COMPOSITE INDICATORS AS A USEFUL TOOL FOR INTERNATIONAL COMPARISON: THE EUROPE 2020 EXAMPLE

Lenka Hudrliková*

Abstract:

Composite indicators as a tool for a ranking become more and more popular, because they illustrate a comprehensive view on a phenomenon that cannot be captured by only one single indicator. Indicators for Europe 2020 are set of indicators used for monitoring targets defined by the European Commission in the Strategy of Smart, Sustainable and Inclusive Growth. The main objective of this paper is the comparison of performance of the EU Member States using the composite indicator principles. Within constructing composite indicators several steps have to be made and corresponding methods have to be chosen. There is not only one correct method how to develop a composite indicator. Of course, the choice of the methods manipulates the results. Primarily, normalisation methods, weighting schemes and aggregation formulas are fundamental but very subjective. This paper deals with two types of normalisation (z-score and min-max) and four weighting and aggregation schemes (equal weighting with linear aggregation, principal components analysis, benefit of doubt method and multi-criteria analysis). European countries ranking is provided according to the seven different scenarios.

Keywords: international comparison, composite indicator, the Europe 2020 indicators, principal component analysis, benefit of doubt analysis, multi-criteria analysis

JEL Classification: C43, O1, C38, C61

1. Introduction

The set of indicators Europe 2020 is created and used by the European Commission to monitor five headline targets of the Strategy for Smart, Sustainable and Inclusive Growth (European Commission; 2010).[1] It is the main strategy which EU Member States had to adopt. The goal of the smart growth is to raise the employment rate of the population aged 20-64 to 75 %, through the greater involvement of women, older workers and the better integration of migrants into the labour force. Furthermore, the investment in R&D should be increased to 3% of GDP in order to help developing an economy based on knowledge and innovations. Targets of the sustainable growth strive for a more resource efficient, greener and more competitive economy. It should be achieved by (i) a reduction of the greenhouse gas emissions by 20% compared to the 1990 levels, by (ii) an increase in the share of renewable energy sources in final

* University of Economics, Prague. The paper was supported by Grant from Internal Grant Agency of University of Economics, Prague, Project No. IGS F4/19/2011 “One-factor and Multi-factor Productivity Analysis in Context of Input-output Analysis and International Comparison”.

1 More on European Commision website http://ec.europa.eu/europe2020.

energy consumption to 20%, and (ii) by the 20% increase in an energy efficiency. The inclusive growth is focused on support a high-employment economy providing economic, social and territorial cohesion. It is associated with two main targets on educational attainment. The former, the current 15% drop out rate of school leavers should decrease to 10 %. The latter, the share of population aged 30-34 with finished education at tertiary level should be at least 40 % by 2020. The European Commission also struggles with poverty by targeting a reduction of the number of people living below poverty thresholds (defined at national level). All of abovementioned targets must be measurable and comparable. Therefore a set of headline indicators, labelled Europe 2020 indicators,[2] was defined:

Employment rate by gender, age group 20-64	(EMP)
Gross domestic expenditure on R&D	(GERD)
Greenhouse gas emissions, base year 1990	(GH)
Share of renewables in gross final energy consumption	(RE)
Energy intensity of the economy	(EN)
Early leavers from education and training	(EL)
Tertiary educational attainment, age group 30-34	(TE)
Population at risk of poverty or exclusion	(POV)

On the one hand, the indicators are capable of reflecting the diversity of countries performance. On the other hand, they measure fulfilling of the targets over time and are used for purposes of comparing countries. Aforementioned targets are not separated; in fact they are closely interrelated. As well as the international comparison of countries by these indicators should be interrelated. The comparison by means of a set of indicators is not easily feasible and convenient for users because of multivariate outputs. The Composite Indicator (CI) represents an overall view on the full set of indicators. Therefore a Composite Indicator approach is a one of the prospective areas of international comparison. Basically the CI is a tool for avoiding a comparison by means of multiple indicators. This approach is very tempting for all users of statistical information (policymakers, academics, experts, journalists, public, *etc*.) because they can operate with only one figure. OECD Glossary[3] defines: "A composite indicator is formed when individual indicators are compiles into a single index, on the basis of an underlying model of the multi-dimensional concept that is being measured". This definition concludes the scope and usage of the CI. Composite indicators play very important role in policymaking and benchmarking as discussed in Saltelli *et al.* (2006). Critical assessment could be found in Freudenberg (2003). Nardo *et al.* (2005) or OECD (2008) summarised pros and cons of using CIs. The Composite Indicator reflects multi-dimensional issues, assesses progress of countries over time, provides benchmarking and ranks countries on a comprehensive phenomenon. CIs could

2 http://epp.eurostat.ec.europa.eu/portal/page/portal/europe_2020_indicators/headline_indicators

3 OCED Glossary http://stats.oecd.org/glossary/detail.asp?ID=6278

facilitate an interpretation of the results because of reduction of the number of indicators without losing information. On contrary, the CIs can give a misleading message if the CI is poorly constructed or misinterpreted. If one dimension is ignored, it may lead to wrong or simplistic policy conclusions. It could be an easy target of the political disputes and speculations. Each step in construction CIs should be as transparent as can be.

2. Methodological Issues

Data come from 2008, because most of the data for the following years were unavailable, especially the ones regarding the environmental dimension. Many missing values cause more difficulties. Therefore the year with only one missing value was chosen. Nardo *et al.* (2005) suggest several different imputation methods based on case deletion, single imputation or multiple imputations. A case deletion means removing either a country or an indicator from the analysis. It could not be used in the case of EU Member States. The other two types are based on statistical estimation, *e.g.* median/mean substitution, regression imputation, hot-and-cold-deck imputation, expectation maximisation, or multiple imputations (Markov Chain Monte Carlo algorithm). The choice of imputation method needs to take into account the type of analysis. The only one missing value (for the indicator GERD from Greece) was replaced by the value of the previous year, because a significant one-year change was not assumed.

2.1 Outlier detection and normalisation

The following step in the data treatment is outlier detection. The finding of outliers is important at the beginning of the construction, because the presence of outliers could wrongly affect the entire construction of the CI. Saisana (2011) suggests easy rules and methods of detection outliers, *e.g.* the detection of values outside the 1.5 interquartile range. Applying formula for lower boundary $P_{25\%} - 1.5\times(P_{75\%} - P_{25\%})$ and formula $P_{75\%} + 1.5\times(P_{75\%} - P_{25\%})$ for an upper boundary, the outlier can be detected. For univariate data, the rules about simultaneous values of a skewness and a kurtosis were used. The skewness greater than 1 (in the absolute value) and the kurtosis greater than 3.5 flag problematic indicators that need to be treated before conducting the final construction, as stated in Groeneveld and Meeden (1984). In accordance with the aforementioned rules, no indicator for any of the compared countries shows these problems. But it is just a basic method in which *e.g.* multivariate techniques to analyse the structure of the data could be used.

Indicators Europe 2020 are not measured in the same units nor have the same direction. Higher values do not always reflect better performance; *i.e.* some higher value of an indicator represents a worse performance. Therefore data transformation is required prior to the next analysis. The goal of the data transformation is adjusted for different ranges, different variances and outliers. There is a wide scale of normalisation methods *e.g.* showed in OECD (2008). Choosing the most appropriate method for normalization depends on the type of data and on weighting and aggregation. Applying normalization results in different outcomes for the CI. This paper deals with the two most common types: min-max method and z-scores. The first, according to original direction of variable is used min-max formula (1) or (2).

$$I_{qc} = \frac{x_{qc} - min_c(x_q)}{max_c(x_{qc}) - min_c(x_q)} \tag{1}$$

$$I_{qc} = \frac{max_c(x_q) - x_{qc}}{max_c(x_{qc}) - min_c(x_q)} \tag{2}$$

Where x_{qc} is the value of an indicator q for country c. The second, normalisation (z-scores) converts data in order to have the same range (from 0 to 1) and the same positive direction. The advantage of this method is that a wide range of raw data is lying within a small interval. For each indicator x_{qc} the average across countries $x_{qc=\overline{c}}$ and standard deviation across countries $\sigma'_{qc=\overline{c}}$ are calculated and used in formula (3):

$$I_{qc} = \frac{x_{qc} - x_{qc=\overline{c}}}{\sigma'_{qc=\overline{c}}} \tag{3}$$

After normalisation, the data have a common scale with a zero mean and standard deviation of one. This method provides no distortion from the mean.

2.2 Weighting and aggregation

Weighting and aggregation systems have a crucial effect on the outcome of the CI. There is not only one proper method. That is why this part of constructing a CI is the most discussed and criticized by opponent of CIs. We can divide methods of weighting into two main groups: statistical approaches and participatory approaches. The list of the most common method is in OECD (2008). One of the participatory methods, called (i) a budget allocation process, is based on a simple idea, to bring together a wide spectrum of experts. They should be concerned, have relevant knowledge and experience. Each of them gets certain 'budget', *e.g.* 100 points and should divide them among indicators according to the weights they should have in the CI. Ii other words they are asked to construct own weighting system. There are some problems which make these methods not so worthwhile. The biggest problem is the selection of experts, their number and background. The experts could face a problem when the number of indicators is higher than 10. They could prefer a field where they are from. (ii) The public opinion method is similar to the budget allocation process but there is a pool of people instead of experts. So the usage of the method is even more difficult. (iii) Conjoint analysis is the next participatory method. It is based on the idea of asking respondents how much importance they give to an individual indicator. In other words, the idea is a "willingness to pay" of the respondents. What they have already in their "basket" have an influence of their willingness. (iv) The analysis hierarchy process is based on using pairwise comparisons. It compares a relative importance of one criterion over another. The main advantages are toleration of inconsistency and pairwise comparisons which can be easily handled by human thinking. But this survey is more time consuming than budget allocation. It is obvious there is a need for (N*(N-1)/2 pairs of comparison. Other disadvantages are the same as in the previous method, i.e. design of survey, less than 10 indicators.

For EU2020 indicators, no results are available for participatory methods. So this paper deals with the first group of weighting methods (*i.e.* statistical methods). Participatory methods are mentioned only for covering all perspectives in constructing weighting systems. The weighting schemes considered in the paper are (i) equal weighting, (ii) principal component analysis and (iii) the benefit of doubt approach. This entire group is based on statistical methods, which are only data driven. There is no need for any subjective value judgment. Using Equal weighting (EW) method, the equal weight is assigned for each indicator:

$$w_q = {}^{1}\!/_{Q} \tag{4}$$

Where w_q is a weight for q^{th} sub-indicator ($q = 1, ..., Q$) for each country. It means all sub-indicators are given the same weight for all countries. There is a risk that a pillar with more indicators will have a higher influence in the CI. Equal weighting may be justified when there is no clear idea what else could be used. The main strength of the method is the simplicity.

A detailed description of Principal component analysis (PCA) can be found in a number of sources *e.g.* Manly (2004), Morrison (2005), textbooks or handbooks about statistical software *e.g.* StatSoft (2011), in a connection with CIs *e.g.* OECD (2008). In this paper only the main idea and short description is presented. The PCA analyses the structure of the data. Before using the PCA as a weighting method, it is useful to use it also as an explanatory analysis. The PCA explains the variance of the data through a few factors which are the linear combinations of the raw data. The original correlated set of indicators is changed into a new smaller set of uncorrelated variables. The motive of using PCA as a weighting method is taking into account correlations among indicators.

The Benefit of Doubt Approach (BOD) is based on the Data Envelopment Analysis (DEA) which is commonly used in production problems. Using the BOD (or DEA) as a weighting method is elaborated in Cherchye *et al*. (2008). There is not one weighting scheme for all countries. Each country has its own weights which are optimal for this country. It guarantees the best possible position for the associated country among all other countries in the sample. With any other weighting scheme, the relative position of that country would be worse. Optimal weights are obtained by solving the following constrained optimisation for each country:[4]

$$CI_c^* = max\sum_{q=1}^{Q} I_{qc} w_q \tag{5}$$

Subject to

$$\sum_{q=1}^{Q} I_{qc} w_q \leq 1$$

$$w_q \geq 0$$

$$\forall k = 1, ..., M; \;\; \forall q = 1, ..., Q$$

Where c means countries and q means sub-indicators. To choose weights that maximize the composite indicator score is a problem of linear programming. The highest relative weights are assigned to the sub-indicators for which the country j achieves the best relative performance in comparison to the other countries. The only restrictions are their non-negativity and normalisation. As a result, we obtain value of CI between 0 and 1, the higher value, the better performance in relative terms. For more countries the value of a composite indicator could be equal to 1. The main disadvantage of this method is that without a setting constraints the weight is given by an indicator in which the country is the best (Cherchye *et al.*, 2009). The results are influenced by the fact that countries which perform very well only in one indicator can be considered successful. On the other hand, a setup of weights constrictions needs to take subjective opinions into account. Cherchye *et al.* (2008) gives an example in the case of the Technological Achievement Index. They set relative weight constrictions called pie share constraints. Setting constraints for each indicator is advantageous if there are some results from a participatory approach. In our case no additional information is available, thus no more constraints for each indicator than those included in (5) are assumed. Another drawback is that it does not provide scores, just ranks. To be more clear, there are BOD scores but they are directly incomparable. For this reason, the paper do not include the BOD scores. The strong point of this method is making an aggregation method redundant! In this paper choosing the aggregation method makes sense only for equal weighting and PCA weights.

In practice, linear aggregation (LIN) is the most widespread. The simplest method is the linear method:

$$CI_c = \sum_{q=1}^{Q} I_{qc} \cdot w_q \tag{6}$$

Subject to

$$\sum_q w_q = 1 \text{ and } 0 \leq w_q \leq 1.$$

Where I_{qc} is a normalized indicator q (q = 1, ..., Q) for country c (c = 1, ..., M) and w_q weight for indicator q (q = 1, ..., Q). The fundamental topic of the aggregation is

4 See also OECD (2008), p. 83.

compensability. Linear aggregation implies full compensability. Poor performance in one indicator can be compensated by sufficiently high values of others indicators. The trade off between indicators is not always desirable. Note that full compensability could be weakened by the weighting scheme. OECD (2008) also suggests different aggregation methods: geometric aggregation and the non-compensatory multi-criteria approach. Geometric aggregation is partially compensable, because it rewards more countries with higher scores. Hence countries with low scores should prefer a linear rather than a geometric aggregation. Geometric aggregation is possible for strictly positive data. It is not this case due to the chosen normalisation methods. Multi-Criteria Analysis (MCA) represents the non-compensatory approach. The drawback of it is the difficulty in computing when the number of countries is high. In this case, 27 European countries count 27! permutations. For further details refer to Munda and Nardo (2009). In order to avoid aforementioned computational problem, Arrow-Raynaud rand Copeland rules were. Both rules are based on Condorchet approach and underlying outranking matrix created by means of pair-wise comparisons. Only ordinal information is relevant which means the method is independent to outliers. The detailed description of the rules can be found in Munda (2010).

2.3 Robustness analysis

This part of building the CI can help to deal with opponents and indeed improve the CI. Uncertainty and sensitivity analysis assess the robustness of the CI. Within constructing the CI, there are several subjective judgements that have to be made: method of normalization, weighting scheme and aggregation. In the paper a comparison of conceivable scenarios according type of the data and the statistical methods is conducted. It can be considered as a part of uncertainty analysis. Average shift in rank (7) is a tool for assessing the robustness.

$$\bar{R}_s = \frac{1}{M} \sum_{c=1}^{M} \left| Rank_{ref}(CI_c) - Rank(CI_c) \right| \tag{7}$$

This measure of the shifts in rankings is calculated as an average of the absolute difference in countries` ranks in respect of a reference ranking over all *M* countries. As the reference ranking is perceived the median rank.

3. Results

3.1 Weighting scheme

There is no general consensus on using one of weighting scheme. The most common is equal weighting due to simplicity and transparency. When there is no clear idea on what method to use, it is possible to prefer a simplicity. In the case of indicators Europe 2020 there are eight indicators. It means that according to formula (4) the weight on each single indicator for each country equals 0.125. According to Saisana (2011) other weighting methods may be justified when there are few indicators (between 3 and 10) and bivariate correlations less than 0.50. As a matter of tact the significance of the weights is given by the correlation between indicators. Table 1 shows moderate or low correlations between Europe 2020 indicators.

Table 1

Correlation Matrix of Europe 2020 Indicators

	EMP	GERD	GH	RE	EN	LE	TE	POV
EMP	1	0.566**	-0.076	0.491**	-0.183	-0.313	0.521**	-0.380
GERD	0.566**	1	-0.001	0.520**	-0.490**	-0.175	0.432*	-0.604**
GH	-0.076	-0.001	1	-0.242	-0.549**	0.440*	0.251	-0.308
RE	0.491**	0.520**	-0.242	1	0.005	-0.072	0.026	0.008
EN	-0.183	-0.490**	-0.549**	0.005	1	-0.215	-0.432*	0.634**
LE	-0.313	-0.175	0.440*	-0.072	-0.215	1	-0.153	0.116
TE	0.521**	0.432*	0.251	0.026	-0.432*	-0.153	1	-0.371
POV	-0.380	-0.604**	-0.308	0.008	0.634**	0.116	-0.371	1

Source: Computation of author

Note: * (**) Correlation is significant at the level 0.05 (0.01), 2-tailed

Nevertheless, there is no unambiguous view on the correlation in the case of CIs. Saltelli (2011) argued two approaches dealing with correlations among indicators. On the one hand, high correlations among indicators could be seen as a problem which should be correct by making appropriate weights. The weight should be set inversely proportional to the strength of the correlation for a given indicator. On the other hand, high correlations should not be corrected because it is a feature of the measured comprehensive phenomenon. Highly correlated indicators could indicate non-compensable different dimensions within the phenomenon. To conclude, correlations imply that indicators measure the same phenomenon, but must be checked whether the correlation could not be caused by the redundancy of the information (double counting of the dimension).

Table 2

Weights for the Europe 2020 Indicators Based on Different Methods

		EMP	GERD	GH	RE	EN	EL	TE	POV
EW		0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
PCA[5]		0.117	0.132	0.123	0.149	0.131	0.138	0.098	0.113
BoD[6]	**Belgium**	0.115	0.128	0.115	0.128	0.186	0.019	0.241	0.217
	Germany	0.182	0.171	0.182	0.171	0.194	0.050	0.236	0.199
	France	0.134	0.129	0.134	0.129	0.169	0.062	0.233	0.200
	Italy	0.162	0.168	0.162	0.168	0.262	0.085	0.381	0.286
	Luxembourg	0.122	0.088	0.122	0.088	0.176	0.012	0.251	0.199
	Netherlands	0.204	0.089	0.204	0.089	0.142	0.015	0.202	0.177
	Denmark	0.190	0.078	0.190	0.078	0.131	0.009	0.187	0.165
	Ireland	0.163	0.082	0.163	0.082	0.125	0.022	0.262	0.213
	United Kingdom	0.168	0.085	0.168	0.085	0.130	0.022	0.268	0.217
	Greece	0.214	0.091	0.214	0.091	0.193	0.045	0.390	0.311
	Spain	0.177	0.116	0.177	0.116	0.142	0.098	0.374	0.077
	Portugal	0.294	0.147	0.294	0.147	0.185	0.233	0.404	0.037
	Austria	0.281	0.255	0.281	0.255	0.208	0.239	0.066	0.316
	Finland	0.137	0.175	0.137	0.175	0.110	0.120	0.151	0.150
	Sweden	0.150	0.149	0.150	0.149	0.106	0.150	0.141	0.117
	Czech Republic	0.187	0.095	0.187	0.095	0.243	0.047	0.143	0.294
	Estonia	0.238	0.075	0.238	0.075	0.274	0.121	0.116	0.206
	Cyprus	0.231	0.000	0.231	0.000	0.000	0.025	0.244	0.209
	Latvia	0.233	0.017	0.233	0.017	0.297	0.200	0.222	0.203
	Lithuania	0.168	0.032	0.168	0.032	0.270	0.095	0.170	0.258
	Hungary	0.063	0.084	0.063	0.084	0.382	0.070	0.299	0.383
	Malta	0.000	0.027	0.000	0.027	0.200	0.000	0.539	0.000
	Poland	0.024	0.032	0.213	0.186	0.389	0.176	0.305	0.314
	Slovenia	0.180	0.103	0.180	0.103	0.145	0.093	0.224	0.275
	Slovakia	0.139	0.029	0.211	0.185	0.382	0.174	0.299	0.304
	Bulgaria	0.181	0.030	0.312	0.009	0.504	0.119	0.000	0.404
	Romania	0.150	0.029	0.261	0.187	0.476	0.195	0.353	0.410

Source: Computation of author

5 PCA could give different results using maximum likelihood method or rotation.

6 BOD could give different results as well with additional constraints.

The weights derived from the PCA take into account the correlation. Two highly correlated indicators will be given higher weights. The need of a correlation between indicators is rooted in the PCA. The correlation structure (see Table 1) does not fully meet the assumption of using the PCA. But result of Kaiser-Meyer-Olkin test statistics 0.608 is in favour of applying the PCA. Barlett's test of sphericity indicated that the hypothesis of uncorrelated indicators can be rejected. Test statistics equals 85.158 (Sig. 0.000), thus the PCA can be applied. The weights derived from the PCA are based on eigenvalues. The optimal numbers of components must be decided. Kaiser criterion suggests selecting all components which are associated to eigenvalues higher than one. Applying that criterion, three components were selected. The weights had to be normalized by squared factor loadings, which are the portions of the variance of the factor explained by the variable. At the end they were scaled to unity sum. The performance of the equal weights, PCA weight and BOD weights results are compared in Table 2. Remind that the weights gained by the BOD method are different for each country.

Even though the BOD weights are presented, they are not applicable for a direct comparison of countries. To sum up the weights, they are not equal one. According to the BOD formula (5) there is no constraint about sum of the weights. If there were one, the results would be an entire weight on only one indicator in most cases. The results obtained by BOD approach show in which indicator is the individual country strong and *vice versa*. For example, the Czech Republic has a good performance at indicator of poverty and social exclusion (POV) and on the other hands the worst performance in EL (*i.e*. early leavers from education and training). In this case, the BOD method without a constraint assigned the indicators values between 0 and 0.54. The constraints were not applied because only the methods based on objective statistical methods were used.

3.2 Ranking countries

The BOD method provides the best weights for each country. It results in the value of the CIs 1 for each country. After the BOD results, a cross efficiency (C-E) is needed for valid ranking because using different weights for each country is inconsistent and lead to the incorrect ranking. Cross efficiency, carried out in Hoolingsworth and Wildman (2002), is based on the comparison of every country with all other countries, applying the weights of others. By performing these comparisons a matrix of cross efficiencies with dimension 27 x 27 is created. For ranking is used the median rank from that matrix. Obviously, any country cannot have the original BOD score lower than the median cross efficiency score. Without using weight the multi-criteria analyses were conducted. The different rules was employed - the first Arrow-Raynaud rule and the second Copeland rule. Table 3 summarizes final rankings by seven different scenarios.

Table 3

EU Country Rankings by Different Weighting and Aggregation Methods

	EW LIN min-max	EW LIN z-score	PCA LIN min-max	PCA LIN z-score	C-E BOD	MCA Arrow-Raynaud rule	MCA Copeland rule	Absolute maximum difference in rank	Median rank
Sweden	1	1	1	1	1	1	1	0	1
Finland	2	2	2	2	3	2	4	2	2
Denmark	3	3	3	3	2	2	3	1	3
Austria	5	4	4	4	6	4	5	2	4
Netherlands	4	5	5	6	4	7	2	5	5
Germany	7	6	6	5	7	5	6	2	6
France	6	7	7	7	5	6	7	2	7
Slovenia	9	8	8	8	12	9	8	4	8
United Kingdom	8	9	9	10	9	8	12	4	9
Estonia	10	10	10	9	13	12	11	4	10
Belgium	12	12	12	13	11	12	9	4	12
Ireland	11	14	13	15	10	12	9	6	12
Lithuania	15	11	15	12	15	12	12	4	12
Luxembourg	13	13	14	14	8	11	12	6	13
Latvia	14	15	11	11	14	12	16	5	14
Czech Republic	16	16	16	16	16	10	12	6	16
Cyprus	17	17	17	18	17	24	19	7	17
Slovakia	20	18	18	17	18	24	19	7	18
Poland	19	19	19	19	19	19	21	2	19
Spain	18	20	20	21	21	18	18	3	20
Portugal	21	21	21	20	24	12	17	12	21
Greece	22	22	22	22	20	20	21	2	22
Italy	23	23	23	23	23	20	24	4	23
Hungary	24	24	24	24	22	22	27	5	24
Romania	25	25	25	25	25	24	16	9	25
Bulgaria	26	26	26	26	27	24	23	4	26
Malta	27	27	27	27	26	23	25	4	27

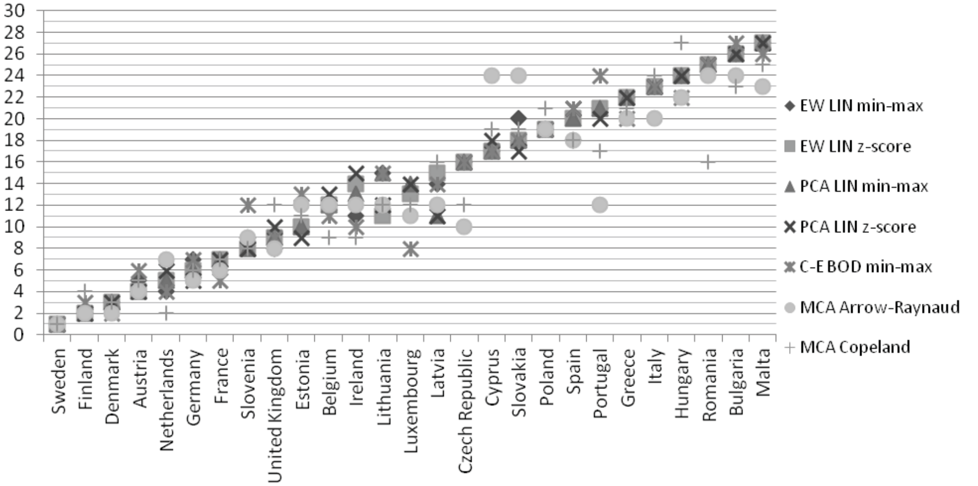
Source: Computation of author.

In the case of equal weighting and the PCA, two different normalisation methods were tested. The next column in the Table 3 belongs to the BOD rank (after applying cross-efficiency). The next two columns show the result of multi-criteria analyses using Arrow-Raynaud rule and Copeland rule, respectively. As stated above, there is no one method how to create the Composite Indicator. One can see a ranking differs under different scenarios. The robustness of CIs is a crucial issue. It can be assessed by means of absolute maximum difference in rank (the lower the better).

Sweden ranks the first under all scenarios. The biggest difference in rankings is assigned to Portugal (from 17th to 24th place). It is caused by different nature of multi-criteria ranks and the other ranks. The differences in rankings can be caused by outliers, very strong performance only in one indicator, correlations *etc*. Each of the used method rewards more some countries, hence the median rank was considered to be "the most correct ranking". Let's have a look on the Czech Republic. In the first five rankings it gained the 16th place. These results are pretty stable. But in the multi-criteria based ranking the Czech Republic improved its position and ranks 10th and 12th positions respectively. Of course, median rank equals 16.

3.3 Robustness analysis

Figure 1
Rankings by Different Method

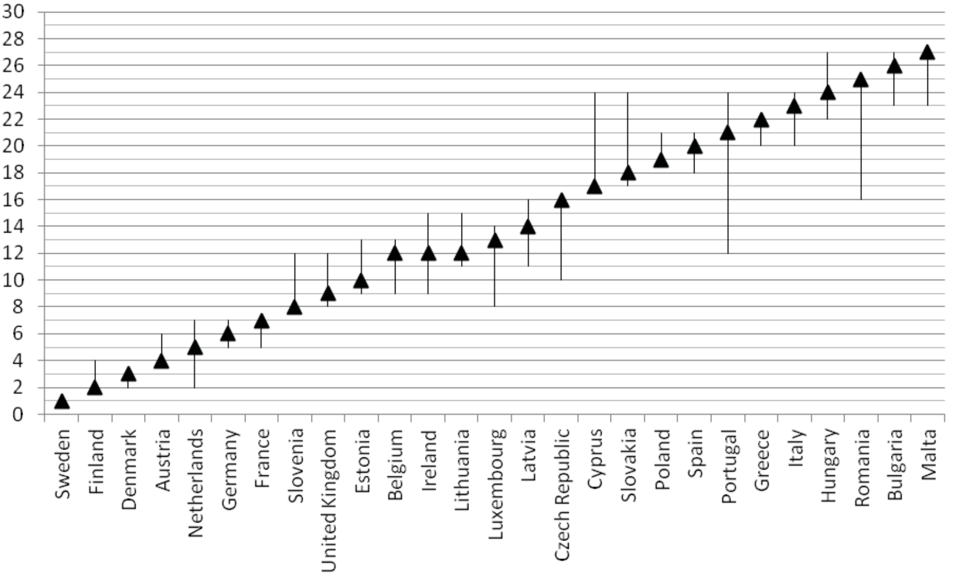


Test of robustness, uncertainty and sensitive analysis could help with interpretation of the results. Uncertainty analysis is focus on source of uncertainty – imputation of missing data, normalisation method, weighting and aggregation. The paper provides rankings only for the combination of the most common method of normalisation, weighting and aggregation which meet the methodological assumptions. In the case of indicators Europe 2020, no indicator could be discarded. One imputed missing value made only approx. 0.5% of the dataset. The seven scenarios were tested explicitly. Therefore conducting of additional uncertainty and/or sensitivity analysis (as described in Saisana *et al*. 2005) would be redundant. However the robustness was assessed and depicted in Figure 1.

Obviously, the more overlaps arise, the more robust rankings are. As well in Figure 2 the shorter line means more robust rankings. Figure 2 shows median and range in rankings.

Figure 2

Range and Median Rankings with Different Method



The rankings should be considered stable. OECD (2008) recommends methods of robustness analysis based on $\overline{R}_S$. Table 3 shows the average shift in country rankings from the median rank.

Table 4

The Average Shift in Rankings

	EW LIN min-max	EW LIN z-score	PCA LIN min-max	PCA LIN z-score	C-E BOD (median) min-max	MCA Arrow-Raynaud rule	MCA Copeland rule
$\overline{\boldsymbol{R}}_s$	0.52	0.15	0.30	0.59	1.33	2.11	1.96

This statistics comprise the relative shift in the in the position of all countries in a single number. The value of $\overline{R}_S$ more close to the zero means the more similar ranking to the median ranking. Multi-criteria approach (for both rules) indicates the largest difference from the median rank. As aforementioned, this is probably caused by very different ideas behind this approach and the other used methods. Not using multi-criteria approach, no countries would shift more than 6 places under any scenarios. There is no high difference between rankings based on the other models excluding MCA models. The equal weights with linear aggregation (based on z-scores) seems to be very similar to the median rank.

4. Conclusion

Rankings are very popular because of their easy interpretations. But it has a hidden drawback. A Composite Indicator combines multi-dimensional concept into a one number and it tempts to misleading conclusion and wrong interpretations. The most important is soundness and transparency of a ranking. Non-aggregators (*i.e.* critics of the concept of CIs) see main shortage of the CIs in subjectivity in their construction. Indeed, the Composite Indicator cannot be fully objective. But this weakness could be also considered to be a strength. Due to the fact that within the construction of the CI subjective judgements cannot be avoided, authors of the CIs usually give all methodological backgrounds. To understand and interpret the results correctly, the de-constructing the CI is very useful. Looking on separate indicators can help to extend the analysis. The Composite Indicator approach develops many tools for international comparisons. This paper shows that composite indicators approach could be perspective, because there is no only one correct method how to create a CI and new methods can be employed. The choice of methods mainly depends on data type. Different method suits different empirical case. In general, the construction of a CI should be guided by OECD (2008). The strengths and the weakness of CIs are mentioned there. In general, the CIs facilitate the ranking of countries on a multidimensional phenomenon. In this case it is policy domain defined in the Strategy for Smart, Sustainable and Inclusive Growth and indicators Europe 2020 which should measure a fulfilling of the targets of the Strategy. Note that the comparison of countries can have a beneficial effect on actual fulfilling the targets. No country wants to be on the tail.

Papers dealt with several methods used in a process of building a ranking. Median ranks derived from different scenarios seem to give a reasonable ranking. Sweden is the best according to all scenarios. It indicates that some countries were not markedly affected by the choice of methods. In order to assess the impact of different methods on ranking, the robustness was also tested. To conclude, a median rank seems to be well-founded and robust.

References

Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T. (2009), "Constructing a Knowledge Economy Composite Indicator with Imprecise Data." CES KU Leuven Discussion Paper Series 09.15.

Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T., Saisana, M., Saltelli, A., Liska, R., Tarantola, S. (2008), "Creating Composite Indicators with DEA and Robustness Analysis: the case of the Technology Achievement Index." *Journal of the Operational Research Society,* Vol. 59, pp. 239-251.

European Commission (2010), *Communication from the Commission. EUROPE 2020: A strategy for smart, sustainable and inclusive growth.* Brussels: European Commission

Freudenberg, M. (2003), "Composite Indicators of Country Performance: A Critical Assessment", OECD Science, Technology and Industry Working Papers 2003/16.

Groeneveld, R. A., Meeden, G. (1984), "Measuring Skewness and Kurtosis." *The Journal of the Royal Statistical Society. The Statistician*, Vol. 33, No. 4, pp.391-399.

Hoolingsworth, B., Wildman, J. (2002), "Efficiency and Cross efficiency Measures: A Validation Using OECD Data. " Centre for Health Program Evaluation Working Paper No. 132.

Manly, B. (2004), *Multivariate Statistical Methods: A Primer.* UK: Chapman & Hall.

Morrison, D. F. (2005), *Multivariate Statistical Methods*. California: Thompson Brooks.

Munda, G. (2010), *Social Multi-Criteria Evaluation for a Sustainable Economy*, Springer.

Munda, G., Nardo, M. (2009), "Noncompensatory/Nonlinear Composite Indicators for Ranking Countries: A Defensible Setting." *Applied Economics*, No. 41, pp. 1513-1523.

Nardo M., Saisana, M., Saltelli, A., Tarantola, S. (2005), *Tools for Composite Indicators Building*. Ispra: Joint Research Centre. EUR 21682.

OECD (2008), *Handbook on Constructing Composite Indicators. Methodology and User Guide*. Paris: Organisation for Economic Co-operation and Development.

Saisana, M. (2011), "Weighting Methods II: Statistical Approaches." Lecture at JRC Seminar on Composite Indicators and Rankings, Ispra Italy.

Saisana, M., Saltelli, A., Tarantola, S. (2005), "Uncertainty and Sensitivity Analysis Techniques as Tools for Quality Assessment of Composite Indicators." *The Journal of the Royal Statistical Society.* 168, Part 2, pp.307-323.

Saltelli, A. (2011), "Composite Indicators: An Introduction." Lecture at JRC Seminar on Composite Indicators and Rankings, Ispra Italy.

Saltelli, A., Munda, G., Nardo, M. (2006). "From Complexity to Multi-dimensionality: the Role of Composite Indicators for Advocacy of EU Reform." *Tijdchrift vor Economie en Management*. Vol. LI, 3.

StatSoft, Inc. (2011), *Electronic Statistics Textbook.* (Electronic Version). StatSoft. www.statsoft.com/textbook.