EUROPEAN COMMISSION
JOINT RESEARCH CENTRE

Institute for the Protection and the Security of the Citizen
Technological and Economic Risk Management Unit
I-21020 Ispra (VA) Italy

State-of-the-Art Report on Current Methodologies and Practices for Composite Indicator Development

Prepared by

Michaela Saisana and Stefano Tarantola

(Applied Statistics Group)

2002 **EUR 20408 EN**

EUROPEAN COMMISSION
JOINT RESEARCH CENTRE

Institute for the Protection and the Security of the Citizen
Technological and Economic Risk Management Unit
I-21020 Ispra (VA) Italy

State-of-the-Art Report on Current Methodologies and Practices for Composite Indicator Development

Prepared by

Michaela Saisana and Stefano Tarantola

(Applied Statistics Group)

2002

EUR 20408 EN

A great deal of additional information on the
European Union is available on the Internet.
It can be accessed through the Europa server
(http://europa.eu.int).

EUR 20408 EN

Table of contents

1 Objectives and Contents

Composite indicators (composite indicators) have received substantial attention in recent years and various methodologies have been developed to handle different aspects of the issue. This report examines a number of methodologies with a view to clarifying how they relate to the development of composite indicator. Several methods are investigated such as

- Aggregation systems,
- Multiple linear regression models,
- Principal components analysis and factor analysis,
- Cronbach alpha,
- Neutralization of correlation effect,
- Efficiency frontier,
- Distance to targets,
- Experts opinion (budget allocation),
- Public opinion, and
- Analytic Hierarchy Process.

The report further examines twenty-four published studies on this topic in a number of fields such as environment, economy, research, technology and health, including practices from the Directorates General of the European Commission. In this report, we offer for each composite indicator reviewed general information on the number and type of sub-indicators, on the preliminary treatment (normalisation, detrending etc.) and on the weighting system considered. Furthermore, each composite indicator is briefly commented.

2 *Introduction*

Indicators are pieces of information that summarize the characteristics of a system or highlight what is happening in a system. They are often a compromise between scientific accuracy and the information available at a reasonable cost. A mathematical combination (or aggregation as it is termed) of a set of indicators is most often called an 'index' or a 'composite indicator'. The definition of composite indicators used in this report is adopted by the Note on composite indicators presented as background paper at the Inter-Service consultation meeting of the European Commission held in Brussels on March 14th 2002. Specifically,

Composite indicators are based on ***sub-indicators that have no common meaningful unit of measurement*** *and there is* ***no obvious way of weighting these sub-indicators.***

A list of pros and cons on composite indicators was reported in the same Note on Composite Indicators. The following points are, in our opinion, worth summarising from the document:

Pros

- Composite indicators can be used to **summarise complex or multi-dimensional issues,** in view of supporting decision-makers.
- Composite indicators provide the **big picture.** They can be easier to interpret than trying to find a trend in many separate indicators. They facilitate the task of ranking countries on complex issues.
- Composite indicators can help **attracting public interest** by providing a summary figure with which to compare the performance across Countries and their progress over time.
- Composite indicators could help to **reduce the size** of a list of indicators **or to include more information** within the existing size limit

Cons

- Composite indicators **may send misleading, non-robust policy messages** if they are poorly constructed or misinterpreted. Sensitivity analysis can be used to test composite indicators for robustness.
- The simple "big picture" results which composite indicators show may invite politicians to draw **simplistic policy conclusions.** Composite indicators should be used in combination with the sub-indicators to draw sophisticated policy conclusions.
- The construction of composite indicators involves stages where **judgement** has to be made: the selection of sub-indicators, choice of model, weighting indicators and treatment of missing values etc. These judgements should be transparent and based on sound statistical principles.
- There could be **more scope for Member States** about composite indicators than on individual indicators. The selection of sub-indicators and weights could be the target of political challenge
- The composite indicators increase the **quantity of data** needed because data are required for all the sub-indicators and for a statistically significant analysis.

The experience shows that disputes over the appropriate method of establishing weights cannot be easily resolved. *Cox et al. (1992)* summarise the difficulties that are commonly encountered when proposing weights to combine indicators to a single measure, and conclude that many published weighting schemes are either arbitrary (e.g. based upon too complex multivariate methods) or unreliable (e.g. have a little social meaning). *Wall et al. (1995)* note that "the development of highly aggregated indicators is confronted with the dilemma that, although a high level of aggregation is necessary in order to intensify the awareness of problems, the existence of disaggregated values is essential in order to draw conclusion for possible courses of action". In spite of these purported shortfalls, composite indicators are nevertheless useful to provide experts, stakeholders and decision-makers with:

- the direction of developments;
- comparison across places, situations and countries;
- assessment of state and trend in relation to goals and targets;
- early warning;
- identification of areas for action;
- anticipation of future conditions and trends; and
- communication channel for general public and decision-makers.

Although science cannot provide an objective method for developing the one-and-only true composite indicator to summarise a complex system, it can help significantly in assuring that the processes of aggregation are as sound and transparent as possible. The status of the matter is perhaps concisely summarised by the Communication from the Commission on Structural Indicators (COM(2001) 619 final, 30.10.2001):

«41. The Commission is also considering an approach based on composite indicators. Such indicators are already used in some of the more detailed sectoral processes such as the Innovation Scoreboard. Composite indicators are calculated by weighting together a set of well chosen constituent indicators to provide a summary of each Member State's progress in a particular policy domain. Composite indicators would have the advantage of providing a broader coverage of information than can be included in the current list of structural indicators and they would also allow for a reduction in the number of indicators presented in the list. However, because composite indicators invite strong policy messages to be concluded they need to be robust and based on a sound methodology. More work is therefore necessary to develop such indicators, to examine how they could be integrated into the list of structural indicators and to assess the consistency of the policy messages they send.»

The debate on whether to condense sets of indicators into composite indicators for the various policy domains (Employment, Innovation and Research, Economic Reform, Social Cohesion, Environment) will likely gain momentum. In this, the Commission is setting the pace by adopting composite indicators such as, e.g., the Business Climate Index, the Internal Market Scoreboard, the Innovation Scoreboard, the Indicators of Investment and Performance in the knowledge-based economy. These and other examples are reviewed in the present report.

The following sections of the document are organized as follows: ***Section 3*** describes briefly the various stages for the construction of a composite indicator. ***Section 4*** critically presents a number of methods for aggregating indicators (normalisation techniques, data analysis, weighting systems). ***Section 5*** is more practically oriented. It reviews 24 representative studies on the topic of composite indicator construction and evaluation. For each study presented, general information on the scope of the composite indicator, the set of sub-indicators, the preliminary treatment and the considered aggregation methodologies is offered. Also described are the ways the composite indicators are communicated to the public. Finally, a brief assessment of each composite indicator regarding the appropriateness of the selected aggregation methodology, in terms of the complexity and inter-correlations of the data, is offered. ***Section 6*** delineates the main conclusions. It is dedicated to suggesting the most appropriate aggregation method for different settings. ***Section 7*** is the concluding part of the report. A case study tackled by JRC is described. It is based on the Technology Achievement Index (TAI) reported in the United Nations' Human Development Report. JRC has applied a series of uncertainty and sensitivity analyses in the TAI values for 50 countries with a view to increase the transparency in the building of the composite indicator. The Annex at the end of the document gives the overall information of the case studies on composite indicators in a concise way.

3 *General scheme of building composite indicators*

Any useful composite indicator has to be based on a sound methodology, which should be easy to understand by non-experts. There are several stages for the construction of composite indicators, such as:

- **Deciding on the phenomenon to be measured** – and whether it would benefit from the use of composite indicators.
- **Selection of sub-indicators** – A clear political idea is needed of which sub-indicators are relevant to the phenomenon to be measured. There is no fully objective way of selecting the relevant sub-indicators.
- **Assessing the quality of the data** – There needs to be high quality data for all the sub-indicators, otherwise the analyst has to decide whether to drop the data or find ways of constructing the missing data points. In case of data gaps, alternative methods could be applied, e.g. mean substitution, correlation results, time series, and assess how the selection of the method can affect the final result.
- **Assessing the relationships between the sub-indicators** – Methods such as Principal Components Analysis can provide insight into the relationships between the sub-indicators. It can be considered as prerequisite for the preliminary analysis of the sub-indicators.
- **Normalising and weighting of the indicators** – Many methods for normalising and weighting the sub-indicators are reported in the literature. The selection of the appropriate methods depends on the data and the analyst.
- **Testing for Robustness and Sensitivity** – Inevitably changes in the weighting system and the choice of sub-indicators will affect the results the composite indicator shows. However it is important to test the degree of sensitivity of the country rankings to avoid basing policy messages on rankings which are highly sensitive to small changes in the construction of the composite indicator. The values of the composite indicator should be displayed in the form of confidence bounds.

4 *Statistical treatment and weighting strategies for building composite indicators*

4.1 Introduction

A number of techniques are being analyzed herein and on the basis of their advantages and drawbacks a comparative presentation is given. These include: aggregation techniques, multiple linear regression analysis, principal components analysis and factor analysis, cronbach alpha, neutralization of correlation effect, efficiency frontier, experts opinion (budget allocation), distance to targets, public opinion and analytic hierarchy process.

4.2 Methodological approaches

4.2.1 Aggregation techniques

Before computing a composite indicator, the sub-indicators that are measured in different units must be transformed into the same unit. This is the simple part of calculating a composite indicator. The more difficult task is to select the proper weights. Table 1 gives the equations for six different methods of calculating a composite indicator [41]. These range from the simplest (Method 1) to the most complex (Method 6). Table 1 does not cover all possible methods of calculating a composite indicator. Several variations on each method exist. However, they were chosen since, in our opinion, they are the most representatives of the philosophy underlying the development of composite indicators as well as the most established in the literature.

Method 1
This is the simplest aggregation method. It entails ranking the countries for each sub-indicator and then summing the country rankings (e.g. Information and Communication Technologies index [11]). Method 1 is therefore based on ordinal levels. Its advantages are its simplicity and the independence to outliers. The disadvantage of this method is that it loses absolute level information.

Method 2
This method only uses nominal level data for each indicator. It simply takes the difference between the number of indicators that are above and below an arbitrarily defined threshold around the mean. This method is used in the 2001 Innovation Scoreboard of DG Enterprise [8]. Its advantages are its simplicity and the fact that this method is unaffected by outliers. The disadvantage of this method is that it loses interval level information. For example, assume that the value of indicator x for country A is 300% above the mean and the value for country B is 25% above the mean, with a threshold of 20% above the mean. Both country A and B are then counted equally as 'above average'.

Method 3
This method essentially takes the average of the ratios (or percentages) around the EU mean for each indicator. For example, assume that the EU mean for indicator x is 4, and the value is 6 for country A, 16 for country B, and 1 for country C. The ratios are: country A = 1.5, country B = 4, country C = 0.25. The ratios for all countries are then summed and divided by the number of indicators (if all weights = 1). The advantage of this method is that it can be used for calculating

changes in the composite indicator over time. However, this method has one important disadvantage. It is less robust when there are outliers.

Table 1. Methods for calculating composite indicators (CIs)

Method	Equation
1. Sum of country rankings	$CI_c^t = \sum_{i=1}^{N} Rank_{ic}^t$
2. Number of indicators above the mean minus the number below the mean.	$CI_c^t = \sum_{i=1}^{N} \cdot \operatorname{sgn}\left[\frac{x_{ic}^t}{x_{EUi}^t} - (1+p)\right]$
3. Ratio or percentage differences from the mean.	$CI_c^t = \frac{\sum_{i=1}^{N} w_i \cdot y_{ic}^t}{\sum_{i=1}^{N} w_i}$, where $y_{ic}^t = \frac{x_{ic}^t}{x_{EUi}^t}$
4. Percentage of annual differences over consecutive years	$CI_c^t = \frac{\sum_{i=1}^{N} w_i \cdot y_{ic}^t}{\sum_{i=1}^{N} w_i}$, where $y_{ic}^t = \frac{x_{ic}^t - x_{ic}^{t-1}}{x_{ic}^t}$
5. Standardized values	$CI_c^t = \frac{\sum_{i=1}^{N} w_i \cdot y_{ic}^t}{\sum_{i=1}^{N} w_i}$, where $y_{ic}^t = \frac{x_{ic}^t - x_{EUi}^t}{\sigma_{EUi}^t}$
6. Re-scaled values	$CI_c^t = \frac{\sum_{i=1}^{N} w_i \cdot y_{ic}^t}{\sum_{i=1}^{N} w_i}$, where $y_{ic}^t = \frac{x_{ic}^t - \min(x_i^t)}{range(x_i^t)}$.

Notes: x_{ic}^t is the value of indicator i for country c at time t. w_i is the weight given to indicator i in the composite index. In Method 2, p = an arbitrarily chosen threshold above and below the mean.

Method 4
The method has been applied for example by the DG MARKT for the development of the Internal Market Index (Scoreboard version 9). The values of the sub-indicators are substituted by the differences in the values between the year in question and the previous year and divided by the value at the previous year.

Method 5
This method has been widely used in other composite indicators (e.g. Environmental Sustainability Index [40]). The composite indicator is based on the standardised scores for each indicator which equal the difference in the indicator for each country and the EU mean, divided by the standard error. This method is more robust when dealing with outliers than Method 3, but it does not entirely solve the problem. This is because the range between the minimum and maximum observed standardised scores will vary for each indicator. This characteristic of

Method 5 is not necessarily undesirable. The method gives greater weight to an indicator in those countries with extreme values. This could be a desirable property if we wish to reward exceptional behaviour, for example if we believe that a few exceptional indicators are worth more than a lot of average scores. With a view to allow comparisons between years, an alternative to this method is to calculate the composite indicator for each year using the values of the EU mean and standard deviation for a reference year.

Method 6

Method 6 is similar to Method 5, except that it uses re-scaled values of the constituent indicators. The result is that the standardised scores for all indicators have an identical range. This makes this method more robust when there are outliers. However, this characteristic introduces the opposite problem - the range for indicators with very little variation are increased. These indicators will therefore contribute more to the composite indicator than they would using Method 5. The result is that Method 6 is more dependent on the value of the weightings for each indicator than methods 3 and 5, where the contribution of each indicator to the composite indicator depends on both the weighting and the variance in the indicator.

4.2.2 Multiple linear regression analysis

One approach that has been used to combine a number of sub-indicators is to compute correlation coefficients between all of the sub-indicators. Linear regression models can tell us something about the 'linkages' between a large number of indicators X_1, X_2, ..., X_n and a single output indicator $\hat{Y}$, but they deal only with linear correlation per se. Regression models can, however, stimulate research into new forms of conceptual models. In regression models, the set of indicators X_1, X_2,,..., X_n is combined on the one hand and an indicator $\hat{Y}$ representing the objective to be attained on the other (e.g. National Innovation Capacity index [27]). A multiple regression model is then constructed to calculate the relative weights of the sub-indicators. Such models are essentially linear,

$$\hat{Y} = a + b_1 X_1 + ... + b_n X_n \tag{1}$$

where $\hat{Y}$ is the indicator, a is a constant, and b_1 to b_n are the regression coefficients (weights) of the associated sub-indicators X_1, X_2,,..., X_n.

These models, although they can handle a large number of variables of different types, there is always the assumption of linear behavior and the uncertainty that the relations, captured by the regression model for a given range of inputs and outputs, may not be valid for different ranges. It is further argued that if the concepts to be measured could be represented by a single indicator $\hat{Y}$, then there would be no need for developing a composite indicator (*Muldur 2001*). However, the set of sub-indicators considered as input in the regression model could be related to various policy actions. The regression model, thereafter, could quantify the relative effect of each policy action on the target, i.e. a suitable output performance indicator identified on a case-by-case basis. In a more general case where a set of input indicators of performance is sought to be related simultaneously with a set of output indicators, then canonical correlation analysis, that is a generalization of multiple regression, could be applied [19].

4.2.3 *Principal Components Analysis & Factor Analysis*

Applications of principal components analysis (PCA) and factor analysis (FA) related to the development of composite indicators are:

(1) to *identify the dimensionality* of the phenomenon (e.g. Environmental Sustainability Index [40]),

(2) to *cluster* the indicators (General Indicator of Science & Technology [23], and

(3) to *define the weights* (e.g. Internal Market Index [9]).

These techniques are broadly explained below with a view to provide an intuitive understanding of the processes and results. For a more detailed explanation the reader is referred to *Manly (1994).*

Principal components analysis

The technique of PCA was first described by Karl Pearson in 1901. A description of practical computing methods came much later from Hotelling in 1933. The objective of the analysis is to take p variables $X_1, X_2, \ldots, X_p$ and find linear combinations of these to produce principal components $Z_1, Z_2, \ldots, Z_p$ that are uncorrelated, following

$$Z_j = \sum_{i=1}^{p} a_{ij} X_i \,, \; j = 1,2,\ldots,p \qquad (2)$$

The lack of correlation is a useful property because it means that the principal components are measuring different "statistical dimensions" in the data. When doing a PCA there is always the hope that some degree of economy can be achieved if the variation in the p original X variables can be accounted for by a small number of Z variables. It must be stressed that PCA does not always work in the sense that a large number of original variables are reduced to a small number of transformed variables. Indeed, if the original variables are uncorrelated then the analysis does absolutely nothing. The best results are obtained when the original variables are very highly correlated, positively or negatively.

The weights a_{ij} applied to the variables X in Eq.2 are chosen so that the principal components Z satisfy the following conditions:

(i) they are uncorrelated (orthogonal),

(ii) the first principal component accounts for the maximum possible proportion of the variance of the set of X's, the second principal component accounts for the maximum of the remaining variance and so on until the last of the principal component absorbs all the remaining variance not accounted for by the preceding components, and

(iii) $a^2_{1j} + a^2_{2j} + \ldots + a^2_{pj} = 1,$, $j = 1,2,\ldots,p$

In brief, PCA just involves finding the eigenvalues λ_j of the sample covariance matrix **C**,

$$C = \begin{bmatrix} c_{11} \; c_{12} \ldots c_{1p} \\ c_{21} \; c_{22} \ldots c_{2p} \\ \ldots \\ c_{p1} \; c_{p2} \ldots c_{pp} \end{bmatrix} \qquad (3)$$

where the diagonal element c_{ii} is the variance of X_i and c_{ij} is the covariance of variables X_i and X_j. The eigenvalues of the matrix **C** are the variances of the principal components. There are p eigenvalues, some of which may be negligible. Negative eigenvalues are not possible for a covariance matrix. An important property of the eigenvalues is that they add up to the sum of the diagonal elements of **C**. This means that the sum of the variances of the principal components is equal to the sum of the variances of the original variables,

$$\lambda_1 + \lambda_2 + ... + \lambda_p = c_{11} + c_{22} + ... + c_{pp} \tag{4}$$

In order to avoid one variable having an undue influence on the principal components it is common to standardize the variables X to have means of zero and unit variances at the start of the analysis. The matrix **C** then takes the form of the correlation matrix. In that case, the sum of the diagonal terms, and hence the sum of the eigenvalues, is equal to p, the number of variables.

The correlation coefficients of the principal components Z with the variables X are called ***loadings***, $r(Z_j,X_i)$. In case of uncorrelated variables X, the loadings are equal to the weights a_{ij} given in Eq.2.

Looking at PCA in a more concrete form, let us consider the case of two variables X_1 and X_2 and n situations that are expressed by the two variables. A distribution diagram of n situations is shown in Figure 1a. The variance of variable X_1 is 60% and the variance of X_2 is 40%. From the distribution of n points, it can be seen that there is some form of correlation between variables X_1 and X_2. If there is a proportional relationship between two variables, n points will be distributed along a straight line, and in this case one variable is sufficient. In Figure 1a, the relationship is not perfectly proportional, although it is nearly proportional, so in approximations a single variable is sufficient.

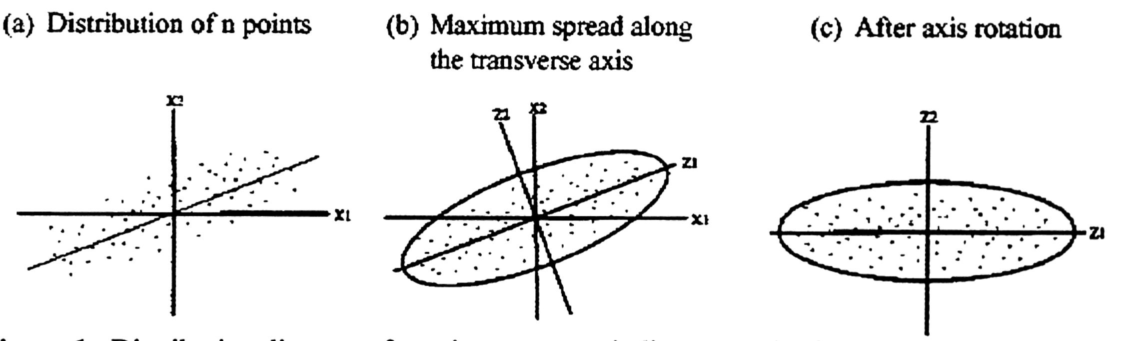

Figure 1. Distribution diagram of n points over two indicators and axis rotation

In Figure 1b, an ellipse is drawn around the circumference of n points to show the shape of their distribution. In this case, a new variable Z_1 is inserted along the transverse axis, and Z_2 is inserted along the conjugate axis (right angles to the transverse axis). This corresponds to a change of coordinates. Here, the variance of Z_1 is 95% and the variance of Z_2 is 5%, that means that Z_1 is the first principal component and Z_2 is the second principal component. A rotation is applied to describe better the situation (Figure 1c). At this point the following characteristics can be observed:

1) There is greater variance of n points on the Z_1 axis than on any other straight line drawn on this plane.
2) There is no correlation regarding the Z_1, Z_2 coordinates of n points.

In the distribution shown in the figure, n points are greatly dispersed along the Z_1 axis, so when observing data on n situations (samples), a considerable proportion can be understood solely through Z_1. Therefore if the information shown by the Z_2 axis is disregarded, the information contained in the two variables X_1 and X_2 can be summarized in Z_1. In the opposite case where the variables X_1 and X_2 are completely independent of the data on *n* situations, then the *n* points are distributed in the shape of a circle and not an ellipse, regardless of the direction of the new coordinate axes. In that case, Z_1 and Z_2 both contain an equal amount of information, so neither can be disregarded.

The PCA method has been widely used in the construction of composite indicators from large sets of sub-indicators, on the basis of correlation among the sub-indicators (e.g. Internal Market Index [9], Science and Technology Indicator [23]). In such cases, principal components have been used with the objective of combining sub-indicators into composite indicators to reflect the maximum possible proportion of the total variation in the set. The first principal component should usually captures sufficient variation to be an adequate representation of the original set (e.g. Business Climate Indicator [7]). However, in other cases the first principal component alone does not explain more than 80% of the total variance of the sub-indicators and several principal components are combined together to create the composite indicator (e.g Success of software process implementation [10], Internal Market Index [9]). As with the other techniques discussed here that are based on correlations, PCA has the disadvantage that correlations do not necessarily represent the *real (or even statistical!) influence* of those sub-indicators on the phenomenon the composite indicator is measuring.

Factor analysis

Factor analysis (FA) has similar aims to PCA. The basic idea is still that it may be possible to describe a set of *p* variables $X_1, X_2, ..., X_p$ in terms of a smaller number of *m* factors, and hence elucidate the relationship between these variables. There is however, one important difference: PCA is not based on any particular statistical model, but FA is based on a rather special model.

The early development of factor analysis was due to Charles Spearman. He studied the correlations between test scores of various types and noted that many observations could be accounted for by a simple model for the scores (*Manly, 1994*). For example, in one case he obtained the following matrix of correlations for boys in a preparatory school for their scores on tests in Classics (C), French (F), English (E), Mathematics (M), Discrimination of pitch (D), and Music (Mu):

	C	F	E	M	D	Mu
C	1.00	0.83	0.78	0.70	0.66	0.63
F	0.83	1.00	0.67	0.67	0.65	0.57
E	0.78	0.67	1.00	0.64	0.54	0.51
M	0.70	0.67	0.64	1.00	0.45	0.51
D	0.66	0.65	0.54	0.45	1.00	0.40
Mu	0.63	0.57	0.51	0.51	0.40	1.00

He noted that this matrix has the interesting property that any two rows are almost proportional if the diagonals are ignored. Thus for rows C and E there are ratios:

$$\frac{0.83}{0.67} \cong \frac{0.70}{0.64} \cong \frac{0.66}{0.54} \cong \frac{0.63}{0.51} \cong 1.2.$$

Spearmann proposed the idea that the six test scores are all of the form $X_i = a_i F + e_i$, where X_i is the ith standardised score with a mean of zero and a standard deviation of one, a_i is a constant,

F is a 'factor' value, which has mean zero and standard deviation of one, and e_i is the part of X_i that is specific to the i^{th} test only. He showed that a constant ratio between rows of a correlation matrix follows as a consequence of these assumptions and that therefore there is a plausible model for the data. In a general form this model is given by:

$$
\begin{aligned}
X_1 &= \alpha_{11}F_1 + \alpha_{12}F_2 + \ldots + \alpha_{1m}F_m + e_1 \\
X_2 &= \alpha_{21}F_1 + \alpha_{22}F_2 + \ldots + \alpha_{2m}F_m + e_2 \\
&\ldots \\
X_p &= \alpha_{p1}F_1 + \alpha_{p2}F_2 + \ldots + \alpha_{pm}F_m + e_p
\end{aligned}
\tag{5}
$$

where X_i is a variable with zero mean and unit variance; $\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{im}$ are the factor loadings related to the variable X_i; $F_1, F_2, \ldots, F_m$ are m uncorrelated common factors, each with zero mean and unit variance; and e_i is the specific factor related only to the variable X_i, has zero mean, and it is uncorrelated with any of the common factors and the specific factors. The first stage to a FA is to determine provisional factor loadings α_{ij}. One way to do this is to do PCA and consider only the first m principal components, which are themselves taken to be the m factors. It is noted that there is an infinite number of alternative solutions for the factor analysis model.

4.2.4 Cronbach alpha

Another way to investigate the degree of the correlations among a set of sub-indicators is to use a coefficient of reliability (or consistency) called Cronbach alpha α. This coefficient measures how well a set of variables (or indicators) measures the same underlying construct. Cronbach alpha can be written as a function of the number p of indicators and the average inter-correlation $\bar{r}$ among the indicators:

$$\alpha = \frac{p \cdot \bar{r}}{1 + (p-1) \cdot \bar{r}} \tag{6}$$

It can be seen that an increase in the number of indicators is associated with an increase in α. Additionally, if the average inter-item correlation is low, alpha will be low. In fact, the coefficient α can vary from zero to one. A coefficient of $\alpha = 0.80$ or higher is considered in most applications as "evidence" that the indicators are measuring the same underlying construct. If α is low for a given set of indicators, this implies that the data are actually multi-dimensional. Cronbach alpha has been considered for example for the index of "Success of software process improvement" [10].

4.2.5 Neutralization of correlation effect

This method was applied for the aggregation of three sub-indicators into a composite indicator measuring the "Relative intensity of regional problems of the Community" by the European Communities in 1984 [4]. The sub-indicators measure:

a. GDP per employed in ECU,
b. GDP per head in PPS, and
c. unemployment rate.

The first two sub-indicators are highly correlated (they are different forms of the same issue). The first step of the method is to standardise the sub-indicators by subtracting the mean and dividing by the standard deviation. The standardised indices are marked as X_1, X_2 (the correlated ones) and Y. A sub-index X is computed as an average of the X_1 and X_2, by

$$X = [2 \cdot (1+r)]^{-1/2}(X_1 + X_2) \tag{7}$$

where r is the correlation coefficient between X_1 and X_2. The sub-index X and the indicator Y are finally combined into a composite indicator via:

$$Z = [2 \cdot (1+r)]^{-1/2}(X + Y) \tag{8}$$

where r is the correlation coefficient between X and Y.

This procedure, illustrated for three sub-indicators, can be extended in principle to any number of sub-indicators. The basic idea of the correlation of pairs of indicators remains.

4.2.6 Efficiency frontier

A thorough description of the methodology may be found in Storrie and Bjurek [34]. The authors believe that this approach is extremely parsimonious as regards the weighting, because it lets the data decide on the weighting issue. The following paragraphs present the essence of the method [35], including a description of the necessary assumptions, and their implications, by creating a composite indicator of two sub-indicators for several countries.

Figure 2 plots the two indicators, unemployment rate and employment rate. Best performance is found as we move towards the origin in both dimensions. We say that a country dominates another when it is best in both indicators. This is the first assumption of the methodology. Dominance is illustrated graphically by drawing an L-shape with the country in question at the intersection of the L (see dashed line). The country dominates all countries above and to the right of the L. For example Denmark (dk) dominates Ireland (ie), United Kingdom (uk), and so on, but not Portugal (pt) and Austria (at). Luxembourg (lu) dominates France (fr), Belgium (be), Greece (gr), Italy (it) and Spain (es). The Netherlands (nl) dominate all countries except Luxembourg and Denmark. These three countries are not dominated by any other and constitute thus the frontier or the multi-dimensional benchmark, which passes through lu-nl-dk. A further assumption is that a linear combination of two countries on the frontier is also on the frontier, i.e. convexity. The frontier is drawn with a solid line in Figure 2.

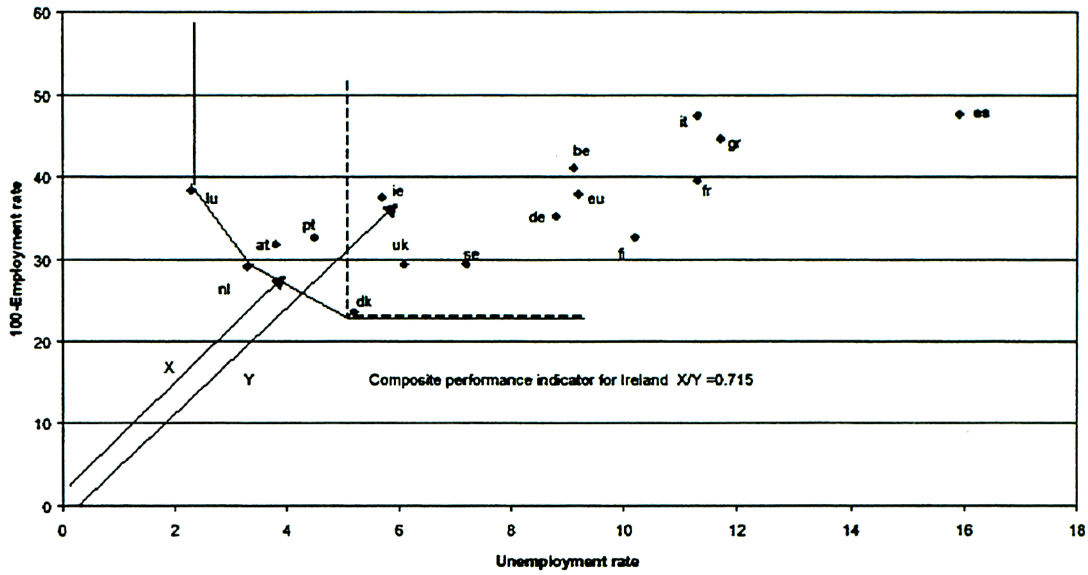

Figure 2. The construction of a frontier

What remains now is to measure the extent to which the other countries deviate from the frontier. The procedure is exemplified with Ireland. The length of the ray from the origin to Ireland is marked as Y. The distance from the origin in the direction of Ireland up to the frontier is denoted as X. The composite indicator for Ireland is then equal to $X/Y = 0.715$. The values of the composite indicator for the remaining countries are calculated in a similar way. It is obvious that for countries on the frontier the composite indicator is equal to unit. Thus, in this methodology it is the benchmark countries that determine the weights. It is emphasized that different countries will be weighted differently depending upon where they are located in relation to the frontier. The idea, illustrated graphically in two dimensions, may be extended in principle to any number of dimensions. The basic idea of a frontier and the distance to a particular segment remains.

4.2.7 Distance to targets

One way to avoid the immediate selection of weights is to measure the need for political intervention and the "urgency" of a problem by the distance to target approach. The urgency is high if we are far away from the goal, and low if the goal is almost reached. The weighting itself is realised by dividing the sub-indicator values by the corresponding target values, both expressed in the same units. The dimensionless parameters that are obtained in this way can be summarized by a simple average to produce the composite indicator.

Using policy goals as targets (e.g. "Environmental Policy Performance Indicator" [1]) convinces the policy makers for the "soundness" of the weighting method, as long as those policy makers have defined the policy targets themselves. This approach is technically feasible when there is a well-defined basis for a certain policy, such as a National Policy Plan or similar reference documents. For international comparisons, such references are often not available, or they deliver contradictory results. Another counter-argument for the use of policy goals as targets is that the benefits of a given policy must be valued independently of the existing policy goals. Alternatively to policy goals, sustainability levels, quantified effects on the environment, or best performance countries can be used as goalposts (e.g. Human Development Index [36])

4.2.8 *Experts opinion (budget allocation)*

A commonly used method is the assignment of weights to sub-indicators based on personal judgment (participatory method). This method, however, reaches its limits when some indicators have little (or no) meaning to the interviewed person. For example, while an ordinary citizen might have a feeling about the importance of cleaner air or a quieter environment, weight assignment is likely to fail if the same person is asked to judge upon the relative importance of oxides of nitrogen versus sulfur dioxide emissions. Obviously, in such cases the opinion of experts is sought. In some policy fields, there is consensus among experts on how to judge at least the relative contribution of physical indicators to the overall problem. There are certain cases, though, where opinions diverge. It is essential to bring together experts that have a wide spectrum of knowledge, experience and concerns, so as to ensure that a proper weighting system is found for a given application [6].

Budget allocation is a participatory method in which experts are given a "budget" of *N* points, to be distributed over a number of sub-indicators, "paying" more for those indicators whose importance they want to stress. The budget allocation method can be divided in four different phases:

- Selection of experts for the valuation;
- Allocation of budget to the sub-indicators;
- Calculation of the weights;
- Iteration of the budget allocation until convergence is reached (optional).

A case study in which 400 German experts in 1991 were asked to allocate a budget to several environmental indicators related to an air pollution problem showed very consistent results, in spite of the fact that the experts came from opposing social spheres like the industrial sector and the environmental sector [21].

A counter argument against the use of the experts opinion is on the weighting reliability. Local intervention cannot be evaluated without considering local strategies, so expert weighting may not be transferable from one area to another. Furthermore, allocating a certain budget over a too large number of indicators can give serious cognitive stress to the experts, as it implies circular thinking. The method is optimal for a maximum number of 10 indicators. Special care should be given in the identification of the population of experts from which to draw a sample, stratified or otherwise.

4.2.9 *Public opinion*

Instead of letting experts determine the weights of the indicators in an index, one could ask the general public. *Parker (1991, p. 95-98)* argues that "public opinion polls have been extensively employed for many years for many purposes, including the setting of weights and they are easy to carry out and inexpensive". In public opinion polls, issues are selected which are already on the public agenda, and thus enjoy roughly the same attention in the media. From a methodological point of view, opinion polls focus on the notion of "concern", that is people are asked to express "much" or "little concern" about certain problems measured by the sub-indicators (e.g "Concern about environmental problems" Index [26]). As with expert assessments, the budget allocation method could also be applied in public opinion polls, however it is more difficult to ask the public to allocate a hundred points to several sub-indicators than to express a degree of concern about the problems that the indicators represent.

4.2.10 Analytic Hierarchy Process

The Analytic Hierarchy Process (AHP) was proposed by Thomas Saaty in the 1970s and is a widely used technique for multi-attribute decision making [30]. It enables decomposition of a problem into hierarchy and assures that both qualitative and quantitative aspects of a problem are incorporated in the evaluation process, during which opinion is systematically extracted by means of pairwise comparisons. The best description of this method is probably the one given by Forman et al. (1983): "AHP is a compensatory decision methodology because alternatives that are efficient with respect to one or more objectives can compensate by their performance with respect to other objectives. AHP allows for the application of data, experience, insight, and intuition in a logical and thorough way within a hierarchy as a whole. In particular, AHP as weighting method enables decision-maker to derive weights as opposed to arbitrarily assign them."

<u>Methodology in brief</u>

The core of AHP is an ordinal pair-wise comparison of attributes, sub-indicators in this context, in which preference statements are addressed. For a given objective, the comparisons are made per pairs of sub-indicators by firstly posing the question "Which of the two is the more important?" and secondly "By how much?". The strength of preference is expressed on a semantic scale of 1-9, which keeps measurement within the same order of magnitude. A preference of 1 indicates equality between two sub-indicators while a preference of 9 indicates that one sub-indicator is 9 times larger or more important than the one to which it is being compared. In this way comparisons are being made between pairs of sub-indicators where perception is sensitive enough to make a distinction. These comparisons result in a comparison matrix A (see Table 2) where $A_{ii} = 1$ and $A_{ij} = 1 / A_{ji}$.

Table 2. Comparison matrix A of three sub-indicators (semantic scale)

Objective	Indicator A	Indicator B	Indicator C
Indicator A	1	3	1
Indicator B	1 / 3	1	1 / 5
Indicator C	1	5	1

For the example shown in Table 2, Indicator A is three times more important than Indicator B, and consequently Indicator B has one-third the importance of Indicator A. Each judgement reflects, in reality, the perception of the ratio of the relative contributions (weights) of the two indicators to the overall objective being assessed as shown in Table 3.

Table 3. Comparison matrix A of three sub-indicators (weights)

Objective	Indicator A	Indicator B	Indicator C
Indicator A	w_A/w_A	w_A/w_B	w_A/w_C
Indicator B	w_B/w_A	w_B/w_B	w_B/w_C
Indicator C	w_C/w_A	w_C/w_B	w_C/w_C

The relative weights of the sub-indicators are calculated using an eigenvector technique. One of the advantages of this method is that it is able to check the consistency of the comparison matrix through the calculation of the eigenvalues.

Consistency

It is often the case that people's thinking is not always consistent. For example, if one claims that A is much more important than B, B slightly more important than C, and C slightly more important than A, judgment is inconsistent and decisions made are less trustworthy. Inconsistency, however, is part of the human nature and therefore in reality it is enough just to measure somehow the degree of inconsistency. This appears to be the only way so results could be defended and justified in front of public.

AHP tolerates inconsistency through the amount of redundancy. For a matrix of size $n \times n$ only $n-1$ comparisons are required to establish weights for n indicators. The actual number of comparisons performed in AHP is $n(n-1)/2$. This redundancy is a useful feature as it is analogous to estimating a number by calculating the average of repeated observations. This results in a set of weights that are less sensitive to errors of judgement. In addition, this redundancy allows for a measure of these judgement errors by providing a means of calculating an inconsistency ratio (*Saaty, 1980; Karlsson, 1998*). According to Saaty small inconsistency ratios (less than 0.1 is the suggested rule-of-thumb, although even 0.2 is often cited) do no drastically affect the weights.

5 *Selected case studies*

5.1 Introduction

In the previous section, the most common methods for aggregating indicators into composites were briefly described. In the following paragraphs, concrete examples showing the practices of developing composite indicators in a wide (although not exhaustive) variety of application areas are described. Twenty-four representative studies in fields such as environment, economy, society, research and technology are reviewed, including practices from the Directorates General of the European Commission. For each study presented, general information on the scope of the composite indicator, the set of sub-indicators, the preliminary treatment and the considered aggregation methodologies is offered. Additionally, a brief assessment of each composite indicator and the appropriateness of the selected aggregation methodology, in terms of the complexity and inter-correlation of the data is further given.

5.2 Description of the selected studies

For this review work, we have selected five composite indicators in the area of *Economy*, six composite indicators related to *Environment*, four composite indicators on *Research and Innovation*, four indices measuring on *Scientific, Technological and Information* aspects and finally five composite indicators that express *Social* concerns (see Table 4).

Several aggregation methodologies have been used at the reviewed studies, including arithmetic average, regression models, principal component analysis and factor analysis, neutralisation of correlation effect, efficiency frontier, experts opinion, distance to targets, public opinion and Analytic Hierarchy Process. More than one method is most often considered at each study. In total, twenty studies have considered statistical methods for the development of a composite indicator, while twelve studies investigated the use of participatory approaches.

Table 4. Selected case studies on composite indicators

FIELD	NAME OF COMPOSITE INDICATOR
Economy	• Economic Sentiment Indicator (by the European Commission) • Composite Leading Indicators (by OECD) • Internal Market Index (by DG MARKT) • Business climate indicator (by DG ECFIN) • Index of sustainable and economic welfare (by CES and NEF)
Environment	• Environmental Sustainability Index (by World Economic Forum) • Synthetic Environmental Indices (by Isla M.,) • Eco-Indicator 99 (by Pre Consultants, the Netherlands) • Concern about Environmental Problems (by Parker) • Index of Environmental Friendliness (by Puolamaa et al., Finland) • Environmental Policy Performance Index (by Adriaanse, the Netherlands)
Society	• Human Development Index (by the United Nations) • Relative intensity of regional problems in the Community (by the European Commission) • Employment (by Storrie and Bjurek) • Overall Health System Attainment (by WHO) • National Health Care Systems Performance (by King's Fund, England)
Research and Innovation	• Summary Innovation Index (by DG ENTR) • National Innovation Capacity (by Porter and Stern) • Investment in the knowledge based economy (by DG RTD) • Performance in the knowledge based economy (by DG RTD)
Science, Technology and Information	• Technology Achievement Index (by the United Nations) • General Indicator of Science and Technology (by NISTEP, Japan) • Information & Communication Technologies (by Fagerberg) • Success of software process improvement (by Emam et al.)

5.2.1 Summary Innovation Index, by DG ENTR

Scope of the Index

The Summary Innovation Index (SII) [8] is part of the Innovation scoreboard, which depicts achievements and trends, highlights strengths and weaknesses of Member States' performances, and examines European convergence in innovation. It is one of the benchmarking exercises of the European Commission that were launched in response to the Lisbon European Council.

Description of sub-indicators

The innovation scoreboard builds on the "structural indicators" [3]. It uses additionally some indicators that apply more restricted definitions to fulfil the purpose of the scoreboard to "zoom" into the area of innovation policy. To minimize statistical burden, the innovation scoreboard mainly uses official ESTAT data, or private data of sufficient reliability if official data are not available. The innovation scoreboard analyses 17 indicators studied between 1995/1997 and 1999/2000 in four areas: a. human resources; b. knowledge creation; c. transmission and application of new knowledge; and d. innovation finance, output and markets.

Preliminary treatment

The differences of the sub-indicators' values from the corresponding European average are used.

Aggregation method

The SII for a given country is equal to the number of indicators that are more than 20 % above the EU overall mean, minus the number of indicators that are more than 20 % below. The SII is adjusted for differences in the number of available indicators for each country. The index can vary between + 10 (all indicators are above average) to -10 (all indicators are below average).

Comments

- Due to sampling, definitional, and other errors for many of the indicators, the authors assume that indicators within +20 % and – 20 % of the overall EU mean do not differ in any meaningful way from the average. The choice of a 20 % boundary is largely arbitrary. Sensitivity analysis found a high correlation ($R^2 = 0.98$) between the SII using a 20 % boundary and those for a 15 % and 25 % boundary.
- A generally applicable model describing how each indicator influences innovation is not available. The authors selected thereafter to give equal weights to all sub-indicators in calculating the SII.
- A different calculation approach for a summary index was tested based on the average percentage by which each indicator varied from the overall EU average. This indicator was strongly correlated with the retained SII ($R^2 = 0.89$). The retained SII was finally preferred over the percentage index because it ignored minor differences from the EU average, which may not be meaningful.
- The SII is correlated ($R^2 = 0.64$) with the Economic Creativity Index from the Global Competitiveness Report 2000 of the World Economic Forum. No statistically significant ($p < 0.05$) correlations were found between the SII and several employment- and GDP-based indicators, which illustrates that there are different ways for a country to achieve a high living standard. On the other hand, statistically significant negative correlations were found between the SII and two indicators of social exclusion: the percentage of the population living below the

poverty line for three consecutive years ($R^2 = 0.52$ with $p = 0.013$) and the skewness of the income distribution ($R^2 = 0.43$ with $p = 0.014$).

- Two issues further handled in the study were a) how much the current innovation performances of the Member States vary (coefficient of variation), and b) if these performances have converged over recent years (percentage change in the standard deviation). It was found that the indicators with the least variation among the EU Member States are strongly influenced by public policy, such as education or public R&D investments. In contrast, there is more variability between countries for indicators directly influenced by private decision.

Presentation of the Index

For an immediate overview, the SII and overall country trends are calculated (Figure 3). Countries above the horizontal axis have an SII above the EU average value, while countries to the right of the vertical axis show an overall trend above the EU average. These two axes divide Figure 3 into four quadrants. Countries in the upper right quadrant are 'moving ahead' because both their summary innovation index and their past rate of change for the trend indicators are above the EU averages. Conversely, countries in the bottom left quadrant are 'falling further behind' because they are below the EU average for both variables. A "snapshot" of present country performances in terms of the SII is given in Figure 4.

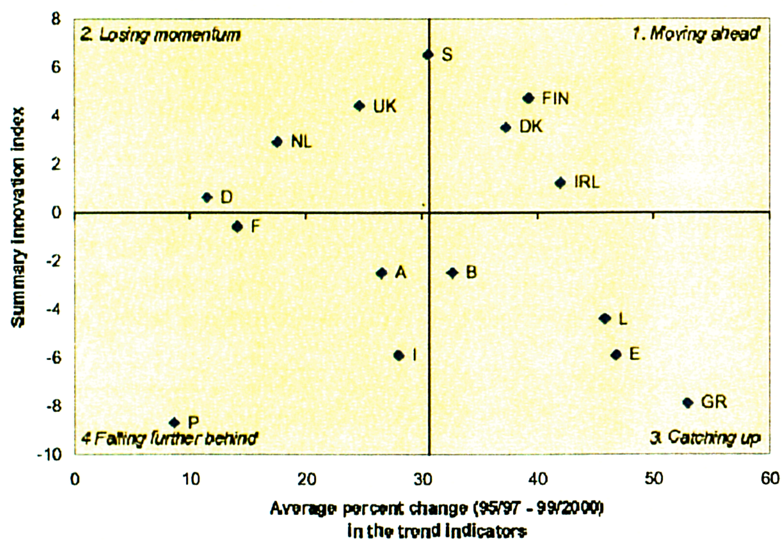

Figure 3. Overall country trends by Summary Innovation Index

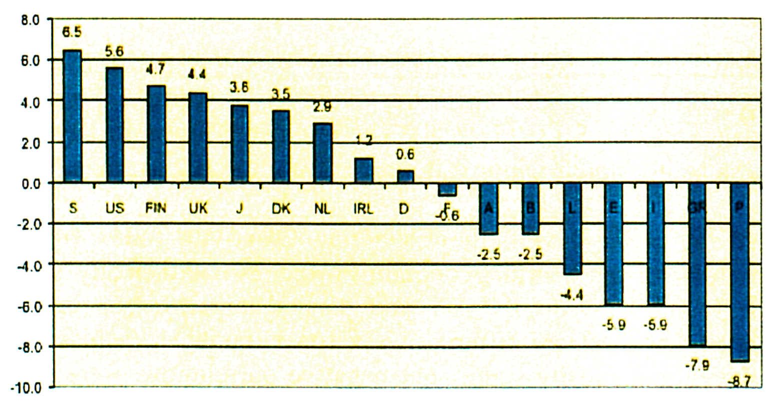

Figure 4. Summary Innovation Index

5.2.2 *"Internal Market Index" by DG MARKT*

Scope of the Index

The objective of the Internal Market Index [9] is to measure whether the "real world" benefits, that the Internal Market Strategy attempts to bring to the citizens and companies, are effectively delivered. These benefits are multiple and of a very different nature: higher incomes, better social cohesion, lower prices, increased possibilities to work and live abroad, a cleaner environment, easier access to capital, etc.

Description of sub-indicators

Nineteen variables are synthesized in the Index, including growth in per-capita income, long-term unemployment, price dispersion, growth in intra-EU trade, prices of utilities services, availability of venture capital, energy intensity and greenhouse gas emissions.

Preliminary treatment

The percentage year-to-year differences for each sub-indicator are used at the aggregation step.

Aggregation method

The resulting Index synthesizes 19 variables, using the statistical method of Principal Components Analysis (PCA) in order to weigh the influence of each variable on the final score. The 19 main Principal Components (number equal to the number of sub-indicators) are further weighted by the % variation of the total information explained by each Principal Component.

Comments

- The authors selected PCA because they considered it as an "objective" weighting scheme between variables, which deals appropriately with correlation between variables. However, the inter-correlation between the sub-indicators is very low (there is no r>0.7) to justify the use of PCA for the weighting of the sub-indicators.
- Furthermore, the number of Principal Components is equal to the number of the sub-indicators (i.e. 19), which implies that the total information of the sub-indicators can not be summarised by PCA and that a simpler weighting system might be more efficient (such as equal weighting or a participatory method, e.g. budget allocation).

Presentation of the Index

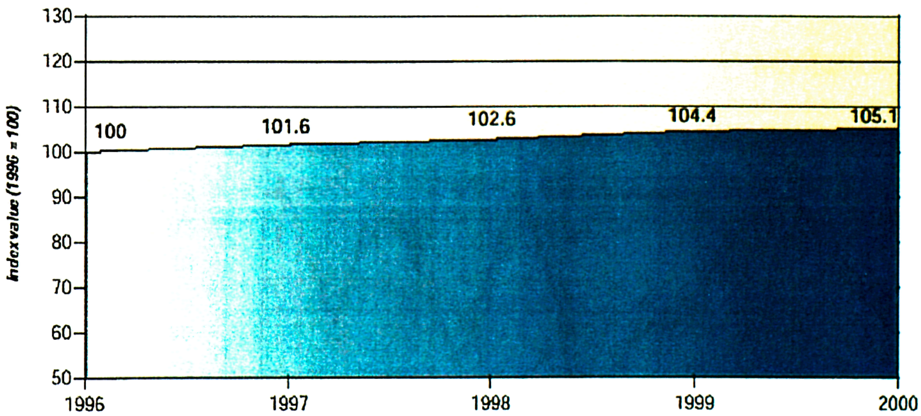

Figure 5. The Internal Market Index presented as a surface graph over several years and over all EU-15 Member States.

5.2.3 *"Business climate indicator" (by DG ECFIN)*

Scope of the Index

To improve the understanding of the business cycle in the European area as a whole DG ECFIN has formulated a composite indicator based on business surveys designed to deliver a clear and early assessment of the cyclical situation within the area [7]. The publication of this indicator is part of a larger project launched by Pedro Solbes, EU Commissioner for Economic and Monetary Affairs, to improve and complete the quality of euro-zone and EU statistics.

Description of sub-indicators

The five sub-indicators are related to the responses of national business surveys and available only from 1985 onwards: production trends in recent past, order books, export order books, stocks and production expectations. The series are drawn from survey data that correspond to the answers that industrial managers give to a question concerning the business climate. They can give a qualitative assessment by choosing between three types of statement, namely that the situation has improved, deteriorated or not changed in comparison with the preceding period.

Preliminary treatment

The answers from the surveys are then aggregated having due regard to the size of the company in question and the sector in which it operates. The information on each question is presented in the form of the difference (hence the term " balance of opinion") between the percentage of firms which have noted an improvement and those which have reported a deterioration. Each series therefore varies by construction between –100 (indicating that all firms have reported a deterioration) and +100 (all firms have noted an improvement).

Aggregation method

PCA identified one main principal component. The weights of the sub-indicators have subsequently been estimated using factor analysis with maximum likelihood estimation.

Comments

- The common factor alone explains 92% of the total variance of the opinion balances, which, according to the authors, provides a statistical justification for the choice of summarizing a priori the information by means of a single composite indicator.
- The PCA requirements of stationarity of the series are confirmed by the empirical autocorrelograms (significant up to the 8th order) and the augmented Dickey-Fuller test.

Presentation of the Index

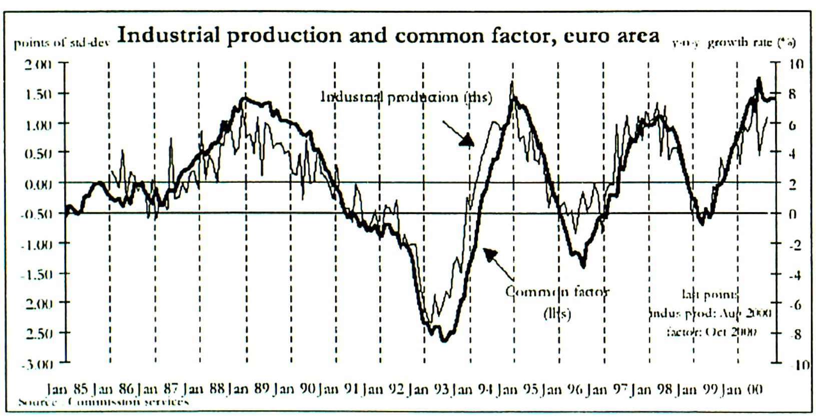

Figure 6. The business climate indicator is presented as a time-series over several years and in parallel to the industrial production, which is the benchmark indicator.

5.2.4 *"Investment in the knowledge based economy" (by DG RTD)*

Scope of the Index

This composite indicator, built by DG RTD, aims to summarize various indicators of national investment in highly qualified human resources in science, technology, research and education, so as to measure a country's capacity to create knowledge. Two composite indicators "investment in the knowledge-based economy" and "performance in the knowledge-based economy" have been developed by DG RTD following the Commission's decision to finalise these two indicators for the 2003 synthesis report, (COM (2001), 619, 30 October, 2001).

Description of sub-indicators

The composite indicator, in its preliminary version examined here, combines seven indicators related to the number of researchers, the number of new doctors in science and technology (annual influx), domestic expenditure on R&D, expenditure on information technologies and imports of high-tech products. All sub-indicators are measured per capita to neutralize the effect of the size of the countries. More sub-indicators are forecast to be included in the list, such as: the number of new higher-education graduates in science and technology, expenditure on education, and the imports side of the technological balance of payments.

Preliminary treatment

DG RTD has selected a simple method for the pre-treatment of the sub-indicators, so that all sub-indicators are brought into line with the same standard before they are combined. This method is similar to the one used by the International Institute for Management Development (Lausanne) in The World Competitiveness Yearbook. Briefly,

- the mean and a dispersion index (standard deviation) are calculated for each indicator for a reference year for all the European Union countries;
- country by country and year by year the original values for each indicator are converted by "centering" them on this mean and dividing them by this dispersion index. This puts all the indicators into harmonized form, so that the distances between two countries can be compared for every indicator and for different years.

Aggregation method

The value of the composite indicator for each country and for each year is the weighted average of the values of all the indicators. An instrument in Excel leaves the choice of the weights up to the user (*Muldur 2001*).

Comments

- JRC carried out uncertainty analysis to identify how much the variation of the weights of the sub-indicators affects the values of the composite indicator for the various countries. A further sensitivity analysis helped to identify which weights are influential to the differences of close ranked countries.
- The composite indicator is in progress.

Presentation of the Index

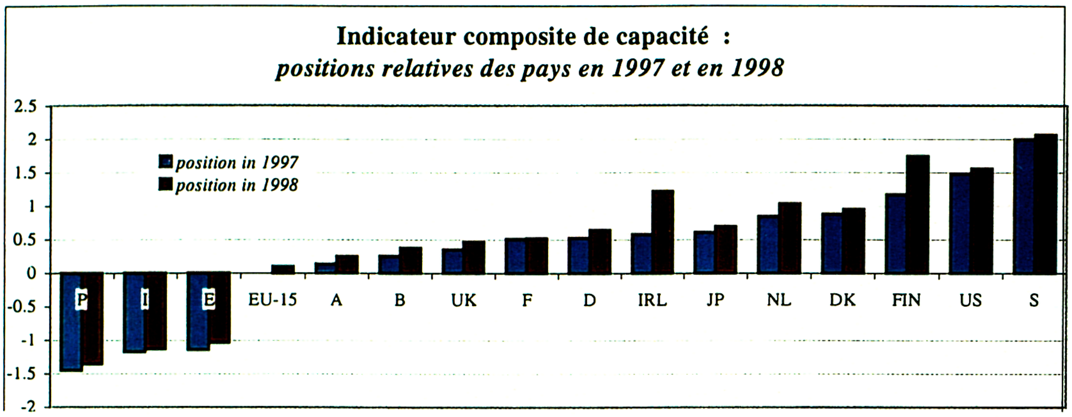

Figure 7. The "Investment in the knowledge-base economy" composite indicator for 13 EU-countries (where data were available), the United States and Japan.

5.2.5 *"Performance in the knowledge based economy" (by DG RTD)*

Scope of the Index

The second index related to the knowledge-based economy aims to measure a country's performance in converting the new knowledge into economic and technological progress to increase both a country's competitiveness and the well being of its citizens.

Description of sub-indicators

The index combines six indicators: the number of EPO and USPTO patents, the number of publications, production of high-tech exports, the employment in the high-tech production and GDP per capita.

Preliminary treatment

Normalised indicators are calculated by mean subtraction and division by the standard deviation

Aggregation method

The composite indicator is calculated as a weighted average of the normalized indicators, where the choice of weights is left up to the users (the procedure is the same as in Section 5.2.4).

Comments

- The composite indicator is in progress.

Presentation of the Index

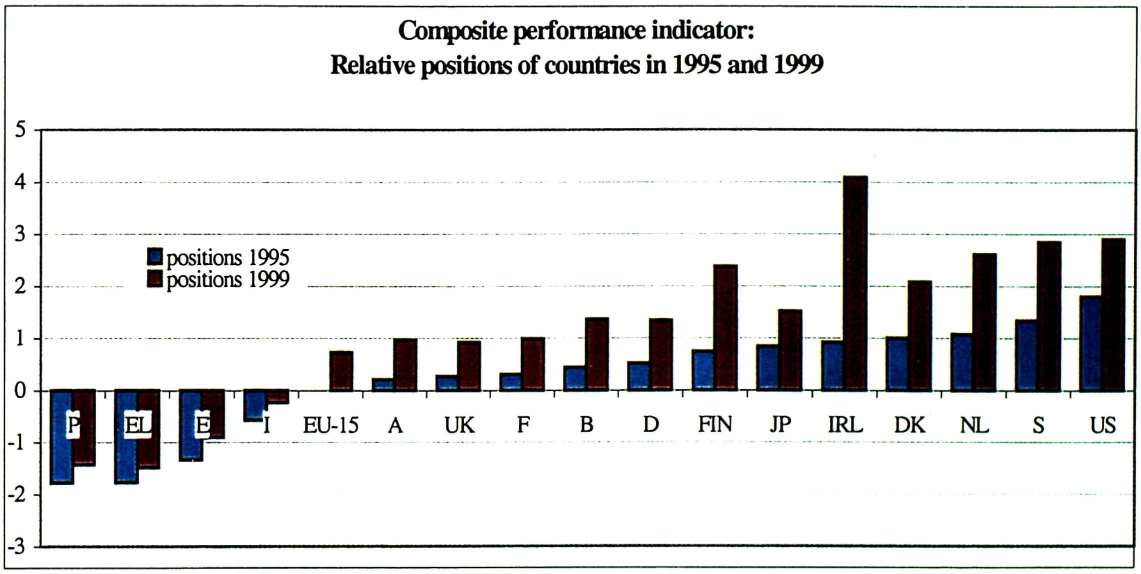

Figure 8. The "Performance in the knowledge-based economy" composite indicator for 13 EU-countries (where data were available), the United States and Japan.

5.2.6 *"Relative intensity of regional problems in the Community" (by the EC)*

Scope of the Index

Back in 1984 the European Commission (EC) constructed a synthetic index measuring the "relative intensity of regional problems in the Community" [4]. The objective of the Index is to assist the Community regional policy to focus on strengthening the economic performance of regions experiencing delayed development. It further aims to shed light on declining industrial regions, and on creating lasting jobs, so as to contribute to the improvement of the competitiveness of the European economy as a whole.

Description of sub-indicators

The Commission, with a view to measure the relative intensity of the regional problems at Community level in a global and synthetic way, uses three sub-indicators: GDP per employed in ECU, GDP per head in PPS, and unemployment rate.

Preliminary treatment

The three sub-indicators are standardized by subtracting the corresponding mean and dividing by the standard deviation.

Aggregation method

This method attempts to neutralise the effects of correlation between variables taken two by two. Empirical weights are determined considering the degree of correlation of sub-indicators taken in pairs through a formula: $w = \left[2 \cdot (1 + r\ )\right]^{-1/2}$, where r is the bivariate correlation coefficient.

Comments

- The authors examined more thoroughly the interrelationship of socio-economic phenomena, expressed in terms of the three indicators by using factor analysis.

Presentation of the Index

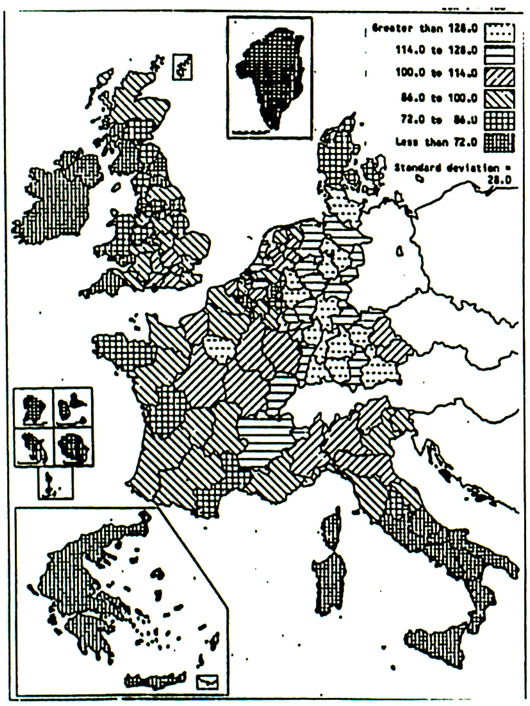

Figure 9. Representation of the composite indicator measuring the "Relative intensity of regional problems in the Communities" in the form of a map.

5.2.7 "Economic Sentiment Indicator" (by the European Commission)

Scope of the Index

The Economic Sentiment Indicator of the European Commission (EC ESI) combines business tendency surveys into a single cyclical composite or confidence indicator, with a view to reduce the risk of false signals and to provide a cyclical indicator with better forecasting and tracking qualities than any of its individual components [24].

Description of sub-indicators

A standard set of four components is used, mainly based on qualitative data from business or consumer tendency surveys. The EC ESI combines the following component series: (a) industrial confidence indicator; (b) construction confidence indicator; (c) consumer confidence indicator; and (d) share price index.

Preliminary treatment

Before the calculation of the composite index, normalised values for the sub-indicators are calculated by dividing the month-to-month changes with the average change.

Aggregation method

The components of the ESI are divided into two groups with equal weights to components in each group. The first group contains the industrial confidence indicator and the consumer confidence indicator, and the second group includes the construction confidence indicator and the share price index. The components in the second group are given half the weight of the components in the first group.

Comments

- The performance of the Index is evaluated both at turning points and over the whole cycle (cross-correlation) against total industrial production as a proxy for economic activity.
- The results of the study of *Nilsson (2000)* show that the forecasting performance of the EC ESI could be improved in almost all the investigated countries, if only the components with the longest lead in this indicator were combined.
- The authors consider that applying PCA to choose the weights would minimise the contribution of the indicators which do not move with the other indicators. This may reduce the reliability of the composite indicator because some indicators perform better in one cycle and others in a different cycle.

Presentation of the Index

See Figure 10 at the following section.

5.2.8 *"Composite Leading Indicators" (by OECD)*

Scope of the Index

The OECD Composite Leading Indicators (CLI) are based on individually selected leading indicators for each country and are calculated for 22 Member States [24]. They aim at providing a cyclical business indicator with better forecasting and tracking qualities than any of its individual components.

Description of sub-indicators

The OECD CLI's are based on individually selected leading indicators for each country. For example the CLI for France is composed of 11 component series, the CLI for United Kingdom combines 9 leading indicators, while the CLI for Germany and Italy are composed of 6 indicators.

Preliminary treatment

1. Periodicity. CLI data are published with a monthly periodicity. Quarterly component series are converted to a monthly series using linear interpolation;
2. Trend estimation via Phase average trend method;
3. Smoothing. In order to reduce the irregularity of the final composite indicator, component series are smoothed according to their "Months for cyclical dominance moving average";
4. Standardisation. The method used is first to subtract the mean and then to divide by the mean of the absolute differences from the mean. The normalised series are then converted into index form by adding 100.

Aggregation method

In the OECD system, the composite indicator is obtained by averaging the normalised indices. This implies that equal weighting has been selected.

Comments

- The performance of the Index is evaluated both at turning points and over the whole cycle (cross-correlation) against total industrial production as a proxy for economic activity.
- *Nillson (2000)* suggests that PCA could help to select the weights. However, the authors argue that such a method would minimise the contribution of indicators, which do not move with the other indicators. This may reduce the reliability of the composite indicator because some indicators perform better in one cycle and others in a different cycle. Therefore, most such indicator systems in operation use an equal weighting system.

Presentation of the Index

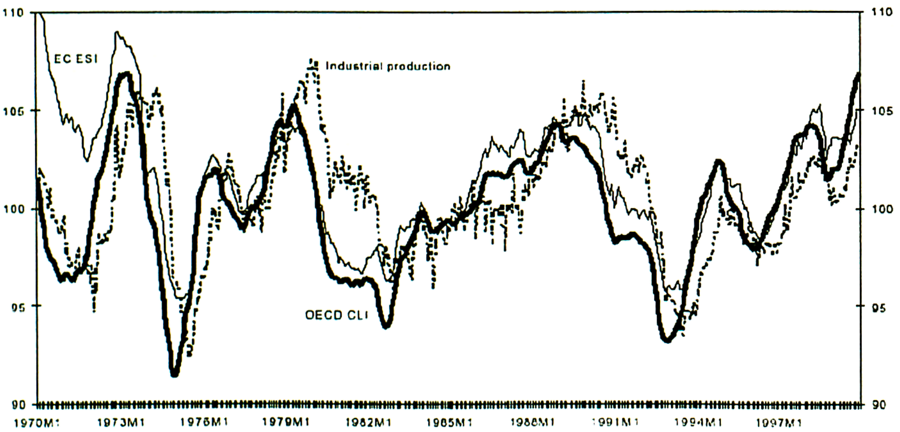

Figure 10. The EC ESI and the OECD CLI for the EU-15 are plotted against the total industrial production.

5.2.9 *"Information and communication technologies" (by Fagerberg J.)*

Scope of the Index

The index aims at providing an overall picture of a country's situation regarding development and application of information and communication technologies [11].

Description of sub-indicators

Five simple indicators (number of mobile telephones, number of Internet users, etc.) are used as components for the development of the composite indicator.

Preliminary treatment

The countries are ranked according to each indicator and the rankings are used as values for the sub-indicators.

Aggregation method

The composite indicator is calculated as the sum of the rankings.

Comments

- The authors preferred the simplicity by using this methodology, however the cardinal distances in the values of the indicators are not considered this way. This method can therefore "hide" how close two countries might be.

5.2.10 "Environmental Sustainability Index" (by the World Economic Forum)

Scope of the Index

The Environmental Sustainability Index (ESI) presented at the World Economic Forum 2001 is a measure of overall progress towards environmental sustainability developed for 122 countries [40]. A high ESI rank indicates that a country has achieved a higher level of environmental sustainability than most other countries; a low ESI rank signals that a country is facing substantial problems in achieving environmental sustainability along multiple dimensions.

Description of sub-indicators

The ESI scores are based upon a set of 22 sub-indicators, each of which combines two to six variables. In total 67 variables are considered. The indicators and variables were chosen through careful review of the environmental literature and available data combined with extensive consultation and analysis.

Preliminary treatment

The steps undertaken before the construction of the Index included:

- Making variables comparable, when necessary, by dividing by population, income or populated land area (area within each country with a population density ≥ 5 persons/km^2)
- Missing data imputation using bivariate correlations (where appropriate).
- Taking logarithms of highly skewed variables (this was done for 14 variables having a skewness measure greater than 5).
- Setting substantive thresholds where appropriate (e.g caloric intake, projected population growth rate).
- Truncating distributions to 95 % range, to account for inaccuracy of data at the extremes and to avoid that very extreme cases become benchmarks for the entire population.
- Standardizing variables by subtracting the mean and dividing by the standard deviation. The sign was changed for those variables where high observed values corresponded to low levels of environmental sustainability (e.g. pollution levels).

Aggregation method

The ESI is calculated by taking the average of the 22 standardized indicators and then converting this value to a standard normal percentile.

Comments

- The Index is highly correlated with the 1990-1998 GDP per capita growth (r =0.76), Human Development Index (r = 0.67) and WEF Current Competitiveness Index (r = 0.65).
- The authors state that if there was an independent, accepted measure of environmental sustainability then weights could be assigned based on the indicators' ability to match that measure efficiently and accurately. But environmental priorities and values vary widely based on development status, existing pollution levels, population density, and other factors.
- PCA was applied, but the set of principal components did not discriminate efficiently among the indicators and more problematically, it assigned negative weights to many indicators. For this reason aggregation via equal weighting was preferred.
- Alternatively to the equal weighting scheme, weights for the sub-indicators were defined following a survey process. A simple sensitivity analysis suggested that the weighting

methodology would not have changed the ranking in any appreciable fashion. In fact, the average shift in rank was only 1.7 places out of 122.

Presentation of the Index

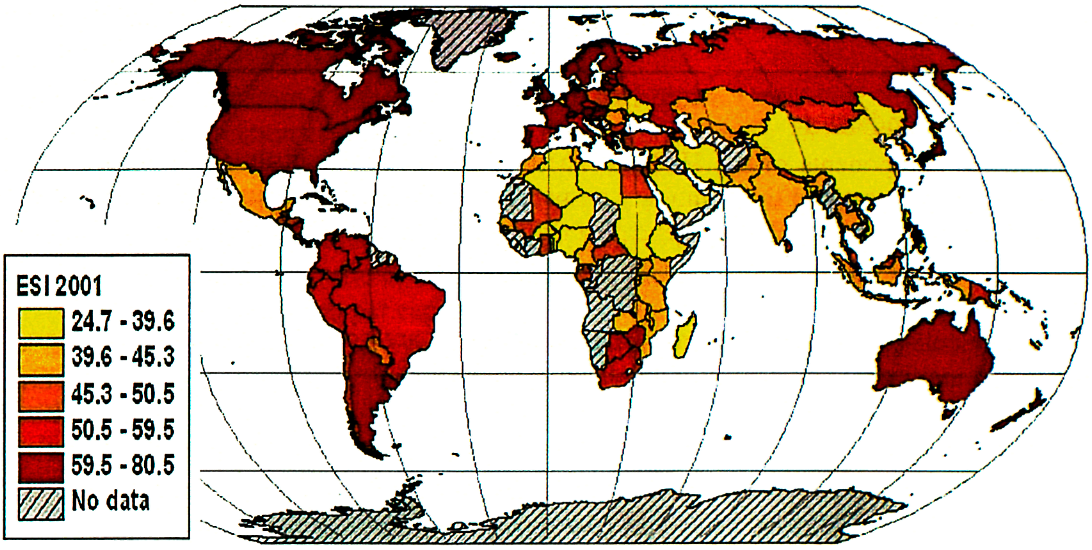

Figure 11. The Environmental Sustainability Index calculated for 122 countries around the world and presented in the form of a colored map.

5.2.11 *"Human Development Index" (by the United Nations)*

Scope of the Index

The Human Development Index (HDI) of the United Nations is a summary measure of human development in three basic dimensions: a long and healthy life, knowledge and GDP per capita [36].

Description of sub-indicators

The three base components of the HDI are: (a) life expectancy at birth, (b) adult literacy rate (with two-thirds weight) and the combined primary, secondary and tertiary gross enrolment ratio (with one-third weight) and (c) GDP per capita (PPP US$).

Preliminary treatment

Before the calculation of the HDI, the base components are scaled between [0, 1], using the minimum and maximum values for each indicator (goalposts).

Aggregation method

The HDI is calculated as the arithmetic average of the three scaled indicators.

Comments

- *Ogwang and Abdou (2000)* after applying PCA, they argue that there is a statistical justification for selecting only one of the three components of the index, i.e. life expectancy, without loss of too much information.

Presentation of the Index

1 Human development index

MONITORING HUMAN DEVELOPMENT: ENLARGING PEOPLE'S CHOICES

HDI rank [a]	Life expectancy at birth (years) 1999	Adult literacy rate (% age 15 and above) 1999	Combined primary, secondary and tertiary gross enrolment ratio (%) [b] 1999	GDP per capita (PPP US$) 1999	Life expectancy index 1999	Education index 1999	GDP index 1999	Human development index (HDI) value 1999	GDP per capita (PPP US$) rank minus HDI rank [c]
High human development									
1 Norway	78.4	.. [d]	97	28,433	0.89	0.98	0.94	0.939	2
2 Australia	78.8	.. [d]	116 [e]	24,574	0.90	0.99	0.92	0.936	10
3 Canada	78.7	.. [d]	97	26,251	0.89	0.98	0.93	0.936	3
4 Sweden	79.6	.. [d]	101 [e]	22,636	0.91	0.99	0.90	0.936	13
5 Belgium	78.2	.. [d]	109 [e]	25,443	0.89	0.99	0.92	0.935	4
6 United States	76.8	.. [d]	95	31,872	0.86	0.98	0.96	0.934	-4
7 Iceland	79.1	.. [d]	89	27,835	0.90	0.96	0.94	0.932	-3
8 Netherlands	78.0	.. [d]	102 [e]	24,215	0.88	0.99	0.92	0.931	5
9 Japan	80.8	.. [d]	82	24,898	0.93	0.93	0.92	0.928	2
10 Finland	77.4	.. [d]	103 [e]	23,096	0.87	0.99	0.91	0.925	5
11 Switzerland	78.8	.. [d]	84	27,171	0.90	0.94	0.94	0.924	-6
12 Luxembourg	77.2	.. [d]	73 [f]	42,769 [g]	0.87	0.90	1.00	0.924	-11
13 France	78.4	.. [d]	94	22,897	0.89	0.97	0.91	0.924	3
14 United Kingdom	77.5	.. [d]	106 [e]	22,093	0.87	0.99	0.90	0.923	5
15 Denmark	76.1	.. [d]	97	25,869	0.85	0.98	0.93	0.921	-7
16 Austria	77.9	.. [d]	90	25,089	0.88	0.96	0.92	0.921	-6
17 Germany	77.6	.. [d]	94	23,742	0.88	0.97	0.91	0.921	-3
18 Ireland	76.4	.. [d]	91	25,918	0.86	0.96	0.93	0.916	-11
19 New Zealand	77.4	.. [d]	99	19,104	0.87	0.99	0.88	0.913	3
20 Italy	78.4	98.4	84	22,172	0.89	0.94	0.90	0.909	-2
21 Spain	78.3	97.6	95	18,079	0.89	0.97	0.87	0.908	6
22 Israel	78.6	95.8	83	18,440	0.89	0.91	0.87	0.893	3
23 Greece	78.1	97.1	81	15,414	0.89	0.92	0.84	0.881	10
24 Hong Kong, China (SAR)	79.4	93.3	63	22,090	0.91	0.83	0.90	0.880	-4
25 Cyprus	77.9	96.9	69 [h]	19,006	0.88	0.87	0.88	0.877	-2

Figure 12. The Human Development Index is given in table format, along with the values of the sub-indicators.

5.2.12 *"Technology Achievement Index" (by the United Nations)*

Scope of the Index

The Technology Achievement Index (TAI) is designed to capture the performance of countries in creating and diffusing technology and in building a human skills base [36].

Description of sub-indicators

The index uses data from 8 indicators grouped in four dimensions:

- Technology creation as measured by the number of patents granted to residents per capita and by receipts of royalties and license fees from abroad per capita.
- Diffusion of recent innovations, as measured by the number of Internet hosts per capita and the share of high-and medium-technology exports in total goods exports.
- Diffusion of old innovations, as measured by telephones (mainline and cellular) per capita and electricity consumption per capita.
- Human skills, as measured by mean years of schooling in the population aged 15 and above and the gross tertiary science enrolment ratio.

Preliminary treatment

The observed minimum and maximum values for each indicator are chosen as goalposts and the performance in terms of each indicator is expressed as a value between 0 and 1. The sub-index for each dimension is then calculated as the simple average of the indicators in that dimension.

Aggregation method

The TAI is the simple average of these four sub-indices.

Comments

- No correlation analysis is presented for the sub-indicators or the sub-indices.

Presentation of the Index

The presentation of the Technology Achievement Index is given in a table format, similar to the one of the Human Development Index (see Figure 12).

5.2.13 *"Overall Health System Attainment" (by the World Health Organization)*

Scope of the Index

The World Health Organization has developed a composite index that summarizes the performance of health systems in 191 countries, in terms of both the overall level of goal achievement and the distribution of that achievement, giving equal weight to these two aspects [39].

Description of sub-indicators

Five components make up the index: overall good health, distribution of good health, overall responsiveness, distribution of responsiveness and fairness in financial contributions. Good health is measured by disability-adjusted life expectancy and the distribution of good health by an equality of child survival index. The overall responsiveness of the health system and the distribution of responsiveness are measured on the basis of survey responses relating to respect for patients and client orientation. Finally, fairness in financial contributions is estimated using the ratio of households' total spending on health to their permanent income above subsistence.

Preliminary treatment

The sub-indicators are scaled in a range between [0, 100].

Aggregation method

The choice of the weights has been based on a survey of preferences of informed individuals for these five components.

Comments

- A notable feature of the index is that all the main results are reported with uncertainty intervals (due to data and weight uncertainty) in order to communicate to the user the plausible range of estimates for each country on each measure. The WHO believe that it is the first time that an international agency has extended this good analytical practice to its official publications. The rank changes induced by variation of weights within plausible limits are found to be much less important than those implied by measurement errors in the data.

Presentation of the Index

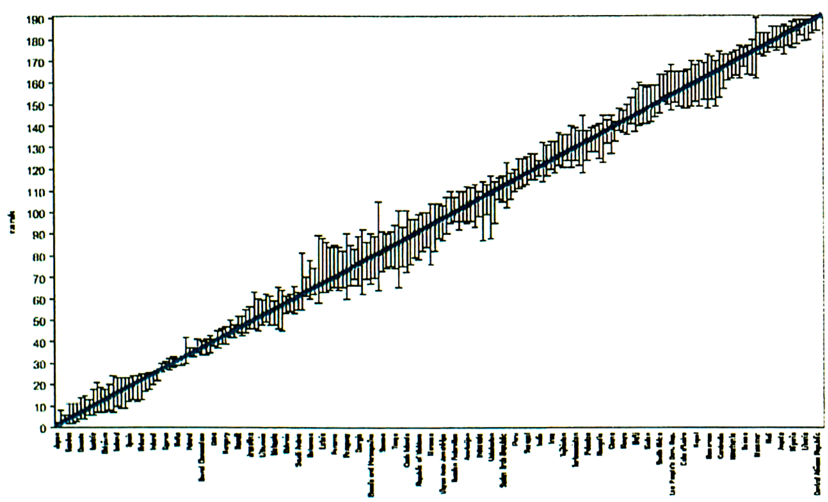

Figure 13. The WHO Overall Health System Attainment Index is given with uncertainty bounds (due to the inherent uncertainty in the data and the weights).

5.2.14 Two "Synthetic environmental indices" (by Isla M.)

Scope of the Index

In the review paper of Isla M., two composite indicators (one structural and one functional) are developed aiming to assist the local municipalities of Barcelona to monitor and evaluate their environmental performance [14].

Description of sub-indicators

Twenty-two sub-indicators for environment are combined into 2 synthetic indices, a structural and a functional one.

Preliminary treatment

No method of preliminary treatment is reported.

Aggregation method

The two composite indicators are calculated as the arithmetic average of the sub-indicators.

Comments

- Factor Analysis revealed a very low correlation among the sub-indicators. The authors considered thereafter that this canceled any attempt to use this statistical approach to obtain synthesis results.
- Experts participation to provide weights for each sub-indicator was dismissed because the authors claimed that the procedure involved a lot of judgment when the diversity of the areas could not guarantee better performance.

Presentation of the Index

TYPE	AVERAGE	VARIANCE
STRUCTURAL	77,69	39,87
FUNCTIONAL	31,97	175,04

Figure 14. The values for the structural and functional composite indicators are given as averages and variances among the municipalities of a town (e.g. Barcelona)

5.2.15 "National innovation capacity" (by Porter and Stern)

Scope of the Index

The central objective of the Index is to create a quantitative benchmark of national innovative capacity, which highlights the resource commitments and policy choices that most affect innovative output in the long run [27].

Description of sub-indicators

Eight sub-indicators are selected: personnel employed in R&D, expenditures on R&D, openness to International Trade & Investment, strength of protection for intellectual property, share of GDP spent on secondary and tertiary education, GDP Per Capita, percentage of R&D Funded by Private Industry and percentage of R&D Performed by Universities.

Preliminary treatment

Consistent with a rich and long empirical literature in economics and technology policy [16], the logarithmic values of the indicators are considered. This form emphasizes the interaction between international patent production and the elements of national innovative capacity.

Aggregation method

The Index uses statistical modeling to distinguish the relative importance of these contributors to national innovative capacity. Regression analysis is employed across a set of 17 OECD countries over a 25-year period from the 1970s through the mid-1990s to link these contributors to an internationally comparable and revealing measure of national innovative output—per capita "international" patenting. Thus, the composite indicator is a combination of the eight indicators, weighted by their contribution to building up this capacity calculated by the multiple regression model. This analysis provides a consistent and comparable way to assign relative weights to the different influences on national innovation capacity.

Comments

- The multiple linear regression model is used for forecasting.

Presentation of the Index

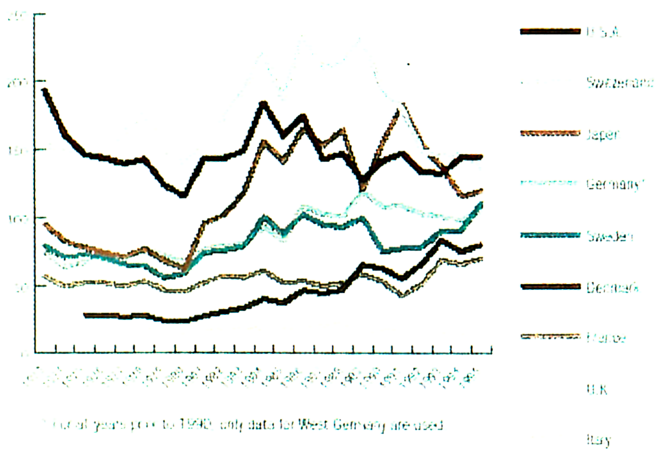

Figure 15. Historical National Innovation Capacity Index for selected countries, 1973-1995.

5.2.16 "General Indicator of Science & Technology" (by NISTEP, Japan)

Scope of the Index

The National Institute of Science and Technology Policy of Japan (NISTEP) created the General Indicator of Science and Technology (GIST) with a view to grasp major trends in Japan's Science and Technology activities and make possible comprehensive international comparisons and time-series analysis [23].

Description of sub-indicators

NISTEP starts with 13 indicators, five of which are classified as "input" and eight as "output". The cluster of inputs includes: "R&D expenditure", "R&D scientists/engineers", "Bachelor's of Science degrees conferred", "Bachelor's of Engineering degrees conferred", and "technology imports". As output are considered: "scientific papers", "scientific paper citations", "domestic patents", "external patents", "patent citations", "product output", "high tech product output" and "technology exports".

Preliminary treatment

Factor analysis has been used to analyze the structure of the two indicators sets, which is quantified after observation and interpretation of the meaning of the factor axes.

Aggregation method

PCA was employed to combine these indicators in two ways. First one general composite indicator was developed, the GIST, and then two additional composite indicators, one based on the set of the "input indicators" and one on the set of the "output indicators". The primary principal component of each set was adopted as the composite indicator.

Comments

- The FA did not cluster the indicators of the input set together and neither those of the output set. However, it was mentioned that this classification was judgmental.
- The plot of the composite indicator for input vs. the one of output reveals a strong quantitative relationship between them, which implies that higher effort (input) for S&T in a given country is accompanied by higher performance (output) and vice versa.

Presentation of the Index

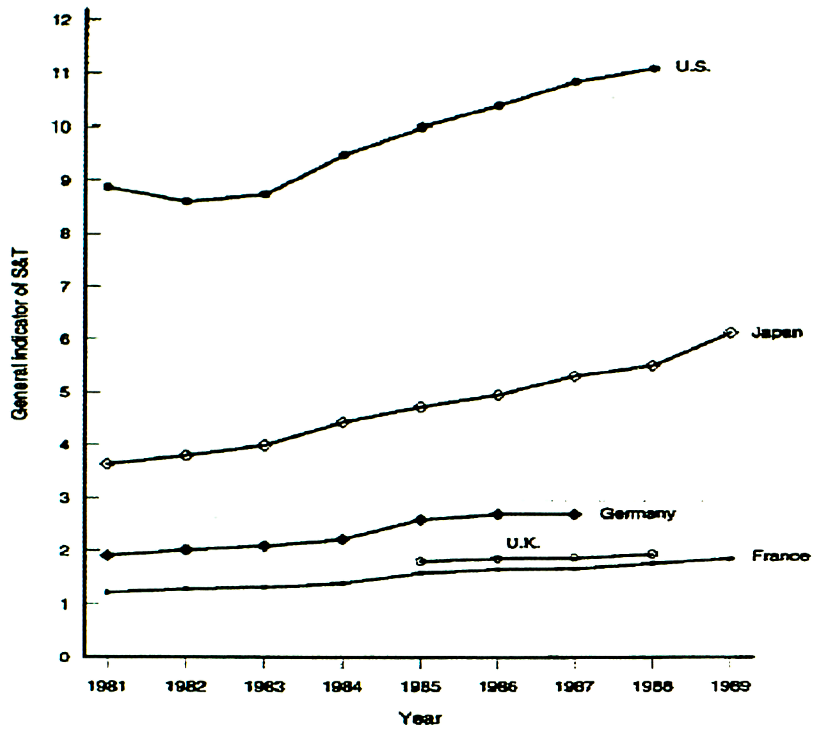

Figure 16. Trend in General Indicator of S&T of selected countries.

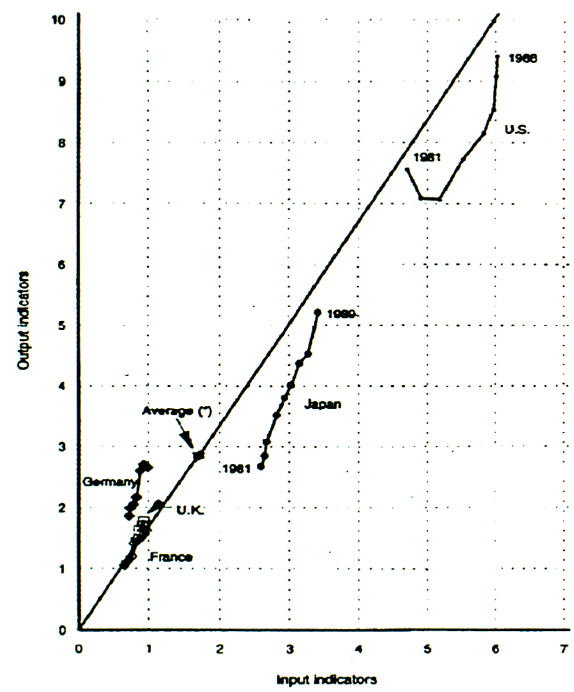

Figure 17. Composite indicator for inputs vs. composite indicator for outputs.

5.2.17 *"Success of software process improvement" (by Emam et al.)*

Scope of the Index

The Index aims at combining the conditions (e.g., organization and funding of improvement efforts) that can explain the successes and failures of software process improvement efforts [10].

Description of sub-indicators

A set of 14 variables is considered as having a significant impact on the software process improvement.

Preliminary treatment

The sub-indicators are qualitative (subjective scales) and no normalisation is applied.

Aggregation method

The authors expect that some of the variables are measuring the same construct, and therefore should be combined into one dimension. Principal components analysis provides a systematic way for performing this reduction, taking into account interactions among several explanatory variables. A composite indicator is therefore calculated as the sum of the 5 main principal components which account for 67% of the total variance.

Comments

- Each principal component is given an interpretation. Furthermore, the Cronbach alpha coefficient (α) for each PC is calculated, which ranges between 0.62 to 0.732. The α measures how well a set of variables represents a single unidimensional latent construct. *Nunnally (1978)* suggested that for the early stages of research a Cronbach alpha approaching the value 0.7 is acceptable. Given that the study is at a formative stage and exploratory in nature, it seemed reasonable to the authors to use this as a general guideline to judge the reliability of the composite variables.
- A second approach for weighting the variables is undertaken by means of a classification and regression tree (CART) algorithm. Constructing classification trees with CART requires that the dependent variable be dichotomized. In this case the composite indicator exists already. It is the dependant variable (unidimensional phenomenon).

Presentation of the Index

No figure for the composite indicator is presented.

5.2.18 *"European Labour Market Performance" (by Storrie and Bjurek)*

Scope of the Index

The objective of the composite index is to monitor labour market performance using many of the Basic Performance Indicators that are used in the benchmarking process according to the Amsterdam Treaty [35].

Description of sub-indicators

For illustration purposes, three measures of unemployment from the Commission's Basic Performance Indicators are benchmarked. The three single indicators are: the unemployment rate, the long-term unemployment rate and the youth unemployment ratio.

Preliminary treatment

The variables are scaled between [0, 100], with 0 corresponding to best performance.

Aggregation method

The efficiency frontier approach (objective method) is used for the calculation of the composite index. The first assumption of the methodology is that a country dominates another when it is best in both indicators. The weight assignment is not based on some value judgement but on the data. More precisely, after the frontier is identified (countries that perform best), the weighting depends upon on the location of the various countries relative to the countries that lie on the performance frontier and that exhibit a similar mix of the indicators. Different countries are weighted differently depending upon where they are located in relation to the frontier.

Comments

- The method is proposed by the authors as extremely parsimonious with regard to the weighting assumptions because it lets the data decide on the weighting issue.
- McCarthy [20] expresses however concern that such an empirical construct might not indicate the appropriate direction of a policy for a given country in order to improve its situation. She adds further that, since employment guidelines exist, one cannot use a completely objective approach to evaluate overall labour market performance, but instead should attach different weights to different indicators.

Presentation of the Index

The values of the index are given in the form of a table.

5.2.19 "Eco-indicator 99" (by Pre Consultants, Netherlands)

Scope of the Index

The Eco-indicator 99 is a state of the art, damage oriented impact assessment method for materials and processes, which has been developed by a large team of experts from 1997 to 1999 [28]. The composite index, which is calculated via as a user-friendly tool, aims to assist designers and product managers to improve products.

Description of sub-indicators

The Eco-indicator 99 addresses three damage categories (endpoints): (a) human health, (b) ecosystem quality and (c) resources, minerals and fossil fuels. Damages to human health are expressed as DALY (Disability Adjusted Life Years). Models have been developed for respiratory and carcinogenic effects, the effects of climate change, ozone layer depletion and ionising radiation. Damages to Ecosystem Quality are expressed as the percentage of species that have disappeared in a certain area due to the environmental load. Resource extraction is related to the quality of the remaining mineral and fossil resources.

Preliminary treatment

The results of the models for the three damage categories have different units. In order to render them dimensionless, they are divided by a reference ("normal") value. As the Eco-indicator is developed for Europe, they use the European reference values.

Aggregation method

Weighting has been considered to be the most controversial step in the life cycle impact assessment. The finally selected weighting scheme is based on expert judgement.

Comments

- The authors believe that the weights should represent the views of society or a group of stakeholders.
- Weighting based on target values was not selected because it is often hard to interpret the basic values that are underlying the decisions of society. For instance policy targets set by governments are often a compromise between the need to reduce loads and the preparedness to make the necessary sacrifices.
- Significant differences in the weighting sets due to cultural perspectives were found. However, the overall result is that Human Health and Ecosystem Quality are considered to be of almost equal importance, while Resources are considered to be half as important.

Presentation of the Index

No figure for the composite indicator is presented.

5.2.20 *"Concern about environmental problems" (by Parker)*

Scope of the Index

The index proposed by Parker aims to measure the concern of the public on certain environmental problems [26]. Three countries are chosen as case studies: Italy, France and the UK. The European Community as a whole is additionally considered.

Description of sub-indicators

Eleven indicators are considered, four related to air problems (nitrogen oxides, sulphur dioxide, carbon dioxide and particulates), two indicators associated with water problems (bathing and fertilizers) and five landscape-related indicators (population change, new dwellings, tourism, traffic and waste).

Preliminary treatment

Each indicator is normalized by dividing its value in each year by its value for the year for which it is first available.

Aggregation method

Weights for the sub-indicators are derived from public opinion polls. The measure of concern for each indicator (problem) is calculated by multiplying the proportion of people who said they worried a "great deal" by 3, a "fair amount" by 2 and "not very much" by 1.The multiplied proportions are added together and finally divided by 100. The weights are further normalised to sum up to 1. The composite indicator is the sum of the normalised weights multiplied by the corresponding normalised indicators. If a sub-indicator is missing, then the weight for this indicator is not applied and the remaining weights are normalised so that they always sum up to 1.

Comments

- In view of the potentially crucial importance of the weights, the effect of five sets of weights derived from several public opinion polls in different countries and years were used for the UK. Sensitivity analysis indicated similar behavior of the composite indicators produced by the different weighting schemes, which indicates that public opinion about the main threats to the environment is remarkably stable across both space and time. Therefore fears that the public evaluates environmental issues on an irrational basis, and therefore weights base upon public opinion will produce instability, appear to be unfounded.
- To improve the comparability of the composite indicator values across countries, the author considered whether division of the sub-indicators' by country's population, area or GNP (Gross National Product) would be useful. He found that dividing for example the nitrogen oxide indicator by population, would make France seem much worse than Italy, whereas dividing it by surface area or GNP would improve the position of France compared to Italy. This conclusion, in a general context, is very important and should be considered in any aggregation approach.

Presentation of the Index

The composite indicator is presented as a time-series.

5.2.21 "National Health Care systems performance" (by King's Fund, England)

Scope of the Index

The composite index aims to measure the performance of all 120 Health Authorities in England, Scotland and Wales with a view to reveal whether there is (a) variation in health care standards across the country, (b) gulf in the health of town and city dwellers, and (c) an important impact of poverty on the health service [18].

Description of sub-indicators

Six sub-indicators were selected which cover various aspects of the performance of the National Health Care System: deaths from cancer, deaths from heart disease, total number of people on hospital waiting lists, percentage of people on waiting lists waiting over 12 months, number of hip operations and deaths from 'avoidable' diseases (e.g. TB, asthma etc).

Preliminary treatment

No standardisation method is applied to the sub-indicators.

Aggregation method

The approach for weighting the sub-indicators that was finally adopted was the "budget allocation" technique, which emphasises that resources for health care are not infinite, and that therefore choices must be made about how much should be spent on different health care services. A polling organization carried out a "budget allocation" survey of 1,000 people across the country. Respondents were asked to distribute a fixed sum of 60 points to some or all of the six sub-indicators. More points spent in an area, the greater the improvement on that sub-indicator. The composite indicator produced was the weighted average of the six sub-indicators.

Comments

- A first way of finding out the relative importance of the sub-indicators was to ask a sample of the public to simply rank the six measures – the most desired at the top, and the least favoured at the bottom. However, this option was not selected because it does not give people the chance to say how much more they value one measure over another.
- Applying the "budget allocation" method it was found that there was little difference in the distribution of the chips depending on respondents' sex, social class, age or the area of the country where they lived.

Presentation of the Index

Position	Health Authority / Health Board
1	Oxfordshire
2	North & East Devon
3	Herefordshire
4	Somerset
5	Dorset
...	
117	St Helens & Knowsley
118	East London & The City
119	Liverpool
120	Manchester

Figure 18. The composite indicator is used for ranking the Health Authorities/Boards on a national basis.

5.2.22 "Index of sustainable and economic welfare" (by CES and NEF)

Scope of the Index

The Index of sustainable and economic welfare (ISEW) is one of the most advanced attempts to create an indicator of economic welfare, developed by the Centre for Environmental Strategy (CES) and the New Economics Foundation (NEF). The main objective is to measure the portion of economic activity that delivers welfare to people. It aims further to replace GDP as an indicator of progress, because GDP is likely to lead in the wrong direction given that it does not distinguish between activities that improve or directly damage the quality of life [2].

Description of sub-indicators

The set of twenty sub-indicators includes seven economic activities that deliver welfare to people, such as adjusted consumer expenditure, services from domestic labour, from consumer durables, from streets and highways, public expenditure on health and education, net capital growth and net change in international position. On the other hand, the thirteen indicators that "reduce" the welfare are: consumer durables (difference between expenditure and value of services), defensive private expenditures on health and education, costs of commuting, of personal pollution control, of automobile accidents, of water pollution, of air pollution, of noise pollution, loss of natural habitats, loss of farmlands, depletion of non-renewable resources, costs of climate change and costs of ozone depletion.

Preliminary treatment

All sub-indicators are expressed in monetary terms. The seven indicators that are related to activities delivering welfare to people are assigned a positive sign, while the thirteen indicators that are associated with a reduction of the welfare are assigned a negative sign.

Aggregation method

The composite indicator is calculated as the simple arithmetic average of the indicators considering the positive and negative signs.

Comments

- The Index has been criticized as subjective and too susceptible to changes in the assumptions that underpin it. To answer these arguments, a relevant website has been created to allow the user to change the weights and assumptions used in the index. According to the authors, this site shows that the index is a good basis on which to construct alternative indicators

Presentation of the Index

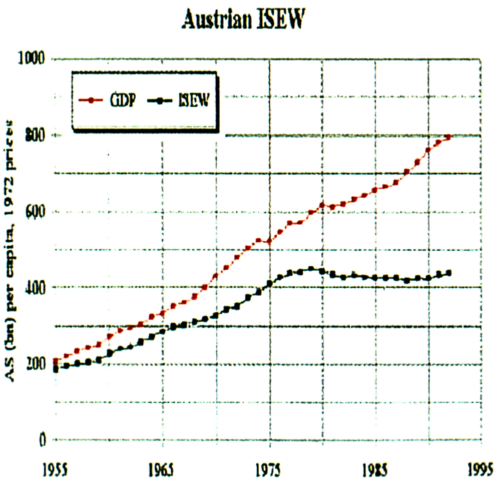

Figure 19. The Index of Sustainable and Economic Welfare (ISEW) for Austria for 1955-1997 and the corresponding GDP values.

5.2.23 *"Index of Environmental Friendliness" (by Puolamaa et al., Eurostat)*

Scope of the Index

The objective of the Index of Environmental Friendliness is to provide diversified quantitative information for environmental and decision-making and overall environmental discussion and communication [29].

Description of sub-indicators

The index is designed to cover the key environmental problems of greenhouse effect, ozone depletion, acidification of soil and water, eutrophication, ecotoxicological effect, resource depletion, photo-oxidation, biodiversity, radiation and noise.

Preliminary treatment

The initial data, expressed as pressure data, are first converted to problem indices using equivalency factors. The problem indices are then normalized on the basis of national total pressures.

Aggregation method

In order to determine the weights for the 11 components of the index, a comprehensive pilot study on societal preferences was conducted in Finland. The valuation of environmental concerns was done by 52 representatives of stakeholders of environmental decision-making. The pair-wise comparison method was used for the valuation study and the results were analyzed with the Analytic Hierarchy Process (AHP). The overall Index of Environmental Friendliness is the weighted sum of the eleven normalized problem indices.

Comments

- AHP is well suited to the type of complex decision-making problems involved and to the multiple goals related to the decision-making. The main advantage of AHP is that it is based on pair-wise comparison; the human mind can easily handle two distinct problems and examine their differences. Another advantage of AHP is that unlike many other methods based on Utility Theory, its use for purposes of comparisons does not require a universal scale.
- High inconsistency values in the weights given by some respondents could have been avoided if an interactive research design had been followed.
- An analysis of uncertainties of the Index was undertaken, integrating uncertainties of individual pressure data, equivalency factors and inconsistencies of the valuation of environmental concerns to an overall assessment. However, the authors pointed out that uncertainties are very seldom documented or assessed in the existing databases.

Presentation of the Index

Valuation of env. concerns	Greenhouse Effect Index 15,6		Acidification Index 5.5_{Aqua}, $6.5_{Terr.}$		Eutrophication Index 5,8		Intermediate Score for Potential Environmental Damage	
	Value weighted normalised problem indices							
	direct	total	direct	total	direct	total	direct	total
Mining of metal ores	0,04	0,06	0,02	0,02	0,01	0,01	0,07	0,09
Other mining and quarring	0,28	0,35	0,19	0,20	0,01	0,01	0,47	0,56
Mnf. of food products	1,53	2,65	0,49	0,71	0,13	0,15	2,15	3,51
Mnf. of textiles	0,11	0,16	0,02	0,03	0,01	0,01	0,14	0,20
Mnf. of leather	0,02	0,21	0,01	0,02	0,02	0,02	0,04	0,25
Mnf. of wood	0,33	2,30	0,18	0,39	0,07	0,09	0,58	2,78
Mnf. of pulp and paper	5,44	13,06	2,32	4,36	5,62	5,78	13,38	23,20
Publishing and printing	0,03	0,31	0,01	0,07	0,00	0,00	0,04	0,39
Mineral oil and chemical industry	6,42	7,18	2,96	3,36	0,62	0,65	10,00	11,18
Mnf. of rubber and plastic products	0,07	0,31	0,09	0,14	0,01	0,01	0,16	0,46
Mnf. of non-metallic mineral products	2,38	1,90	0,61	0,54	0,08	0,07	3,07	2,51
Mnf. of basic metals	10,82	5,85	2,02	1,69	0,39	0,38	13,23	7,92
Mnf. of fabricated metal products	0,08	0,27	0,02	0,06	0,01	0,01	0,11	0,35
Mnf. of machinery and equipment	0,09	0,52	0,06	0,12	0,01	0,01	0,16	0,65
Mnf. of electrical machinery	0,03	0,28	0,00	0,06	0,00	0,01	0,04	0,35
Mnf. of motor vehicles	0,11	0,26	0,03	0,07	0,00	0,00	0,15	0,33
Other mnf.	0,01	0,09	0,00	0,01	0,00	0,00	0,02	0,11
Energy supply	62,14	8,00	14,20	2,11	1,05	0,18	77,40	10,29
Fresh water supply	0,06	0,14	0,01	0,03	0,00	0,00	0,07	0,17
Mnf. total	89,98	43,88	23,26	14,00	8,07	7,43	121,31	65,30
Finland total	156,00	156,00	64,00	64,00	58,00	58,00	278,00	278,00

Figure 20. The Index of Environmental Friendliness is presented in a table format and calculated for several manufacturing companies in Finland. For illustration purposes only 4 sub-indicators are used for the calculation. The values of the sub-indicators corresponding to "direct" are the unweighted ones, while those under column "total" correspond to the weighted values. Weights are 15.6 for the Greenhouse effect index, 5.5 for the water acidification index, 6.5 for the land acidification index and 5.8 for the eutrophication index.

5.2.24 "Environmental Policy Performance Indicator" (by Adriaanse A., the Netherlands)

Scope of the Index

The composite indicator aims to monitor the trend in the total environmental pressure in the Netherlands and indicate whether the environmental policy is heading in the right direction or not [1].

Description of sub-indicators

Six theme indicators (composed of several simple indicators) are combined, including: (a) change of climate, (b) acidification, (c) eutrophication, (d) dispersion of toxic substances, (e) disposal of solid waste, and (f) odour and noise disturbance.

Preliminary treatment

Two independent approaches are followed for the scaling of the theme indicators: division by the corresponding a) sustainability levels and b) policy targets for each theme indicator. This results in a common unit (i.e. environmental pressure equivalents- EPeq) for expressing environmental pressure within themes.

Aggregation method

The composite indicator for each year is calculated as the sum of the six dimensionless theme equivalents.

Comments

- The alternative use of target values and sustainability levels for the scaling of the theme indicators results in images depicting similar trend for the composite indicator.

Presentation of the Index

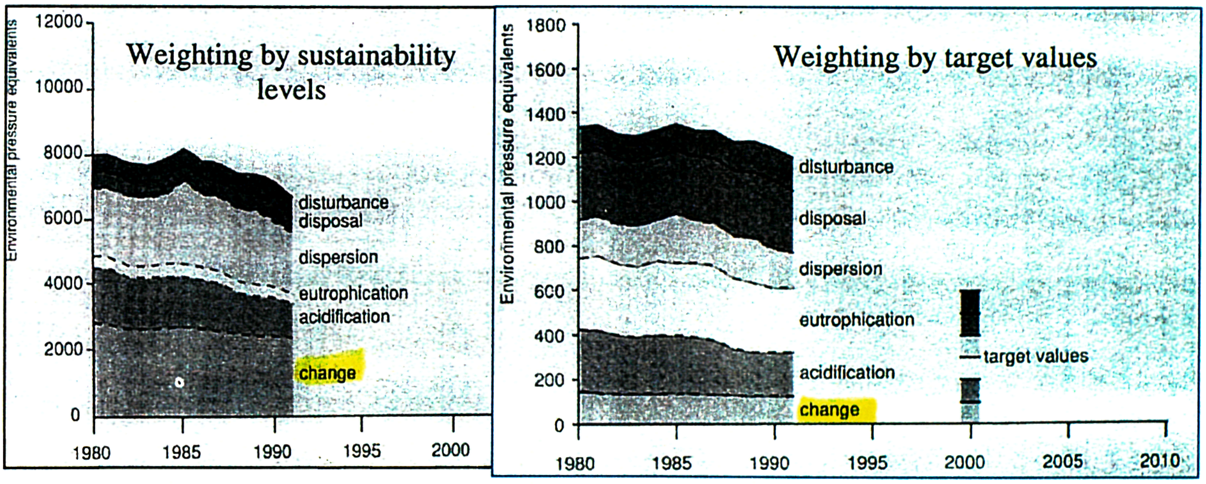

Figure 21. The Environmental Policy Performance Indicator for the Netherlands 1980-1992, presented as the sum of the six theme indicators weighted by sustainability levels (left graph) and target values (right graph). The target values for 2000 are also given.

6 *Conclusions of the review*

6.1 Overall Conclusions

Composite indicators have received a substantial amount of attention in recent years and various methodologies have been adopted to handle the issue. The methods (statistical and participatory) and the case studies on aggregating indicators that have been presented are only a small subset of those available in the literature. They were chosen since, in our opinion, they are the most representatives of the philosophy underlying the development of composite indicators as well as the most established in the literature.

Our conclusion at this stage of the review is that there are several methods, which are applied for developing composite indicators, depending on the knowledge of the developers, or the complexity of the data. Participatory methods, in the form of experts opinion or public opinion polls, are often preferred for the evaluation of the importance of the indicators in respect to purely statistical methods, so that the composite indicator will be accepted by the public and the policy-makers. As Brundtland [13] argues "...All political decisions must, in fact, be supported by, or have the potential support of the people". An equal weighting system is adopted in various case studies, given that, as the developers of the index claim, there is no evidence of the sound procedure to follow.

The consideration of the uncertainty associated to the data and the weights in building composite indicators is cited in very few studies. The Human Development Index produced annually since 1990 by the United Nations Development Programme has encouraged improvement in the indicators used in its formulation. "No index can be better than the data it uses. But this is an argument for improving the data, not abandoning the index." [37]. A good analytical practice is to report the values of the composite indicator with uncertainty intervals (due to data and weight uncertainty) in order to communicate to the user the plausible range of estimates and extend this practice to the official publications.

A typical feature of some weighting procedures is the identification of correlations in the set of indicators. However, the weights assigned by methods based on correlations bear no relation to the underlying analytical model that the composite indicator is trying to represent. As a result, a good strategy is to use several different analytical techniques to explore groups of indicators. The analysis of correlation patterns, through the use of some of the techniques presented here like PCA, is indeed extremely interesting. However, the final choice of weights should be made on the basis of sound qualitative and theoretical understanding of the phenomena in question. The results of the composite indicator can further be analysed through the implementation of uncertainty and sensitivity analysis. These types of analyses could well assist the defensibility of the results.

6.2 When to use what

On the basis of the examples reviewed so far, building a composite indicator includes consideration of the

- degree of correlation among the sub-indicators;
- number of sub-indicators; and
- diverse nature of sub-indicators.

There is no "universal recipe" for aggregating composite indicators. In the following paragraphs we present a number of considerations that should help an investigator in making such a choice. Primary to the construction of a composite indicator several considerations should be made involving:

- Making variables comparable, when necessary, by dividing by population, income, land area etc. Care should be given if the choice of the denominator variable has a great influence on the results;
- Missing data imputation and how different methods affect the results;
- Taking logarithms of highly skewed variables;
- Truncating distributions to 95 % range, to account for inaccuracy of data at the extremes and to avoid that very extreme cases become benchmarks for the entire population;
- Standardizing variables, for example using minimum and maximum values or by subtracting the mean and dividing by the standard deviation. The advantage of the latter scaling is that it avoids attaching excessive importance to extreme values.

The choice of which method of aggregation to select is difficult, since each method has strengths and weaknesses. Such a choice depends on the objective of the composite indicator, the characteristics of the sub-indicators and also on the computational cost that the investigator can afford. A few issues to take into account are:

(a) The ***equal weighting*** can be applied after a proper scaling of the sub-indicators. Equal weighting works well if all sub-indicators are uncorrelated, or they are all highly correlated. However, when a few highly correlated indicators are involved, this method, albeit simple, may not provide the best means of aggregation.

(b) ***Multiple regression models*** can handle a large number of indicators. This approach can be applied in cases where the sub-indicators considered as input to the model are related to various policy actions and the output of the model is the target. The regression model, thereafter, could quantify the relative effect of each policy action on the output, i.e. the single indicator. However, this means that there must be a "dependent variable" that accurately (and satisfactorily) measures the target in question. Measuring the influence of a number of independent variables on this policy target is a reasonable question. However, in such cases the dependent variable is not a composite indicator, as this is defined by the Note on Composite Indicators. Alternatively such an approach could be used for forecasting purposes. In a more general case of multiple output indicators, *canonical correlation analysis* that is a generalization of multiple regression could be applied. However, in any case, there is always the uncertainty that the relations, captured by the regression model for a given range of inputs and output, may not be valid for different ranges.

(c) The ***Principal Components Analysis*** is a very interesting exploratory technique to examine the correlation structure of groups of variables. In the development of composite indicators, it has been argued to apply PCA to identify the dimensions of the data and/or define the weights for the base indicators, while ***Factor Analysis*** is usually employed as a supplementary method with a view to examine thoroughly the interrelationsips of the base indicators. However, there are two crucial problems with these arguments. First, weights assigned to sub-indicators in both of these techniques are based on correlations which do not necessarily correspond to the underlying relationships between the sub indicators and the phenomena being measured. In other words there is confusion between correlation and causality. It is not possible to know (or estimate) the *real* weights since we would need a dependent variable. If there were a satisfactory dependent variable there would be no need for a composite indicator. It is further not advisable to use PCA when the base indicators have different cycles, as this would reduce the reliability of the composite indicator because some indicators perform better in one cycle and others in a different cycle [24].

(d) The ***Cronbach alpha*** is used to verify whether a set of indicators measures a single dimension. A coefficient of $\alpha = 0.80$ or higher is considered in most applications as "evidence" that the indicators are measuring the same underlying construct. When the indicators have a multidimensional structure, Cronbach alpha will usually be low. In such case, PCA or FA could be undertaken to see which indicators load highest on which dimensions, and then calculate the alpha of each subset of indicators separately to verify the reliability of the analysis.

(e) The ***neutralization of correlation effect*** method works well when there are a few indicators, some of which are highly correlated or are different forms of the same issue. This is an empirical method, which has received limited application.

(f) The ***efficiency frontier*** approach is extremely parsimonious as regards the weighting assumptions, because it lets the data decide on the weighting issue. It is argued, though, that such an empirical approach might not indicate the appropriate direction of a policy for a given country in order to improve its situation.

(g) A way to avoid the immediate selection of weights is to measure the need for political intervention and the "urgency" of a problem by the ***distance to target*** approach. Using policy goals as targets convinces the policy makers for the "soundness" of the weighting method, as long as those policy makers have defined the policy targets themselves. This approach is technically feasible when there is a well-defined basis for a certain policy, such as a National Policy Plan or similar reference documents. For international comparisons, such references are often not available, or they deliver contradictory results. Another counter-argument for the use of policy goals as targets is that the benefits of a given policy must be valued independently of the existing policy goals.

(h) ***Expert judgement*** is adopted when a participatory method of evaluating the weights is sought. It is essential to bring together experts that have a wide spectrum of knowledge, experience and concerns, so as to ensure that a proper weighting system is found for a given application. The ***budget allocation*** is optimal for a maximum number of 10-12 indicators. If a too large number of indicators is involved, this method can give serious cognitive stress to the experts who are asked to allocate the budget.

(i) ***Public opinion polls*** have been extensively employed for many years for the setting of weights. In public opinion polls, issues are selected which are already on the public agenda, and thus receive roughly the same attention by the media. In many case studies, public opinion polls in different countries and years resulted in similar weighting schemes for certain environmental problems, which indicates that public opinion about the main threats to the environment is remarkably stable across both space and time. Therefore fears that the public evaluates environmental issues on an irrational basis, and therefore weights base upon public opinion will produce instability, appear to be unfounded.

(j) The ***Analytic Hierarchy Process*** is a widely used technique for multi-attribute decision making and as weighting method enables the decision-maker to derive weights as opposed to arbitrarily assign them. An advantage of AHP is that unlike many other methods based on Utility Theory, its use for purposes of comparisons does not require a universal scale. Furthermore, AHP tolerates inconsistency in the way people think through the amount of redundancy (more equations are available than the number of weights to be defined). This redundancy is a useful feature as it is analogous to estimating a number by calculating the average of repeated observations. The resulting weights are less sensitive to errors of judgement. These advantages render the weights derived from AHP defended and justified in front of public.

7 *Uncertainty and sensitivity analysis for composite indicators*

7.1 Introduction

Uncertainty analysis (UA) allows the analyst to assess the uncertainty associated with the composite indicator values (or model in a more general context) as the result of the propagation through the errors in the sub-indicators data, and uncertainties in the weights of the sub-indicators. Sensitivity analysis (SA) studies how the variation in the values of a composite indicator can be apportioned, qualitatively or quantitatively, to different sources of variation, and of how the given composite indicator depends upon the information fed into it. On this basis, we contend that UA and SA are prerequisites for building composite indicators.

An illustration of how UA & SA can contribute to the transparency during the development of a composite indicator is given by exemplifying the technology achievement index (TAI), developed by the United Nations and described in detail in the Human Development Report 2001 [36]. The TAI is intended to help policy-makers propose better technology investment and help countries situate themselves relative to others, especially those farther ahead. It is designed to capture the performance of countries in creating and diffusing technology and in building a human-skills base. The index measures achievements in four dimensions:

- Technology creation, as measured by the ***number of patents*** granted to residents per capita and by ***receipts of royalties and license fees*** from abroad per capita.
- Diffusion of recent innovations, as measured by the ***number of Internet hosts*** per capita and the share of ***high-and medium-technology exports*** in total goods exports.
- Diffusion of old innovations, as measured by ***telephones*** (mainline and cellular) per capita and ***electricity consumption*** per capita.
- Human skills, as measured by ***mean years of schooling*** in the population aged 15 and above and the ***gross tertiary science enrolment ratio***.

In our current analysis, each of the eight sub-indicators has been normalised using the mean and the standard deviation of each series, for the first 50 countries. The TAI is then calculated as the weighted average of the normalised sub-indicators. The weights have been derived from two pilot surveys (internal JRC surveys) using the methods of budget allocation [6] and analytic hierarchy process [31]. For comparison purposes, TAI has also been calculated as the simple average of the normalized indicators.

7.2 Uncertainty analysis of the composite indicator

In this section, the weights of the sub-indicators are considered as uncertain, due to the plurality of perspectives of the various stakeholders. Two surveys, each of 22 individuals informed about the objective of the TAI and the various sub-indicators composing it, resulted in 39 sets of weights, which were calculated using budget allocation and analytic hierarchy process. For the purposes of the uncertainty analysis, the weights of the sub-indicators are assumed uniformly distributed and sampled in their entire acceptable range, determined herein between the 10^{th} and 90^{th} percentiles of the 39 weights. The 39 sets of weights and the acceptable range for each weight are given in Table 5.

Table 5. Thirty-nine sets of weights for the eight sub-indicators composing TAI. The weights have been derived from two internal JRC surveys using budget allocation and analytic hierarchy process. Last two rows indicate the 10^{th} and 90^{th} percentiles, which define the acceptable range for each weight.

No. of set	Patents	Royalties	Internet hosts	Technology exports	Telephones	Electricity	Schooling	University students
1	0.025	0.056	0.056	0.445	0.052	0.015	0.138	0.213
2	0.038	0.045	0.029	0.179	0.023	0.023	0.274	0.389
3	0.045	0.040	0.357	0.070	0.019	0.012	0.176	0.280
4	0.024	0.041	0.133	0.225	0.062	0.031	0.242	0.242
5	0.076	0.202	0.092	0.279	0.054	0.056	0.127	0.115
6	0.024	0.054	0.066	0.120	0.085	0.116	0.245	0.290
7	0.094	0.024	0.065	0.171	0.041	0.019	0.353	0.234
8	0.065	0.265	0.038	0.128	0.027	0.017	0.229	0.229
9	0.080	0.065	0.067	0.121	0.036	0.020	0.249	0.362
10	0.123	0.028	0.033	0.440	0.020	0.013	0.083	0.260
11	0.176	0.099	0.103	0.227	0.033	0.053	0.047	0.262
12	0.114	0.153	0.013	0.154	0.013	0.141	0.094	0.317
13	0.090	0.049	0.011	0.030	0.018	0.418	0.138	0.246
14	0.116	0.288	0.026	0.302	0.030	0.028	0.055	0.155
15	0.078	0.110	0.051	0.182	0.066	0.088	0.165	0.259
16	0.229	0.135	0.059	0.176	0.039	0.039	0.252	0.072
17	0.109	0.103	0.029	0.117	0.030	0.014	0.301	0.297
18	0.029	0.081	0.070	0.396	0.165	0.029	0.026	0.204
19	0.050	0.050	0.050	0.200	0.100	0.050	0.200	0.300
20	0.150	0.150	0.100	0.150	0.100	0.050	0.200	0.100
21	0.200	0.000	0.100	0.200	0.000	0.000	0.200	0.300
22	0.070	0.070	0.100	0.150	0.130	0.060	0.230	0.190
23	0.100	0.150	0.150	0.300	0.050	0.050	0.200	0.000
24	0.100	0.050	0.200	0.100	0.150	0.050	0.150	0.200
25	0.200	0.200	0.050	0.100	0.100	0.150	0.150	0.050
26	0.100	0.050	0.150	0.200	0.100	0.100	0.150	0.150
27	0.100	0.100	0.150	0.150	0.100	0.050	0.200	0.150
28	0.200	0.050	0.050	0.300	0.050	0.000	0.050	0.300
29	0.120	0.150	0.120	0.150	0.100	0.120	0.080	0.160
30	0.050	0.050	0.200	0.100	0.050	0.050	0.250	0.250
31	0.100	0.050	0.100	0.250	0.050	0.150	0.100	0.200
32	0.050	0.050	0.050	0.100	0.200	0.050	0.200	0.300
33	0.100	0.300	0.020	0.330	0.030	0.020	0.050	0.150
34	0.100	0.150	0.100	0.200	0.150	0.100	0.100	0.100
35	0.150	0.130	0.110	0.140	0.100	0.100	0.150	0.120
36	0.100	0.150	0.100	0.200	0.100	0.000	0.150	0.200
37	0.120	0.220	0.090	0.090	0.090	0.090	0.100	0.200
38	0.050	0.020	0.200	0.150	0.150	0.030	0.150	0.250
39	0.100	0.100	0.100	0.250	0.150	0.050	0.050	0.200
10^{th} prc.	*0.036*	*0.037*	*0.028*	*0.100*	*0.020*	*0.012*	*0.050*	*0.100*
90^{th} prc.	*0.180*	*0.205*	*0.160*	*0.307*	*0.150*	*0.124*	*0.250*	*0.300*

The values of the composite indicator for each country were calculated 2000 times, using 2000 random sets of weights, each weight sampled within its acceptable range in a Monte Carlo-like procedure, applying the extended FAST sampling method [32]. FAST has been used as it allows the analyst to perform both uncertainty and sensitivity analysis. The results of the uncertainty analysis for the 50 countries are displayed in the form of bars (the median, i.e. 50^{th} percentile of the TAI values) and the associated confidence bounds, corresponding to the 5th and 95th percentile of the TAI values, as shown in **Figure 22** (left graph). The right graph of **Figure 22** presents the TAI values without UA for the same countries as before, however this time only the mean values of the weights that derived from the analytic hierarchy process, the budget allocation and the simple average are used. In the right graph of **Figure 22** the countries are ordered according to the TAI-equal weights.

The analysis shows that some countries perform clearly better than others, but a few countries show significant overlap and therefore the ranking is not clear. For instance, Sweden performs better than Norway for any combination of weights. However, this is not the case for Sweden and United States for example, where in half of the cases the one country performs better than the other. This can be explained by the fact that Sweden and United States have similar values for the sub-indicators and thereafter for the composite indicator, so the change of the weights favors slightly one country over the other. It can further be noticed that the TAI values for Norway, Australia and New Zealand span over a wide range (wide uncertainty bounds). This could be attributed to the fact that the values of the sub-indicators for those countries are very high for some sub-indicators and very low for others.

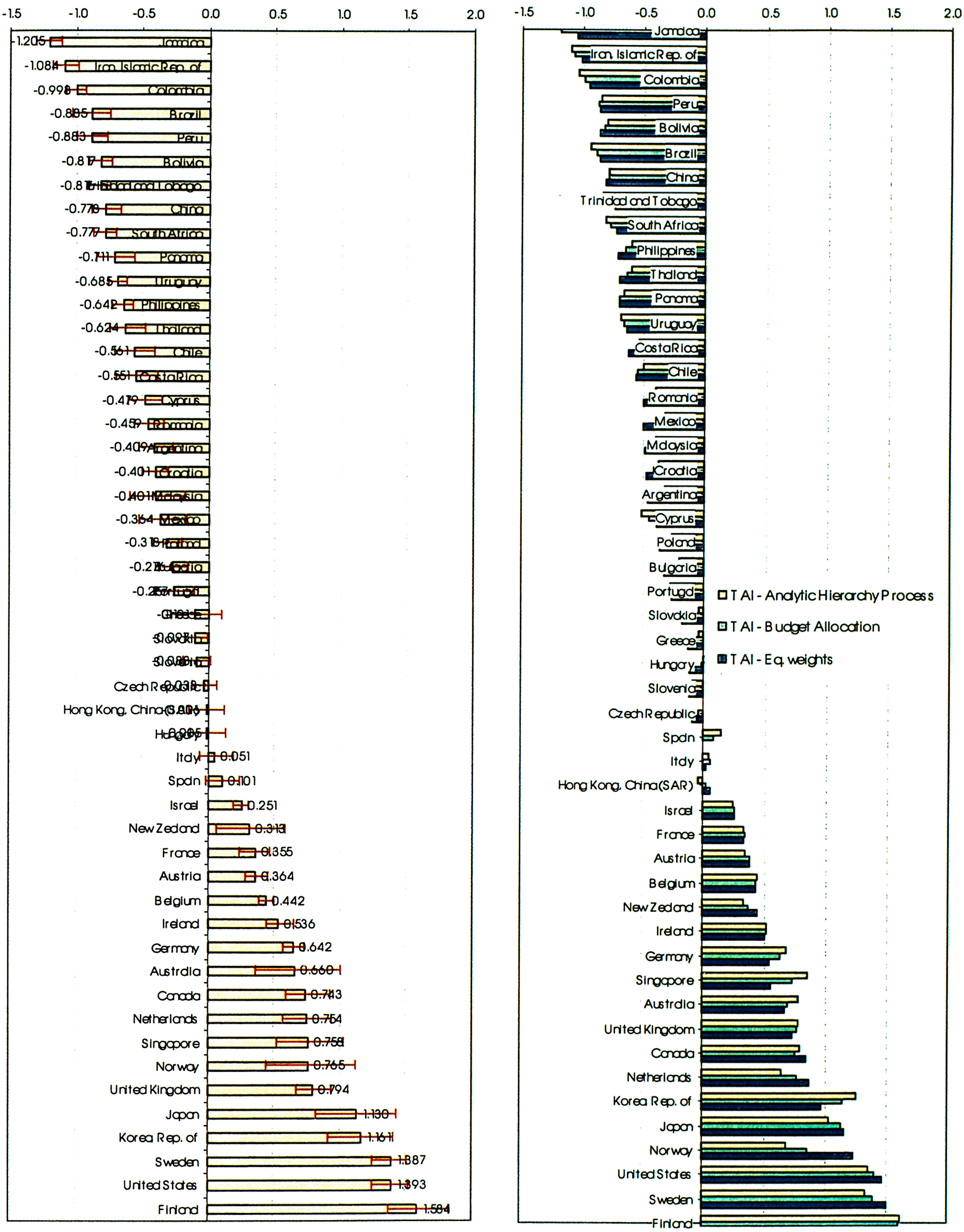

Figure 22. UA showing median TAI values for 50 countries and respective uncertainty bounds (left graph). TAI values for 50 countries calculated using weights based on analytic hierarchy process, budget allocation and simple mean of the eight sub-indicators (right graph).

7.3 Sensitivity of the composite indicator to the weights

Once the system of sub-indicators is determined and used to obtain the composite indicator it is important to analyze how much the composite indicator values are influenced by uncertainty in the source data and/or uncertainty in the weights (due to the stakeholders' plurality of perspectives).

The uncertainty analysis results show the composite indicator values per country in the form of histograms. **Figure 23** illustrates the histograms of the TAI values for Finland, the United States and the United Kingdom. It can be noticed that significant overlapping between Finland and the United States occurs, therefore the relative position of those countries cannot be distinguished efficiently. In such cases, where partial overlapping between two countries occurs, the difference in the TAI values for that pair of countries can be further analyzed in a sensitivity framework.

Sensitivity analysis (SA) complements uncertainty analysis in that it attempts to apportion quantitatively the variations in the TAI values to different sources of variation (e.g. weights, sub-indicator values). At a first stage, it is interesting to identify which weights are mostly responsible for the overlapping of countries, assuming that the values of the sub-indicators are error free.

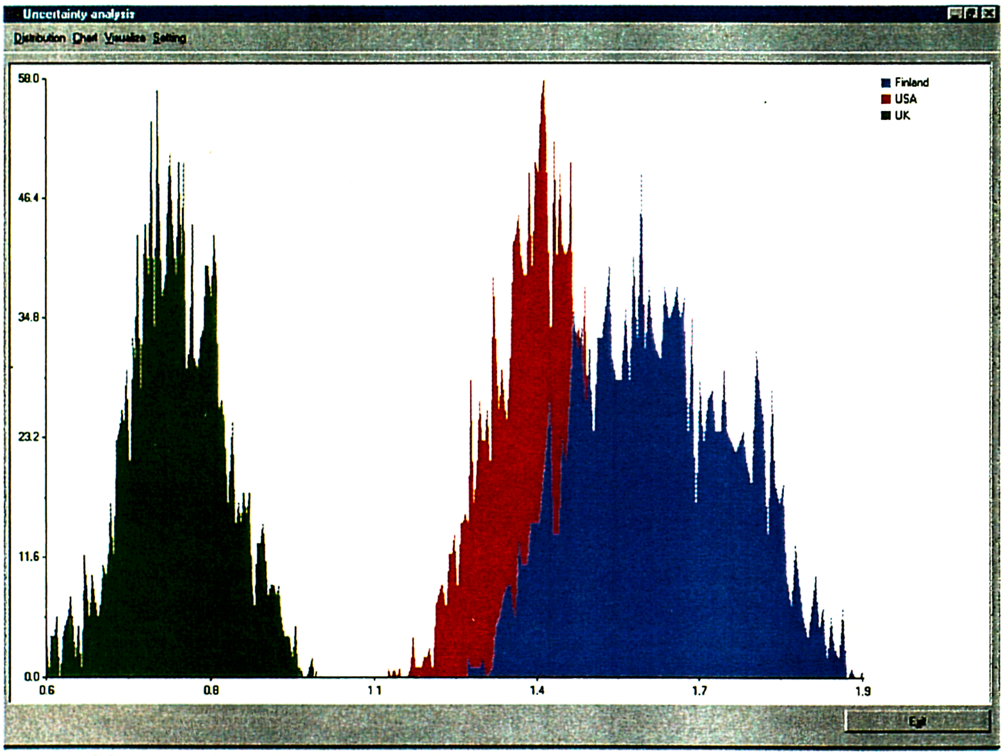

Figure 23. Empirical distribution of TAI values per country for three countries calculated using 2000 random combinations of weights for the sub-indicators.

The extended FAST sampling method, as implemented in the SIMLAB software [33], allows for both uncertainty and sensitivity analysis. Sensitivity indices are calculated regarding the contribution of each weight to the *difference in the TAI values between two countries*. The higher the value of the sensitivity index for a given weight, the more sensitive the output to the variation of that weight. **Figure 24** shows the histogram of the differences in the TAI values between the

United States and Sweden. These two countries are ranked 2nd and 3rd in terms of the median and vice versa, depending on the values of the weights used. It can be seen that this difference is half of the times positive (i.e. the United States performs better than Sweden) and half of the times negative (i.e. Sweden performs better than the United States). By decomposing the variance of this difference, it is shown (up right side of **Figure 24**) that out of the eight sub-indicators, the impact of the weight of the 2nd ("Royalties") and 3rd ("Internet hosts") sub-indicators on the difference is quite large (almost 50%).

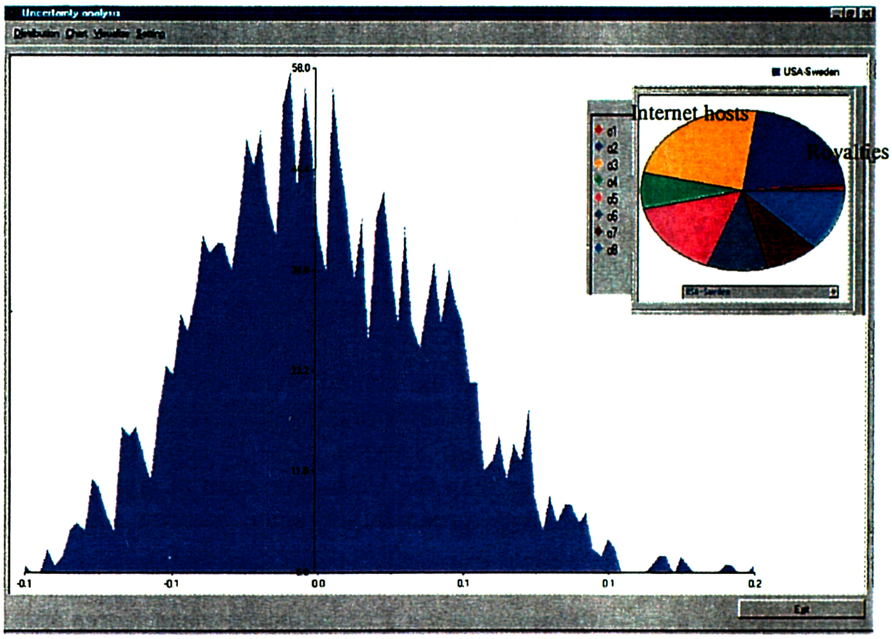

Figure 24. Variance decomposition of the difference in the TAI values between the United States and Sweden, due to the variation of the weights (a1, a2,...,a8) of the eight sub-indicators. Weights of the 2nd (Royalties) and 3rd indicator (Internet) have the highest impact on the difference.

The identification of the most important weighting factors is a useful information that sensitivity analysis can supply to start a convergence process among the experts (if possible). **Figure 25** plots the difference of the values of the composite indicator for the two countries in the space defined by the two most important weights (corresponding to the indicators "Royalties" and "Internet"). One can see what combination of weights can produce a positive (pink color), close to zero (blue) and negative (green) difference among the two TAI values for the two countries. Such a scatterplot is one step further towards improving the understanding of the composite indicator and increasing transparency in the decision and could be used as a tool to monitor the evolution of the discussion among the stakeholders.

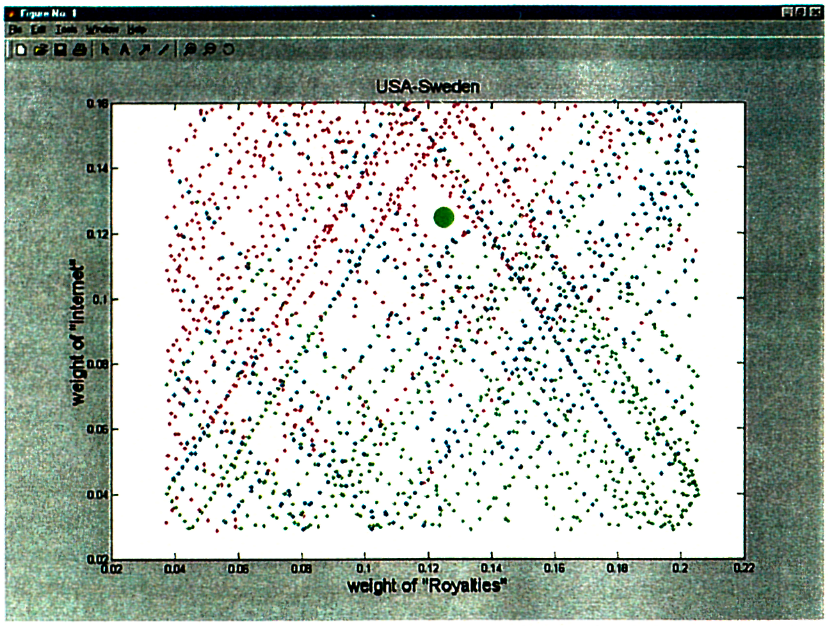

Figure 25. Differences in the TAI values between the United States and Sweden plotted in the space of the weights of Royalties and Internet sub-indicators. The differences are positive (pink color) indicating that the USA performs better than Sweden, close to zero (blue) and negative (green) indicating that Sweden performs better than the USA. The case of equal weights in the two sub-indicators (0.125) is marked with large green bullet, since Sweden seems to perform better than the United States.

Similar results are given below for the difference in the TAI values between Korea and Norway. Norway is ranked 7th using the median of the 2000 TAI values, while it is in the 4th position if the TAI is calculated using equally weighted sub-indicators. On the other hand, Korea is at the 4th position using the median of the 2000 TAI values. It seemed therefore, that a comparison of the TAI values between these two countries would give more insight into the situation.

Figure 26 shows the histogram of the TAI value for Korea minus the corresponding TAI value for Norway. More than 80% of the times (for the different sets of weights) Korea performs better than Norway, as verified by the positive sign of the difference. By decomposing the variance of this difference, it is shown that the impact of the weight of the 6th sub-indicator ("Electricity") on the difference is quite large, followed by the impact of the weight of the 3rd indicator ("Internet").

Figure 27 presents a plot of the difference of the TAI values for these countries in the space defined by the two most important weights corresponding to the "Electricity" and "Internet" sub-indicators. The case of equal weighting, which assigns a 0.125 weight to each sub-indicator (8 indicators in total), is located at the top left right side of the plot (large green bullet). In that area of weights, Norway only occasionally seems to have a higher TAI value than Korea, thus the sign of the difference is negative (green color). However, the analytic hierarchy process and budget allocation give lower values than 0.125 to the weights for Electricity and Internet, and therefore the TAI value for Korea confirms a better performance for this country over Norway.

As an overall remark, it can be stated that uncertainty and sensitivity analysis can be used as tools to monitor the evolution of the discussion among the stakeholders. UA & SA can provide useful information on the identification of the most important weighting factors, which could guide a convergence process among the experts focusing on the important weights.

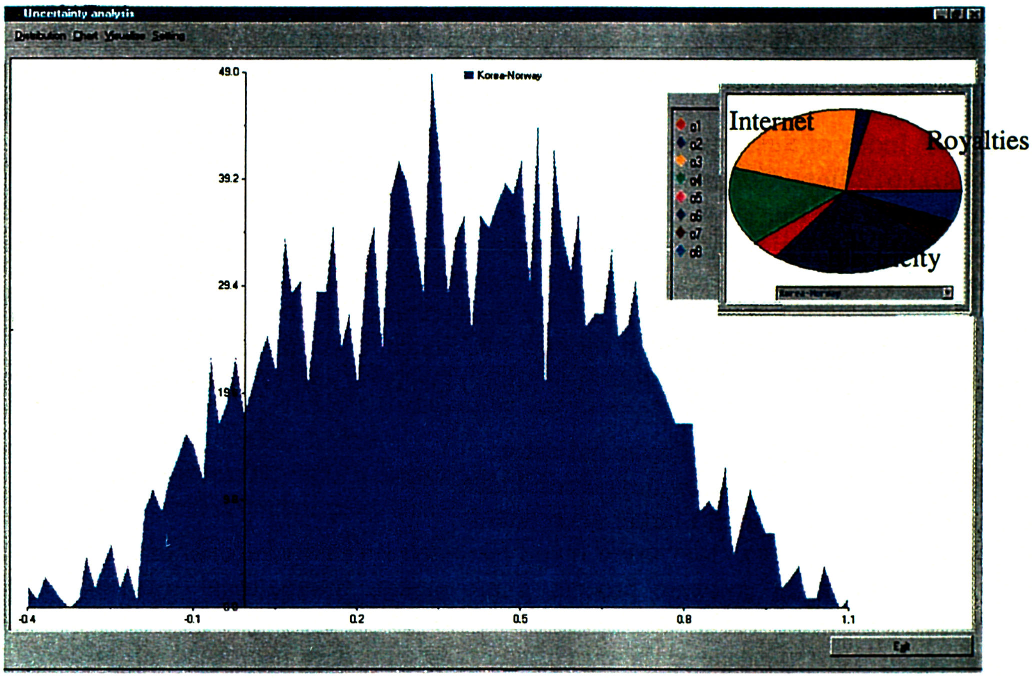

Figure 26. Variance decomposition of the difference in the TAI values between Korea and Norway, due to the variation of the weights (a1, a2,...,a8) of the eight sub-indicators. Weights of the 6th (Electricity), 3rd (Internet) and 2nd (Royalties) indicator have the highest impact on the difference.

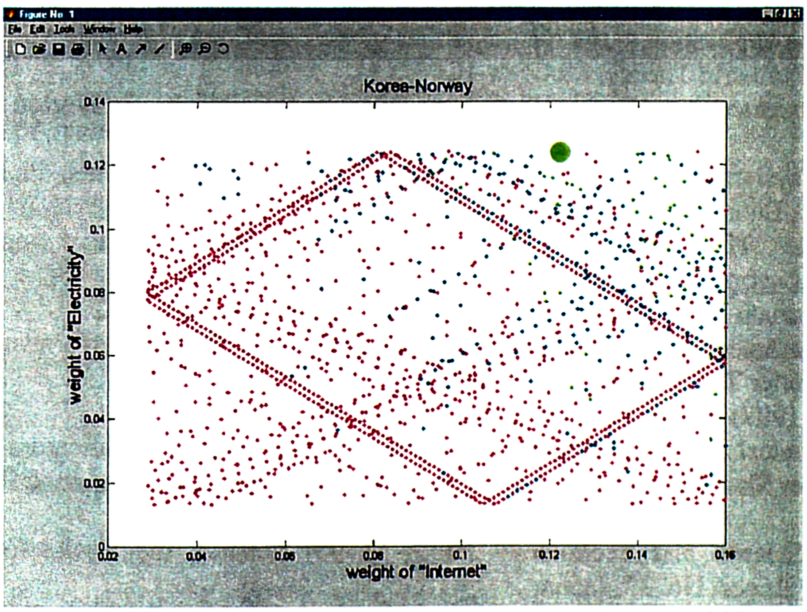

Figure 27. Differences in the TAI values between Korea and Norway plotted in the space of the weights of Internet and Electricity indicators. The differences are positive (pink color) indicating that Korea performs better than Norway, close to zero (blue) and negative (green) indicating that Norway performs better than Korea. The case of equal weights in the two sub-indicators (0.125) is marked with large green bullet, since Norway seems to perform better than Korea.

8 References

[1] Adriaanse A. (1993) Environmental policy performance indicators. A study on the development of indicators for environmental policy in the Netherlands. SDV Publishers, The Hague.

[2] CES (2000) Index of Sustainable and Economic Welfare.

[3] COM(2001) 619 final, Communication from the Commission, Structural indicators.

[4] Commission of the European Communities (1984) The regions of Europe: Second periodic report on the social and economic situation of the regions of the Community, together with a statement of the regional policy committee, OPOCE, Luxembourg.

[5] Cox D.; Fitzpatrick R.; Fletcher A.; Gore S.; Spiegelhalter D. and Jones D. (1992) Quality-of-life assessment: can we keep it simple? J.R. Statist. Soc. 155 (3): 353-393.

[6] Detlof von Winterfeld and Ward Edwards (1986) Decision analysis and Behavioral Research, Cambridge U. press.

[7] DG ECFIN (2000) Business Climate Indicator

[8] DG Enterprise (2001) Summary Innovation Index

[9] DG MARKT (2001) Internal Market Scoreboard.

[10] Emam K.; Goldenson D.; McCurley J. and Herbsleb J. (1998) Success or failure? Modeling the likelihood of software process improvement. International software engineering research network. Technical Report ISERN-98-15.

[11] Fagerberg J. (2001) Europe at the crossroads: The challenge from innovation-based growth in the Globalising Learning Economy, B. Lundvall and D. Archibugi eds., Oxford Press.

[12] Forman, E.H. (1983) The Analytic Hierarchy Process as a decision support system, Proceedings of the IEEE Computer Society. .

[13] Gro brundtland, Environmental Challenges of the 1990s, The Tanner lecture on Human Values, Clare Hall, Cambridge, UK,1991, p.8

[14] Isla M. (1997) A review of the urban indicators experience and a proposal to overcome current situation. The application to the municipalities of the Barcelona province.

[15] J. Nunnally (1978) Psychometric Theory, McGraw-Hill.

[16] Jones, Chad (1998). *Introduction to Economic Growth.* New York (NY): W.W. Norton & Company.

[17] Karlsson J. (1998) A systematic approach for prioritizing software requirements, Ph.D. dissertation No. 526, Linkoping, Sverige.

[18] King's Fund (2001) The sick list 2000, the NHS from best to worst. http://www.fulcrumtv.com/sick%20list.htm

[19] Manly B. (1994) Multivariate Statistical Methods. A Primer, Chapman & Hall, UK.

[20] McCarthy M. (2001) Comments on "Benchmarking European Labour Market Performance with Efficiency Frontier Techniques", DG ECFIN.

[21] Moldan B.; Billharz S. (1997) Sustainability indicators. Scope 58, Wiley Pub.

[22] Muldur U. (2001) Technical Annex on Structural Indicators. Two composite indicators to assess the progress of Member States in their transition towards a knowledge-based economy. DG RTD.

[23] National Institute of Science and Technology Policy (1995) Science and Technology Indicators, NISTEP Report No. 37, Japan.

[24] Nilsson R. (2000) "Calculation of Composite leading Indicators: A comparison of two different methods", Paper presented at the CIRET conference, Paris, October 2000.

[25] Ogwang T. and Abdou A. (2000) The choice of principal component analysis for computing some measures of human well-being. 26th General conference of the international association for research in income and wealth, Cracow, Poland.

[26] Parker J. (1991) Environmental reporting and environmental indices. Ph.D Thesis, Cambridge, United Kingdom.
[27] Porter M. and Stern S. (1999) The new challenge to America's prosperity: Findings from the Innovation Index, Council on Competitiveness, Washington D.C.
[28] Pré Consultants (2000) The Eco-indicator 99. A damage oriented method for life cycle impact assessment.
[29] Puolamaa M.; Kaplas M.; Reinikainen T. (1996) Index of Environmental Friendliness. A methodological study, Eurostat.
[30] Saaty R.W. (1987) The analytic hierarchy process- what it is and how it is used. Mathematical Modelling, 9, 161-176.
[31] Saaty, T. L. (1980) *The Analytic Hierarchy Process*, McGraw-Hill Book Co., N.Y.
[32] Saltelli A., Tarantola S., and Chan K. (1999) A quantitative, model independent method for global sensitivity analysis of model output, *Technometrics* **41(1),** 39-56.
[33] SIMLAB software 1.1 (1999) Simulation environment for uncertainty and sensitivity analysis, POLIS ScaRL, Milan.
[34] Storrie D. and Bjurek H. (2000) Benchmarking European labour market performance with efficiency frontier techniques. Discussion paper FS I 00-211.
[35] Storrie D., and Bjurek H., (1999) "Benchmarking the Basic Performance Indicators Using Efficiency Frontier Techniques", Report Presented to the European Commission, Employment and Social Affairs DG.
[36] United Nations (1990, 2001) Human Development Report. http://www.undp.org
[37] UN Development Programme (1992) Human Development Report, Oxford University Press, UK, 1992, p.21
[38] Wall R., Ostertag, K. and Block, N. (1995) Synopsis of selected indicator systems for sustainable development. Report for the research project, 'Further development of indicator systems for reporting on the environment' of the Federal Ministry of the Environment. Fraünhofer Institute for Systems and Innovation Research, Karlsruhe.
[39] WHO (2000) Overall Health System attainment.
[40] World Economic Forum (2001) Environmental Sustainability Index http://www.ciesin_org/indicators/ESI/index.html
[41] Arundel A. and Bordoy C. (2002) Methodological evaluation of DG Research's composite indicators for the knowledge based economy. Document presented by DG RTD at the Inter-service consultation meeting on Structural Indicators on July 11th 2002.

ANNEX. Summary Table of the Composite Indicators Review

	Name	Number of sub-indicators	Preliminary treatment of sub-indicators	Aggregation methodology	Comments
1	"Summary Innovation Index" (by DG ENTR)	17	Mean subtraction	Number of indicators that are more than 20% above the European average minus the number of indicators which are more than 20% below and division by the total number of available indicators for each country.	• The index varies between +10 (all indicators are above average) to -10 (all indicators are below average). • The figure 20% was chosen considering the accuracy of the data. A sensitivity analysis showed a high correlation ($R^2 = 0.98$) between the index using a 20% boundary and those for a 15% and 25% boundary. • Given that a generally applicable model describing how each indicator influences innovation is not available, all indicators are given equal importance by the authors. • A different calculation approach for a summary index was tested based on the average percentage by which each indicator varied from the overall EU average. This indicator was strongly correlated with the retained SII ($R^2 = 0.89$). The retained SII was finally preferred over the percentage index because it ignored minor differences from the EU average, which may not be meaningful.
2	"Internal Market Index" (by DG MARKT)	19	Percentage annual differences.	PCA used to define weights for the sub-indicators. The 19 Principal Components are further weighted by the % variation of the total information explained by each Principal Component.	• The weights were derived from PCA because the authors considered that this method deals appropriately with correlation between variables. However, the inter-correlations between the sub-indicators are very low (there is no r>0.7). This implies that a simpler weighting system might have been more efficient (same case as composite indicator developed by Isla). • The number of Principal Components is equal to the number of the sub-indicators (i.e. 19), which implies that the total information of the sub-indicators could not be summarised in the first place by PCA.

3	"Business climate indicator" (by DG ECFIN)	5	Scaling in [-100, 100]	PCA applied to define weights. One principal component adopted as the composite indicator.	• The PCA requirements of stationarity of the series are confirmed by the empirical autocorrelograms (significant up to the 8th order) and the augmented Dickey-Fuller test. • PCA indicated that a single factor is sufficient to explain the bulk of the common information (92% of the variance). This gave a statistical justification to the authors for the choice of summarizing a priori the information by means of a single composite indicator. In this case the phenomenon that the composite indicator aims to measure has one statistical dimension.
4	"Investment in the knowledge based economy" (by DG RTD)	7	Normalised indicators calculated by mean subtraction and division by the standard deviation	Choice of weights is up to the user. Composite indicator is in progress.	• An instrument in Excel leaves the choice of the weights up to the user (Muldur, 2001). • High correlation between pairs of sub-indicators is identified, which suggests that PCA could be used to identify the statistical dimensions of the data. • Two principal components are identified by PCA, one consisting of four indicators and one of two. One indicator is not loaded on any principal component, thus, on a purely statistical basis, this indicator could be omitted. However, this indicator is of political interest, therefore it is worth to keep it. • JRC carried out uncertainty analysis to identify how much the variation of the weights can affect the values of the composite indicator for the various countries. A further sensitivity analysis helped to identify which weights are influential to the differences of close ranked countries.
5	"Performance in the knowledge based economy" (by DG RTD)	7	Normalised indicators calculated by mean subtraction and division by the standard deviation	Choice of weights is up to the user. Composite indicator is in progress.	• The correlation between the sub-indicators is not high in general. PCA reveals that there are indeed 5 main principal components.
6	"Relative intensity of regional problems in the Community" (by the European Commission)	3	Normalised indicators calculated by mean subtraction and division by the standard deviation	Empirical weights are determined considering the degree of correlation between two sub-indicators	• A thorough examination of the interrelationships of the sub-indicators was undertaken by using Factor Analysis.

7.	"Economic Sentiment Indicator" (by the European Commission)	4	1. Trend estimation by using month-to-month changes. 2. No smoothing. 3. Normalisation by dividing the month-to –month changes with the average month to month change.	Sub-indicators are divided in two groups with equal weights within group. The second group is given ½ the weight of the first group.	• The performance of the Index is evaluated against total industrial production as a proxy for economic activity. • The authors consider that applying principal components analysis to choose the weights would minimise the contribution of the indicators which do not move with the other indicators. This may reduce the reliability of the composite indicator because some indicators perform better in one cycle and others in a different cycle.
8.	"Composite Leading Indicators" (by OECD)	Number varies across Member States (e.g. France:11, Germany and Italy: 6, UK: 9)	1. Trend estimation via Phase average trend method 2. Smoothing via "Months for cyclical dominance moving average". 3. Normalisation via mean subtraction and division by the mean of the absolute differences from the mean.	Arithmetic average of the normalised indicators	• The performance of the Index is evaluated against total industrial production as a proxy for economic activity. • Nillson [2000] suggests that PCA could help to select the weights. However, the authors argue that such a method would minimise the contribution of indicators, which do not move with the other indicators. This may reduce the reliability of the composite indicator because some indicators perform better in one cycle and others in a different cycle. Therefore, most such indicator systems in operation use an equal weighting system.
9.	"Information and communication technologies" (by J. Fagerberg)	5	Country rankings for each indicator	Sum of rankings	• The authors preferred the simplicity by using this methodology, however the cardinal distances in the values of the indicators are not considered this way. This method can therefore "hide" how close two countries might be.

10.	"Environmental Sustainability Index" (by the World Economic Forum)	22	1. Division by population or income or populated land area. 2. Missing data imputation via bivariate correlations. 3. Logarithms for highly skewed variables. 4. Truncating distributions to 95% range (account for inaccuracy of data). 5. Normalisation by mean subtraction and division by the standard deviation. The sign was changed for those indicators where high observed values corresponded to low levels of sustainability.	Arithmetic average of the normalised indicators	• PCA was initially applied, but the set of principal components did not discriminate efficiently among the observations and more problematically, it assigned negative weights to many variables. Thus, aggregation via equal weighting was preferred. • Alternatively, weights were defined following a survey process. A simple sensitivity analysis suggested that the weighting methodology would not have changed the ranking in any appreciable fashion. In particular, the Index score was calculated using the survey-generated weights. The average shift in rank was only 1.7 places out of 122. • The Index is highly correlated with the 1990-1998 GDP per capita growth (r =0.76), the Human Development Index (r = 0.67) and the WEF Current Competitiveness Index (r = 0.65).
11.	"Human Development Index" (by the United Nations)	3	Scaling in [0, 1]	Arithmetic average of the scaled indicators	• Ogwang and Abdou (2000) applied PCA, and they argue that there is a statistical justification for selecting only one of the three components of the index, i.e life expectancy, without loss of too much information.
12.	"Technology Achievement Index" (by the United Nations).	8 (grouped in 4 sub-indices)	Scaling in [0, 1]	Arithmetic average of the 4 sub-indices	• No correlation analysis is presented for the sub-indicators or the sub-indices.
13.	"Overall Health System Attainment" (by the World Health Organization)	5	Scaling in [0, 100]	Weights based on survey of preferences of informed individuals	• A notable feature is that all the main results are reported with uncertainty intervals to communicate the plausible range of estimates for each country on each measure. The WHO believes that it is the first time that an international agency has extended this analytical practice to its official publications. The rank changes induced by variation of weights within plausible limits are found to be much less important than those implied by measurement errors in the data.

14.	Two "Synthetic environmental indices" (by Isla M.)	22	-	Arithmetic average of the indicators	• FA revealed a very low correlation among indicators, thus it canceled any attempt to use this approach to obtain synthesis results. • Experts' participation to provide weights for each sub-indicator was dismissed because the procedure involved a lot of judgment when the diversity of the areas could not guarantee better performance.
15.	"National innovation capacity" (by Porter and Stern)	8	The logarithmic values of the sub-indicators are considered	Weights determined by a multiple regression model	• Multiple regression model, where sub-indicators are used as input and patent indicator as the output. The selection of an output variable helps in identifying statistically the weights for the input indicators. This model is used for forecasting purposes.
16.	"General Indicator of Science&Technology" (by NISTEP, Japan)	13	Judgmental classification of five sub-indicators as "input" and eight as "output"	PCA was applied to define weights. The main principal component is adopted as the composite indicator.	• FA is used to analyze the structure of the indicators sets. This analysis did not cluster the indicators of the input set together and neither those of the output set. However, it was emphasized that this classification was subjective in the first place.
17.	"Success of software process improvement" (by Emam et al. 1998)	14	Initial indicators are expressed on subjective scales and then transformed on quantitative scales. No normalisation is performed.	PCA applied to identify the main components. The overall index is calculated as the sum of the 5 main principal components.	• For each principal component (PC) the cronbach alpha coefficients are calculated (0.62 to 0.732), which verify that the five PCs represent each a single construct. • A second approach for weighting the variables is undertaken by means of a classification and regression tree (CART) algorithm.
18.	"European Labour Market Performance" (by Storrie and Bjurek, 1999)	3	Scaling between 0 (best performance) and 100 (worst performance)	Efficiency frontier (objective method)	• The method is proposed by the authors as extremely parsimonious with regard to the weighting assumptions, because it lets the data decide on the weighting issue. • M. McCarthy [] expresses however concern that such an empirical construct might not indicate the appropriate direction of a policy for a given country in order to improve its situation. She adds further that, since employment guidelines exist, one cannot use such an approach to evaluate overall labour market performance, but instead should attach different weights to different indicators.

19.	"Eco-indicator 99" (by Pre Consultants, Netherlands)	3	Division by a reference value for each indicator	Weighting scheme is selected by a panel of experts	• The authors believe that the weights should represent the views of society or a group of stakeholders. • Weighting based on target values was not selected because it is often hard to interpret the basic values that are underlying the decisions of society. For instance policy targets set by governments are often a compromise between the need to reduce loads and the preparedness to make the necessary sacrifices. • Significant differences in the weighting sets due to cultural perspectives were found.
20.	"Concern about environmental problems" (by Parker)	11	Indicators are normalized by dividing the value in each year by the value for the year for which each indicator is first available.	Weights derived from public opinion polls	• Sensitivity analysis indicated similar behavior of the composite indicators produced by the different weighting schemes (based on public opinion polls over different years and different countries), which indicates that public opinion about the main threats to the environment is remarkably stable across both space and time. • To improve the comparability of the composite indicator values across countries, the author considered whether division of the sub-indicators' by country's population, area or GNP (Gross National Product) would be useful. He found that dividing for example the nitrogen oxide indicator by population, would make France seem much worse than Italy, whereas dividing it by surface area or GNP would improve the position of France compared to Italy. This conclusion, in a general context, is very important and should be considered in any aggregation approach.
21.	"National Health Care systems performance" (by King's Fund, England)	6	No standardisation	'Budget allocation' survey of 1,000 people across the UK defined the weights for the indicators.	• Respondents were asked to distribute a fixed sum of 60 'chips' to some or all of the six performance indicators – more chips spent in an area, the greater the improvement on that indicator. It was found that there was little difference in the distribution of the chips depending on respondents' sex, social class, age or the area of the country where they lived.

22.	"Index of sustainable and economic welfare" (by CES and NEF)	20	Sub-indicators are expressed in monetary terms. Negative sign assigned to the indicators that are associated with a reduction of the welfare.	Arithmetic average of the indicators (7 indicators are +, 13 indicators are -)	• The authors state that the index has been criticized as subjective and too susceptible to changes in the assumptions that underpin it. Therefore, a relevant website has been created to allow the user to change the weightings and assumptions used in the index. According to the authors, this site shows that the index is a good basis on which to construct alternative indicators.
23.	"Index of Environmental Friendliness" (by Puolamaa, Kaplas, Reinikainen, Eurostat)	11	1. Aggregation into problem indices by equivalency factors. 2. Normalisation of problem indices by dividing the sectoral problem index by the value of the national problem index.	Subjective weights for the normalised problem indices are determined from experts by means of the Analytic Hierarchy Process.	• The main advantage of the Analytic Hierarchy Process (AHP) is that it is based on pair-wise comparison; the human mind can easily handle two distinct problems and examine their differences. Another advantage of AHP is that unlike many other methods based on Utility Theory, its use for purposes of comparisons does not require a universal scale. • High inconsistency values in some respondents could have been avoided by interaction in the research design. • Although an uncertainty analysis assessment was undertaken, the authors pointed out that uncertainties are very seldom documented or assessed in the existing databases.
24.	"Environmental Policy Performance Indicator" (by Adriaanse, the Netherlands)	6 theme indicators (composed of several simple indicators)	Scaling of the theme indicators: division by the corresponding a) sustainability levels, and b) policy targets	Sum of the six theme indicators scaled by the: a) sustainability levels, b) policy targets	• The alternative use of sustainability levels and target values for the scaling of the theme indicators results in composite indicators showing similar trend.

Mission of the JRC

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

EUROPEAN COMMISSION
JOINT RESEARCH CENTRE

EUNA20408ENC