Workpackage 5

State-of-the-Art Report on Composite Indicators for the Knowledge-based Economy

Deliverable 5.1

2005

List of contributors:

Michaela Saisana, Stefano Tarantola, Niels Schulze, Joint Research Centre, Ispra
Laurens Cherchye, Wim Moesen, Tom Van Puyenbroeck, KU Leuven.

Main responsibility:

Michaela Saisana, Joint Research Centre, Ispra

CIS8–CT–2004–502529 KEI

The project is supported by European Commission by funding from the Sixth Framework Programme for Research.

http://europa.eu.int/comm/research/index_en.cfm

http://europa.eu.int/comm/research/fp6/ssp/kei_en.htm

http://www.cordis.lu/citizens/kick_off3.htm

http://kei.publicstatistics.net/

Preface

This report is the first deliverable of the work-package 5 (WP 5, composite indicators) of the KEI-project (Knowledge Economy Indicators: Development of Innovative and Reliable Indicator Systems). KEI (`http://kei.publicstatistics.net`) is part of the *Policy Orientated Research* section of the specific programme *Integrating and Strengthening the European Research Area* in the context of the Sixth Framework Programme of the European Commission.

The first part of the report offers an update of the 2002 State-of-the-art report (Saisana and Tarantola, 2002), focusing on various aspects of composite indicator construction. Composite indicators (CI) construction is today a rapidly evolving field in terms of theory and practical applications. Therefore, adding some ideas can be useful for practitioners. The lack of a common accepted methodology for CI building is essentially due to their breadth of application and to the latitude of opinions formulated by experts from different disciplines. Taking inspiration from a series of workshops held in cooperation with the OECD, we try to identify best practices and to introduce some useful recommendations.

The last part of the report provides a succinct overview of existing composite indicators that have been proposed to summarize phenomena related to the knowledge based economy (e.g. research, innovation,...)

Contents

List of Figures

List of Tables

Chapter 1

Objectives and Contents

Since JRC's 'State-of-the-Art Report' on composite indicators appeared in 2002, the interest in the subject has gained further momentum, as can be gauged by the expanded list of references of the present revision. Furthermore two Workshops on Composite Indicators of Country Performance have been held, the first hosted by JRC in May 2003, the second by the OECD in February 2004, and a web-based information centre exclusively devoted to the topic has been developed. Composite indicators still do stir controversy. Yet one cannot escape today's reality that they *are* used, and that the country rankings which they typically generate often flair public debate. Hence, whether one is highly sympathetic or staunchly opposed to composite indicators, one has to live with them, while practitioners probe deeper into their meaning and seek to identify best practices..

The purpose of Saisana and Tarantola (2002) was to examine a number of methodologies with a view to clarifying how they relate to the development of composite indicators, providing investigations of several methods such as

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

Next to that, the report examined twenty-four published studies on this topic in fields such as environment, economy, research, technology and health, including practices from the Directorates General of the European Commission. For each composite indicator, the authors also reviewed general information on the number and type of sub-indicators, on the preliminary treatment (normalisation, de-trending etc.) and on the presentational tools considered.

A main aim of the current paper is to offer an update of Saisana and Tarantola (2002), while expanding the chapter of composite indicators of the knowledge-based economy. Many of the items in the above list will thus reappear, with comments and references added whenever appropriate. Since composite indicator construction *is* today a rapidly evolving field in terms of theory, as well as in terms of practical applications, adding some of the newer ideas is surely warranted. But precisely because the subject is currently in a state of flux, the reader is warned that the current *status quaestionis* does not even seek to provide a 'conclusive' in-depth treatment of many methodological aspects of composite indicator construction.[1]

In fact, looking at some critical and interesting surveys that appeared since Saisana and Tarantola (2002) (e.g. Booysen, 2002; Freudenberg, 2003; Salzman, 2003), there is today not such a thing as a commonly accepted methodology. In our view, this is likely to remain the case the near future. This is essentially due to the *intrinsic* 'vagueness' or ambiguity of composite indicators. Inspired by participants' discussions during the two workshops already mentioned, we will attempt to introduce a somewhat broader, interpretative, unifying perspective on the concept of composite indicators.

The following sections of this report are organized as follows: ***Section 2*** will directly address the 'contentious nature' of composite indicators. ***Section 3*** describes briefly the various stages for the construction of a composite indicator. ***Section 4*** focuses on the normalisation problem and, in particular, on potential problems associated with the hitherto 'standard' aggregation procedure which is to take a weighted sum of normalised sub-indicators. ***Section 5*** surveys the most common procedures to provide weights in the aggregation stage. ***Section 6*** provides a succinct overview of methods commonly used to present results from (inter-country) comparisons based on a composite indicator. Finally, ***section 7*** presents an overview of existing applications (related) to the knowledge-based economy.

[1] A thorough analysis of composite indicators will be the subject of the 'Handbook on Composite Indicators', a joint JRC-OECD project that is currently under development.

Chapter 2

The pros and cons of Composite Indicators

Indicators are often a compromise between scientific accuracy and the information available at a reasonable cost.

Clearly, even single indicators need to respect some principles in order not to compromise the purposes for which they are collected. A recent list of such principles is provided in Atkinson *et al.* (2002a, p. 21 ff.). We recall some of these desiderata:

- an indicator should identify the essence of the problem and have a clear and accepted normative interpretation
- an indicator should be robust and statistically validated
- it should be responsive to effective policy but not subject to manipulation

But still, in view of the reasonable cost-argument:

- its measurement should not place too high a burden on statistical services, enterprises or citizens.

Also, since they are as a rule used for comparisons between different entities (e.g. countries):

- it should be measurable in a sufficiently comparable way, and comparable as far as practicable with internationally applied standards

Giovannini (2004) presents quality frameworks developed and implemented by several international organisations. As an example, Eurostat adopted a "Quality declaration of the European statistical system" based on seven dimensions: *relevance* (are the data what the user expects?), *accuracy* (is the figure reliable?), *comparability* (are the data in all necessary respects comparable across countries?), *coherence*(are the data coherent with other data?), *timeliness and punctuality* (does the user get the data on time and according to pre-established dates?), and *accessibility and clarity* (is the figure accessible and understandable?). The following table lists the dimensions used by OECD, IMF and Eurostat:

Table 2.1: Relevant dimensions for composite indicators (Giovannini, 2004)

OECD	IMF/Eurostat
Relevance	Relevance/Serviceability
Accuracy	Accuracy/reliability/meth. soundness
Credibility	Integrity
Timeliness	Timeliness
Accessibility	Accessibility
Interpretability	Clarity
Coherence	Consistency/Comparability

The past century has seen massive efforts in statistical capacity building, broadly defined, which indisputably enlarged the informational basis of policy decisions. The qualitative issues displayed in the previous box have surely benefited from these efforts. However, in quantitative terms, there are some signs that this basis has come close to the point of information overload. At least, one may infer this from the fact that many organisations such as the European Commission, the OECD, and others, have recently either singled out some "key indicators" or have constructed "composite indicators" in which several single indicators are aggregated into a single index. While both answers can be viewed as antithetic to one another, they are both possible answers to the information overload problem.

If one is interested in assessing a particular complex phenomenon, there is indeed a danger when many single indicators are listed next to each other and subsequently used for an 'overall' comparison between countries. Micklewright (1991) made the point that, lacking a *good* synthetic index, there is always the danger that eventually none of the indicators gets sufficient attention, or even that excessive public attention is again focused on just one dimension, thus abolishing the original basic desideratum of respecting the multidimensional nature of the problem under consideration. Especially when such a list is used to compare a set of countries, one may add to Micklewright's list the potential loss of credibility associated with a plethora of single indicators: every stakeholder understands that the larger the list, the easier it is for any country to be 'excellent' (in comparative terms) in one or two of the many sub-indicators.

Two important issues in the construction of composite indicators concern the role of possible *correlation among input variables* and *compensability*. The first point addresses the question if and how the influence of correlated variables should be adjusted properly (JRC-OECD, 2005). The second term refers to the implicit ratio of substitution inherent in the weighting of different sub-indicators. This kind of trade off might not be seen as adequate and can be circumvented by the application of a multi criteria approach (Munda and Nardo, 2003).

Another problem is the fact that it is as a rule impossible to *completely order* all observations if each of these observations in fact stands for an array of distinct numbers. In such settings, it may be uncontroversial to state that a country is only demonstrably outperformed if it is dominated in *all* dimensions of the indicator-set by at least one other country.[1] But it is clear that the power of this dominance criterion may well quickly decrease as more indicator-dimensions are considered. Partial (hence incomplete) orders are the best one could hope for. Note however that incompleteness is not solely the result of adding more indicators. The example below, taken from Cherchye *et al.* (2004) shows that just two indicators can already be sufficient to yield serious comparison problems. Indeed, there is no partial ordering between *any* pair of countries in the table.

Table 2.2: A hypothetical example

Country	I	II	III	IV	V	VI	VII	VIII	IX	X
Poverty rate	10	9	7	12	18	19.9	20	6.9	5	4.9
Long term Unemployment rate	6	9	10	5.9	4	2.1	2	14	15	19

Composite indicators are intended to cope with situations such as the one in the table (and, indeed, usually with more complex ones). By an 'index' or a 'composite indicator', we mean a mathematical combination (or aggregation as it is termed) of a set of single indicators. Intuitively, if one can reduce the many dimensions of each observation into one number, then this may allow overcoming the overall comparison problem.

Strictly speaking, composite indicators are no new measures. Well-known gauges such as GDP, the CPI, the Gini coefficient, and so on, also merge information about different markets or individuals into a single number. Each of these examples is even firmly entrenched as a policy instrument, despite the fact that they continue to be criticized as adequate measures of the underlying phenomenon they purport to quantify. To mention but a few familiar criticisms: GDP is not an adequate indicator of a country's economic activity (let alone of its citizens well-being) if only because it neglects the underground economy per definition, the CPI is at best (i.e. if one assumes homothetic utility functions) just a lower bound to changes in the true cost-of-living, the Gini coefficient is rooted in a rather distinct welfarist framework when comparing individual incomes, etc.

Nonetheless, it is safe to say that such traditional aggregates are presently far less controversial than their more recent siblings. This is also evident in the definition of composite indicators as it can be found in many places (e.g. in Saisana and Tarantola (2002), or on the composite indicator website `http://farmweb.jrc.cec.eu.int/ci/`)

> *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****.*

[1]Sen labelled this an 'intersection quasi-ordering', see Foster and Sen (1997).

A list of pros and cons on composite indicators was reported in the same Note on Composite Indicators (see also Saisana, Saltelli and Tarantola, 2005). The following points are, in our opinion, still 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.

Again, it is immediately clear that exactly the same list is in large measure appropriate for traditional composite indicators as well. But there is undeniably a particular *epistemological* sense in which the "new" composite indicators differ from their nowadays less contested precursors. The newcomers tend to lack the necessary degree of scientific consensus about an appropriate theoretical model that should, in principle, describe how the sub-indicators contribute to the underlying 'composite' phenomenon.[2] Indeed, one

[2]Thus, whereas the traditional indices are not totally free of criticism either, the assessment of their qualities as composite indicators notably goes back to an analysis of their underlying *well-defined* theo-

often observes that an agreement emerges about the choice of "key" single indicators that should play a role in constructing the index. But while there may (or may not) be a broad consent that all these single indicators can be 'associated with' the comprehensive phenomenon at hand, the hard question remains how and by how much. In fact, an alternative characterisation of a composite indicator may precisely reflect the lack of such consensus.

When stated in such terms, a good deal of the critical remarks raised towards composite indicators immediately become obvious. Indeed, as Booysen (2002, p. 131) rightly noted, today "not one single element of the methodology of composite indexing is above criticism". Or, as in the laconic warning of Freudenberg (2003, p. 29): "Composite indicators risk becoming exercises in measurement without a theoretical underpinning." We agree with Freudenberg, at least to the extent that we too are wary about the possibility that 'exact numbers' (and 'precise country rankings') may be all too easily fabricated to pop up as media headlines. Yet, we are also quite receptive to the thought that 'number-fetishism' in itself cannot lead to the conclusion that composite indicators are inexorably void, neither as regards their content, nor, *a fortiori*, as a basis for relative comparisons. Particularly when partial orderings of several indicators leave its users with (hardly) any clue, they may well be needed, at least if one agrees that the overall phenomenon one alleges to portray is not void itself.

Composite indicator construction is an exercise which is entirely and *inescapably* permeated with uncertainties. If anything, establishing a good practice for their construction must hence recognize such uncertainties as being part and parcel of the exercise. As noted by Foster and Sen (1997, p. 121), „if a concept has some basic ambiguity (. . .), then a *precise* representation of that ambiguous concept must *preserve* that ambiguity, rather than try to remove it through some arbitrary complete ordering." (italics in original).

Very similar conclusions have been reached by Science in the field of mathematical modelling of systems, in the sense that assessing the uncertainty in model based inference is considered as a prerequisite for the inference itself (Rosen, 1991, Saltelli *et al.*, 2004).

Looking back at the list of pros and cons, one may discern a second reason which renders 'specialists' *a priori* suspect as far as their instinctive attitude towards composite indicators is concerned. From their perspective, it is quite obvious that a composite indicator is by no means a substitute for detailed policy analysis. Being a summary, *any* index is likely to conceal information embedded in the disaggregate data it seeks to abstract.[3] But it is too easy to criticize a composite indicator merely for such reasons.

In fact, 'specialists' are usually not considered as the primary audience of composite indices. Osberg (2004) contends that, from an idealist perspective on public decision-making, "affecting public policy is the whole point of constructing an index", that "communicability is therefore key" and. . . "that the whole point of [. . .] an index is lost if it is only used by specialist researchers". In respect of our earlier remarks, it is also interesting to mention Osberg's concomitant observation that, from a more cynical point of view on politicians' incentives, "the fact that a 'report card' on many dissimilar incommensurable

retical framework. Similarly, one could say that the public acceptance of such indices is fostered by the broader (scientific) acceptance of the theories from which they are derived.

[3]In fact, they could be considered as 'second-order summaries', since they are a summary measure of several indicators, which themselves are often imperfect gauges.

indicators cannot be clearly aggregated to a summative evaluation is a positive advantage."[4] That is, policy makers may perhaps prefer to point at those indicators which are favourable to them, and simultaneously defer further action on the other, less favourable, dimensions of the report card.

Of course then, aggregation is a double-edged sword. 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". From a more general perspective, it is clear that *all* public-decision making has to cope somehow with the problem of trying to locate the optimal degree of informational (dis)aggregation. (Yet again, this also holds for the more traditional summary indicators as well as for indicators *tout court*. E.g. unemployment data at national level may be non-informative for designing labour policies when significant regional divides exist.). As we just made clear, however, that degree may be highly dependent on the specific aim for which the information is gathered.

Composite indicators stir scientific controversies, and some experts may even dislike the very idea of summarizing complex phenomena into one number. Recognizing this, CI scores should ideally be such as to preserve the lack of scientific consensus as well as the likely lack of consensus among stakeholders.

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 steps underlying its development are as sound and transparent as possible. In particular, science can help significantly in assuring that the processes of composite indicator construction are as transparent as to facilitate the debate among the legitimate institutional actors.

[4]For related comments from a public choice perspective, see Cherchye *et al.* (2004, p. 924).

Chapter 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. We list them with some succinct comments. We clarify at the outset that distinguishing several stages is above all a matter of representational convenience. In practice, there may well be many feedbacks between the stages, a feature to which we return later on.

- **Deciding on the phenomenon to be measured** – and whether it would benefit from the use of composite indicators.
- **Selection of sub-indicators** –There is no fully objective, formalised way of selecting relevant sub-indicators. Although the problem is not genuine to composite indicators, one should avoid basing the selection exclusively on the *availability* of data series. Clearly, data availability imposes a real pragmatic constraint on all information gathering, but one should address the question whether and which better sub-indicators are lacking (at the time of analysis).[1] Note that already at this stage a balance must be struck between simplification and grasping the full core of the measured phenomenon. The greatest threat to simplicity is the tendency to keep on adding variables and components (Booysen, 2002, p. 121). Also, one must be cautious and explicit about the value judgements associated with the choices.
- **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 (with a feedback to the previous stage) 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, complemented by an assessment of how the selection of the method can affect the final result.

[1] A good practice on the discussion about the choice of indicators that all partly contribute to the same phenomenon can be found in the field of the European Union's Social Inclusion Policy. See in particular the book by Atkinson, Cantillon, Marlier and Nolan (2002a), and the special issue of Atkinson, Cantillon, Marlier and Nolan (2002b)

- **Assessing the relationships between the sub-indicators** – Methods such as Principal Components Analysis or correlation analysis can provide insight into the relationships between the sub-indicators. It can be considered as a prerequisite for the preliminary analysis of the sub-indicators. Note that this stage may also provide a feedback into the sub-indicator selection exercise: the best summary index may not always be the one which is based on a 'comprehensive' but highly intercorrelated set of commonly agreed indicators. That is: when the aim is to provide a better *understanding* of the underlying causes and processes of a multi-faceted phenomenon, high correlations may indeed be helpful. Conversely though, this implies that there is little information lost in terms of *overall* explanatory power when one of two highly correlated indicators is neglected.

- **Normalising and weighting of the indicators** – Recalling the definition of composite indicators, this stage is crucial in the construction process. Several methods for normalising and weighting the sub-indicators are reported in the literature. One should note that we present this stage as one in which normalising *and* weighting (or more generally aggregating) the sub-indicators figure simultaneously. This is not a coincidence since both steps are as a rule not unrelated. We will return to this issue at length in the following section.

- **Testing for Robustness and Sensitivity** –Ample room should be given to an assessment of an indicator's robustness to changes in many of the foregoing steps. In particular, it is known that changes in the specific weighting system, a switch to different normalisation procedures, or the choice of sub-indicators will often affect the results the composite indicator shows. If only therefore, 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. In short: uncertainty and sensitivity analysis are indispensable in the construction of composite indicators. They will however not be systematically treated in this report (see, next to the state-of-the-art-report under workpackage 7 of the KEI-project, Saisana, Saltelli and Tarantola, 2005 and Saltelli, Tarantola, Campolongo and Ratto, 2004).

Chapter 4

Rescaling raw data and the generic problem of measurement

4.1 A brief reminder of standard approaches

Before computing a composite indicator, the sub-indicators that are measured in different units must normally be transformed into the same unit.[1] Table 4.1 gives the equations for six common different methods of calculating a composite indicator. These range from the simplest (Method 1) to the most complex (Method 7). Table 4.1 does not cover all possible methods of calculating a composite indicator. Several variations on each method exist. However, they were chosen since they are rather representative of the philosophy underlying the development of composite indicators as well as the most established in the literature. In particular, all methods described in the table have in common the fact that they use a weighted sum of normalised indicators. Stated differently, their mutual difference pertains to the normalisation procedures only. (In the following subsection, we will look at the question whether a 'weighted sum' is the most appropriate aggregation method).

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, Fagerberg, 2001). 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 (2001). 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

[1] This is the normal procedure, but it is not always necessary (see e.g. subsection 5.8). See also the next subsection, esp. footnote 9.

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'.

Table 4.1: 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.	$CI_c^t = \sum_{i=1}^{N} \cdot \mathrm{sgn}\left[\frac{x_{ic}^t}{x_{EUi}^t} - (1+p)\right]$
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 $y_{ic}^t = \frac{x_{ic}^t - \min(x_i^t)}{range(x_i^t)}$
7. Achievement-level sensitive indices	e.g. $CI_c^t = \frac{\sum_{i=1}^{N} \cdot (y_{ic}^t)^r}{N}$, where y_{ic}^t as in 6., and $0 < r < 1$

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 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.

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, World Economic Forum, 2001). The composite indicator is based on the standardized scores for each indicator which equal the difference in the indicator for each country and the 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 standardized 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 behavior, 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 countries 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 standardized 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.

Method 7

Method 7 is akin to Method 6, but adds the property that an increase in the value of a particular sub-indicator represents a greater increase at lower levels than equivalent increases at higher levels. In fact, if r =1 in the formula of the table, this level sensitivity disappears and one is again left with method 6. This is not the most general formulation conceivable for such an index. What is essentially required is that the normalized sub-indicators are an increasing concave function of the normalized achievement levels (see Chakravarty, 2003).

4.2 Aggregation: a caveat

All methods listed above comply with the way Desai (1994, p. 34-35), in his discussion of the Human Development Index, defined the *problem of measurement*, viz. as one of reducing a vector of variables to a scalar by means of a weighted sum. Freudenberg (2003, p. 7) likewise focuses on 'typical composite indicators' of the weighted sum-form

$$CI = \sum_{i=1}^{N} w_i y_i \tag{4.1}$$

First, it is clear that several other ways are conceivable to perform such a reduction. Indeed, physicists, economists, and others have dwelt in measurement theory, and have put forward other sensible aggregation methods.[2] A more general description of the generic measurement problem is therefore that it addresses the *interdependency* of quantitatively meaningful representations of "raw data" on the one hand, and the precise method of aggregating these representations into a scalar on the other hand.

We offer here a succinct summary of some salient aspects that are needed for the present analysis (see e.g. Roberts (1979) for a good introduction, or Aczél (1988), for a discussion about its relevance for economics in general). The basic elements of that theory are the scales (ordinal, cardinal, ...) used as a numerical representation for each of the *individual* sub-indicators. Each scale is associated with a set of admissible transformations, which in turn define what kind of numerical statements are meaningful. For example, if one observes that the price for sending a standard letter is twice as high in France than in Germany, then this remains true regardless of the currency in which both prices are denominated (i.e., one can apply the same ratio scale transformation on both original figures without compromising the 'truth' that sending a letter is twice as cheap in Germany). The other sub-indicators may be classified according to their (possibly different) measurement scales as well.

Measurement theory studies what kind of *aggregation* function can be applied to a given set of data, each having its associated measurement scale, and what kind of meaningful statements can be associated with the 'aggregate values' produced by this function. Of course, each aggregation method presupposes specific measure-theoretic qualities of the original data in order to be meaningful.

In fact, this interdependency is *indirectly* recognized in composite indicator development, as witnessed by the fact that one frequently assesses the sensitivity of eventual country rankings to the preliminary normalisation method used for the raw data (e.g. Freudenberg, 2003; Saisana *et al.*, 2005). But direct applications of measurement theory have only recently found their way into the field (Munda and Nardo, 2003). Measurement theory insights are particularly relevant for composite indicators. It can help to tackle the strong link between the 'normalization' problem ("sub-indicators have no common meaningful

[2] In economics, the field of social choice theory has been particularly concerned with the measurement problem. Measurement theoretic issues are also relevant in composite indicator construction.

measurement unit") and the 'aggregation problem' ("there is no obvious way of weighting these sub-indicators").[3]

Recent papers apply some of this theory's insights to composite indicators, and have in that way raised important criticisms against the standard aggregation/reduction of using a weighted sum. Munda and Nardo (2003) stress that a function such as (4.1) *implicitly* imposes hidden value judgments on the nature of the aggregation process. Specifically, a linear aggregation rule such as (4.1) is consistent with an interpretation of the weights as *trade-offs* (substitution rates). Stated differently: what *really* matters in the linear index are the relative weights (i.e. the $-w_i/w_j$, which directly refer to the substitutability among the different dimensions) rather than the absolute weights. This puts into perspective the 'requirement' that "the weights should add up to one". On measurement theoretic grounds this is a superfluous requirement.[4] More importantly however, Munda and Nardo (2003, p. 5-6) indicate "a theoretical inconsistency in the way weights are actually used and their real theoretical meaning" if the linear aggregation rule (4.1) is coupled with the conventional interpretation of absolute weights as indicators of the *intrinsic importance of the indicators* (i.e. as factors 'contributing to' the composite phenomenon at hand). As soon as one opts for the reduction method (4.1), and hence, the notion of substitution rates, one implicitly introduces the judgment of compensability of the sub-indicators. A lesser performance in one dimension may be outweighed by better performance in another one. This may be a value judgment that is not shared by everyone as far as particular composite phenomena are considered.[5] Munda and Nardo propose an ordering procedure, based on the 'ordinal' Condorcet criterion as used in voting theory, which sidesteps the problems associated with the implicitly assumed possibility of trading off sub-indicator values in (4.1).

The idea to use concepts or aggregation/ranking methods stemming from voting theory or social choice theory in composite indicator construction is attractive. At the abstract level, there is a parallel between e.g. ranking 'candidates' on the basis of individuals'

[3] Once again, the issue here is the lack of consensus about the right 'model' which relates sub-indicators to the composite phenomenon. To see this, note that the aggregation of apples and oranges (or of apples and scientific journals and many other goods that are hardly commensurable) is in fact a rather uncontroversial problem when constructing GDP. The trick to render them so is of course by multiplying with market prices, i.e. to work with monetary values. Trivial as this example may be, it proves the point made above, and also by Ebert and Welsch (2004, p. 271) that "arbitrary choices of measurement units can be accommodated on the basis of known scientific relationships". In the GDP-example, the 'known scientific relationship' requires a sufficient consensus that prices are sensible weights (e.g. because they are taken to represent relative factor productivities or marginal utilities).

[4] Or more rigorously: any ordering of $n$different $m-$vectors of sub-indicators is unique up to a similarity transformation of the weights (i.e. $\sum_{i=1}^{m} w_i(y_i^k/y_i^0)$and $\sum_{i=1}^{m} \omega_i(y_i^k/y_i^0)$convey the same meaning iff $\omega_i = \lambda w_i, \forall \lambda > 0$: weights could add up to any number –as long as their ratios are unaffected– without altering the results).

[5] For example, in his discussion of social inclusion, Brandolini (2002) states: " For the sake of simplicity, -but the observation carries over to more complicated formulations- suppose that the summary index equals the arithmetic mean of the selected indicators. In adopting such an index, we are implicitly assuming that one unit more of indicator A can be substituted for one unit less of indicator B or vice versa. If A is the unemployment rate and B the proportion of people failing to reach 65, our summary index would suggest that the valuation of the social situation is unchanged when the unemployment rate is reduced by 1 percentage point at the same time as the proportion of people dying before 65 is raised by 1 percentage point. I do not think that this conclusion is acceptable, nor is it likely to gain wide acceptance."

performance measures of these candidates and 'ranking countries' on the basis of a multi-dimensional set of attributes. For example, Panigrahi and Sivramkrishna (2002) apply a similar idea (although a different method, viz. the Borda-rule) to the Human Development Index.

The parallel can indeed be exploited still further: if one *is* willing to accept more than ordinal comparisons, then other aggregation procedures can become available. In principle at least, expression (4.1) can be thought off as a limiting case. However, it remains true that assumptions on measurability and on inter-comparability of normalized sub-indicators usually lead up to *specific* aggregation methods. This is the most important point of Ebert and Welsch (2004), and they list various examples of the contingency of specific aggregation methods on underlying assumptions. More specifically, they too raise the point that a weighted-sum (4.1) may often be chosen superficially, although their critique is different from that of Munda and Nardo (2003). Ebert and Welsch (2004) in fact point out that several of the normalization methods listed in table 4.1 may be inadmissible transformations of the raw data when coupled with a weighted sum, because the meaning of the original numbers may get lost in the transformation.

If the role of science in composite indicator construction is to assure that the processes of aggregation are as sound and transparent as possible, then the insights which measurement theory can offer for composite indicator construction deserve further analysis. The remarks that have been raised against the standard aggregation method (4.1) should not be interpreted as a condemnation of that method. If one opts for an aggregation of the form (4.1), it seems advisable to analyze the trade-offs in more detail than is hitherto usual (see e.g. Lind (2004), for a discussion of the Human Development Index along these lines).

Composite indicator construction can only benefit from the kind of axiomatic approach that is e.g. displayed in Kakwani (1993), Chakravarty (2003), Munda and Nardo (2003), and Ebert and Welsch (2004). By rendering the underlying assumptions and mathematical properties of a specific proposal explicit, the problem of their conceptual transparency (or the possible lack of such transparency) can be tackled at a fundamental level. This does not imply that such an approach will be able to tackle *all* problems associated with composite indicators. Measurement theory alone will usually not be sufficient to provide 'the' suitable composite indicator. First, because it will probably only strengthen the thought that "the choice of any particular index must be guided by its intended use".[6] And second, because if one opts for a 'weighted aggregation approach', even when an appropriate class of indices is chosen one still has to decide on the magnitude of weights. We turn to the weighting issue in the following section.

[6] Maasoumi and Nickelsburg (1988, p. 328), from which the quote is taken, consider a whole class of entropy measures, which are quite conventional in inequality measurement but have hitherto received little attention in the composite indicator literature. Still, they may indeed be considered as composite indicators if one is concerned with measuring the 'relative inequality in the distribution of the composite measure' (i.e. multi-dimensional inequality).

Chapter 5

Weighting strategies for building composite indicators

5.1 Introduction

A number of techniques are analysed in this section, and a comparative analysis is offered of their advantages and drawbacks. The techniques are: 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. The use of some of these techniques for composite indicator construction extends beyond the issue of weighting. For example, as discussed below, a method such as principal components analysis can also be instrumental in the 'variable selection'-stage identified earlier. Also, the techniques surveyed below are not always used in direct conjunction with the 'standard approach' summarized in table 4.1.

5.2 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 $\widehat{Y}$, 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 and an indicator $\widehat{Y}$ representing the desired objective (e.g. National Innovation Capacity index, Porter and Stern, 1999). A multiple regression model is then constructed to calculate the relative weights of the sub-indicators. Such models are essentially linear,

$$\widehat{Y} = a + b_1 X_1 + ... + b_n X_n \tag{5.1}$$

where $\widehat{Y}$ 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 can handle a large number of variables of different types, within the assumption of linear behaviour, 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. This critique could be overcome by considering other (non-linear) functional specifications.

It is further argued that if the concept to be measured could be represented by a single indicator $\widehat{Y}$, then there would be no need for developing a composite indicator in the first place (Muldur, 2001). This indeed stands to reason in a case where 'one' $\widehat{Y}$ would be considered as a valuable proxy for the underlying phenomenon, but such an assumption sits uneasily with the general idea of a composite, i.e. presumably multi-faceted phenomenon. Slottje (1991) considers several latent variables (and, indeed, a logarithmic alternative for expression 5.1) for measuring the quality of life. Of course, each equation may then well lead to a distinct country ranking, which ultimately again requires some form of averaging over the rankings.[1]

Dowrick, Dunlop and Quiggins (2003) also use regression results (viz. on the 'production function' for life expectancy) as a basic element of the analysis, but in their case it is used to overcome the commensurability problem (i.e. the problem of different measurement units) with food and educational expenditures. Their specific ranking method is based on the well-known revealed preference argument in economics, which in this specific case amounts to test whether country X can afford to devote more resources to increase life expectancy than country Y, under the assumption that country X consumes at least as much goods (i.e. food and education) as country Y, and given the price structure and budget constraint faced by X.[2] The revealed preference argument has been used in another context as well, as shown later.

5.3 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, World Economic Forum, 2001),
2. to *cluster* the indicators (General Indicator of Science & Technology, (National Institute of Science and Technology Policy, 1995), and

[1]Alternatively, the set of sub-indicators considered as input in the regression model (5.1) 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, a generalization of multiple regression, could be applied, (Manly, 1994).

[2]Stated as such, the concomitant answer in a revealed preference test is either 'yes', 'no', or inconclusive. Dowrick *et al.* (2003) first present pairwise comparisons of this kind. However, in a further step they proceed by imposing more structure (viz. homotheticity) on the underlying preference relationship, which allows a ranking on the basis of index numbers. Adding more assumptions about preferences to get more fine-grained results clearly is reminiscent of the discussion in section 4.

3. to *define the weights* (e.g. Internal Market Index, (DG MARKT, 2001, Noorbakhsh, 1988, Maasoumi and Nickelsburg, 1988, Slottje, 1991).

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 $X_1, X_2, ..., X_p$ and find linear combinations of these to produce principal components $Z_1, Z_2, ..., Z_p$ that are uncorrelated, following

$$Z_j = \sum_{i=1}^{p} a_{ij} X_i \quad , \quad j = 1, 2, ..., p \tag{5.2}$$

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.5.2 are chosen so that the principal components Z satisfy the following conditions:

1. they are uncorrelated (orthogonal),
2. 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
3. $a_{1j}^2 + a_{2j}^2 + ... + a_{pj}^2 = 1$, j = 1,2,...,p

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

$$C = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \cdots & & & \\ c_{p1} & c_{p2} & \cdots & c_{pp} \end{bmatrix} \tag{5.3}$$

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,

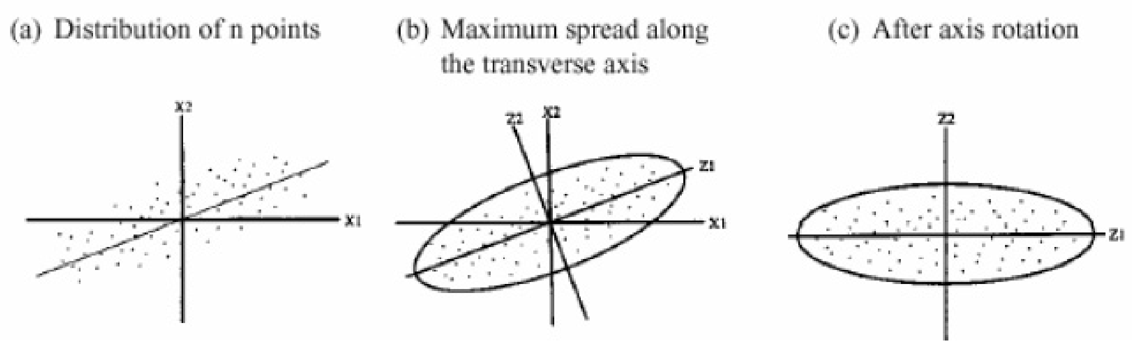
$$\lambda_1 + \lambda_2 + ... + \lambda_p = c_{11} + c_{22} + ... + c_{pp} \tag{5.4}$$

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.5.2.

In order to show graphically PCA at work, 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 points is shown in Figure1a. 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 5.1a, the relationship is not perfectly proportional, although it is nearly proportional, so in approximations a single variable is sufficient.

Figure 5.1: Distribution diagram of n points over two indicators and axis rotation



In Figure 5.1b, 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 the situation (figure 5.1c). 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_2coordinates of n points.

In the distribution shown in the figure, n points are greatly dispersed along the Z_1 axis, so when observing the data, a considerable proportion of the information content of the data can be understood 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 points, then the $n$points are distributed in the shape of a circle, 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 DG MARKT, 2001, Science and Technology Indicator National Institute of Science and Technology Policy, 1995). 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 capture sufficient variation to be an adequate representation of the original set (e.g. Business Climate Indicator DG ECFIN, 2000). 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 Emam *et al.*, 1998, Internal Market Index DG MARKT, 2001).

Factor analysis

Factor analysis (FA) has similar aims as 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 ith 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:

$$\begin{aligned} X_1 &= \alpha_{11}F_1 + \alpha_{12}F_2 + ... + \alpha_{1m}F_m + e_1 \\ X_2 &= \alpha_{21}F_1 + \alpha_{22}F_2 + ... + \alpha_{2m}F_m + e_2 \\ \ldots &= \ldots \\ X_p &= \alpha_{p1}F_1 + \alpha_{p2}F_2 + ... + \alpha_{pm}F_m + e_p \end{aligned} \tag{5.5}$$

where X_i is a variable with zero mean and unit variance; α_{i1}, α_{i2}, ..., α_{im} are the factor loadings related to the variable X_i; $F_{1,}F_2, ..., F_m$ 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 to be noted that there is an infinite number of alternative solutions for the factor analysis model. The standard practice is to choose factors that: (i) have associated eigenvalues larger than one; (ii) individually contribute to the explanation of overall variance by more than 10%; (iii) cumulatively contribute to the explanation of the overall variance by more than 60%.

5.4 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 $\overline{r}$ among the indicators:

$$\alpha = \frac{p \cdot \overline{r}}{1 + (p-1) \cdot \overline{r}} \tag{5.6}$$

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"(Emam *et al.*, 1998).

5.5 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 (Commission of the European Communities, 1984). 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

$$X = [2 \cdot (1+r)]^{-1/2} (X_1 + X_2) \tag{5.7}$$

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:

$$Z = [2 \cdot (1+\rho)]^{-1/2} (X + Y) \tag{5.8}$$

where ρ 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.

The techniques hitherto discussed are based on correlations or other 'statistical' information and thus seem to be neutral in the sense that they extract information from the numerical values of sub-indicators themselves. However, the information extracted from data by using a regression may be dependent on the particular functional form that has been used to estimate the relationship (and it is not always possible to test the 'truth' of such a choice). Similarly, 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. Booysen (2002, p. 141) also points at this '*relative*' objectivity of multivariate techniques. In his words, "indices are subject to subjectivity despite the objectivity of the methods employed in composite indexing".

Since the weighting approaches that are discussed hereafter are sometimes criticized precisely on account of their subjectivity, it may be good to remind that 'objective' weighting schemes of the kind just surveyed are in the end also prone to similar critiques.

5.6 Distance to targets

One way to avoid the explicit selection of weights is to measure the need for political intervention and the "urgency" of a problem by the distance to target. 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.[3]

Using policy goals as targets (e.g. "Environmental Policy Performance Indicator" Adriaanse, 1993) may appeal to policy makers for the "soundness" of the weighting method, as long as the policy makers agree with the policy targets. 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 United Nations, 2001)

5.7 Experts opinion (budget allocation)

A commonly used method is the assignment of weights to sub-indicators based on personal judgment of stakeholders (participatory method). This method, however, has limits when the 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 becomes problematic if the same person is asked

[3] The soundness of using a simple *arithmetic* average may be questioned on measurement theory grounds. A distance-to-target normalization is one particular *ratio-scale* transformation, which is naturally linked with a (simple) *geometric* average.

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 (values), so as to ensure that a proper weighting system is found for a given application (von Winterfeld and Ward, 1986).

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 (Moldan and Billharz, 1997).

A counter argument against the use of expert 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 (although problems of circular thinking or inconsistencies are more pertinent to the AHP technique we discuss below). 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.

Two other caveats are relevant here (although they also hold for other methods based on expert opinion). First, care must be taken to ensure that experts do not misunderstand the nature of the weights they are asked to provide (they must, for example, be aware of the fact that they are giving *trade-off* values, and in some cases these trade-off values may pertain to normalised variables rather than to raw data). Second, divergent expert opinions are likely to re-introduce problems of aggregation (in this case: of the opinions). Should one, for example, take the experts' average weights as the final ones? While the latter option is a common practice, it does not go without criticisms.[4]

[4]One way to rationalize this option is to consider the selected expert opinions as nothing more than a sample extracted from a population, to which the laws of probability can be applied. However, if the experts are truly experts, i.e. when each one of them is endowed with knowledge of the phenomenon under consideration, then "decision makers who whish to base choices on the advice of the panel have no way to objectively assign probabilities to the alternatives". Or still: "Since experts' opinions vary because of underlying theories, in many circumstances the relative number of experts that hold a particular position tells us little about the likelihood that that perspective will be correct" (see notably Woodward and Bishop (1997), who use this observation to justify techniques for decision-making under (Knightian) uncertainty.

5.8 A note on 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" (Parker, 1991)). 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. Furthermore, all problems already mentioned are aggravated if one turns from a small group (of experts) to a larger audience.

5.9 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 (Saaty, 1987). It enables decomposition of a problem into a 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. According to Forman (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."

Methodology in brief

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 5.1) where $A_{ii} = 1$ and $A_{ij} = 1 / A_{ji}$.

For the example shown in Table 5.1, Indicator A is three times more important than Indicator B, and consequently Indicator B has one-third the importance of Indicator A.

Table 5.1: 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

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 the table below.

Table 5.2: 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 more important than B, B more important than C, and C 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.

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 is less sensitive to errors of judgement. In addition, this redundancy allows for a measure of these judgement errors by 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. Finally, experts can be themselves rated based on their inconsistency rate.

5.10 Benefit-of-the-doubt weighting

A problem originally encountered in micro-economics is the assessment of the performance of a firm or agent relative to similar firms or agents on the basis of multiple input and/or output data, but without knowledge of prices. A technique known as 'data envelopment analysis', can be applied when knowledge of trade offs is lacking. Data envelopment analysis has proven sufficiently flexible to be applied to a vast array of problems (e.g. for assessments of the macro-economy, as a voting procedure, ...). A by-product of its popularity in management science and economics is the availability of different software packages (or add-ins, e.g. for MS excel) that facilitate computations of benefit-of-the-doubt composite indicators.

A recent introduction to the theory and practice of data envelopment analysis, as well as an overview of available software, is provided by Cooper *et al.* (2004). Here we only point at some essential ideas as applied to composite indicators, see e.g. Melyn and Moesen (1991) or Cherchye (2001) for a methodological discussion. Specific applications to composite indicators include Storrie and Bjurek (1999),2000 (labour market performance); Mahlberg and Obersteiner (2001), Despotis (2004) (HDI), Lauer *et al.* (2004) (WHO country performance ranking); Cherchye and Kuosmanen (2004) (sustainable development) and Cherchye, Moesen and Van Puyenbroeck (2004) (social inclusion).[5]

The starting point of all these analyses is the observation that there is usually no (expert) consensus on the weights used to aggregate the (possibly normalised) sub-indicators. Moreover, any specific choice of a weighting vector is by definition imposed upon the evaluated country, which may not always be received positively. For example, some authors have argued that differential weighting may be desirable in composite indicators, e.g. because of different environments or political attitudes (Veenhoven, 1996) or because the very idea of imposing weights may be inconsistent with the subsidiarity principle (Cherchye, Moesen and Van Puyenbroeck, 2004). These worries can be overcome by rendering the weight selection problem *endogenous* for each observation. That is, the relative weight accorded to each sub-indicator is endogenously determined in this type of performance evaluation models, so as to reflect the associated relative performance for the country under evaluation. Thus, a good relative performance in a particular dimension is seen as 'revealed evidence' of high national policy priority to that dimension, which explains the 'benefit-of-the-doubt'-terminology that has alternatively been used for this method. (Melyn and Moesen, 1991). Note also that the resulting index number is a gauge of *relative* performance: using its own benefit-of-the-doubt weights, a country's sub-indicators are compared with those of the other countries in the sample.

The fact that weights are observation-specific, hence flexible, may well lead to identification of benchmark observations that are different for each evaluated country.

The following formula concisely recaptures the ideas just discussed: each country's com-

[5] All the papers listed above have in common that they focus on the aggregation of performance indicators (i.e. outputs). It should however be noted that the method is particularly suited for comparisons of multiple inputs *and* outputs. When undertaken in the latter terms, the analysis is similar in spirit as the one of Dowrick *et al.* (2003) already mentioned. See especially Zaim, Färe and Grosskopf (2001) for a 'non-parametric' (endogenous weighting) revealed preference analysis of achievement indices.

posite indicator value will be found by solving:

$$\max_{w_i \geqslant 0, i=1,\ldots,m} \frac{\sum_{i=1}^{m} y_{ij}^n w_i}{\max_{y_k : k \in \{1,\ldots,n\}} \sum_{i=1}^{m} y_{ik}^n w_i} \tag{5.9}$$

As the formula reveals, the denominator of the index value, i.e. the benchmark observation value, is itself obtained from an optimization problem. It is in fact the observation that, by employing the 'most favorable weights' for the *evaluated* observation, obtains the maximal weighted sum of all observations in the sample. Consequently, this benchmark is endogenous too. Literally, it either demonstrably outperforms the evaluated observation in terms of the latter's most flattering weighting scheme or, if such superior benchmark doesn't exist, the evaluated observation serves as its proper benchmark.

Clearly, as $\sum_{i=1}^{m} y_{ij}^n w_i \leqslant \max_{y_k : k \in \{1,\ldots,n\}} \sum_{i=1}^{m} y_{ik}^n w_i$ for each weighting scheme w_i $(i = 1, ..., m)$,the value of the composite indicator will be a figure between 0 and 1. This implies that the endogenous benchmark value is normalized at that '100%'-value. Note that this goes against a common practice in composite indicator construction, where as a rule the weights themselves are directly restricted to add up to one (see however section 4, footnote 10).

These features imply that national policy makers can hardly complain about an unfair weighting scheme (as any other weight profile would only worsen the position *vis-à-vis* the other countries in the sample). Furthermore, –at least in the basic formulation of these models- the normalization problem can be entirely left aside, in view of the method's unit invariance property (see Cherchye, Moesen and Van Puyenbroeck (2004), for more details). However, the apparent judgmental relativism which the basic, 'unconstrained' approach entails may be criticised.[6]

At the same time, one should note that many of the methods described above *do* introduce additional value judgments in the optimization problem. In particular, while experts in practice often fail to agree on a *specific* set of weights, there are ways in which some form of broad , mutually agreed judgments can be linked to endogenous weighting. This can be achieved by applying the experts' stated weight vectors as constraints in a weight optimization problem that seeks to maximize the aggregate performance index for each particular observation. Formally, one can append requirements $(w_1, .., w_i, .., w_m) \in \mathbf{W} \subseteq \mathbf{R}_+^{\mathbf{m}}$ to the above optimization problem. These can e.g. capture agreement on 'ordinal' weight bounds, on upper and lower bounds for trade-off values (i.e. on relative weights), etc. Hence, in a more general setting, for each observation, the weights leading up to its index value are *to some degree* endogenous, the degree of endogeneity depending on the extent of disagreement between the expert panel.

[6] On a more fundamental level, since this method hinges on differential weighting, it should be pointed out that the resulting set of composite indicators cannot be considered as a conventional *ordering* of countries. However, referring to the remark on weight bounds in the following paragraph, it should also be pointed out that one can precisely restrict endogenous weights to be 'similar' or even 'the same' for all observations. See e.g. Cherchye and Kuosmanen (2004), Kuosmanen *et al.* (2004), Despotis (2004).

5.11 Equal weighting

Despite the many weighting techniques that where hitherto listed, the one most often used in composite indicator construction is that of equal weighting after data normalisation. Babbie (1995, p. 171) goes as far as to recommend this approach as a standard, stating that "items be weighted equally unless there are compelling reasons for differential weighting. That is, the burden of the proof should be on differential weighting; equal weighting should be the norm".

Yet several criticisms can be raised against equal weighting. First, equal weighting is often only 'apparently' simple. In fact the results of an equally weighted index are still conditioned by the choice of the normalization method, a troublesome feature that is well-documented in the literature on composite indicators. Second, the most prominent substantive justification for equal weighting goes back to Occam's razor. For example, Hopkins (1991, p. 1471) states that "Since it is probably impossible to obtain agreement on weights, the simplest arrangement is the best choice." But the principle of parsimony may be inadequate here. As far as composite indicators are concerned, taking the simplest weighting scheme in fact does not imply choosing the simplest model from a set of *otherwise equivalent* models of a given phenomenon. As a rule, the problem in this context is rather that there are at best partially conflicting opinions about 'underlying models' available. Stated differently, equal weighting may sometimes not even be an adequate *description* of the debate in composite indicator construction.

Chapter 6

Visualization

The presentation of composite indicators is not a trivial issue. Given their eventual purpose, composite indicators must be able to communicate the picture to decision-makers and users quickly and accurately. Visual models of these composites must provide signals, in particular warning signals that flag for decision-makers those areas requiring policy intervention.

Hereafter we list some interesting ways to display and visualize composite indicators. We accompany each type by a brief commentary of the pros and cons. We start from the simplest tools and explore their modifications. Again, we will not be exhaustive here, but confine ourselves to those tools that have hitherto been among the most popular to be employed in this area.

6.1 Tabular format

This is the simplest format whereby, for each country, the composite indicator and its underlying indicators are presented as a table of values. Usually countries are displayed in decreasing ranking order. This is a comprehensive approach to display results, yet not particularly visually appealing. The approach could be adapted to show targeted information for sets of countries grouped, for example, by location, GDP, etc.

A tabular format can also be used when country rankings, rather than composite indicator values, are reported. When such a table shows the rankings of countries for two or more consecutive years, it can be used to track changes of country performance over time. However, as stated before, the limitation of ranks is that one loses the information on the difference between countries performances.

In several cases one provides both levels and country rankings, for both the sub-indicators and the composite one. For example, the British Office of National Statistics has produced indices of economic deprivation in six domains (income, employment, health deprivation and disability; education; skills and training; housing; and access to services) for the all the districts in 2000. The composite is the average of scores out of a 100 for each sub-indicator. The rank is the average of ranks for each sub-indicator; ranks go from 1 to approximately 8,000 (the total number of districts).

6.2 Bar charts

Bar charts, with countries on the vertical axis and the values of the composite on the horizontal axis can also be used. One refinement is to add a top bar indicating the average performance of all countries (in the sample, in the world, ...), which enables the reader to quickly identify how a country is performing with regards to the average. The top bar could alternatively be thought as a target to be reached by countries. Colors can make the graph more visually appealing and highlight countries performing well or bad, or showing either growth or slow down, or, finally, to highlight countries having reached an average or mandatory standard.

6.3 Line charts

Line charts are used to show performance across time. Performance can be displayed using a) absolute levels; b) absolute growths (in percentage points with respect to the previous year or a number of past years); c) indexed levels and d) indexed growths.[1] A number of lines are usually superimposed in the same chart to allow comparisons between countries. An example of the latter practice is given by the Internal Market Index 2004, published on the Internal Market Scoreboard N. 13 (DG MARKT, 2001). Here, groups of countries with similar performance (better, similar or worse than the EU) have been displayed in the same chart. All the countries have been indexed to 100 in the starting year (1994).

The above methods are all relatively simple as regards their set-up. More sophisticated presentational tools also exist. The following one is a well-known example:

6.4 Dashboards

The Dashboard of Sustainability (see `http://esl.jrc.it/envind/`) is a non-commercial software which allows to present complex relationships between economic, social and environmental issues in a highly communicative format aimed at decision-makers and citizens interested in Sustainable Development. It is also particularly recommended to students, university lecturers, researchers and indicator experts.

The Dashboard includes maps of all continents and can be developed using one's own dataset. A vast collection of dashboards already exist. To make some examples, on the internet site one can find the "ecological footprint", a pure environmental composite, the "environment sustainability index", presented by the World Economic Forum annual meetings, the "European Environmental Agency's EEA Environmental Signals". The "From Rio to Johannesburg" and the "Millennium Development Goals" versions are recommended for introductory courses on Sustainable Development.

The Dashboard can help answering some typical questions as:

[1]The word 'indexed' means that the values of the indicator are linearly transformed so that their indexed value at a given base year is 100.

1. What is the situation of my country compared to others?
2. What are specific strengths and weaknesses of my continent/my country?
3. How are certain indicators linked to each other?

6.5 Levels vs. growth rates

When a composite indicator is available for a set of countries for at least two different time points, one is commonly interested not only in the levels at a given time point, but also in the growth rates between the available years.

Examples are given in the next section, e.g. the presentation of the European Innovation Scoreboard. The scoreboard includes the Summary Innovation Index to track relative performance of Member Countries in Innovation. Overall country trends are reported on the X-axis and levels are given on the Y-axis. A horizontal axis gives the EU average value and a vertical axis gives the EU trend. The two axes divide the area into four quadrants. Countries in the upper quadrant are "moving ahead", because both their value and their trend are above the EU average. Countries in the bottom left quadrant are "falling further behind" because they are below the EU average for both variables.

Other examples of this presentational tool in the following section are the composite indicators of investment and performance in the knowledge-based economy, also developed by the European Commission in the framework of the Lisbon agenda.

Chapter 7

Selected Applications

This section gives an overview of existing applications of composite indicators in the areas of research, innovation and S&T. First, a recent approach by the European Commission to measure the knowledge based economy is described and some illustrative results are presented. Next, several other contributions to the field are shortly discussed.

7.1 Two Composite Indicators of the Knowledge-Based Economy (by the European Commission)

Scope of the Index

In the framework of the European Commission's Structural Indicators exercise, it was decided that it would be useful for the Commission services to investigate and develop composite indicators of the knowledge economy. A number of Commission services have been involved and consulted during the development work, including DG Education, Eurostat, DG Information Society and DG Enterprise. External technical assistance with the refinement of the methodology was provided by Anthony Arundel and Catalina Bordoy of MERIT. The Applied Statistics Group of the Joint Research Centre also contributed significantly to reviewing different approaches and testing the sensitivity of the chosen method. This cooperation resulted in the production of two new indicators: a composite indicator of investment in the knowledge-based economy, and a composite indicator of performance in the transition to the knowledge-based economy. This subsection presents some preliminary results emerging from this work on composite indicators.

The two indicators attempt to capture the complex, multidimensional nature of the knowledge-based economy by aggregating a number of key variables, and expressing the result in the form of an overall index. The two composite indicators refer to the overall investment and performance in the transition to the knowledge-based economy. They focus on the 'knowledge dimension' of that transition and, therefore, do not take into account the other dimensions (e.g. employment, sustainable development, etc.) of the Lisbon Agenda.

7.1.1 Investment in the knowledge-based economy

Description of sub-indicators

In order to advance effectively towards the knowledge-based economy, countries need to invest in both the creation and the diffusion of new knowledge. The composite indicator of investment in the knowledge-based economy addresses these two crucial dimensions of investment. It includes key indicators relating to R&D effort, investment in highly-skilled human capital (researchers and PhDs), the capacity and quality of education systems (education spending and life-long learning), purchase of new capital equipment that may contain new technology, and the modernisation of public services (e-government). The following table (taken from the "Third European Report on S&T Indicators") shows the sub-indicators of this composite indicator.

Table 7.1: Component indicators and weightings for the composite indicator on investment in the knowledge-based economy

Sub-indicators	**Type of knowledge indicator**	**Weight**
Total R&D (GERD) per capita	Knowledge creation	2/24
Number of researchers per capita	Knowledge creation	2/24
New S&T PhDs per capita	Knowledge creation	4/24
Total Education Spending per capita	Knowledge creation and Knowledge diffusion	4/24 3/24
Life-long learning	Knowledge diffusion : human capital	3/24
E-government	Knowledge diffusion: information infrastructure	3/24
Gross fixed capital formation (excluding construction)	Knowledge diffusion new embedded technology	3/24

Source: DG Research
Data: Key Figures, 2002 Third European Report on S&T Indicators, 2003

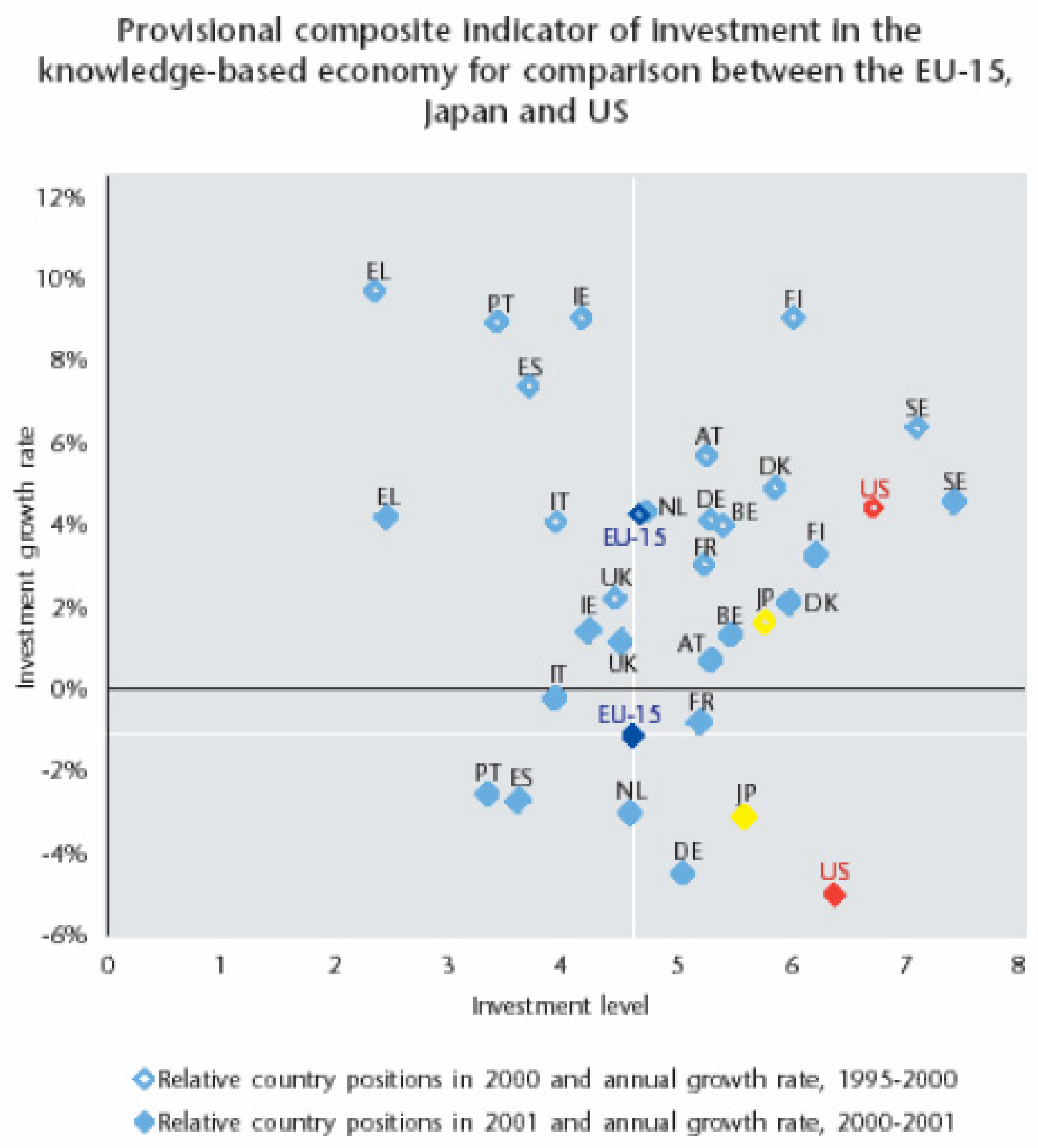
Presentation of the Index

The following figure (taken from the Commission's "Key Figures 2003-2004") maps the values of the indicator for all EU-15 countries except for Luxembourg. On the horizontal axis, the level of the investment indicator is plotted for 2000 and 2001 (framed and solid). Vertically, the graph shows the growth rate (1995-2000 and 2000-2001, respectively) of the investment in the knowledge-based economy. It can be seen that for the solidly marked country positions, roughly three different groups of states can be identified.

Figure 7.1: Visual presentation of KBE-Investment Composite Indicator

The next graph compares the EU-15 with the United States and Japan. Due to data limitations, only four (GERD per capita, number of researchers per capita, new S&T PhDs per capita, gross fixed capital formation) out of the seven sub-indicators were included. The figure reveals that the US faced a negative investment growth during the 2000-2001 period, but still had the second-highest investment level after Sweden.

Figure 7.2: KBE-Investment Comparisons



7.1.2 Performance in the knowledge-based economy

Description of sub-indicators

Investing more in knowledge is, however, only half the story. Investment also needs to be allocated in the most effective way in order to increase productivity, competitiveness and economic growth. For this to happen, and to be sustainable, investment in knowledge thus has to induce a higher performance in research and innovation and increased labour productivity, an effective use of the information infrastructure and a successful implementation of the education system. This relationship between investment and performance, however, is very complex and certainly not linear. It depends in part on favourable framework conditions and policies. Moreover, there is always a time-lag between investment and a recorded increase in performance.

The second composite indicator regroups the four most important elements of the 'performance in the transition to the knowledge-based economy': overall labour productivity, scientific and technological performance, usage of the information infrastructure and effectiveness of the education system (see the following table, taken from the "Third European Report on S&T Indicators").

Table 7.2: Component indicators and weightings for the composite indicator on investment in the knowledge-based economy

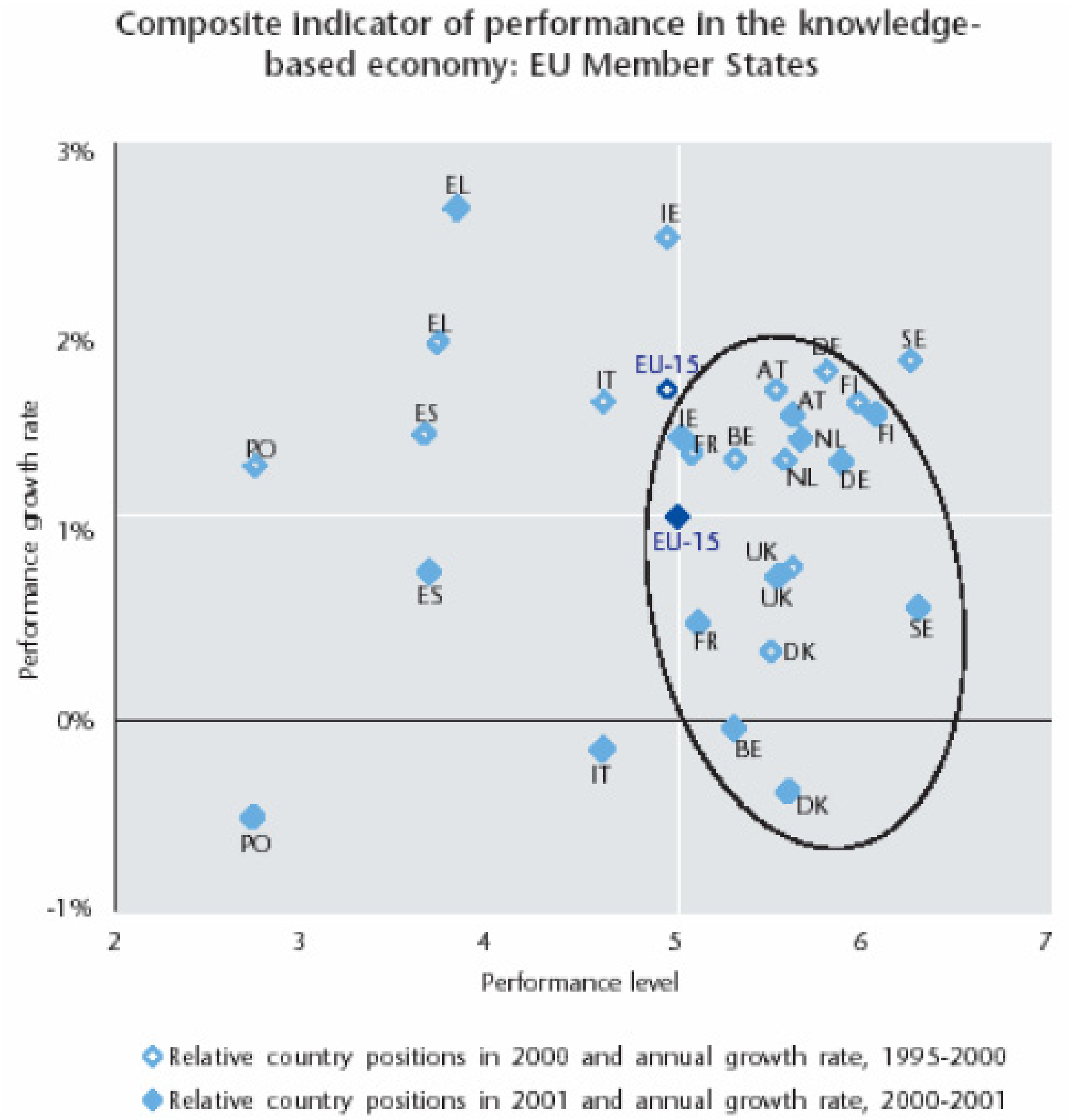
Component indicators	**Conceptual group**	**Weight**
GDP per hour worked	Productivity	4/16
European and US patents per capita	S&T performance	2/16
Scientific publications per capita	S&T performance	2/16
E-commerce	Output of the information infrastructure	4/16
Schooling success rate	Effectiveness of the education system	4/16

Sourcce: DG Research
Data: Key Figures,2002 Third European Report on S&T indicators, 2003

Presentation of the Index

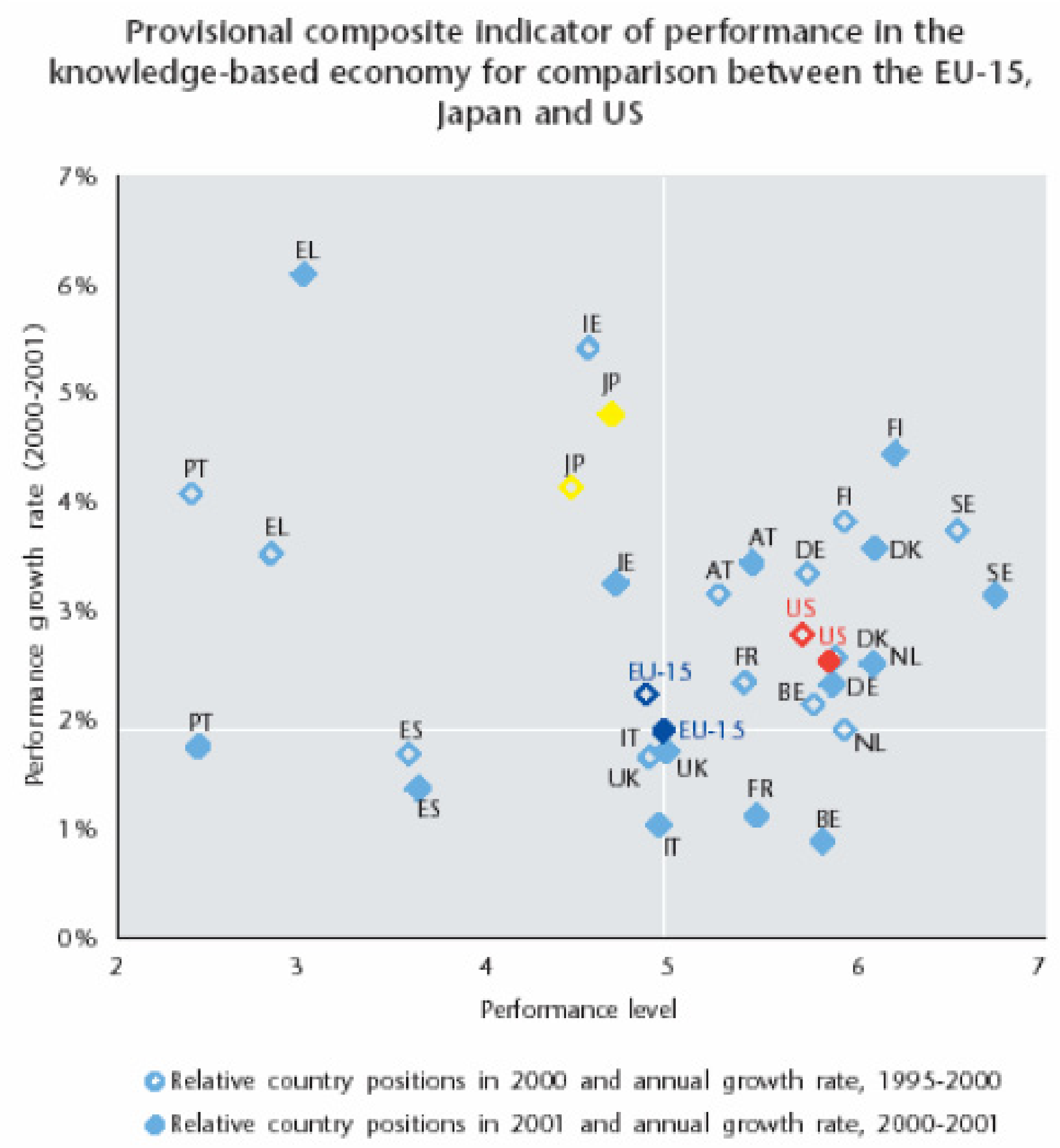
Also for the performance of the knowledge-based economy, the indicator values can be mapped in a level/growth rate graph. The graph (taken from the "Key Figures 2003-2004") shows that in 2001, ten EU-15 countries were comparable in terms of performance level, while Greece, Portugal, Spain and Italy were lagging behind.

Figure 7.3: Visual presentation of KBE-performance Composite Indicator



For the comparison with the United States and Japan, only three (GDP per hour worked, patents per capita and publications per capita) could be considered. The following figure indicates that in 2001, the three Nordic countries Denmark, Finland and Sweden had a better value of their performance in the knowledge-based economy than the US, both in level and growth rate.

Figure 7.4: KBE-Performance Comparisons



7.2 Summary Innovation Index (by DG ENTR)

Scope of the Index

The Summary Innovation Index (SII) is part of the European Innovation Scoreboard (EIS), 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". It additionally uses 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 EUROSTAT data, or private data of sufficient reliability if official data are not available. The innovation scoreboard analyses 20 indicators in four areas: (i) human resources; (ii) knowledge creation; (iii) transmission and application of new knowledge; and (iv) innovation finance, output and markets.

Aggregation method

The SII is calculated using re-scaled values of the indicator data, where the highest value within the group of EU25 countries is set to 1 and the lowest value within the group of EU25 countries to 0. The SII is then calculated as the average value of all re-scaled values and is by definition between 0 and 1 for the EU25 countries. The weights for the sub-indicators are either 0.5 or 1.

Comments

- Since 2004, the innovation scoreboard also includes an analysis of innovation performance by sector, called Innovation Sector Index (ISI).

- Another new indicator concerning "non-technical change" has been integrated to measure changes in organisational structures, management techniques and product design.

Presentation of the Index

Figure 7.5 shows the results for the 2004 SII. It can be seen that Sweden and Finland remain the innovative leaders within the EU. Figure 7.6 graphs the current performance as shown by the SII (vertical axis) against the medium-term trend performance (horizontal axis) which is calculated as the percentage change between the last available year and the average over the preceding three years, after a one-year lag.

Figure 7.5: The 2004 Summary Innovation Index (SII)

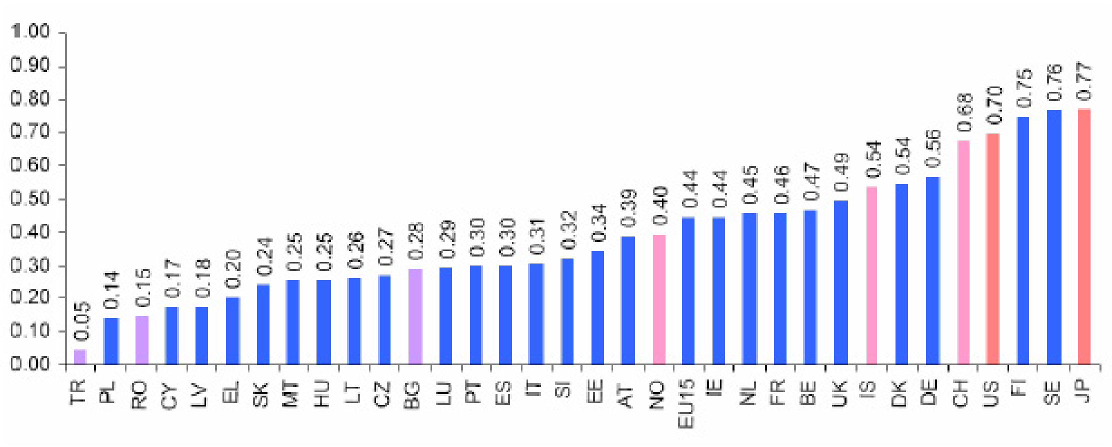
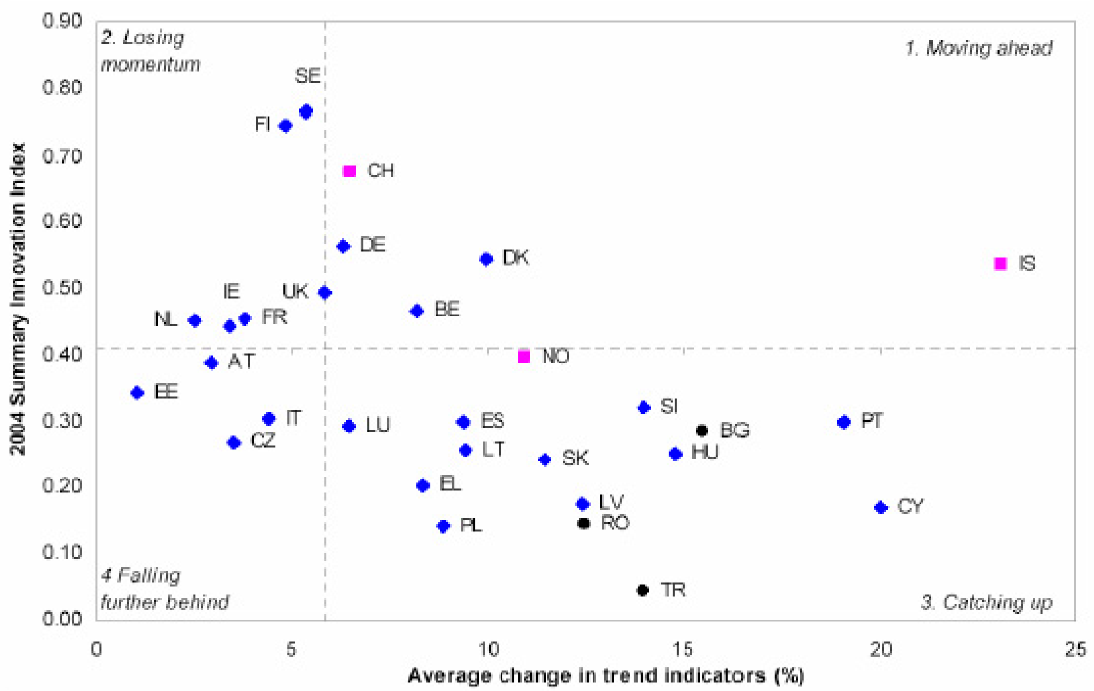


Figure 7.6: Average country trend by SII



7.3 *Welfare of Nations* (by Hans-Olof Hagén)

Scope of the Index

Welfare has consistently been described as a multifaceted concept whose components cannot be measured with the same yardstick. In order to render these components comparable, a composite indicator is created which measures the welfare in OECD countries.

Description of sub-indicators

The welfare measure is composed of four sub-indicators: (i) economic standard, expressed in gross national income (GNI); (ii) leisure time; (iii) health, measured by life expectancy of males/females and infant mortality; (iv) environment, derived from the emission of pollutants containing sulphur, nitrogen and carbon dioxide.

Aggregation method

All variables are standardised to be 0 for the country with the lowest value and 100 for the country with the highest value. Both the sub-indicators for health and environment and the final welfare indicator are aggregated using equal weights of the components, respectively.

Comments

- A variety of sensitivity and robustness analyses are conducted. As an example, a Monte Carlo simulation with random weights reveals that Norway stays at first place in more than 60 percent of the cases.
- Another composite indicator in the study measures the “input factor”. It consists of labour quantity welfare, labour quality, research and IT.
- Several correlation and regression analyses are conducted to assess in more detail the interrelations between the (sub-)indicators.

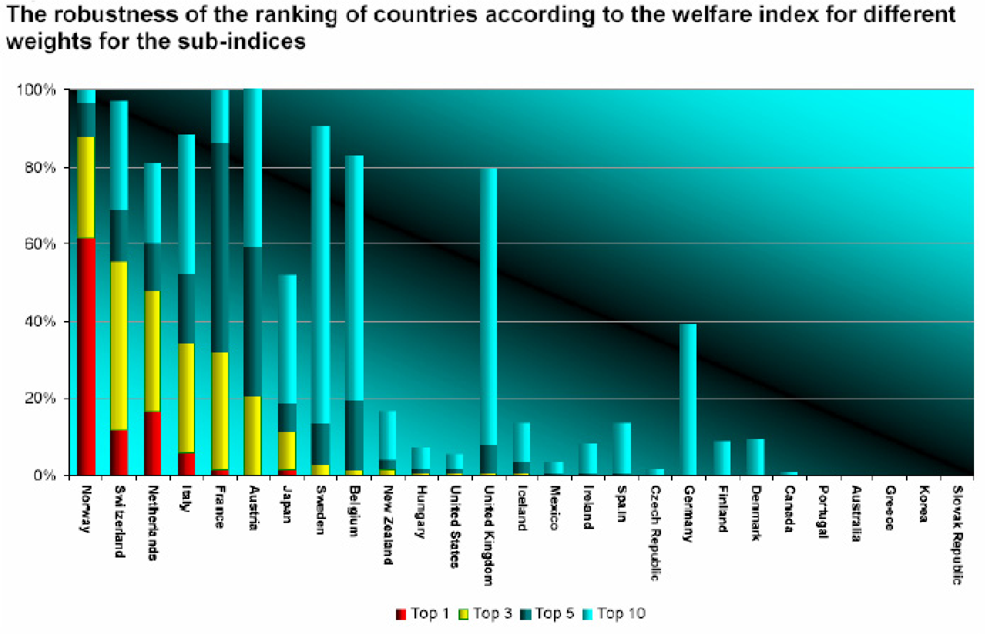
Presentation of the Index

The composite welfare index is given in a table format. The mentioned example of a robustness analysis is presented graphically in Figure 7.7.

Table 7.3: Welfare and its components

	Economic standard	Leisure time	Health	Environ-ment	Welfare
Norway	100	67	74	82	81
Switzerland	84	43	83	94	76
France	67	72	78	82	75
Netherlands	68	100	66	60	74
Austria	68	63	73	89	74
Italy	58	85	83	66	73
Sweden	63	55	83	80	70
Belgium	68	77	64	72	70
United Kingdom	75	54	65	80	69
Japan	70	41	100	60	68
Germany	59	64	64	75	66
Spain	46	49	79	79	63
Ireland	60	60	53	76	62
New Zealand	46	35	66	100	62
Denmark	76	48	54	69	62
Finland	61	26	70	83	60
Iceland	71	3	88	68	58
Greece	33	54	64	72	56
Portugal	32	56	57	77	56
United States	91	44	52	23	52
Canada	71	33	77	23	51
Czech Republic	27	33	45	97	50
Hungary	18	79	12	89	50
Australia	61	33	80	0	44
Mexico	0	86	0	79	41
Korea	28	0	42	87	39
Slovak Republic	15	38	26	78	39

Figure 7.7: Visualisation of Robustness Check for 'Welfare of Nations'-CI



7.4 *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.

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

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

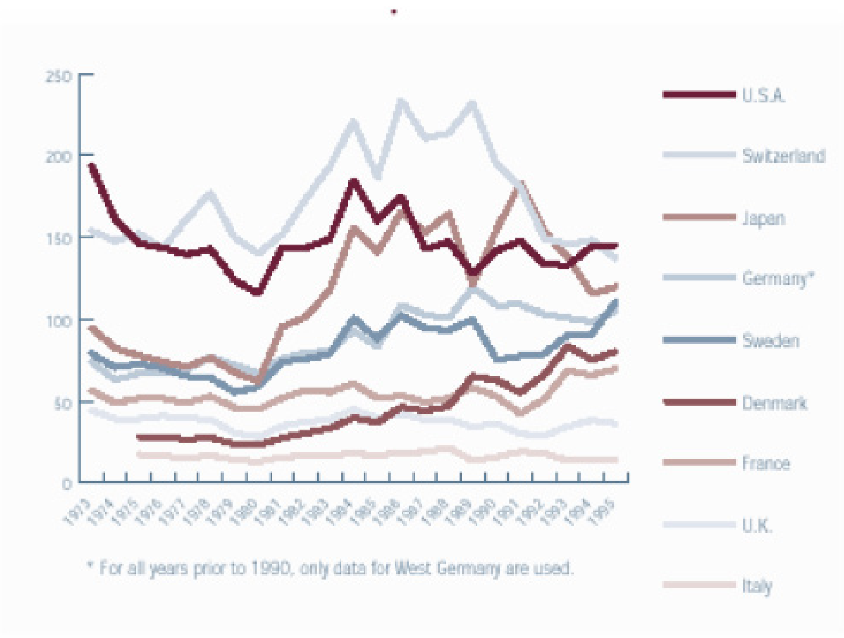
The Index uses statistical modelling to distinguish the relative importance of these contributors to national innovative capacity. Regression analysis is employed across a set of 17 OECD countries over a 25-year period from the 1970s through the mid-1990s to link these contributors to an internationally 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 7.8: Historical National Innovation Capacity Index for selected countries, 1973-1995.



7.5 *Information and communication technologies* (by J. Fagerberg)

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.

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, as mentioned in section 4 of this report, 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.

7.6 *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.

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.

Presentation of the Index

The presentation of the Technology Achievement Index is given in a table format.

7.7 *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 to enable comprehensive international comparisons and time-series analysis.

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.

Bibliography

Aczél, J. (1988) Cheaper by the Dozen : Twelve Functional Equations and their Applications to the "law of science" and to Measurement in Economics. In *Measurement in Economics: Theory and Applications of Economic Indices*, W. Eichorn (ed.), pp. 3–17. Heidelberg: Physica-Verlag.

Adriaanse, A. (1993) *Environmental policy performance indicators. A study on the development of indicators for environmental policy in the Netherlands.* The Hague: SDV Publisher.

Atkinson, A., Cantillon, B., Marlier, E. and Nolan, B. (2002a) *Social Indicators. The EU and Social Inclusion.* Oxford University Press.

Atkinson, T., Cantillon, B., Marlier, E. and Nolan, B. (2002b) Indicators for Social Inclusion. *Politica Economica* **18**, 7–28.

Babbie, E. (1995) *The Practice of Social Research.* Belmont: Wadsworth Publishing Company.

Booysen, F. (2002) An Overview and Evaluation of Composite Indices of Development. *Social Indicators Research* **59**, 115–151.

Brandolini, A. (2002) Education and Employment Indicators for the EU Social Agenda. *Politica Economica* **18**, 55–62.

Brundtland, B. (1991) *Environmental Challenges of the 1990s, The Tanner lecture on Human Values.* UK: Clare Hall. p.8.

CES (2000) Index of Sustainable and Economic Welfare.

Chakravarty, S. R. (2003) A Generalized Human Development Index. *Review of Development Economics* **7**, 99–114.

Cherchye, L. (2001) Using Data Envelopment Analysis to Assess Macroeconomics Policy Performance. *Applied Economics* **33**, 407–416.

Cherchye, L. and Kuosmanen, T. (2004) Benchmarking Sustainable Development: A Synthetic Meta-Index Approach. In *Perspectives on Inequality, Poverty and Human Well Being*, forthcoming in M. McGillivray (ed.). United Nations University Press.

Cherchye, L., Moesen, W. and Van Puyenbroeck, T. (2004) Legitimately diverse, yet comparable: On synthesising social inclusion performance in the EU. *Journal of Common Market Studies* **42**, 919–955.

COM.619 final (2001) Communication from the Commission, Structural Indicators.

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.* Luxembourg: OPOCE.

Cooper, W. W., Seiford, L. and Zhu, J. (2004) *Handbook on Data Envelopment Analysis.* Kluwer Academic Press.

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.

Desai, M. (1994) The Measurement Problem in Economics. *Scottish Journal of Political Economy* **41**, 34–42.

Despotis, D. (2004) A reassessment of the human development index via data envelopment analysis. *Journal of the Operational Research Society* pp. 1–12.

DG ECFIN (2000) Business Climate Indicator.

DG Enterprise (2001) Summary Innovation Index.

DG MARKT (2001) Internal Market Scoreboard.

Dowrick, S., Dunlop, Y. and Quiggins, J. (2003) Social Indicators and Comparisons of Living Standards. *Journal of Development Economics* **70**, 501–529.

Ebert, U. and Welsch, H. (2004) Meaningful Enviromental Indices: A Social Choice Approach. *Journal of Environmental Economics and Management* **47**, 270–283.

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

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

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

Foster, J. and Sen, A. (1997) *On Economic Inequality.* Second edition. Clarendon Press.

Freudenberg, M. (2003) Composite Indicators of Country Performance: a Critical Assessment. STI working paper 2003/16, OECD, Paris.

Giovannini, E. (2004) *Towards a common strategy of international organizations to improve the quality of international statistics.* United Nations Committee for the Coordination of Statistical Activities. SA/2004/6/Add.1.

Hopkins, M. (1991) Human Development Revisited: A New UNDP Report. *World Development* **19**, 1469–1473.

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.

JRC-OECD (2005) *Handbook of Good Practices in Composite Indicators Building*, forthcoming edition.

Kakwani, N. C. (1993) *Income Inequality and Policy Applications.* Oxford University Press.

Karlsson, J. (1998) *A systematic approach for prioritizing software requirements.* Ph.D. thesis, No.526, Lonkoping, Sverige.

Kuosmanen, T., Cherchye, L. and Spilainen, T. (2004) The law of one price in data enevelopment analysis: restricting weight flexibility across firms. In *European Journal of Operational Research*.

Lauer, J. A., Christopher, C. A. K. and Evans, D. B. (2004) World health system performance revisited: the impact of varying the relative importance of health system goals. Technical report, Health Services Research.

Lind, N. (2004) Values reflected in the Human Development Index. *Social Indicators Research* **66**, 283–293.

Maasoumi, E. and Nickelsburg, G. (1988) Multivariate Measures of Well-Being and an Analysis of Inequality in the Michigan Data. *Journal of Business and Economic Statistics* **6**, 327–334.

Mahlberg, B. and Obersteiner, M. (2001) Remeasuring the hdi by data envelopment analysis. Technical report, International Institute for Applied Systems Analysis Interim Report 01-069.

Manly, B. (1994) *Multivariate Statistical Methods. A Primer.* UK: Chapman & Hall.

McCarthy, M. (2001) Comments on "bechmarking european labour market performance with efficiency frontier technique". DG ECFIN.

Melyn, W. and Moesen, W. (1991) Towards a synthetic indicator of macroeconomic performance: Unequal weighting when limited information is available. Public Economics Research paper 17, CES, KU Leuven.

Micklewright, J. (1991) Should the UK Government Measure Poverty and Social Exclusion with a Composite Index? In *CASE, Indicators of Progress: A Discussion of Approaches to Monitor the Government's Strategy to Tackle Poverty and Social Exclusion.* CASE Report 13, London School of Economics.

Moldan, B. and Billharz, S. (1997) *Sustainability indicators: Report of the Project on Indicators of Sustainable Development.* Scope 58. Wiley Pub.

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.

Munda, G. and Nardo, M. (2003) *On the Construction of Composite Indicators for Ranking Countries.* mimeo, Universitat Autonoma de Barcelona.

National Institute of Science and Technology Policy (1995) Science and technology indicators. Technical report, NISTEP Report No. 37, Japan.

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

Noorbakhsh, F. (1988) A Modified Human Development Index. *World Development* **25**, 517–528.

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.

Osberg, L. (2004) The Relevance of Objective Indicators of Well-Being. Paper prepared for the Annual Meeting of the Canadian Economics Association.

Panigrahi, R. and Sivramkrishna, S. (2002) An Adjusted Human Development Index: Robust Country Rankings with Respect to the Choice of Fixed Maximum Indicator Values. *Journal of Human Development* **3(2)**, 301–311.

Parker, J. (1991) *Environmental reporting and environmental indices.* Ph.D. thesis, Cambridge, United Kingdom.

Porter, M. and Stern, S. (1999) *The new challenge to America's prosperity: Findings from the Innovation Index.* Council of Competitiveness, Washington D.C.

Pré Consultants (2000) The Eco-indicator 99. A damage oriented method for life cycle impact assessment.

Puolamaa, M., Kaplas, M. and Reinikainen, T. (1976) *Index of Environmental Friendliness. A methodological study.* Eurostat.

Roberts, F. S. (1979) *Measurement Theory.* London: Addison-Wesley.

Rosen, R. (1991) *Life Itself - A Comprehensive Inquiry into Nature, Origin, and Fabrication of Life.* Columbia University Press.

Saaty, R. (1987) The analytic hierarchy process - what it is and how it is used. *Mathematical Modelling* **9**, 161–176.

Saaty, T. (1980) *The Analytic Hierarchy Process.* N.Y.: McGraw - Hill Book Co.

Saisana, M., Saltelli, A. and Tarantola, S. (2005) Uncertainty and Sensitivity Analysis as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society Series A* **168**, 1–17.

Saisana, M. and Tarantola, S. (2002) State-of-the-art Report on Current Methodologies and Practices for Composite Indicator Development. EUR Report 20408 EN, Institute for the Protection and Security of the Citizen, JRC Ispra, Italy, 72p.

Saltelli, A., Tarantola, S., Campolongo, F. and Ratto, M. (2004) *Sensitivity Analysis in practice, a guide to assessing scientific models.* New York: John Wiley & Sons. A software for sensitivity analysis is available at `http://www.jrc.cec.eu.int/uasa/prj-sa-soft.asp`.

Salzman, J. (2003) Methodological Choices Encountered in the Construction of Composite Indices of Economic and Social Well-Being. mimeo, Center for the Study of Living Standards, Ontario.

Selvanathan, E. A. and Prasada Rao, D. S. (1994) *Index Numbers: A Stochastic Approach.* University of Michigan Press, Ann Arbor.

SIMLAB software 1.1 (1999) *Simulation environment for uncertainty and sensitivity analysis.* POLIS ScaRL, Milan.

Slottje, D. (1991) Measuring the quality of life across countries. *Review of Economics and Statistics* **73**, 684–693.

Storrie, D. and Bjurek, H. (1999) Benchmarking the basic performance indicators using efficiency frontier techniques. In *Report Presented to the European Commission, Employment and Social Affairs DG.*

Storrie, D. and Bjurek, H. (2000) Benchmarking European labour market performance with efficiency frontier techniques. Discussion paper FS I 00-211.

Tarantola, S., Liska, R., Saltelli, A., Leapman, N. and Grant, C. (2004) The internal market index 2004. European Commission: Joint Research Centre.

Tarantola, S., Saisana, M. and Saltelli, A. (2002) Internal Market Index 2002: Technical Details of the Methodology. European Commission: Joint Research Centre.

UN Development Programme (1992) *Human Development Report.* Oxford University Press, UK, p.21.

United Nations (2001) *Human Development Report.* Oxford University Press, UK. `http://www.undp.org`.

Veenhoven, R. (1996) Happy Life Expectancy: A comprehensive measure of quality-of-life in nations. *Social Indicators Research* **39**, 1–58.

von Winterfeld, D. and Ward, E. (1986) *Decision analysis and Behavioral Research.* Cambridge U. press.

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.

WHO (2000) Overall Health System attainment.

Woodward, R. T. and Bishop, R. C. (1997) How to Decide when Experts Disagree: Uncertainty - Based Choice Rules in Environmental Policy. *Land Economics* **73**, 492–507.

World Economic Forum (2001) Environmental sustainability index. http://www.ciesin_org/indicators/ESI/index.html.

Zaim, O., Färe, R. and Grosskopf, S. (2001) An Economic Approach to Achievement and Improvement Indexes. *Social Indicators Research* **56**, 91–118.