

Step 8

Quality control & Robustness

18th JRC Annual training on Composite Indicators and Scoreboards

William Becker

10 STEPS to build a Composite Indicator



Some nice feedback

“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).

Booyesen (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*”

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)

Tools for “serious efforts”

1. **T**ransparency: full record of methodology, reproducible calculations
2. **R**obustness: check the effects of uncertainties on the index
3. **A**ccuracy: automate and try not to make errors
4. **C**redibility: follow established guidelines, check and visualise data
5. **E**ngagement of stakeholders and experts throughout

...better acronym suggestions are welcome

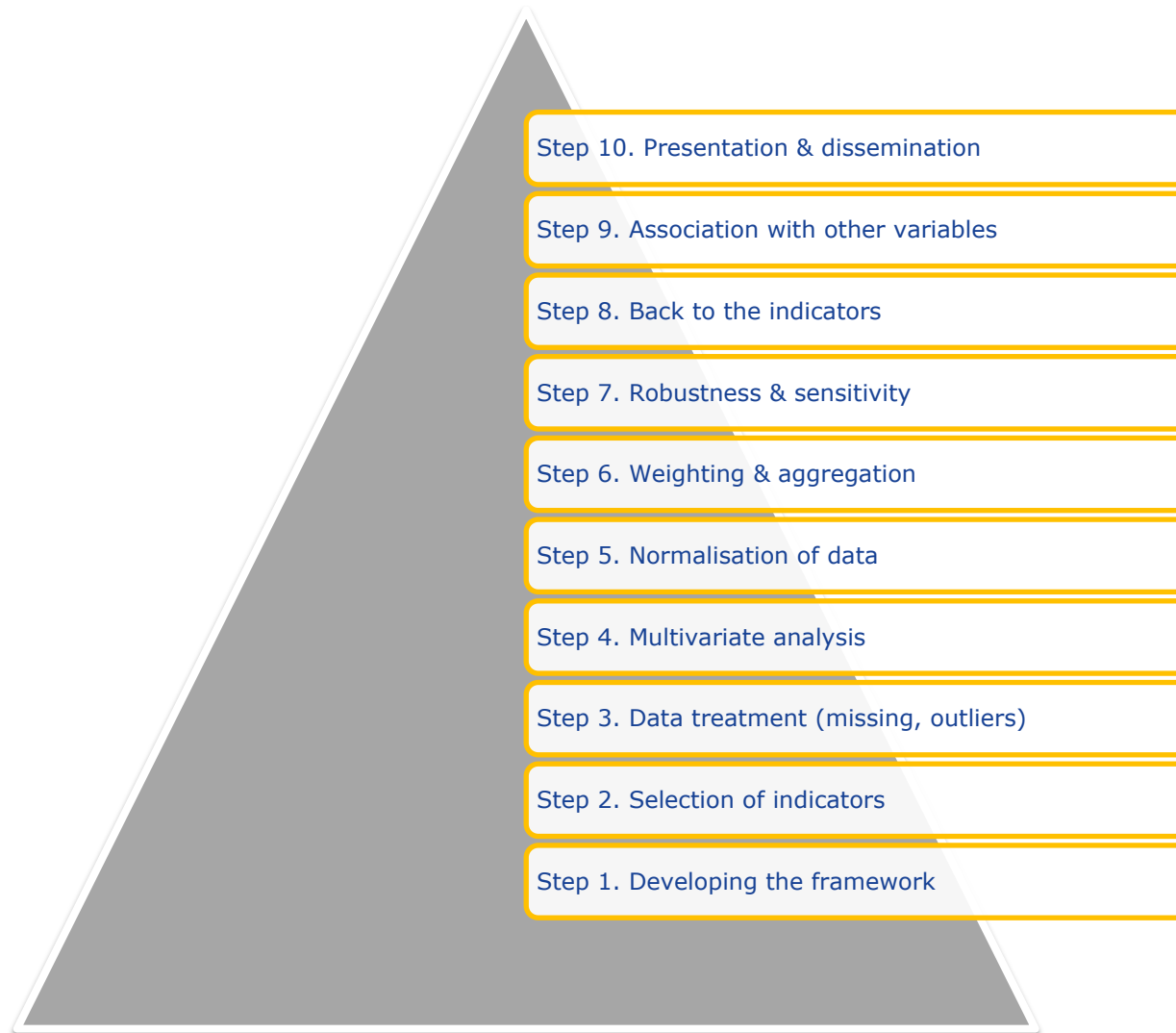
Which steps in the construction of a composite indicator are uncertain?



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



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



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.

Robustness, uncertainty and sensitivity

Uncertainties: assumptions, methodological decisions, parameters, data, etc, used in building the composite indicator, which could plausibly take other values.

Robustness: how stable are the results of the composite indicator, given these uncertainties?

Sensitivity: the opposite of robustness

Uncertainty analysis: what is the uncertainty in the results of the composite indicator, given the uncertainties in its construction?

Sensitivity analysis: how much does each input uncertainty contribute to the output uncertainty?

Example selected uncertainties

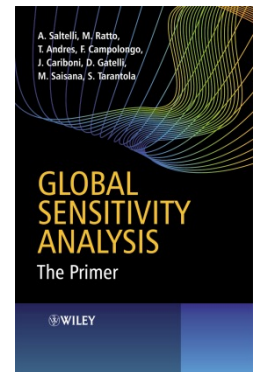
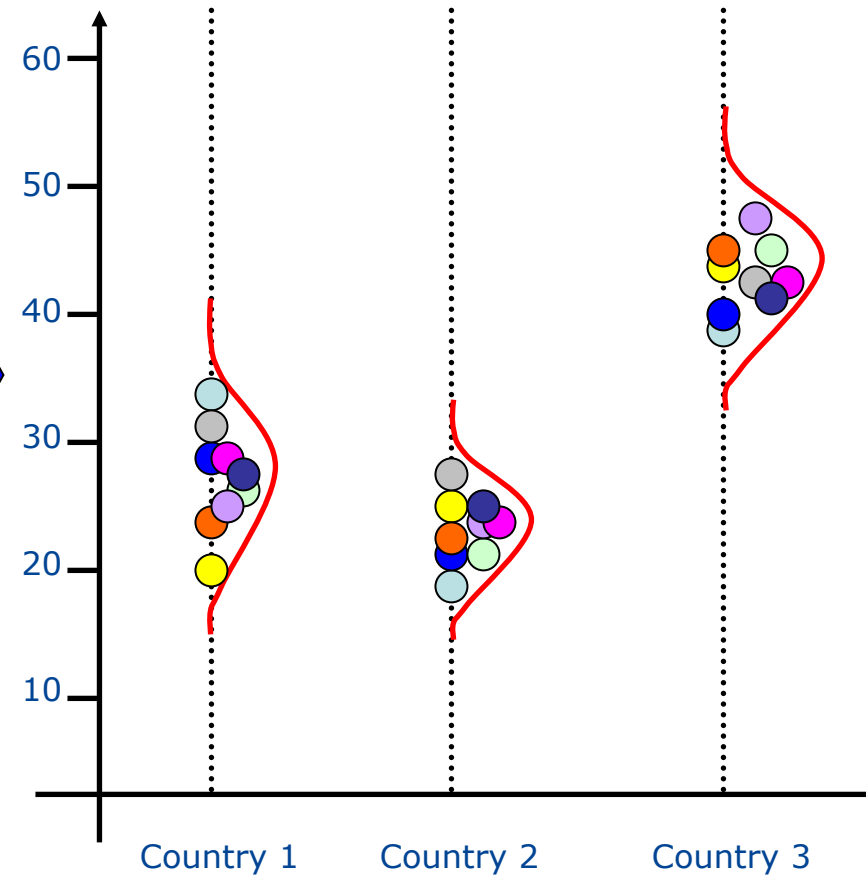
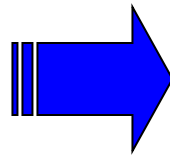
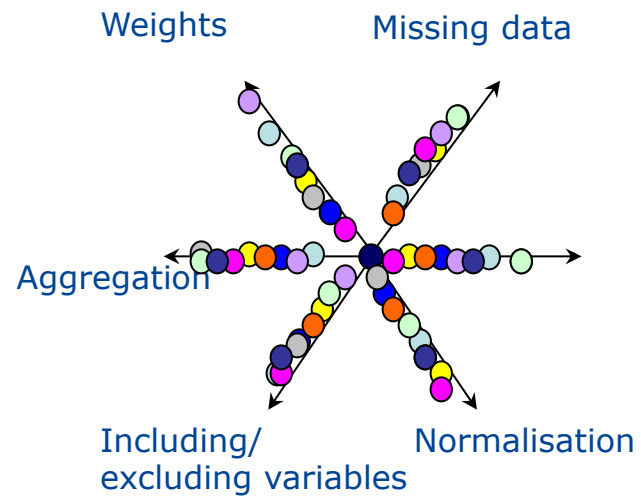
Consider a few assumptions we could test:

Assumption	Nominal	Alternative
Indicator selection	Nominal set	Swap indicator 2 for an alternative data source
Imputation	No imputation (use aggregation group mean)	Indicator group mean
Normalisation	Min-max	Borda scores
Aggregation	Arithmetic mean	Geometric mean at final aggregation level
Weights	Equal weighting	Perturb +/- 25% of nominal value

Note that we can have more than two alternatives for each assumption (only 2 are used here to keep it simple)

Monte Carlo

Space of alternatives



Check your interactions

it is important to vary assumptions simultaneously to fully explore the uncertainty.



Only child 1:
chaos = $f_1(C_1)$



Only child 2:
chaos = $f_2(C_2)$

Child 1 *plus* child 2:

Fighting, crying interactions,
occasional peaceful playing, etc.



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

Testing the behaviour of one child at a time would not at all explore the full space of chaos.

In practice

Running an uncertainty/sensitivity analysis is tricky because it requires rebuilding the composite indicator many times according to an *experimental design*.

Luckily some faster options are available:

1. If you use the **COIN Tool**, you can get a rough idea of uncertainty from the “Scenaria” tab. This is not a strict UA/SA, but a good start.
2. If you use **COINr***, you can perform a full global uncertainty and sensitivity analysis, varying a large number of assumptions.

* COINr is an R package for developing and analysing composite indicators

Scenaria in the COIN Tool

Scenaria

Randomness parameter between [0, 1], default 0.2

Rank summary across all methods				
Unit	Median	Interval	Width	Trimmed
AT	15	[9, 23]	14	[11, 17]
BE	3	[1, 5]	4	[2, 3]
BG	31	[29, 33]	4	[29, 32]
HR	11	[10, 25]	15	[10, 12]
CY	16	[9, 26]	17	[10, 20]
CZ	15	[13, 17]	4	[14, 16]
DK	8	[6, 9]	3	[6, 9]
EE	10	[7, 15]	8	[7, 11]
FI	20	[19, 23]	4	[20, 21]
FR	24	[22, 27]	5	[24, 25]
DE	22	[18, 23]	5	[18, 22]
GR	33	[31, 34]	3	[32, 33]
HU	18	[10, 20]	10	[13, 19]
IE	15	[12, 18]	6	[13, 16]
IT	32	[30, 35]	5	[31, 32]
LV	26	[24, 27]	3	[25, 26]
LT	25	[21, 28]	7	[22, 26]
LU	1	[1, 2]	1	[1, 1]
MT	6	[3, 9]	6	[4, 6]
NL	5	[4, 6]	2	[4, 5]
NO	13	[10, 16]	6	[11, 14]
PL	20	[14, 24]	10	[17, 21]
PT	28	[22, 30]	8	[23, 28]
RO	30	[25, 31]	6	[27, 30]
SK	9	[7, 16]	9	[8, 11]
SI	8	[7, 15]	8	[7, 8]

Indicator ranks		Non-compensatory		
Median	Average	Borda (eq.)	Borda	Copeland
9	16	16	13	31
3	1	1	1	2
24	29	29	28	34
24	22	22	24	14
28	27	27	30	32
16	13	13	14	17
4	6	6	5	15
6	7	7	10	23
11	19	19	21	22
20	25	25	23	21
17	12	12	8	11
29	31	31	31	36
18	18	18	15	18
15	17	17	22	13
29	28	28	25	27
26	26	26	27	25
23	24	24	26	26
1	3	3	3	1
12	10	10	19	9
5	2	2	2	4
9	8	8	6	12
26	20	20	17	10
22	21	21	20	28
32	30	30	29	19
20	15	15	12	5
12	13	13	18	6

NORMALISATION	Minmax					
AGGREGATION	Arithmetic			Geometric		
WEIGHTS	Adjusted	Equal	Random	Adjusted	Equal	Random
	19	11	22	22	14	23
	3	4	3	3	4	3
	32	32	32	32	31	32
	10	12	10	10	10	10
	16	10	23	23	12	24
	14	16	17	14	16	17
	9	7	9	9	7	9
	11	9	13	12	9	15
	23	20	21	21	20	21
	24	24	24	24	24	22
	22	22	18	20	22	18
	33	31	34	33	32	34
	20	19	19	16	19	19
	13	15	12	13	15	12
	31	34	30	31	33	30
	26	26	27	26	26	27
	25	25	26	25	25	26
	1	1	1	1	1	1
	6	5	6	6	5	6
	4	6	4	4	6	4
	12	14	11	11	13	11
	17	21	15	17	21	14
	28	28	29	28	27	29
	27	30	25	27	30	25
	8	13	7	8	11	7
	7	8	8	7	8	8

UA/SA in COINr

Assuming your composite indicator is constructed in COINr...

1. Specify what to vary and alternatives
2. Run the sensitivity analysis

See book chapter:

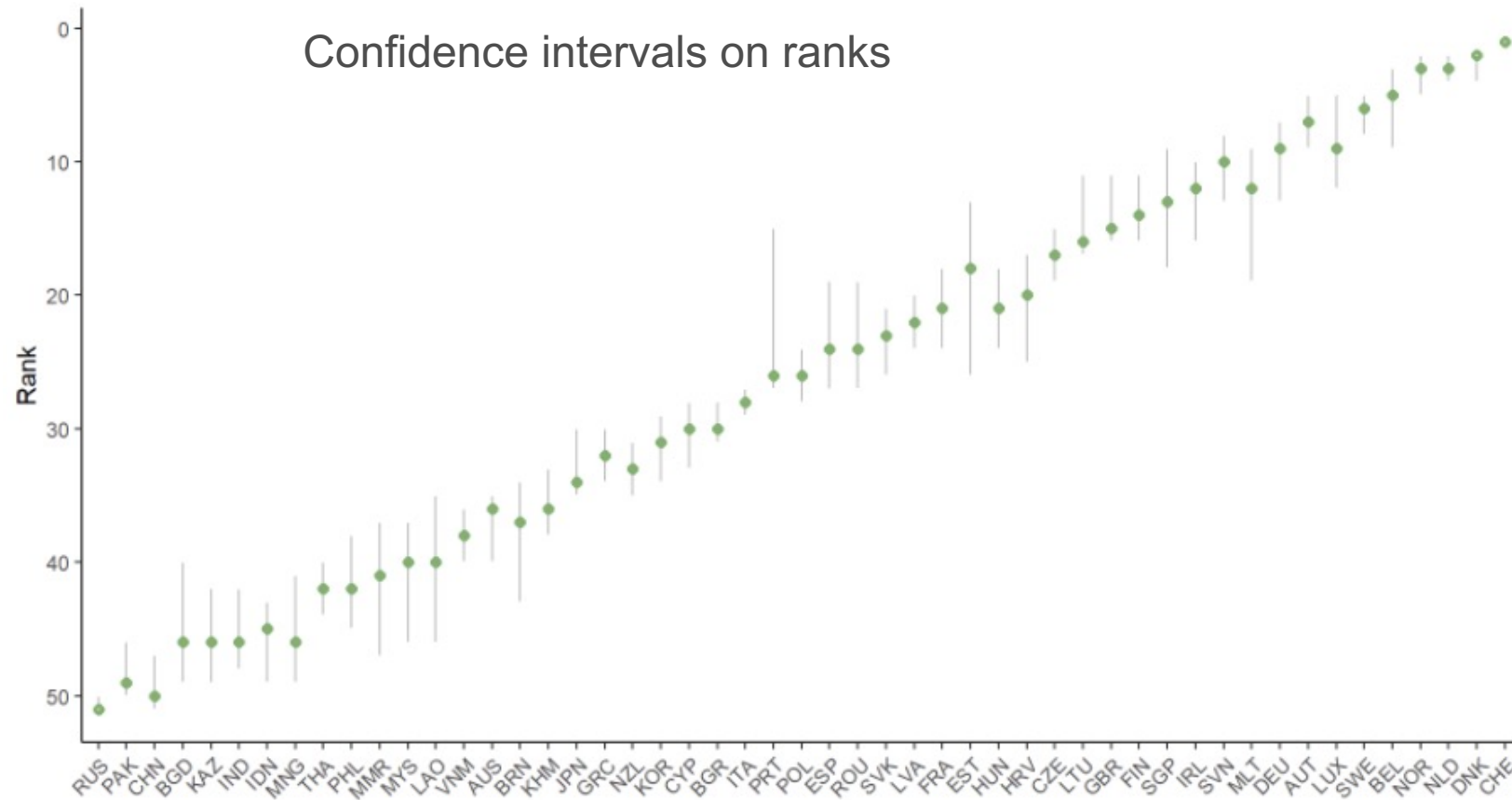
<https://bluefoxr.github.io/COINrDoc/sensitivity-analysis.html#sensitivity-in-coinr>

```
# define noise to be applied to weights
nspecs <- data.frame(AgLevel = c(2,3), NoiseFactor = c(0.25,0.25))

# create list specifying assumptions to vary and alternatives
SAspecs <- list(
  impute = list(imtype = c("indgroup_mean", "ind_mean", "none")),
  normalise = list(ntyte = c("minmax", "rank", "dist2max")),
  weights = list(NoiseSpecs = nspecs, Nominal = "Original")
)
```

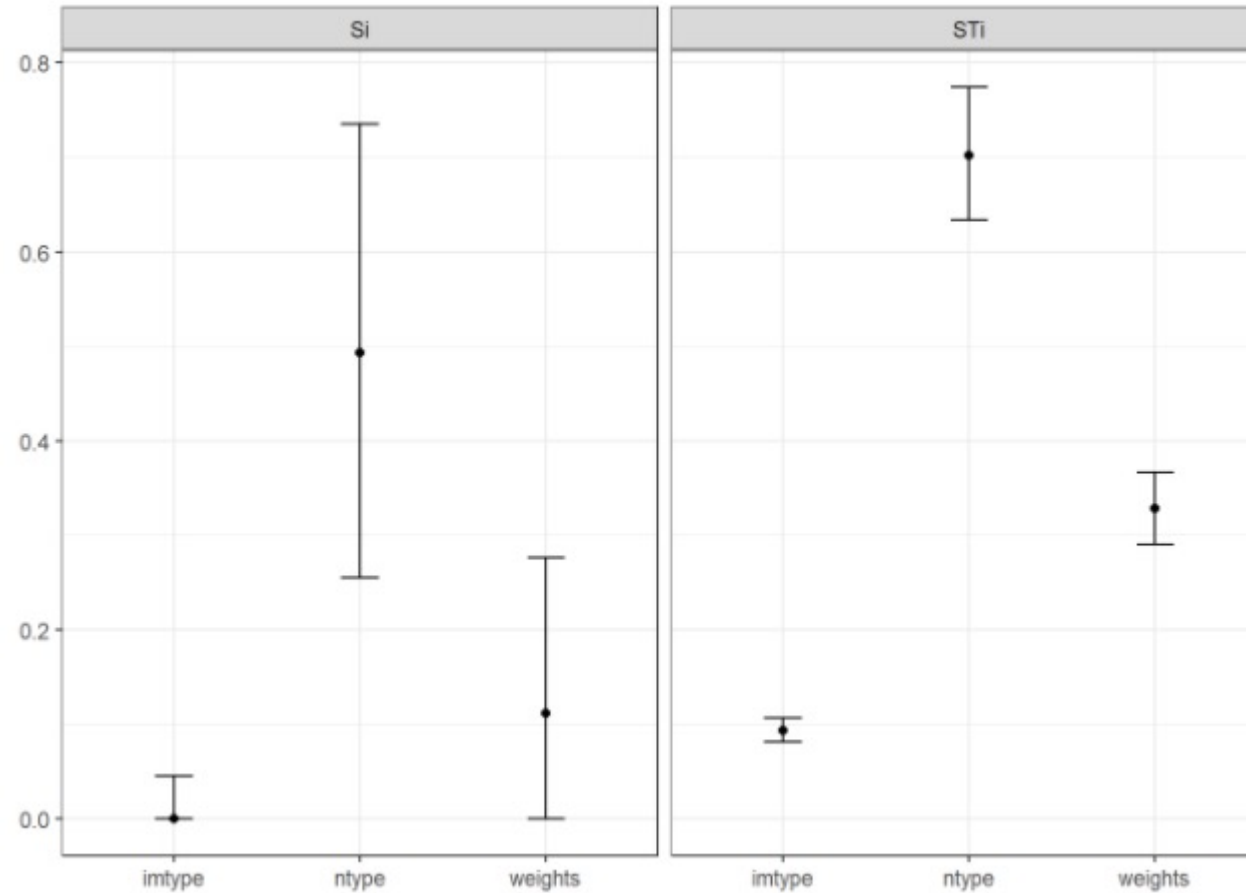
```
# Not actually run here. If you run this, it will take a few minutes
SAResults <- sensitivity(ASEM, v_targ = "Index",
  SA_specs = SAspecs,
  N = 500,
  SA_type = "SA", Nboot = 1000)
```

UA/SA in COINr

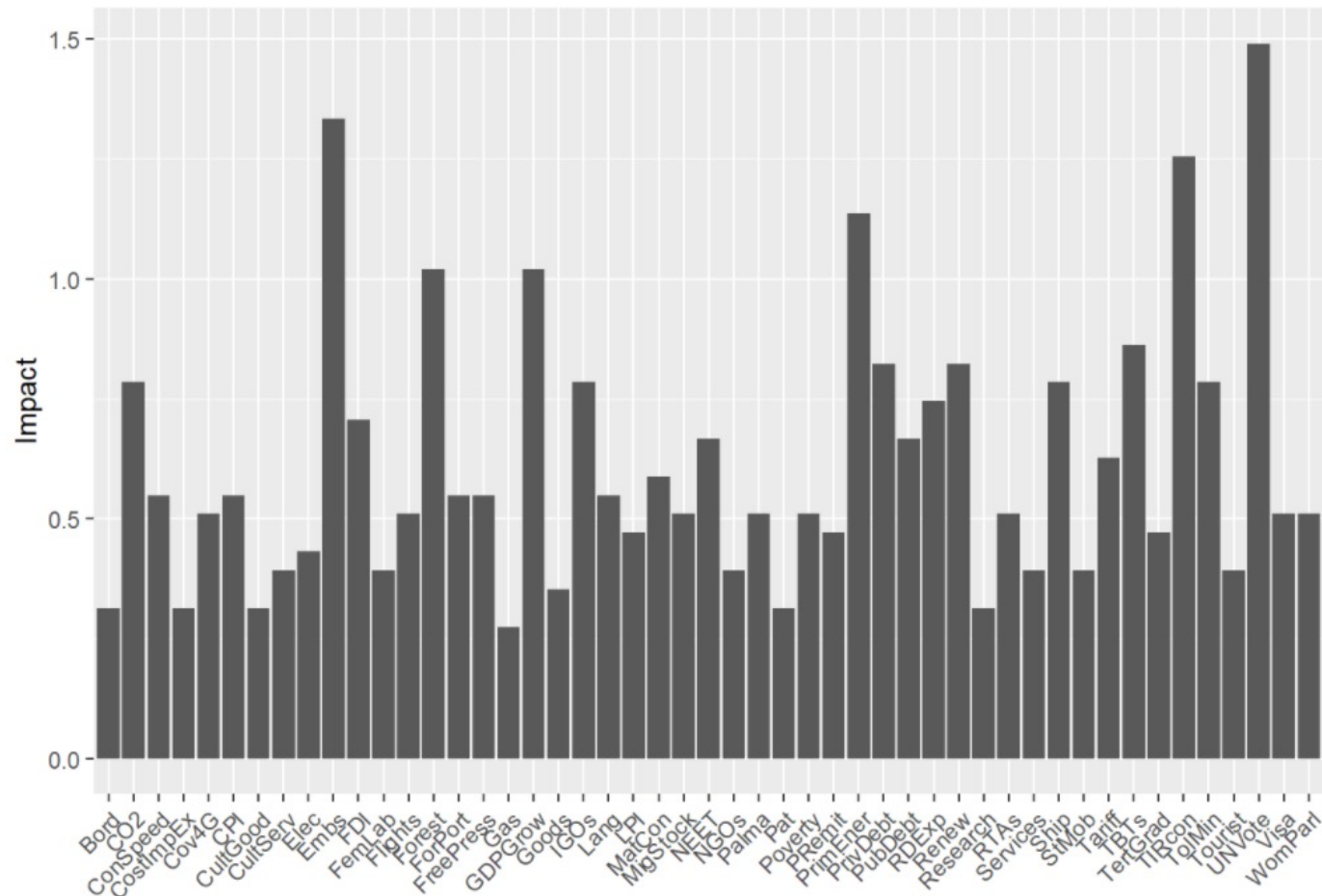


UA/SA in COINr

Bootstrapped sensitivity indices



Indicator sensitivity



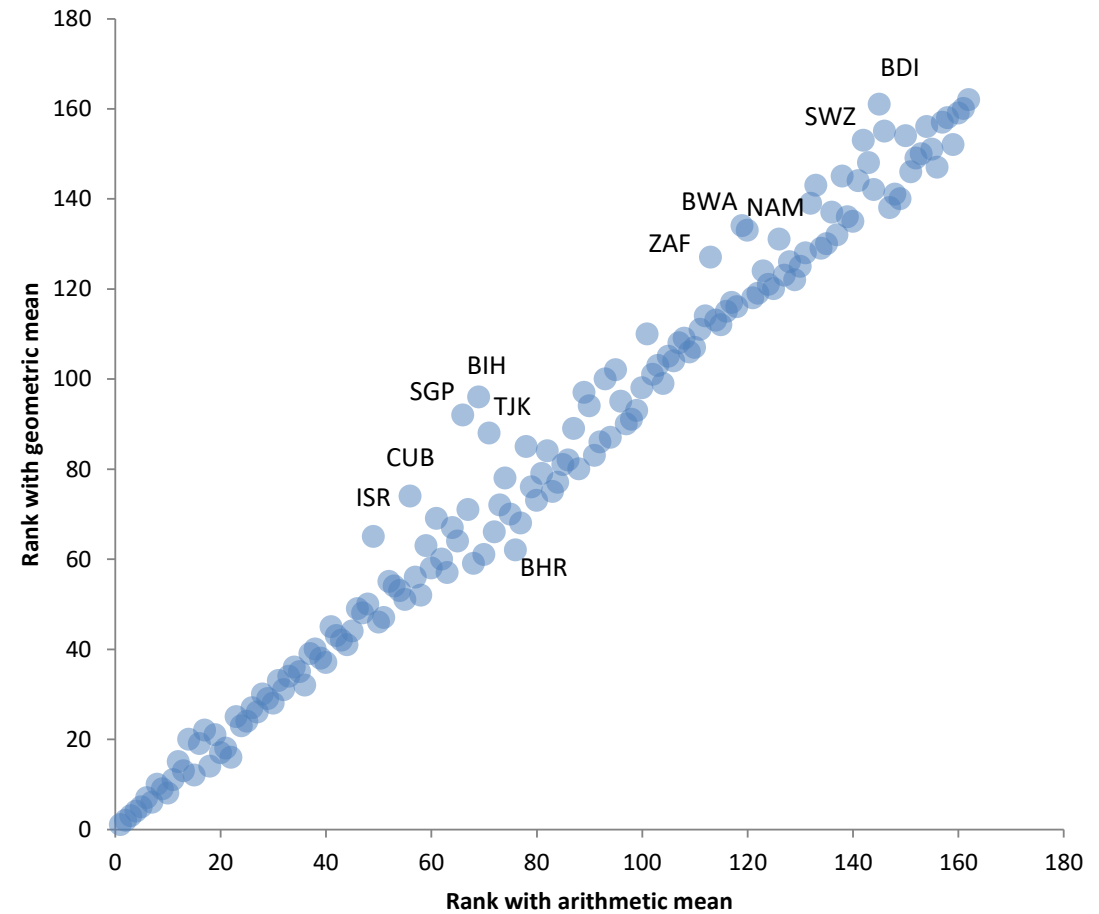
COINr also lets you test the effects of removing indicators and/or whole aggregate groups. This gives an alternative definition of indicator importance (as opposed to correlation).

“What if” experiments

Not strictly “global” but also useful for examining individual cases in detail.

1. Build index
2. Try alternative method
3. Compare the differences

The **scatter** indicates sensitivity. Useful for showing effects on individual countries.



Dominance pairs

Outranking Matrix (pillar level)					
Dominance %:	49.7%				
Dominance pairs:	188				
Total pairs:	378				
Weights	2	1			
Relative weights	0.50	0.25			
	p.01	p.02	p.01	p.02	
	Name	Unit	AT	BE	BG
			CY	CZ	DE
	p.01	p.02	Name	Unit	unit.001
	69.1	84.5	AT	unit.001	-
	52.1	53.4	BE	unit.002	-
	36.3	22.6	BG	unit.003	0.25
	31.8	66.2	CY	unit.004	-
	63.4	70.1	CZ	unit.005	0.25
	59.5	74.7	DE	unit.006	0.25

Another useful metric for robustness is *dominance pairs*, which are calculated from the outranking matrix.

A dominance (robust) pair is a pair of units where one scores better than the other in all indicators or aggregates.

A **higher percentage** of dominance pairs indicates **higher robustness** to weighting and other methodological changes.

COIN Tool gives outranking matrix for pillar level.

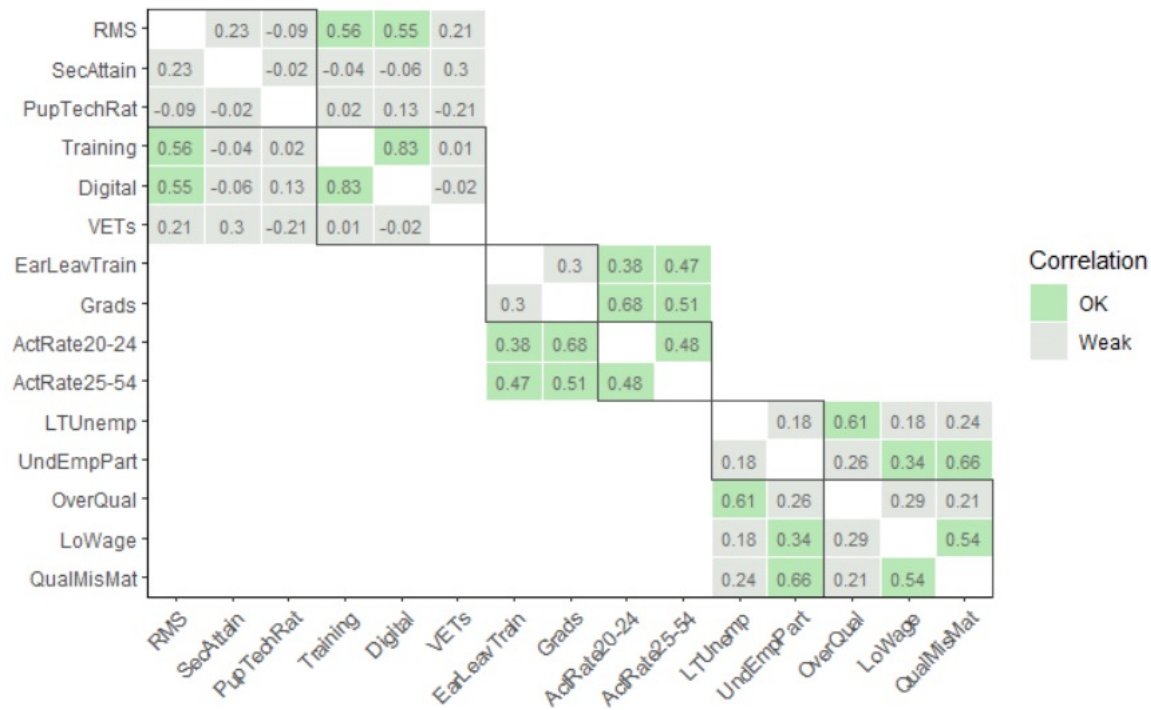
In **COINr**, use the `outrankMatrix()` function for any level and group of indicators/aggregates.

Quality tools: correlations

Correlations are useful because:

1. They help you understand relationships between indicators, spot errors, unexpected features of the data.
2. Correlations between indicators and the index give an idea of how well represented each indicator is in the index.

Quality tools: correlations



Generally it is a good idea to have moderate-high correlations between indicators in aggregation groups. This helps to ensure that:

1. Sufficient information is transferred to the index
2. Each indicator still contributes unique information to the index



Environmental Modelling & Software

Volume 145, November 2021, 105208



A framework based on statistical analysis and stakeholders' preferences to inform weighting in composite indicators

ReW8R v0.2

Correlate this:

Level 1

with this:

Level 2

☐ Separate groups

Correlation type

pearson

Weighting

Select indicator here or by clicking a point on plot.

Goods

Indicator weight

Goods

0

1

Weight sets

Equal weights

Original

Low corr. threshold:

-0.2

High corr. threshold:

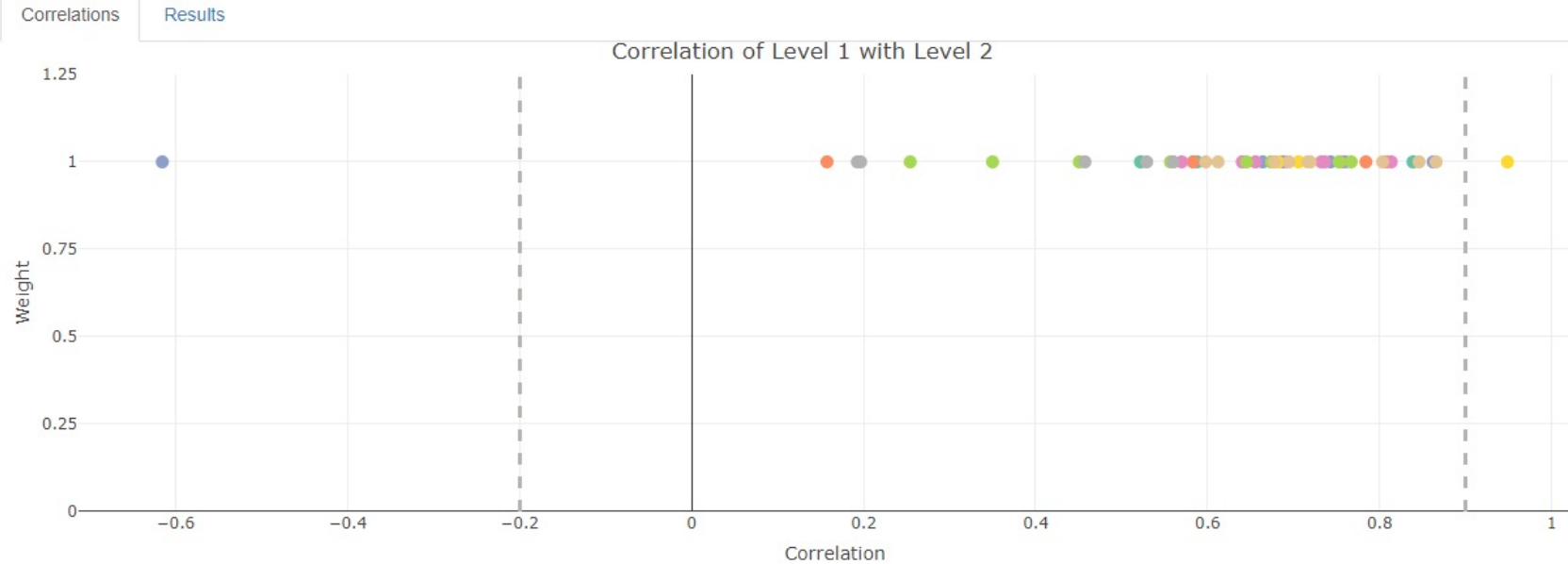
0.9

Weights output

Save as

AltWeights

Close app

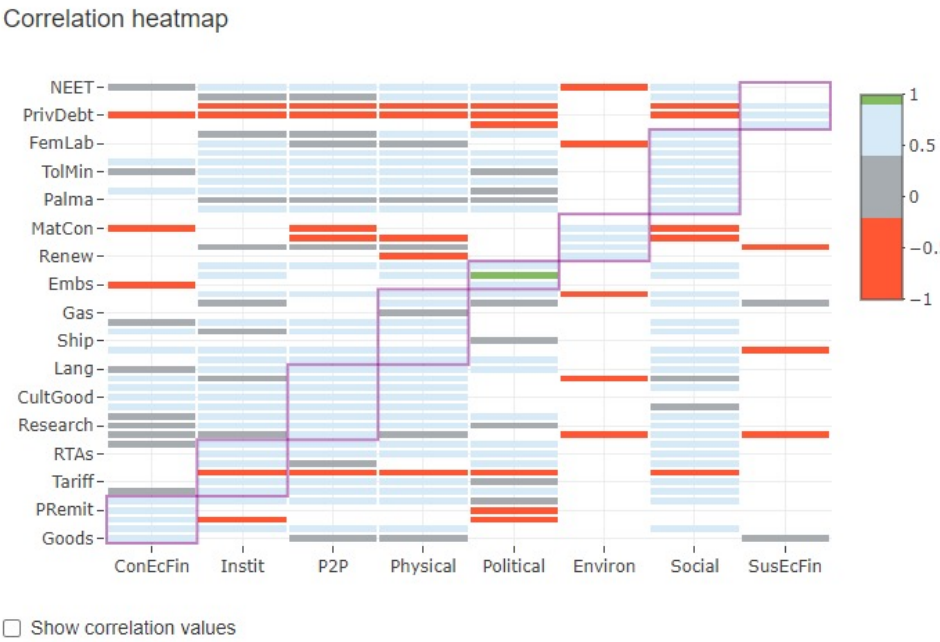


Ind-Ind

Cross-level

Flagged indicators

Indicator	In	Weight	↑ Correlati...	Flag
TBTs	Instit	1	-0.615	Low
IGOs	Political	1	0.949	High



COINr's rew8r app allows you to visualise correlations in detail and interactively adjust weights, immediately seeing the effects.

BUT: don't get too hung up on correlations. They are just one consideration of many.

What about the other four letters?

Transparency

- Keep a record of methodology. If working e.g. in R, build an R Markdown document (or documents). Otherwise take notes in a Word doc or similar.
- Record of indicators included/excluded and why
- Write up methodology and make publicly available
- Make underlying data publicly available
- Aim to make the index reproducible by others

What about the other four letters?

Accuracy

- Automate where possible, avoid cutting and pasting, hand calculations
- Ideally, independently cross-check results in a separate spreadsheet or script
- Clear, logical workflow
- Existing tools can help: COIN Tool, COINr

What about the other four letters?

Credibility

- Follow established guidelines – JRC Handbook, also read methodology of established composite indicators, listen to JRC training *(cough)*
- Investigate and visualise data, pay attention to statistics
- Involve stakeholders (see next)

What about the other four letters?

Engagement

- Understand the purpose of the index, *who* will use, *what for*? Engage end-users from the outset
- Literature review, review of existing indicator frameworks
- Expert workshop for conceptual framework and indicator selection
- “Reality check” workshop – do the results make sense to experts? Is the framework useful? Is it presented helpfully?

From irresponsible musing to serious efforts

Composite indicators

- Are very widely used
- Fill a demand for which there are few or no alternatives

Therefore we should use available tools to increase robustness and credibility

Transparency: full record of methodology

Robustness: check the effects of uncertainties on the index, honestly acknowledge uncertainty.

Accuracy: automate and try not to make errors

Credibility: follow established guidelines, check and visualise data

Engagement of stakeholder and experts throughout

Thank you



© European Union 2021

Unless otherwise noted the reuse of this presentation is authorised under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license. For any use or reproduction of elements that are not owned by the EU, permission may need to be sought directly from the respective right holders.

Robustness, uncertainty and sensitivity

Generally, we want to:

- a) **Quantify** the uncertainty in the composite indicator results (rankings, usually)
- b) See if there is any way to **reduce** the uncertainty (pretty difficult)

Good news:

- Methods are available to achieve this

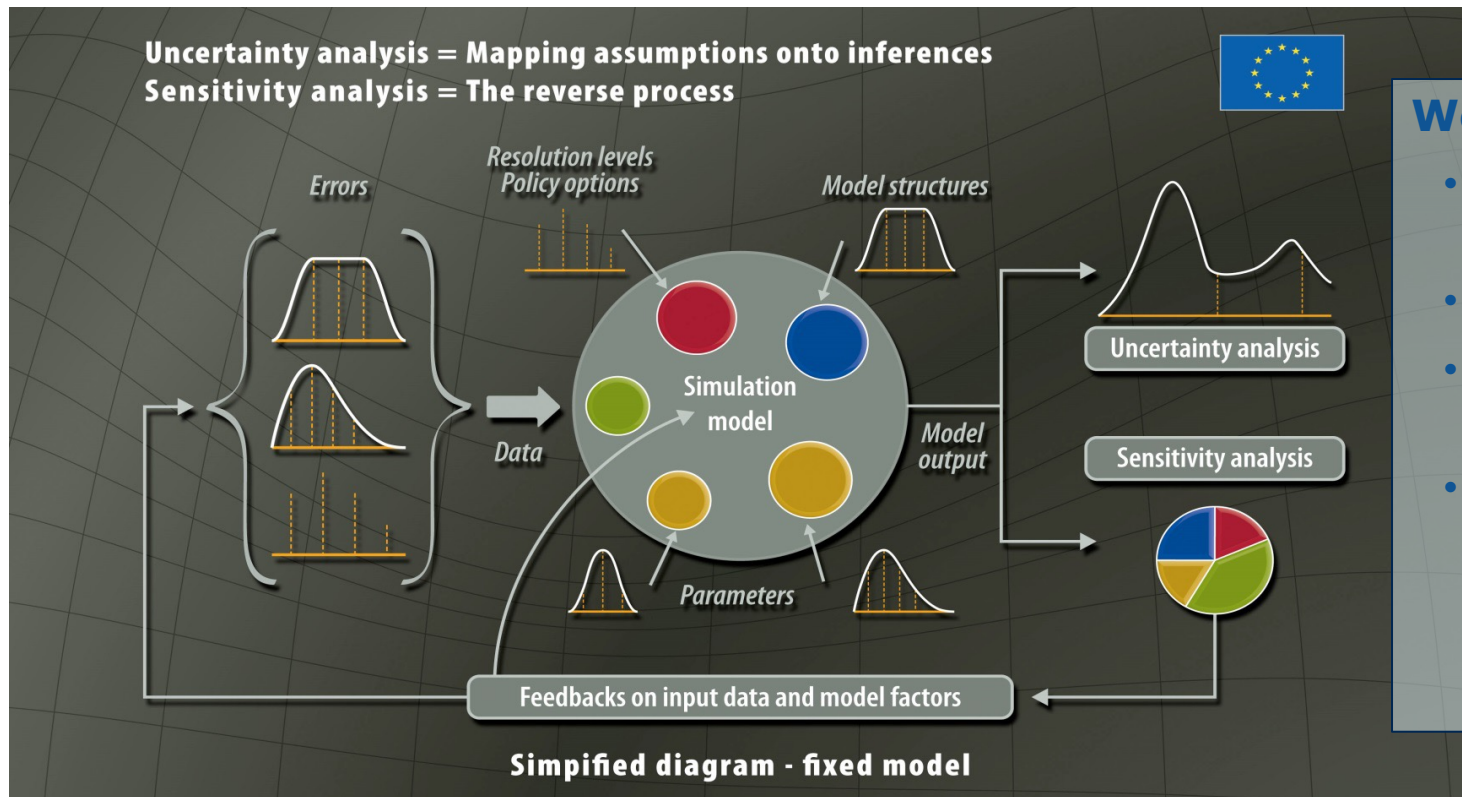
Bad news:

- You will have to get your hands dirty

Good news:

- Software available to help

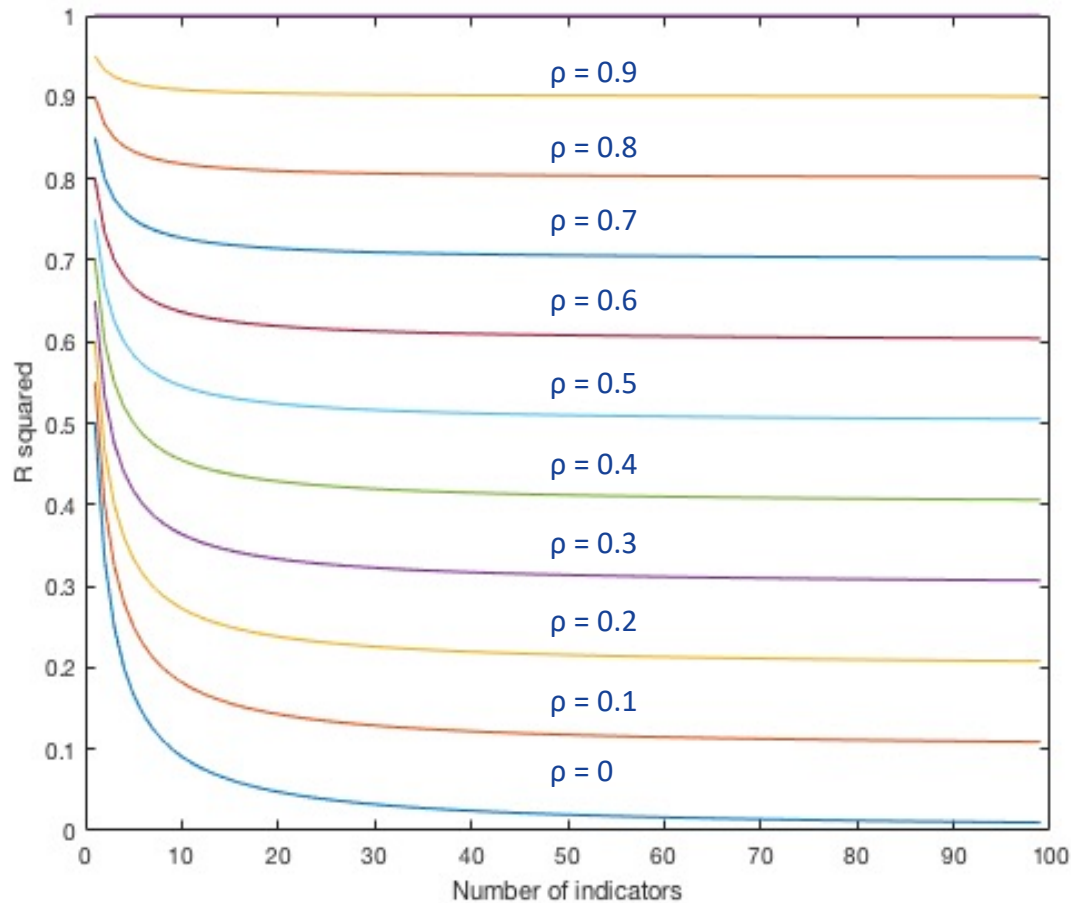
Uncertainty and sensitivity analysis



We need:

- To decide **which assumptions** to test
- To **assign 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)

Quality tools: correlations



Implication: we can only squeeze so much information into the index. Choose indicators carefully.

Av. correlation

between indicators

Information

“transferred”

Unique information of

each indicator