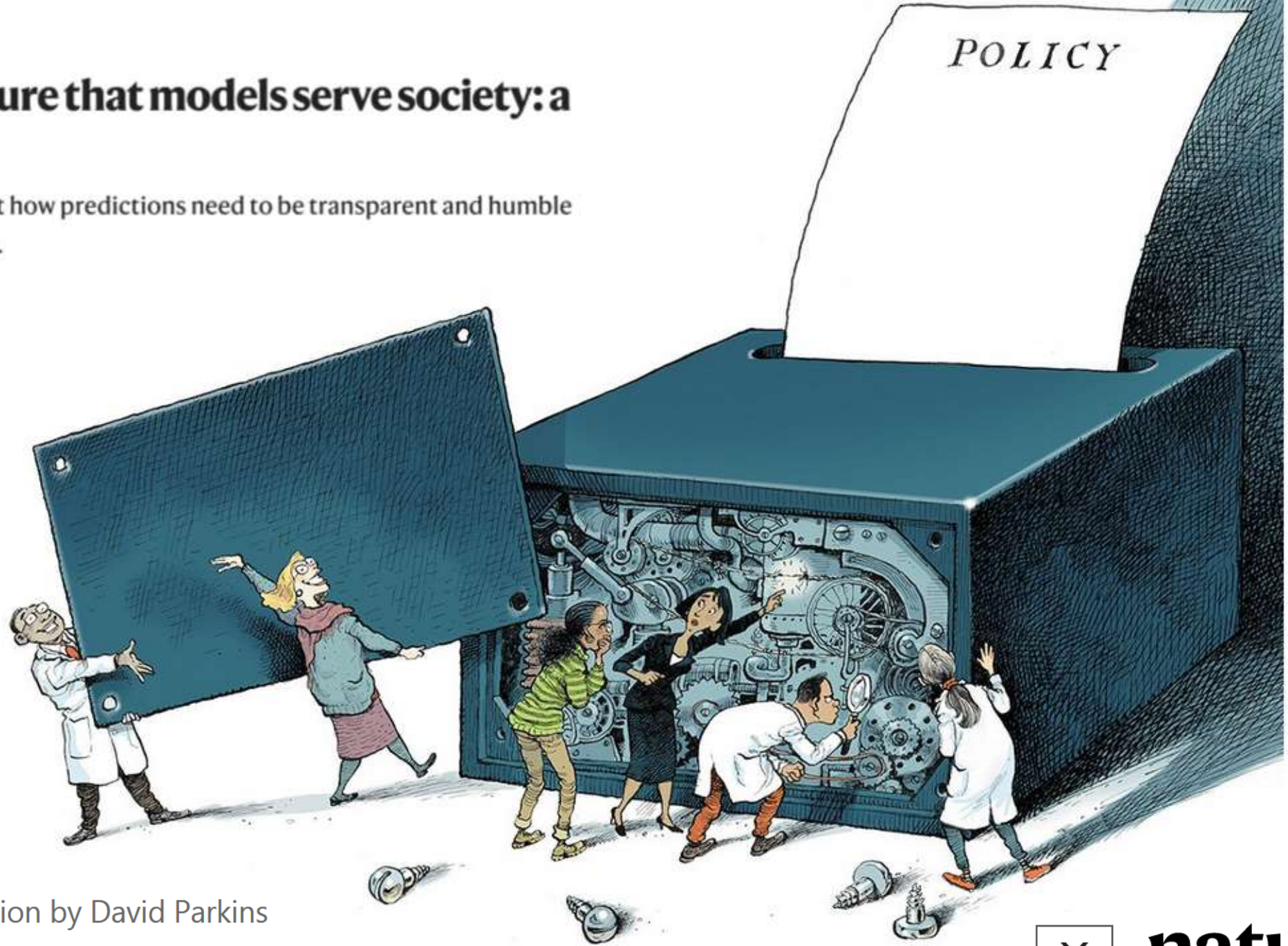


Illustration by David Parkins

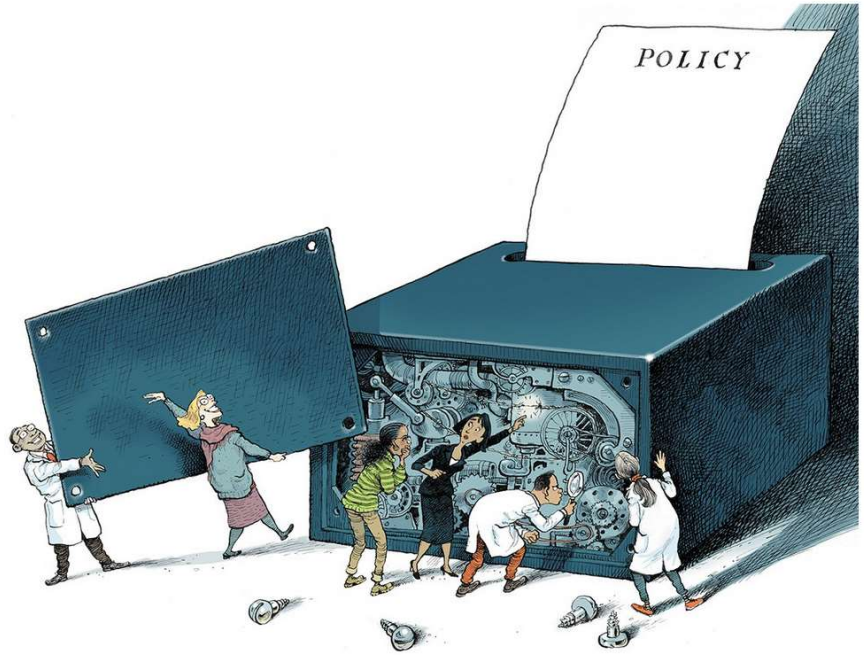
nature

Mind the assumptions

Assess uncertainty and sensitivity

Mind the hubris

Complexity can be the enemy of relevance



Mind the framing

Match purpose and context

Mind the consequences

Quantification can backfire

Mind the unknowns

Acknowledge ignorance

SUPPLEMENTARY INFORMATION

1. Additional information and references >260 references

What if different stakeholders have different preferences? A test case of EU convergence analysis;

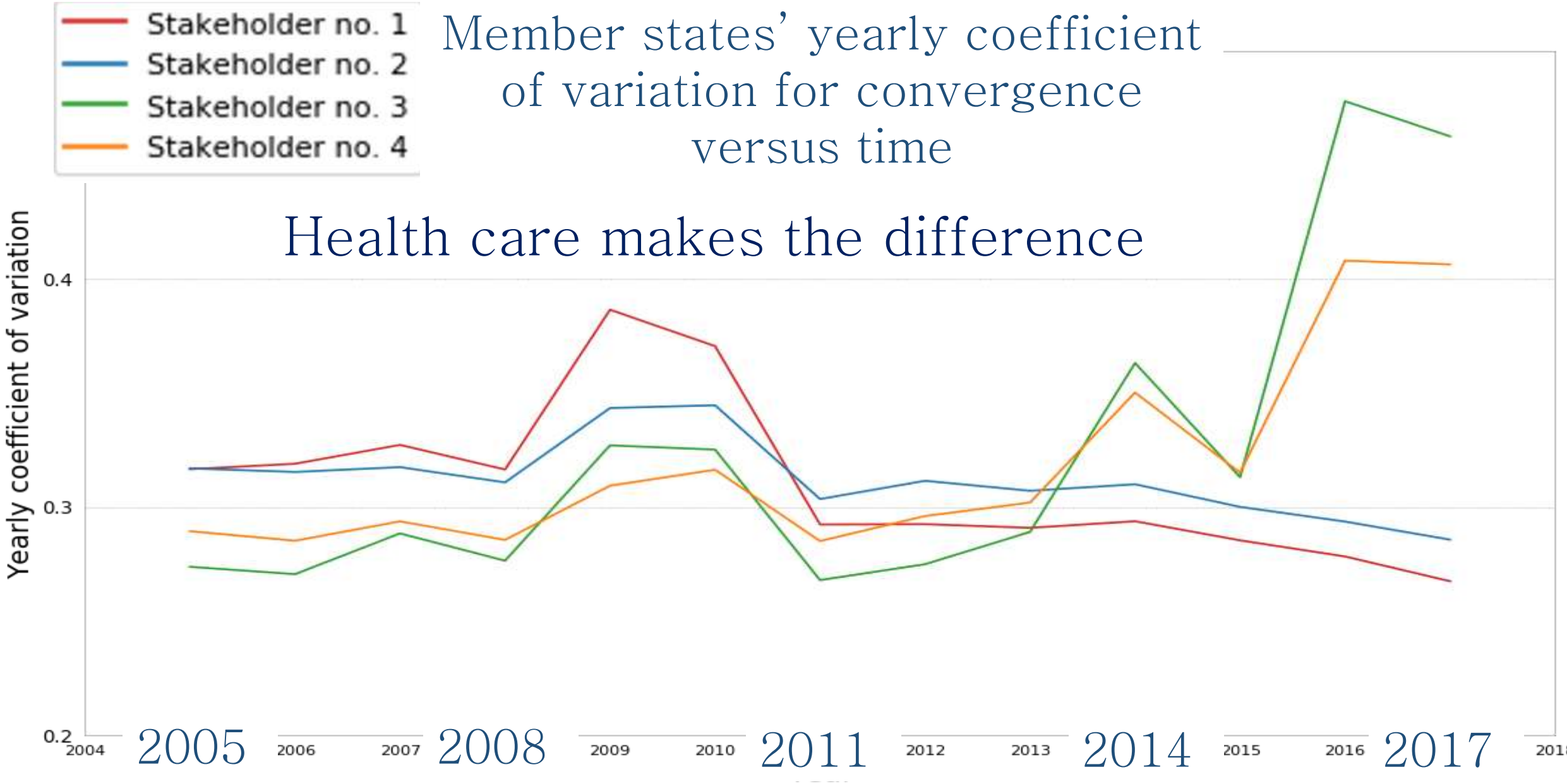
Four different viewpoints are compared

Kuc-Czarnecka, M., Lo Piano, S. and Saltelli, A. (2020) 'Quantitative storytelling in the making of a composite indicator', Social Indicators Research, accepted, http://www.andreasaltelli.eu/file/repository/SIR_TEMP.pdf

Stakeholder 1	Stakeholder 2	Stakeholder 3	Stakeholder 4
Access to labour market	Access to labour market	Access to labour market	Access to labour market
Fair working conditions	Fair working conditions	Fair working conditions	Fair working conditions
Social protection	Social protection	Social protection	Social protection
	Fairness	Health care	Fairness
			Health care

Member states' yearly coefficient of variation for convergence versus time

Health care makes the difference



Deconstructing the implicit normative framing of an indicator

Too much is being read in the OECD-PISA data

IJCED
19,1

20

Received 14 December 2016
Revised 17 February 2017
Accepted 24 February 2017

Do PISA data justify PISA-based education policy?

Luisa Araujo
Department of Human Capital and Employment, European Commission Joint Research Centre Ispra Sector, Ispra, Italy
Andrea Saltelli
University of Bergen, Bergen, Norway and Universitat Autonoma de Barcelona, Barcelona, Spain, and
Sylke V. Schnepf
Competence Centre on Microeconomic Evaluation, European Commission Joint Research Centre Ispra Sector, Ispra, Italy

International Journal of Comparative Education and Development
Vol. 19 No. 1, 2017
pp. 20-34
© Emerald Publishing Limited
2396-7404
DOI 10.1108/IJCED-12-2016-0023

THE CONVERSATION
Academic rigour, journalistic flair

International PISA tests show how evidence-based policy can go wrong

June 12, 2017 3.55pm AEST

Conclusions: CI – instructions for use

Awareness of the imperfections and non-neutrality of measures

Beware damage; mind the interpretant

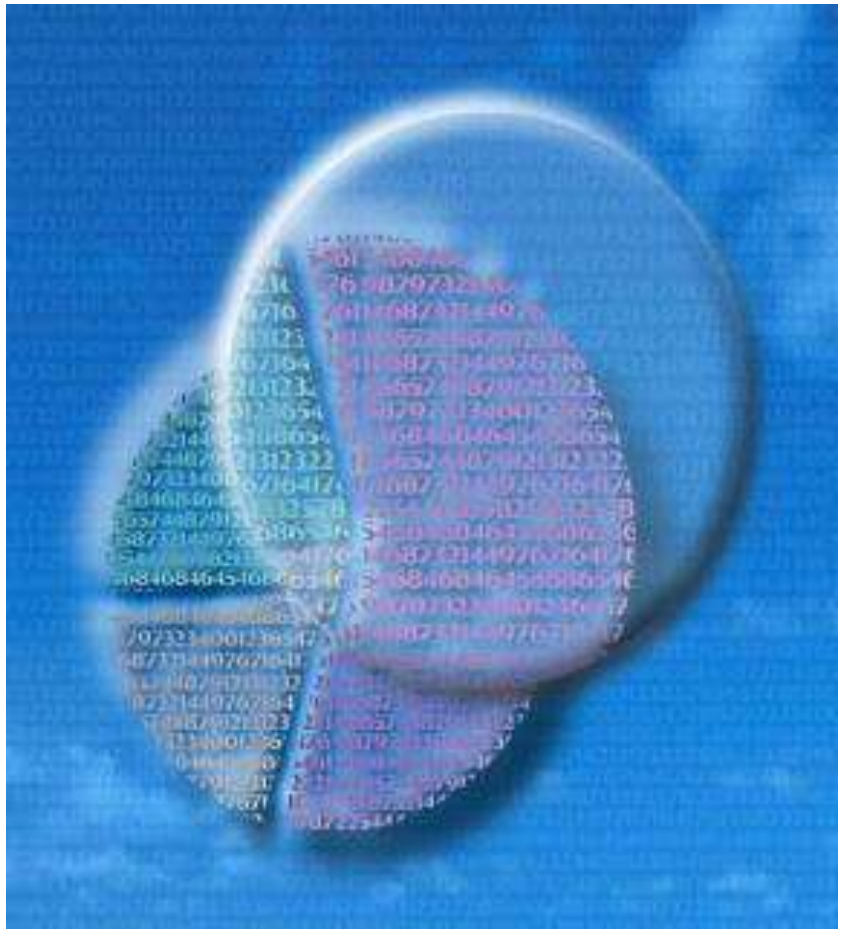
Investigate properties and assumptions (uncertainty and sensitivity analysis, sensitivity auditing)

Use for social discovery; deliberative extended participation; quality as fitness for purpose (interpretant)

Reading material

Becker, W. et al. (2017) 'Weights and Importance in Composite Indicators: Mind the Gap', in Roger Ghanem, David Higdon, H. O. (ed.) Handbook of Uncertainty Quantification. Springer.
http://www.andreasaltelli.eu/file/repository/Full_Copy_CI_Handbook_2017.pdf

Kuc-Czarnecka, M., Lo Piano, S. and Saltelli, A. (2020) 'Quantitative storytelling in the making of a composite indicator', Social Indicators Research, accepted.
http://www.andreasaltelli.eu/file/repository/SIR_TEMP.pdf



END

@andreasaltelli

We shall review some elements of **history** of composite indicators, reasons for and against their use, and available theories for their construction. We shall then touch on some common methodological **conundrums**, with possible solutions based on the theory of sensitivity analysis. After noting the strong normative dimension of these measures—which ultimately aim to 'tell a story', e.g. to promote the social discovery of a particular phenomenon, we inquire whether a less partisan use of a composite indicator can be proposed by allowing more latitude in the framing of its construction. We thus explore whether a composite indicator can be built to tell '**more than one story**' and test this in practical contexts.