**The R&D Indicators in the Knowledge-Based Economy: The Research Paradox**

Asterios Chatziparadeisis

**Abstract**

Out of all R&D indicators, the older yet more preferred one is the Gross Domestic Expenditure on R&D (GERD) most of the times expressed by the “r” indicator, i.e. the percentage of GERD on Gross Domestic Product (GDP). In this paper we prove that the capacity of the GERD or r indicator to convey in a reliable way the actual evolution of the R&D sector and of the economy in general, is limited. Indeed, a variety of contradictions have been observed between the aforementioned indicators, which we will henceforth refer to as the “Research Paradox”. The paper makes recommendations for improving the selection and usage of existing R&D indicators as well as proposes alternative measurements while in the Knowledge-Based Economy (KBE).

**Measuring Research & Development**

The aim of Scientific Research is the generation of new scientific knowledge as well as the correction and integration of previous knowledge, either immediately applicable or not. The roots of scientific research can be traced as far back as the ancient times, boomed during the renaissance period but the massive and systematic engagement in Research and Development (R&D) activities, not only in Academia but also in Enterprises, is only a recent phenomenon of the 20th century. The importance of R&D in the 20th century resulted to a respective need for measurements and indicators

During the 1930s the first measurements of R&D indicators took place in the USA to account for the need of managing industrial laboratories and the respective planning of scientific activities. Canada followed one decade later, while the UK took some additional ten years to carry out such activities. The massive financing of R&D by national governments and enterprises of other developed countries after World War II, generated the need for measurements of similar indicators in those countries as well.

During the early 1960s, OECD undertook a leading role and coordinated the work of measuring R&D performance mainly by developing the first methodological manual for the collection of R&D data, which is known as the Frascati Manual. The Frascati Manual, the first edition of which was issued in the beginning of the ‘60s, was focusing mainly on two input indicators of R&D investments: the Financial Resources spent for R&D activities and the Human Resources invested in R&D activities. The measurements covered four sectors: the sector of Government, which included the

---

1 Asterios Chatziparadeisis, Ministry of Development, Greece, (asteri1@hol.gr)
3 La mesure des activités scientifiques et techniques. Méthode type proposée pour les enquêtes sur la recherche et le développement expérimental “Manuel de Frascati” OECD 1980.
government research laboratories, the sector of Enterprises, the sector of Higher Education and the sector of Private Non-Profit Organizations. The measurements were focused on institutions rather than individuals.

During late '80s and early '90s additional manuals were created by OECD and EUROSTAT on impact indicators, e.g. the Technology Balance of Payments\(^4\) (TBP), on output indicators, e.g. the Oslo Manual for Innovation\(^5\), as well as on a broader spectrum of subjects such as the Canberra Manual for the Human Resources for Science and Technology\(^6\).

Out of all R&D indicators, the older yet more preferred one is the GERD and the \(r\) indicator\(^7\). Since the '60s when the collection of R&D data started in the OECD countries, this organization has developed various analyses and studies based on GERD or \(r\). The countries were grouped in categories such as “big”, “intermediate” and “small”, according to the size of their GERD, or “high-”, “intermediate-” and “low-R&D intensity” according to the \(r\) indicator.

The use of these indicators was diffused rapidly. Through several analyses and categorizations, the positive relation between GERD and economic growth of the countries was supported, e.g. in the USA with an \(r\) at the level of 3% since the beginning of the ‘60s, was considered to be a good example that others should follow.

The importance attributed to GERD continues to be very high even today. In the EU this indicator constitutes one of the central indicators of the Lisbon Strategy; one of the main objectives being the rise of this indicator across EU members from an average of 1,9% in 2000 to 3% of the EU GDP up to 2010.

Furthermore, in a relatively recent survey of the OECD\(^8\) it was presented that GERD and \(r\) continues to be the most popular indicators. More than 80% of the countries value it very high in their preference list, while other indicators like Patents, the Technological Balance of Payments, and the Trade of High Technology are preferred by less than 50%.

**The Research Paradox**

The ability of the GERD or the \(r\) indicators to express in a reliable way the actual growth of R&D and of the economy in general, is in our opinion limited. A variety of contradictions have been observed in the application of the aforementioned indicators, which we will henceforth refer to as “Research Paradox”.

In Figure 1, the evolution of the \(r\) indicator is presented for the total of the OECD countries during the time period 1981 - 2001\(^9\). The histogram reflects what seems to be a remarkable stagnation, which does not keep up with the revolutionary developments that happened during this period. Indeed, from 1,93% of GDP in 1981 it rose to just 2,28% in 2001, and not even in a linear way but by exhibiting fluctuation during that period.

---

\(^4\) Technology Balance of Payments “TBP” Manual, OECD


\(^8\) OECD (1998), How to improve the MSTI: First Suggestions From Users, DSTI / EAS / STP / NESTI / RD (98)9

\(^9\) OECD, Main Science and Technology Indicators
If we examine the EU-15 or the USA independently, during the same period, the picture is not very different. The indicator rose from 1.67% to 1.92% in the EU-15 and from 2.34% to 2.73% in the USA. In the USA the same indicator was at the level of 2.6% in 1959 and at 2.8% in 1962. Thus USA’s $r$ indicator has been hovering at the same level during the last four decades.

The stagnation of the $r$ indicator shows that the growth of GERD in absolute values is approximately at the levels of the growth of the GDP for this period i.e. at the level of 3-5% annually. By examining the above figures, one would say that the promotion of R&D and the production of knowledge didn’t constitute an important priority of national governments and enterprises during the last 20 years, at least not at the level for it to increase at a higher rate than GDP.

If we study other important input indicators for the same period, like the number of researchers per 1,000 workers, we observe that it presents almost the same stagnation (Figure 2): the number of researchers per 1000 workers rose from 4.5/1000 researchers in 1981 to 6.5/1000 researchers in 2001.

On the other hand, if we study output indicators, the picture is totally different: the number of new products increased heavily, the life-cycle of the products shrank intensely, technologies are getting obsolete rapidly, etc. Unfortunately there is no data available on the number of innovations in order to quantify the rapidly rise of this indicator. The number of patents constitutes a proxy of production of innovations as only a part of innovations is patented. Still, despite the partial coverage of innovation by the patents indicator, this indicator presents a growth obviously bigger than that of input indicators. For example, the number of patents registered in the European Patent Office (EPO) by the OECD countries increased by 272% during 1981 - 2001, while the number of patents registered in the US Patent Office (USPTO) increased by 157%. In the knowledge intensity sectors the rise is even
bigger. Furthermore, in the same period the patents in the sectors of Information and Communication Technologies (ICTs) presented a rise of 470% and in Biotechnology 805%.

**The Old (Industrial) and the New (Knowledge) Economy**

In order to interpret this “Research Paradox” (i.e. stagnation in R&D indicators compared to an explosion in the production of knowledge and innovations) we need to look deeper into the changes that came about with the New Knowledge Economy. The main characteristics of the “old” Industrial Economy are the following: Almost the entire generation of knowledge took place in Research Laboratories (RLs), those had almost exclusive access to the knowledge infrastructure (libraries, laboratories, other supervisory instruments etc.) while the rest of the human capital was relatively alienated from this, they were also staffed with human capital that had the required skills and capacities for scientific developments (e.g. people with postgraduate titles and doctorates), and continuous learning processes were implemented. In the enterprises, on the other hand, human capital were in high percentages either semiskilled or even unskilled workers. Furthermore, the life cycle of the products was comparatively big, the Economy was supply-driven and innovation was periodical and linear.\(^\text{10}\)

The catalyst that caused the rapid changes in research but also in the entire economy was the very rapid development and diffusion of the Information and Communication Technologies (ICTs). The heavy rise of the processing power of the computers (doubling every 18 months – Moore’s Law) contributed a lot to the improvement of the quality of the performed R&D. The penetration of the Internet promoted the exchange of ideas between research teams in different geographic areas, thus improving further the quality of R&D through collaborations. Still the big revolution was the implementation and wide-scale adoption of the World Wide Web (WWW).

The creation and very rapid spread of the Web were the catalyst that crumbled the walls and permitted the diffusion of the information and the knowledge to the whole population. Through the Web an enormous volume of data, information and knowledge became available to everyone accessing the Internet. As such, the access to information and knowledge has evolved from being a “privilege” to being a “right”, of everyone instead of only people working in research departments. Each worker, each citizen having a PC connected to the Web, has access to a volume of knowledge inconceivable in the previous decade(s). Therefore each scientist has the possibility to reorganize and treat the existing knowledge and produce new knowledge. In the New Economy each and every scientist can be a potential researcher and inventor.

Briefly the following are other related changes which happened in the New Knowledge-based Economy:

- in big enterprises the production of knowledge does not emanate exclusively from the R&D departments and the researchers but, depending on the sector and the firm, an important part of the scientific staff of the firm participates in this. The

---

\(^{10}\) Anthony Arundel, Wendy Hansen, Rene Kemp, State of the Art of the Knowledge-Based Economy, report prepared for Knowledge Economy Indicators: Development of Innovative and Reliable Indicator Systems (ref. WP1.1, KEI project), DG Research, 2006.
departments of production, marketing, etc. are particularly active in the
generation of information and knowledge for new products. For the organizational
innovations, a significant role is assumed not only by the product departments but
also by the general administration of the firm.

- the rapid development of the technology intensified the training processes of the
personnel of enterprises and created learning mechanisms for the majority -if not
for all- of the staff. Many scientists name the current economy as “Learning
Economy”.

- the education level of the employed human capital is nowadays considerably
higher. The working personnel of enterprises comprise of many more workers
with University, post-university and doctorate degrees, compared to the past.
This personnel is familiar with research and generation of new knowledge, since
it has invested significant time in such activities, and it is capable of generating
new knowledge, despite working outside R&D departments.

- the innovation model has changed from a “linear” to an “interactive” one. The
development of new products in the modern enterprise is a complicated process
consisting of a lot of stages in which many departments of the firm participate by
contributing knowledge and information\(^\text{11}\).

- the goods produced by enterprises become increasingly “individualised” and
oriented towards the needs of the customer. In the previous model, the supply
determined to a large extent the demand and the products had a big degree of
standardisation determined generally by the R&D department. The new model is
customer-driven.

- the life cycle of the products has become very short.

- innovation becomes the basic objective of all enterprises, not only of the
manufacturing ones but also of those in the services sector. There is
consequently production of knowledge in the services enterprises, which is
inadequately measured or not measured at all.

- new forms of R&D are carried out such as the “distributed research” (research in
network) which is hardly measured in its wholeness by the surveys

\begin{center}
\textbf{What to measure in the Knowledge-Based Economy}
\end{center}

It can be therefore entailed that the existing methodology for the measurement of
R&D captures only a small part of the activity and consequently the indicators are
significantly underestimated. It is believed that the knowledge-generating people are
many more than the counted since the existing measurement system records as
such only those officially declared as “researchers”.

This discrepancy between the values of R&D indicators and reality should be
investigated in the changes that took place in the research system and in the
weakness of the measurement system to adapt and measure adequately the new
situation.

The methodology of the Frascati Manual for the measurement of R&D was created
in the beginning of the ’60s and was influenced by work previously done in the USA,
as adapted in the conditions of that period. It is well-known that it is difficult to come

York, Spon Press
up with an all-inclusive and pin-pointed definition of research and separating the R&D activities from other scientific and technical activities is far from being an easy task. For the data-collecting surveys on R&D, it is mandatory to clearly define a unit that will ensure the continuity and cohesion of the measurements.

In the decades of the '60s, '70s and '80s, new knowledge was produced mainly in Research Laboratories belonging to a University, Research Center or Private Enterprise. Consequently the focus of measurements in these institutions had given a satisfactory measure of the countries’ effort for the production of new knowledge. The R&D outside the system was not included in the measurements since it was not important in volume.

This methodology significantly underestimated for example, the R&D performed by the Small and Medium-sized Enterprises (SME's). The R&D activities of the SMEs are generally mixed up with other activities of the enterprise since the SMEs very rarely have personnel exclusively dedicated to the R&D activities. As it is not easy to isolate the total research effort of SMEs, there is an underestimation of the indicators. Kleinknecht\(^{12}\) conducted an assessment of the R&D measurements produced by the official surveys for the enterprises. He carried out a survey and found out that the number of man-months of R&D in SMEs were four times bigger than the one reported in the official statistics. The underestimation of R&D in SMEs in official surveys was globally about at the level of 33%. This underestimation should be even bigger today since the personnel of SMEs is much more active in innovation.

All this scattered production of knowledge, which is very difficult to be measured with satisfactory precision, has very serious impact on the economy. It is used in the processes of decision-making, problem-solving, creation of new products, processes, services etc. A great part of this new knowledge is coded and presented in various web sites or portals and contributes to further diffusion and production of knowledge.

If we consider the R&D conducted by the research laboratories as "formal" and the R&D outside the laboratories as "informal", it becomes obvious that the volume of informal R&D has increased rapidly due to the reasons that were mentioned above. Although effort is being put in the last edition of the Frascati Manual\(^{13}\) to measure the informal R&D, this is still being measured inadequately and is thus the reason why we continue to observe this stagnation of the indicators. Several knowledge-intensive enterprises such as Microsoft and Nokia, have adapted their measurements to the new approach, and their indicators reflect now more accurately the generation of knowledge. For this reason, they exhibit very high percentages for R&D expenditure and personnel compared to other companies, e.g. Nokia Finland states that research personnel accounts for the 1/3 of its total personnel.

If we want to correct to some extent the Research Paradox for the examined period 1981-2001 we should measure and add the "informal" R&D as well. There are also other adjustments that could be applied on R&D expenditure in order to bring it closer to reality. These adjustments could be the use of e.g. “hedonic prices”, since a lot of goods used in R&D (e.g. computers, electronic equipment, communications,


\(^{13}\) Frascati Manual, Proposed Standard Practice for Surveys on Research and Experimental Development. OECD 2002
new materials etc) present enormous improvement in their performance while their prices remain at the same levels or are even decreasing.

Still, these adjustments would not resolve completely the problem which seems to be of structural nature.

The measurement of R&D, as it is described in the methodology of the Frascati Manual, is linked to material economy and particularly to Manufacturing. For that reason the Manufacturing sectors have concentrated the bigger percentage of the measured R&D. We have to have in mind that, before WWII, services were considered as a sector that did not contribute to the development of a country. During 1930s when the GDP indicator and its application was refined in the USA, it was even assumed that services should not be included in the GDP because they do not produce wealth! According to this point of view it was expected that the measurements of R&D, which began around the same period in the USA, included only the Manufacturing sector. As stated also in the last edition of the FM, “The basic definitions in this Manual were originally developed for manufacturing industry and research in the natural sciences and engineering”. In the last editions of the Manual an effort has been put to cover also activities such as software development, certain branches of services, social sciences etc. but the results are not encouraging since the way of measurement continues to be the research laboratory which is not typical of the firms in the services sector.

The same phenomenon is also observed in the measurements of Innovation. The methodology of the Community Innovation Surveys (CIS) seems to work well for manufacturing but not for the services 14.

It is already realizable that the Manufacturing sector shrinks and we are moving rapidly towards a services economy. Consequently the measurements of R&D will continue to present more and more ambiguities and underestimations.

In the Knowledge Based Economy the most important good is Knowledge. Therefore the effort should be oriented towards the measurement and the management of knowledge.

The high importance attributed to knowledge is being verified also by the fact that enterprises are moving to the direction of organizing seminars on the management of knowledge rather than on R&D or technology. International organizations (OECD, EU etc), universities, research institutions etc organize seminars and conferences on the comprehension of mechanisms and processes that affect the creation, organization, interaction of knowledge with the ICTs, the entrepreneurship, the economy and the society. Recent examples are the organization by OECD, EU and NSF of an international conference on “Advancing Knowledge and the Knowledge Economy”, the Conference “Knowledge Economy – Challenges for Measurement” by Eurostat etc.

During the last decade OECD organized important events covering the subjects of policymaking and the measurements as well. Since 1996 in the frame of the “Blue Sky Project” a “Conference on New Indicators for the Knowledge - Based Economy” was organized. Working Groups were created which produced several papers and reports. Many results were incorporated in a series of publications on policies called “Science Technology Industry Scoreboard”.

The approach of OECD produced useful results and concepts but from the “Scoreboard” publications one can conclude that the KBE was used simply as a label. Under the same umbrella a series of well known indicators are gathered which are either already measured for decades or are variants of old ones and are nowadays included as KBE indicators\textsuperscript{15}.

Knowledge is a non-material and hardly definable concept, therefore its measurement is obviously a difficult task. F. Gault states that «Measuring knowledge itself is more challenging, if not impossible. There is no unit of knowledge that corresponds to a currency unit in the System of National Accounts and there is nothing comparable to concepts of current and constant currency units which support comparisons of the economic system over time. There is also nothing comparable to purchasing power parities (PPP) that support comparisons across space».

The above shows how difficult it is to work under the narrow viewpoint of financial measurements. KBE should however be considered with broad mind and we not be limited to the measurements in some sectors which are directly related to the KBE; instead, we have to adopt innovative approaches.

Obviously it is very challenging to formulate definitive proposals and solutions for the measurements and the indicators of the KBE. The following are some initial thoughts on the direction to which further work could be oriented:

A new \textit{Manual for Knowledge} is needed. The production of knowledge, its diffusion, transfer, acquisition, absorption-assimilation, its use in generating added value, are all important parameters that should be measured.

The fundamental task here is the definition of the concept of knowledge as accurately as possible. A second task closely related to the first one is its discrimination in categories. Nowadays, for reasons of tradition and facility, the measurements for the new knowledge are very restrictive. The R&D expenditure is considered as the expenditure for the production of new knowledge while the articles published in reviews used by Science Citation Index (SCI) are considered as outputs.

Undoubtedly these two indicators represent the production of core knowledge. The insufficiencies in the measurements of R&D were already reported above. Regarding publications, the articles in reviews of the SCI should not be considered as the unique source of knowledge; on the contrary, working papers, theses for post-graduate studies, articles in technical reviews, all contribute substantially in the production of new knowledge. This contribution is of different type but equally useful in the generation of new products and services. Thousands of articles and documents are uploaded every day in the Web. Although they lack the high quality of those published in scientific reviews they still provide precious knowledge to scientists, company executives, and everyone else all over the world.

An accurate and functional categorization of knowledge is a basic condition for the comprehension of KBE. The existing categorizations of knowledge as “codified” and “tacit”, productive and theoretical etc. can be a starting point for the creation of new concepts and categories. The measurement of codified knowledge, which spreads with outstanding intensity, could give a measure of the advantages that the KBE provides to the scientist and the executive of the enterprise.

\textsuperscript{15} F. Gault - Measuring Knowledge and its Economic Effects: the Role of Official Statistics, in Conference Advancing Knowledge and the Knowledge Economy Conference, January 2005
The Flows of knowledge are a very important parameter. Measurements so far are limited to the easily recognizable cases: the co-publication of articles, the purchasing by the enterprises of R&D services produced by the Universities, the purchasing of patents etc. Yet, important flowing of knowledge exists also in the collaborations between research institutions and enterprises without purchasing of R&D services as is the case of e.g. the EU Framework Programs or the national programs for the promotion of R&D.

Unprecedented flows of knowledge are created through Internet and the World Wide Web. Unfortunately statistics for the Internet, besides the elementary ones, actually do not exist and this is a very big disadvantage. KBE is the “child” of the ICTs, the Internet and the World Wide Web and the lack of measurements for the Internet is a huge obstacle for the comprehension of the KBE.

The Absorption and assimilation of knowledge from human capital is another important parameter. The indicators of formal education give a picture which is quite different from reality. As an example, the indicators on educational degrees reflect the situation at the moment of the acquisition of the degree. The deeper this moment lies in the past the more dramatically this snapshot changes whatsoever. The histogram of specializations as it is given by the surveys on diplomas of human potential is very different than the one that corresponds to the real employment of this potential.

The Training of the human resources, which leads to new specializations and skills, plays a very important role. The existing indicators for the training are few and very general (number of individuals that had a training, hours of training etc). Without an official system of certified training from which specific qualifications and skills are “produced” it is difficult to measure the actual level and the specialization of the human potential of an enterprise or country at a given moment.

The Utilisation of knowledge is one of the most important indicators. Innovation lies at the heart of the KBE and constitutes the most important economic output. The surveys on innovation began as measurements of the output of the R&D but then changed to measurements of activities. B. Godin’s statement is eloquent 16 “The recent internationalization of innovation surveys was characterized by a conceptual shift from outputs in the 1979s to activities in the 1990s. Without really noticing that they had departed from their original goal, national governments and the OECD ended up measuring innovation the way they measured R&D, i.e.: in term of inputs and activities”. The surveys on innovation should therefore become again measurement of outputs.

The spread of innovation beyond the technological, market and organizational level that has been included in the last revision of the Oslo Manual is towards the right direction. The surveys should cover not only the business sector but also the rest of the sectors of the economic activity.

Finally innovation should be more connected to the production, assimilation and use of knowledge than to R&D. From this point of view the management of knowledge should be studied in relation with the promotion of innovation and knowledge inside the firm or the organization. An initial effort 17 has been put into this but more should be done in a more innovative way.

---

17 Knowledge Management, Measuring Knowledge Managements in the Business Sector, First Steps. OECD and Minister of Industry, Canada 2003
Conclusions

The description and measurement of the New Economic Paradigm is not possible using the concepts, tools and measurements of the old one.

The shifting of the center of gravity from the Manufacturing to the Services sectors, from the Material Economy to the Immaterial and the Knowledge Based Economy, stipulate also the need of upgrading our measurements and indicators: R&D indicators were the appropriate measurement for Manufacturing but hardly anymore for the KBE. Instead, the measurements should be directed to the more important element of KBE: knowledge. The production, distribution, transfer, acquisition, absorption – assimilation, use etc. of knowledge are variables which should interest both the researchers and the policy makers.

KBE is a phenomenon produced by the development and diffusion of the ICTs, the Internet and the World Wide Web. These elements therefore lie in the core of the new system, through which the generation and distribution of the bigger part of information and knowledge is accomplished. As such, coming up with data and indicators for the Internet and the Web is an imperative step, not only for comprehending the revolutionary phenomena of the KBE but also for quantifying its effect and designing future roadmaps.

New tools are needed, which will not be a simple modification or adaptation of the old ones. The gathering of old, known indicators under a new umbrella does not constitute a solution and it does not promote the comprehension of the new system whatsoever. As it has been already stated, adoption of innovative approaches is needed.

KBE is a new reality, which is still in evolution and its characteristics are not yet crystallized. More economic research, pilot studies and surveys are needed as well in order to reveal the main characteristics of the New Economic Paradigm and help the policy makers in their work.
European Conference on Quality in Survey Statistics
(The R&D indicators in the KBE)

The Research Paradox

Asterios Chatziparadeisis
Dr. Statistician
Ministry of Development
GREECE
E-mail: asteri1@hol.gr
The history of the S&T indicators

• The measurement of the R&D activities started in USA the ’30s with:
  – The R&D expenditure
  – The R&D Personnel

• In Canada one decade later …

• In UK the ’50s

• In Europe the ’60s

The data for the OECD countries are comparable from 1981
The history of the S&T indicators the role of the OECD

- The OECD played a key role in the methodology for S&T indicators

- The Frascati Manual in the ’60s
  - Priority to the Input Indicators (R&D Expenditure, R&D Personnel)
  - Sectors of Measurement (Government, Higher Education, Business, PNP)

- Other Manuals produced later at the ’80s (TBP, HRST etc.). In ’90 Manual for Innovation
GERD (Expenditure on R&D): the most cherished indicator! (and “r”=GERD/GDP)

- OECD survey 1963-64: Ranking the countries according to GERD and “r”
- OECD (1984): grouping countries according to GERD in High, Medium, Low, Others
- “r” is key indicator of the Lisbon Strategy (the EU must achieve the 3%)
- “r” is used in a lot of publications for R&D
- In the survey of the OECD for the improvement the MSTI (1998) countries manifested a great preference for this indicator
The evolution of the “r” in OECD countries: Stagnation
“r” in EU countries: Stagnation

<table>
<thead>
<tr>
<th>Year</th>
<th>EU</th>
</tr>
</thead>
<tbody>
<tr>
<td>1981</td>
<td>1.50</td>
</tr>
<tr>
<td>1982</td>
<td>1.60</td>
</tr>
<tr>
<td>1983</td>
<td>1.70</td>
</tr>
<tr>
<td>1984</td>
<td>1.80</td>
</tr>
<tr>
<td>1985</td>
<td>1.90</td>
</tr>
<tr>
<td>1986</td>
<td>2.00</td>
</tr>
<tr>
<td>1987</td>
<td>2.10</td>
</tr>
<tr>
<td>1988</td>
<td>2.20</td>
</tr>
<tr>
<td>1989</td>
<td>2.30</td>
</tr>
<tr>
<td>1990</td>
<td>2.40</td>
</tr>
<tr>
<td>1991</td>
<td>2.50</td>
</tr>
<tr>
<td>1992</td>
<td>2.60</td>
</tr>
<tr>
<td>1993</td>
<td>2.70</td>
</tr>
<tr>
<td>1994</td>
<td>2.80</td>
</tr>
<tr>
<td>1995</td>
<td>2.90</td>
</tr>
<tr>
<td>1996</td>
<td>3.00</td>
</tr>
<tr>
<td>1997</td>
<td>3.10</td>
</tr>
<tr>
<td>1998</td>
<td>3.20</td>
</tr>
<tr>
<td>1999</td>
<td>3.30</td>
</tr>
<tr>
<td>2000</td>
<td>3.40</td>
</tr>
<tr>
<td>2001</td>
<td>3.50</td>
</tr>
</tbody>
</table>
“r” in USA: Stagnation

1959: \( r = 2.6 \)  1962: \( r = 2.8 \)
Researchers per 1000 workers: weak increase
The Research Paradox

- **Stagnation** in R&D versus emergence of the *Knowledge Based Economy*
- **Stagnation** in R&D vs explosion in the creation and the use of Knowledge
- **Stagnation** in the R&D vs *tremendous development* of innovations.
- **Stagnation** in R&D vs *important* scientific discoveries (Human Genome etc)
Explanation of the Paradox

An increasing part of the R&D is not measured by the indicators
The Old Industrial Economy (characteristics)

- Almost all the new knowledge is produced in the Research Laboratories (RL)
- The access to knowledge and the learning processes are limited to the RL
- Innovation is periodic and linear
- Key source of innovation is the R&D
- Workforce of the firm: high proportion of semi-skilled or unskilled people
- Economy: supplier driven
- The product lifecycle is long
The Knowledge Based Economy

• The ICTs (Information and Communication Technologies) were the catalyst for the creation of the KBE.

• The Computer, the Internet and the World Wide Web produced revolutionary changes for the whole economy

• The Web crumbled the walls and powered the knowledge in the whole economy and society. The access of knowledge is free to every scientist

• Every scientist is potentially a producer of new knowledge (Researcher)
The Knowledge Based Economy (characteristics)

- Learning processes are established in firms and the whole economy (Learning Economy)
- Innovation is continuous and systemic
- Several departments of the firm participate in the innovation process producing ideas, information and knowledge
- Workforce of the firm: high proportion of graduates with R&D experience
- Economy: customer driven
- The product lifecycle is short
The Measurements of R&D

- Frascati Manual (FM) is concentrated to measure the (formal) R&D of the Research Laboratories in 4 sectors (Business, Government, Higher Education, Private Non Profit).

- The (informal) R&D performed outside of RL is not measured.

- The informal R&D increases rapidly in the KBE. The measurement of the informal R&D is not an easy task.

- The orientation of the FM is to measure the R&D of the Manufacturing sector (RL). The production of knowledge is different in the Services sectors.

- The Manufacturing sector is shrinking but the Services is increasing.
What to Measure in the KBE

To measure **Knowledge**

- To be innovative in measuring knowledge
- To feel free from the existing concepts and measurements
- Not creating just an umbrella concept for known indicators but create new concepts and indicators
- Not to underestimate the difficulties to measure knowledge
- Create a New (Frascati) Manual for Knowledge
What to Measure in the KBE (examples)

• Clear **definition** of the concept of knowledge

• Create an operational **Classification** of categories of knowledge ex. basic vs applied knowledge, tacit vs codified knowledge etc.

• The **Production** of knowledge: not only limited to the (formal) R&D activities but include the informal

• The **Output** indicators: not only the articles of the SCI. Create several levels of publications: theses for diplomas, working papers, publication in technical reviews, in newspapers, in the Web etc.
What to Measure in the KBE (examples)

• The **Flows** of knowledge: not only the co-publication, co-patenting etc. but also the collaborations in R&D projects (i.e. the Framework Program of the EU), the strategic alliances, the ventures between firms etc.

The flows of knowledge by Internet are the most important and the most voluminous

Start a procedure for the statistics on Internet
What to Measure in the KBE (examples)

• The **management** of the knowledge.

  The acquiring, capturing, diffusion of knowledge inside the firm and generally in the organizations.

• The **absorption** of the knowledge.

  The human resources and their capacity to absorb new knowledge is crucial.

  The indicators on education give an initial picture but the training, the lifelong learning etc. are more important measures.
What to Measure in the KBE  
(examples)

• The use of new knowledge.
  – Innovation is the most important but not the only use of new knowledge. The measures on innovation aimed to measure outputs of R&D but turn out to measure activities. Include all kinds of innovation: product, process, services, organizational, entrepreneurial etc. measured in all sectors (Business, Gov, HE, PNP).
  – Measuring the other uses of new knowledge

    Decisions etc. etc.
What to Measure in the KBE
(examples)

The big modern enterprises give the example:

• Microsoft: a lot of activities outside the R&D departments are declared as R&D activities. In reality there are knowledge production activities.

• NOKIA: more than one third of its personnel in Finland is considered as R&D personnel (knowledge production personnel).
Conclusions

- A New Paradigm can’t be understood with the old concepts, measurements and indicators
- To be innovative in the creation of new ones
- Oriented more to results than to activities
- More to output than to input
- Research with pilot surveys are needed to apply the new concepts
Weighting and Aggregation for Composite Indicators

Giuseppe Munda* and Michela Nardo

Abstract: Composite indicators (or indexes) are used whenever a plurality of variables is needed for an evaluation exercise (e.g. for a macroeconomic dimension). They are widely used for benchmarking the mutual and relative progress of countries in a variety of policy domains such as industrial competitiveness, sustainable development, globalisation and innovation. The proliferation of the production of composite indicators by all the major international organizations is a clear symptom of their political importance and operational relevance in policy-making. As a consequence, improvements in the way these indicators are constructed and used seem to be a very important research issue from both the theoretical and operational points of view. This paper shows that a theoretical inconsistency exists between the real theoretical meaning of weights and the meaning that is generally attributed to them by the standard practice in constructing composite indicators; thus, a recursive important mistake is present in most of the empirical applications. Guidelines to solve this drawback are given.

Key Words: Index Numbers, Measurement Theory, Multi-Criteria Analysis

1 Giuseppe Munda, European Commission, Joint Research Centre, TP 361, 21020 Ispra (VA), Italy and Universitat Autonoma de Barcelona, Dept. of Economics and Economic History, Barcelona, Spain (giuseppe.munda@jrc.it) *Corresponding author

Michela Nardo, European Commission, Joint Research Centre, TP 361, 21020 Ispra (VA), Italy (michela.nardo@jrc.it)
1. Introduction

Composite indicators are very common in fields such as economic and business statistics (e.g., the OECD Composite of Leading Indicators) and are used in a variety of policy domains such as industrial competitiveness, sustainable development, quality of life assessment, globalisation, innovation or academic performance (see Cox and others 1992, Cribari-Neto et al 1999, Färe et al. 1994, Griliches 1990, Forni et al. 2001, Huggins 2003, Grupp and Mogee 2004, Lovell et al. 1995, Munda 2005, Nardo et al. 2005, Saisana and Tarantola 2002, and Wilson and Jones 2002, among others). The proliferation of these indicators is a clear symptom of their importance in policy-making, and operational relevance in economic statistics in general (see e.g. Granger, 2001). All the major international organizations such as OECD, the EU, the World Economic Forum or the IMF are producing composite indicators in a wide variety of fields (Nardo et al., 2005). A general objective of most of these indicators is the ranking of countries and their benchmarking according to some aggregated dimensions (see e.g. Cherchye, 2001, Kleinknecht 2002 and OECD, 2003). As a consequence, the improvement of the way these indicators are constructed and used seems to be a very important research issue from both theoretical and operational points of view. Our main objective here is to contribute to the improvement of the overall quality of composite indicators by looking at one of their technical weaknesses, that is, the consistency between the mathematical aggregation rule used for their construction and the meaning of weights. Along the paper, concepts coming from measurement theory, multi-criteria decision analysis and social choice are used.

2. Linear Aggregation Rules and Meaning of Weights

Although various functional forms for the underlying aggregation rules of a composite indicator have been developed in the literature (e.g. Diewert, 1976, Journal of Economic and Social Measurement, 2002), in the standard practice, a composite indicator, \( I \), can be considered a weighted linear aggregation rule applied to a set of variables (OECD, 2003, p. 5):

\[
I = \sum_{i=1}^{N} w_i x_i
\]

where \( x_i \) is a scale adjusted variable (e.g. GDP per capita) normalised between zero and one, and \( w_i \) a weight attached to \( x_i \), usually with \( \sum_{i=1}^{N} w_i = 1 \) and \( 0 \leq w_i \leq 1 \), \( i = 1, 2, ..., N \). In this framework, a crucial role is played by the concept of weight.

The common practice in attaching weights is well synthesised by a recent OECD document: “Greater weight should be given to components which are considered to be more significant in the context of the particular composite indicator” (OECD, 2003, p. 10). In the decision theory literature, this concept of weights is usually referred to as symmetrical importance, that is “… if we have two non-equal numbers to construct a vector in \( \mathbb{R}^2 \), then it is preferable to place the greatest number in the position corresponding to the most important criterion.” (Podinovskii, 1994, p. 241). Let’s try to put some light on this issue, by proving formally that the concept of
symmetrical importance is incompatible with a linear aggregation rule, given that in a linear aggregation rule, weights can only have the meaning of a trade-off ratio (see also Vincke, 1992, pp. 36-37).

Suppose that country $a$ is evaluated according to some variables $(x_1(a),...,x_n(a))$, then the substitution rate at $a$, of the variable $j$ with respect to the variable $r$ (taken as a reference variable) is the amount $S_{jr}(a)$ such that, country $b$ whose evaluations are: $x_j(a) = x_j(b), \forall l \neq j, r$; $x_j(b) = x_j(a) - 1$; and $x_r(b) = x_r(a) + S_{jr}(a)$ is indifferent to country $a$. Therefore, $S_{jr}(a)$ is the amount which must be added to the variable $r$ in order to compensate the loss of one unit on variable $j$ for country $a$. Consider now a composite indicator $I(x_1, x_2, ..., x_n)$ and suppose that the score of this indicator is the same for the two countries. Let $z(a) = (x_1(a), x_2(a), ..., x_n(a))$ and $z(b) = (x_1(b), x_2(b), ..., x_n(b))$, then as a first approximation one has:

$$0 = I(z_b) - I(z_a) = \sum_{i=1}^{n} \left( \frac{\partial I}{\partial x_i} \right) z_{a} (x_i(b) - x_i(a)) = -\left( \frac{\partial I}{\partial x_j} \right) z_{a} + S_{jr}(a) \left( \frac{\partial I}{\partial x_r} \right) z_{a}$$

and manipulating

$$S_{jr}(a) = \frac{\left( \frac{\partial I}{\partial x_j} \right) (z_a)}{\left( \frac{\partial I}{\partial x_r} \right) (z_a)}$$ \hspace{1cm} (2)

When the function $I$ is a weighted sum of all the normalised variables, i.e.

$$I(x_1, x_2, ..., x_n) = \sum_{i=1}^{n} w_i x_i$$ \hspace{1cm} (3)

from expression (2) one obtains:

$$S_{jr}(a) = \frac{w_j}{w_r} = \text{constant.} \hspace{1cm} (4)$$

This means that in the weighted linear aggregation, the substitution rates equal the weights of the variables up to a multiplicative coefficient. As a consequence, the estimation of weights is equivalent to that of substitution rates, implying a compensatory logic. Compensability refers to the existence of trade-offs, i.e. the possibility of offsetting a disadvantage on some variables by a sufficiently large advantage on another variable. Therefore, the use of weights in combination with intensity of preference (given that variables are always supposed to be measured on an interval or ratio scale) within a linear aggregation rule originates compensatory aggregation conventions and gives the meaning of trade-offs to the weights.

In other words, in a linear aggregation framework, the weights always depend on the value of the trade-off. Such a trade-off holds a constant value, since in this context, the local trade-off (i.e. the marginal rate of substitution) is also the global one, i.e. it does not depend on the values that variable scores may have in a given point. However, one has to note that the trade-off always depends on the
measurement scale used for measuring the variable scores and on the range that the measurements of variable scores may present. To clarify the issue consider the hypothetical example presented in Table 1.

<table>
<thead>
<tr>
<th>GDP (Millions of Euro)</th>
<th>Populations (Number of Inhabitants)</th>
<th>Percentage of Protected Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>A 32,000</td>
<td>1,000,000</td>
<td>60%</td>
</tr>
<tr>
<td>B 80,000</td>
<td>3,000,000</td>
<td>70%</td>
</tr>
<tr>
<td>C 100,000</td>
<td>5,000,000</td>
<td>40%</td>
</tr>
</tbody>
</table>

Table 1. Illustrative Example with Three Countries and Three Variables

Consider first the measurement scale. Suppose that in the construction of a sustainability composite indicator, the trade-off between protected species and GDP is set such that a decrease of 1 point in the percentage of protected species can be compensated by an increase of 100,000,000 Euro of GDP. This trade-off can be expressed as 

\[ \frac{w_{\text{species}}}{w_{\text{GDP}}} = 100,000,000 \]

If instead the measurement scales of GDP is changed and this variable is measured per capita, the same trade-off indicated above now would be modified e.g. in “1% of protected species less can be compensated by 100 Euro of GDP per capita more”. Thus in this case one has 

\[ \frac{w_{\text{species}}}{w_{\text{GDP}}} = 100 \]

Since the measurement scale of the variable protected species has not changed, the only weight that must change value is the one attached to GDP, that in the second case has to increase considerably (since the numerator remain constant and the value of the ratio decrease).

One obvious observation might be that in a composite indicator variables are normalized and thus effects due to measurement scales should disappear. This, however is not true. Consider for example the normalization technique distance from the group leader, which assigns 100 to the leading alternative and other alternatives are ranked as percentage points away from the leader (Saisana and Tarantola, 2002), that is 

\[ \frac{100}{\text{actual value}} \times \frac{\text{maximum value}}{100,000} \]

By applying this normalization technique while keeping the original trade-off “1% decrease in species versus 100 million Euro GDP” one has to standardize the value 100 Mill. Euros according to the new scale. This is equivalent to dividing this value by the score of the country with the highest GDP: 

\[ \frac{100}{101,000} = 10^{-3} \]

When income is expressed as GDP per capita, then the trade-off would now be “1% decrease in species versus \[ \frac{100}{32,000} = 3.125 \times 10^{-3} \]

Again trade-offs and corresponding weights must change according to the range of variation of the measurement scale considered. One may easily check that this kind of consequences apply independently to the normalization.

\[ ^2 \text{A is the country with the highest GDP per capita with 32,000 Euro, followed by B with 26,667 and C with 20,000 Euro.} \]
technique chosen. The conclusion is that in the case of a linear aggregation rule, trade-offs depend on the scales of measurement, and since weights are connected to the values of trade-offs they also depend on the scales of measurement.

Clearly trade-offs can be evaluated only if one knows the quantitative scores of the variables involved without any uncertainty. On the contrary, the concept of importance is connected to the variable itself and NOT with its quantification. If protected species are considered more, equal or less important than GDP, this is a quality of the variables which is independent from any measurement scale one may use. As clearly shown by Anderson and Zalinski (1988), when weights depend on the range of variable scores, such as in the context of a linear aggregation rule, the interpretation of weights as a measurement of the psychological concept of importance is always completely inappropriate.

More formally, to use the compensatory approach in practice, such as the linear aggregation rule, one has to determine for each individual indicator, a mapping \( \phi_i: x_i \rightarrow R \) which provides at least an interval scale of measurement and to assess scaling constants (i.e. weights) in order to specify how the compensability should be accomplished, given the scales \( \phi_i \) between the different individual indicators (Roberts, 1979). Note that the scaling constants which appear in the compensatory approach depend on the scales \( \phi_i \), thus they do not characterise the intrinsic relative importance of individual indicators. The implication is the existence of a theoretical inconsistency in the way weights are actually used and their real theoretical meaning.

An overview of methods to attach weights in a multi-attribute value function framework (the general framework to which the linear aggregation rule belongs) can be found in Beinat (1997) and Keeney and Raiffa (1976). There is unanimous agreement in the literature that the only method where weights are computed as scaling constants and there is no ambiguous interpretation is the so-called trade-off method starting with revealed preferences. No weight importance judgment is required in this method. The trade-off method can be briefly described as follows. Let's consider two countries \( A \) and \( B \), differing only for the scores of variables \( x_k \) and \( x_t \). The problem is then to adjust the score of say \( x_k \) for \( B \), in such a way that \( A \) and \( B \) become indifferent. Formally, it is:

\[
I(A) = I(B) \Leftrightarrow I(x_1, ..., x_k', ..., x_r, ..., x_n) = I(x_1, ..., x_k^*, ..., x_r^*, ..., x_n) \Rightarrow
\]

\[
\sum_{i=1}^{N} w_i x_i + w_k x_k' + w_t x_t' = \sum_{i=1}^{N} w_i x_i + w_k x_k^* + w_t x_t^* \Rightarrow
\]

\[
w_k x_k' + w_t x_t' = w_k x_k^* + w_t x_t^* \]

Equation (7) is an equation in the unknown \( w_k \) and \( w_t \). To compute the \( N \) weights as trade-offs, it is necessary to assess \( N-1 \) equivalence relations which together with the usual normalisation constraint \( w_1 + ... + w_N = 1 \) determine a linear system of \( N \) equations in the \( N \) unknown weights. Of course if some uncertainty on the variable scores exists, this method cannot be applied.

As one can easily understand to assess weights as trade-offs, as it should be always done when using a linear aggregation rule, it is a much harder job than to use weights as importance coefficients. This is probably the main reason why the
standard practice tends to use weights as importance coefficients, but unfortunately this practice is not defensible on theoretical grounds. Vansnick (1990) showed that the two main approaches in multi-criteria aggregation procedures i.e., the compensatory and non-compensatory ones can be directly derived from the seminal work of Borda and Condorcet. If one wants the weights to be interpreted as “importance coefficients” (or equivalently symmetrical importance of variables) non-compensatory aggregation procedures must be used (Bouyssou, 1986; Bouyssou and Vansnick, 1986). From a social choice point of view, these non-compensatory rules are always Condorcet consistent rules; their use in the framework of composite indicators, can be corroborated by referring to a clear result of social choice literature. The majority rule is theoretically the most desirable aggregation rule, but practically often produces undesirable intransitivities, thus “more limited ambitions are compulsory. The next highest ambition for an aggregation algorithm is to be Condorcet” (Arrow and Raynaud, 1986, p. 77).

Thus we can conclude that the use of non-compensatory aggregation rules to construct composite indicators is compulsory for reasons of theoretical consistency when weights with the meaning of importance coefficients are used. Moreover the use of Condorcet consistent rules is also desirable in general as advised by social choice literature. Unfortunately these considerations are completely neglected by the standard practice on composite indicators.

3. Conclusion

The following main conclusions can be drawn:

1. Weights in linear aggregation rules have always the meaning of trade-off ratio. Therefore, given that in the standard practice of constructions of a composite indicator, weights are used as importance coefficients in combination with linear aggregation rules, a theoretical inconsistency exists. This inconsistency applies to most of the empirical applications.

2. When a linear aggregation rule is used, the only method able to derive theoretically consistent weights is the so-called trade-off method. Operationally this method is very complex. Moreover the assumption that the variable scores are measured on an interval or ratio scale of measurement and no uncertainty exists must always apply. Rarely this happens in the practice of composite indicators, where for instance, sometimes quantitative scores are arbitrarily given to variable scores originally measured on an ordinal measurement scale (see e.g. Nicoletti et al., 2000).

3. In standard composite indicators based on the linear aggregation rule, compensability among the different individual indicators must always be assumed; this implies complete substitutability among the various components considered. For example, in a hypothetical sustainability index, economic growth can always substitute any environmental destruction or inside e.g., the environmental dimension, clean air can compensate for a loss of potable water. From a descriptive point of view, such a complete compensability is often not desirable.

4. Whenever weights are used with the meaning of importance coefficients, the aggregation algorithm must be a Condorcet consistent rule. The use of these

---

3The only exception probably being the 2005 Environmental Sustainability Index, where at least some methodological weaknesses are acknowledged. [http://www.yale.edu/esi/a_methodology.pdf](http://www.yale.edu/esi/a_methodology.pdf)
rules is desirable on more general grounds too. In particular, it should be noted that by using Condorcet aggregation rules no limitation on the measurement scale of the variable scores exists. The cost to pay is that information on the intensity of preference may be lost.

ACKNOWLEDGEMENTS: This research has been partially developed in the framework of the European Commission financially supported project “Knowledge Economy Indicators: Development of Innovative and Reliable Indicator Systems (KEI)” (KEI SCS8 2004-502529).

References


European Conference on Quality in Survey Statistics

2006
"Weighting and Aggregation for Composite Indicators: A Non-compensatory Approach"

Giuseppe Munda
DG JRC, G04 and
Universitat Autonoma de Barcelona
Structure of the presentation

• The axiomatic system of linear aggregation rules
• A methodological proposal: a multi-criteria aggregation convention for ranking countries
• Numerical examples
• The issue of quality
A typical composite indicator, I, is built as follows (OECD, 2003, p. 5):

\[ I = \sum_{i=1}^{N} w_i X_i, \]  

where \( X_i \) is a normalised variable and \( w_i \) a weight attached to \( X_i \), \[ \sum_{i=1}^{N} w_i = 1 \text{ and } 0 \leq w_i \leq 1, \ i = 1, 2, ..., N. \]

It is clear that from a mathematical point of view a composite indicator entails a weighted \textit{linear aggregation rule} applied to a set of variables.
Linear aggregation rules and preference independence

The variables $X_1, X_2, ..., X_n$ are *mutually preferentially independent* if every subset $Y$ of these variables is preferentially independent of its complementary set of evaluators.

The following theorem holds:

*given the variables $X_1, X_2, ..., X_n$, an additive aggregation function exists if and only if these variables are mutually preferentially independent.*
Preferential independence implies that the trade-off ratio between two variables is independent of the values of the n-2 other variables, i.e.

$$\frac{\partial S_{x,y}}{\partial z} = 0 \quad \forall x, y \in Y, \forall z \in Z$$

Preferential independence is a very strong condition from both the **operational** and **epistemological** points of view.
The meaning of weights in linear aggregation rules

• “Greater weight should be given to components which are considered to be more significant in the context of the particular composite indicator”. (OECD, 2003, p. 10).

• Weights as symmetrical importance, that is "... if we have two non-equal numbers to construct a vector in R2, then it is preferable to place the greatest number in the position corresponding to the most important criterion." (Podinovskii, 1994, p. 241).
To use the compensatory approach in practice, we have to determine for each normalised variable, a mapping \( \phi_i : X_i \rightarrow R \) which provides an *interval scale of measurement* and to assess *scaling constants* in order to specify how the compensability should be accomplished, given the scales \( \phi_i \) between the different variables. Note that the scaling constants appearing in the compensatory approach depend on the scales \( \phi_i \); thus they *do not characterise the intrinsic relative importance of the indicators* (Roberts, 1979).
Table 1. Illustrative Example with Three Countries and Three Variables

<table>
<thead>
<tr>
<th></th>
<th>GDP (Millions of Euro)</th>
<th>Populations Number of Inhabitants</th>
<th>Percentage of Protected Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>32,000</td>
<td>1,000,000</td>
<td>60%</td>
</tr>
<tr>
<td>B</td>
<td>80,000</td>
<td>3,000,000</td>
<td>70%</td>
</tr>
<tr>
<td>C</td>
<td>100,000</td>
<td>5,000,000</td>
<td>40%</td>
</tr>
</tbody>
</table>
Example

the trade-off between protected species and GDP is set such that a decrease of **1 point in the percentage of protected species** can be compensated by an increase of **100,000,000 Euro of GDP**. If instead the measurement scales of GDP is changed and this variable is measured per capita, the same trade-off indicated above now would be modified e.g. in “1% of protected species less can be compensated by **100 Euro of GDP** per capita more”. Since the measurement scale of the variable protected species has not changed, the only weight that must change value is the one attached to GDP, that in the second case has to increase considerably (since the numerator remain constant and the value of the ratio decrease).
Example

\[ \frac{W_{\text{species}}}{W_{GDP}} = 100,000,000 \]

\[ \frac{W_{\text{species}}}{W_{GDP}} = 100 \]
• **Weights** in linear aggregation rules have always the meaning of trade-off ratio. In all constructions of a composite indicator, weights are used as importance coefficients, as a consequence, *a theoretical inconsistency exists*.

• The assumption of *preference independence* is essential for the existence of a linear aggregation rule. Unfortunately, this assumption has very strong consequences which often *are not desirable* in a composite indicator.

• In standard composite indicators, *compensability* among the different individual indicators is always assumed; this implies complete substitutability among the various components considered. For example, in a sustainability index, economic growth can always substitute any environmental destruction or inside e.g., the environmental dimension, clean air can compensate for a loss of potable water. *From a descriptive point of view, such a complete compensability is often not desirable.*
• Both social choice literature and multi-criteria decision theory agree that whenever the majority rule can be operationalized, it should be applied. However, the majority rule often produces undesirable intransitivities, thus “more limited ambitions are compulsory. The next highest ambition for an aggregation algorithm is to be Condorcet” (Arrow and Raynaud, 1986, p. 77).
The axiomatic system

• **Axiom 1: Diversity.** Each individual indicator is a total order on the finite set $A$ of countries to be ranked, and there is no restriction on the individual indicators; they can be any total order on $A$.

• **Axiom 2: Symmetry.** Since individual indicators have incommensurable scales, the only preference information they provide is the existence of the pair-wise preferences they contain.

• **Axiom 3: Positive Responsiveness.** The degree of preference between two countries $a$ and $b$ is a strictly increasing function of the number and weights of individual indicators that rank $a$ before $b$. 
The axiomatic system

• **Neutrality:** it does not depend on the name of any country, all countries are equally treated.

• **Unanimity** (sometimes called Pareto Optimality): if all individual indicators prefer country a to country b than b should not be chosen.

• **Monotonicity:** if country a is chosen in any pair-wise comparison and only the individual indicator scores (i.e. the variables) of a are improved, then a should be still the winning country.

• **Reinforcement:** if the set A of countries is ranked by 2 subsets G1 and G2 of the individual indicator set G, such that the ranking is the same for both G1 and G2, then should still supply the same ranking. This general consistency requirement is very important in the framework of composite indicators, since one may wish to apply the individual indicators belonging to each single dimension first and then pool them in the general model.
## Sustainability Indicator

<table>
<thead>
<tr>
<th>Country</th>
<th>Indic.</th>
<th>GDP</th>
<th>Unemp. Rate</th>
<th>Solid wastes</th>
<th>Inc. disp.</th>
<th>Crime rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td>25,000</td>
<td>0.15</td>
<td>0.4</td>
<td>9.2</td>
<td>40</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>45,000</td>
<td>0.10</td>
<td>0.7</td>
<td>13.2</td>
<td>52</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>20,000</td>
<td>0.08</td>
<td>0.35</td>
<td>5.3</td>
<td>80</td>
</tr>
<tr>
<td>weights</td>
<td></td>
<td>0.165</td>
<td>0.165</td>
<td>0.333</td>
<td>0.165</td>
<td>0.165</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>0.666</td>
<td>0.333</td>
</tr>
<tr>
<td>B</td>
<td>0.333</td>
<td>0</td>
<td>0.333</td>
</tr>
<tr>
<td>C</td>
<td>0.666</td>
<td>0.666</td>
<td>0</td>
</tr>
</tbody>
</table>

- **ABC** = 0.666 + 0.333 + 0.333 = 1.333
- **BCA** = 0.333 + 0.666 + 0.333 = 1.333
- **CAB** = 0.666 + 0.666 + 0.666 = 2
- **ACB** = 0.333 + 0.666 + 0.666 = 1.666
- **BAC** = 0.333 + 0.333 + 0.333 = 1
- **CBA** = 0.666 + 0.333 + 0.666 = 1.666
The Computational problem

Moulin (1988, p. 312) clearly states that the Kemeny method is "the correct method" for ranking alternatives, and that the "only drawback of this aggregation method is the difficulty in computing it when the number of candidates grows".

One should note that the number of permutations can easily become unmanageable; for example when 10 alternatives are present, it is $10! = 3,628,800$. 
A NP-hard problem

• The complexity class of decision problems that are intrinsically harder than those that can be solved by a nondeterministic Turing machine in polynomial time. When a decision version of a combinatorial optimization problem is proved to belong to the class of NP-complete problems, then the optimization version is NP-hard.

• (definition given by the National Institute of Standards and Technology, http://www.nist.gov/dads/HTML/nphard.html )
• This NP-hardness has discouraged the development of algorithms searching for exact solutions, thus the majority of the algorithms which have been proposed in the literature; are mainly

• *heuristics based on artificial intelligence,*

• *branch and bound approaches and*

• *multi-stage techniques*

(see e.g., Barthelemy et al., 1989; Charon et al., 1997; Cohen et al., 1999; Davenport and Kalagnam, 2004; Dwork et al., 2001; Truchon, 1998b).
• A new numerical algorithm aimed at solving the computational problem connected to linear median orders by finding exact solutions has been proposed by Munda (2005). Main characteristics of this algorithm are that linear median orders are computed by using their theoretical equivalence with maximum likelihood rankings and that outranking matrixes are used as a starting computational step.
\[
\begin{align*}
    a_j P a_k & \iff g_m(a_j) > g_m(a_k) + p \\
    a_j Q a_k & \iff g_m(a_k) + p \geq g_m(a_j) > g_m(a_k) + q \\
    a_j I a_k & \iff \\
    & \begin{cases} 
        g_m(a_k) + q \geq g_m(a_j) \\
        g_m(a_j) + q \geq g_m(a_k)
    \end{cases}
\end{align*}
\]

Taking into account *intensity of preference*
<table>
<thead>
<tr>
<th>Criteria</th>
<th>Budapest</th>
<th>Moscow</th>
<th>Amsterdam</th>
<th>New York</th>
</tr>
</thead>
<tbody>
<tr>
<td>Houses owned (%)</td>
<td>50.5</td>
<td>40.2</td>
<td>2.2</td>
<td>10.3</td>
</tr>
<tr>
<td>Residential density (pers./hectare)</td>
<td>123.3</td>
<td>225.2</td>
<td>152.1</td>
<td>72</td>
</tr>
<tr>
<td>Use of private car (%)</td>
<td>31.1</td>
<td>10</td>
<td>60</td>
<td>32.5</td>
</tr>
<tr>
<td>Mean travel time to work (minutes)</td>
<td>40</td>
<td>62</td>
<td>22</td>
<td>36.5</td>
</tr>
<tr>
<td>Solid waste generated per capita (t/year)</td>
<td>0.2</td>
<td>0.29</td>
<td>0.4</td>
<td>0.61</td>
</tr>
<tr>
<td>City product per person (US$/year)</td>
<td>4750</td>
<td>5100</td>
<td>28251</td>
<td>30952</td>
</tr>
<tr>
<td>Income disparity (Q5/Q1)</td>
<td>9.19</td>
<td>7.61</td>
<td>5.25</td>
<td>14.81</td>
</tr>
<tr>
<td>Households below poverty line (%)</td>
<td>36.6</td>
<td>15</td>
<td>20.5</td>
<td>16.3</td>
</tr>
<tr>
<td>Crime rate per 1000 (theft)</td>
<td>39.4</td>
<td>4.3</td>
<td>144.05</td>
<td>56.7</td>
</tr>
</tbody>
</table>
Normalisation technique used for the different measurement units dealt with.

Scale adjustment used, for example population or GDP of each country considered.

Common measurement unit used (money, energy, space and so on).
### Normalized Impact Matrix

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>78.674</td>
<td>0</td>
<td>16.770</td>
</tr>
<tr>
<td>33.485</td>
<td>100</td>
<td>52.28</td>
<td>0</td>
</tr>
<tr>
<td>42.2</td>
<td>0</td>
<td>100</td>
<td>45</td>
</tr>
<tr>
<td>45</td>
<td>100</td>
<td>0</td>
<td>36.25</td>
</tr>
<tr>
<td>0</td>
<td>21.95</td>
<td>48.78</td>
<td>100</td>
</tr>
<tr>
<td>0</td>
<td>1.335</td>
<td>89.691</td>
<td>100</td>
</tr>
<tr>
<td>41.213</td>
<td>24.686</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>25.462</td>
<td>6.018</td>
</tr>
<tr>
<td>25.116</td>
<td>0</td>
<td>100</td>
<td>37.495</td>
</tr>
<tr>
<td></td>
<td>Impact 1</td>
<td>Impact 2</td>
<td>Impact 3</td>
</tr>
<tr>
<td>----------------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>Impact 1</td>
<td>100</td>
<td>78.674</td>
<td>0</td>
</tr>
<tr>
<td>Impact 2</td>
<td>66.515</td>
<td>0</td>
<td>47.72</td>
</tr>
<tr>
<td>Impact 3</td>
<td>57.8</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Impact 4</td>
<td>55</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Impact 5</td>
<td>100</td>
<td>78.05</td>
<td>51.22</td>
</tr>
<tr>
<td>Impact 6</td>
<td>0</td>
<td>1.335</td>
<td>89.691</td>
</tr>
<tr>
<td>Impact 7</td>
<td>58.787</td>
<td>75.314</td>
<td>100</td>
</tr>
<tr>
<td>Impact 8</td>
<td>0</td>
<td>100</td>
<td>74.538</td>
</tr>
<tr>
<td>Impact 9</td>
<td>74.884</td>
<td>100</td>
<td>0</td>
</tr>
</tbody>
</table>

**Normalised Impact Matrix Accounting for Minimisation Objectives**
Budapest = 512.986
Moscow = 533.373
Amsterdam = 463.169
New York = 492.052
From where are these results coming from?

- Information available
- Indicators chosen
- Direction of each indicator
- Relative importance
- Aggregation Procedure
<table>
<thead>
<tr>
<th></th>
<th>Budapest</th>
<th>Moscow</th>
<th>Amsterdam</th>
<th>New York</th>
</tr>
</thead>
<tbody>
<tr>
<td>Budapest</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Moscow</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Amsterdam</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>New York</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>0</td>
</tr>
</tbody>
</table>

Outranking Matrix of the 4 Cities According to the 9 Indicators
<table>
<thead>
<tr>
<th>B</th>
<th>A</th>
<th>D</th>
<th>C</th>
<th>31</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>D</td>
<td>C</td>
<td>A</td>
<td>31</td>
</tr>
</tbody>
</table>

Where A is Budapest, **B is Moscow**, C is Amsterdam and D is New York.
**Economic dimension**
City product per person

**Environmental dimension**
Use of private car
Solid waste generated per capita

**Social dimension**
Houses owned
Residential density
Mean travel time to work
Income disparity
Households below poverty line
Crime rate
A reasonable decision might be to consider the three dimensions equally important. This would imply to give the same weight to each dimension considered and finally to split this weight among the indicators. That is, each dimension has a weight of 0.333; then the economic indicator has a weight of 0.333, the 2 environmental indicators have a weight of 0.1666 each, and each one of the 6 social indicators receives a weight equal to 0.0555. As one can see, if dimensions are considered, weighting indicators by means of importance coefficients is crucial.
<table>
<thead>
<tr>
<th></th>
<th>Budapest</th>
<th>Moscow</th>
<th>Amsterdam</th>
<th>New York</th>
</tr>
</thead>
<tbody>
<tr>
<td>Budapest</td>
<td>0</td>
<td>0.3</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>Moscow</td>
<td>0.7</td>
<td>0</td>
<td>0.5</td>
<td>0.6</td>
</tr>
<tr>
<td>Amsterdam</td>
<td>0.6</td>
<td>0.5</td>
<td>0</td>
<td>0.3</td>
</tr>
<tr>
<td>New York</td>
<td>0.6</td>
<td>0.4</td>
<td>0.7</td>
<td>0</td>
</tr>
</tbody>
</table>

**Weighted Outranking Matrix**
B D C A

Where A is Budapest, B is Moscow, C is Amsterdam and D is New York.
CONCLUSION:

Results are heavily dependent on the problem structuring step!!
QUALITY OF PRODUCT

PROCEDURAL RATIONALITY

LEARNING HOLARCHIES

MCDA

QUALITY OF “SOCIAL” PROCESS

ETHICS

RESPONSIBILITY

CONSISTENCY

PARTICIPATION

TRANSPARENCY

MULTI/INTER-DISCIPLINARITY
European Conference on Quality in Survey Statistics

2006
Knowledge Economy Indicators: Development of Innovative and Reliable Indicator Systems

European Conference on Quality in Survey Statistics
Cardiff, 08/10/2010
European Commission: DG Research in cooperation with DG ECFIN and DG ESTAT

Framework Programme 6
Integrating and Strengthening the European Research Area
Policy Orientated Research

Project period:
1.9.2004 – 28.2.2007

EC contribution: 1.58 M€
Aims

KEI will focus on indicators and composite indicators for policy purposes

- Review of state-of-the-art methodology
- Policy scenarios
- Main thematic areas in relation to Lisbon and Barcelona objectives
- Forward looking policy and indicator analysis
Aims (continued)

Development of innovative approaches for composite indicators

- Improved statistical methodology
- Multi-criteria methods
- Aggregation and weighting techniques
- Elaboration of sensitivity analysis
- Evaluation of analytical properties of indicators
- Investigation of adequate presentational techniques
- Support by large-scale simulation study
- Scenario analysis in co-operation with Commission services
Participants

- Eberhard Karls University of Tübingen, Germany: Eberhard Schaich, Dominik Ohly, Kersten Magg, Rolf Wiegert
- Joint Research Center, Ispra, Italy: Andrea Saltelli, Niels Schulze, Giuseppe Munda, Stefano Tarantola, Michaela Saisana
- Katholieke Universiteit Leuven, Belgium: Tom Van Puyenbroeck, Laurens Cherchye, Wim Moesen, Nicky Rogge
- University of Maastricht, MERIT, The Netherlands: Anthony Arundel, Wendy Hansen
- Statistics Finland, Finland: Mikael Åkerblom, Tero Luhtala
- University of Trier, Germany: Ralf Münich (CO), Nicole Thees
Contact Details

Ralf Münnich (coordinator)
University of Trier, Faculty IV
Economic and Social Statistics Department
Universitätsring 15
D-54286 Trier
Phone: +49 (0)651 201-2651
Fax: +49 (0)651 201-3931
e-mail: muennich@uni-trier.de
WEB: http://www.kei.publicstatistics.net

Invited Session: Knowledge Economy Indicators

• The R&D Indicators in the Knowledge Based Economy
  Asterios Chatziparadeisis, Greek Ministry of Development

• Quality of Knowledge Economy Indicators
  Nicole Thees, University of Trier

• Weighting and Aggregation for Composite Indicators: A Non-compensatory Approach
  Giuseppe Munda, DG JRC, Ispra, and University Autonoma de Barcelona

• An Introduction to Benefit of the Doubt Composite Indicators
  Nicky Rogge, Catholic University of Leuven
Quality of Knowledge Economy Indicators

Ralf Münnich, Nicole Thees and Rolf Wiegert

Abstract

In the context of measuring performance and development of European economies, one main focus as a target are the indicators of the knowledge-based economy. The following paper will present the project Knowledge Economy Indicators: Development of Innovative and Reliable Indicator System and will outline the first work on the available quality information of the data which become increasingly important when measuring multidimensional aspects.

1. Aims of the KEI project

The project Knowledge Economy Indicators: Development of Innovative and Reliable Indicator Systems (KEI) is in the context of the Sixth Framework Programme of the European Commission with main focus on "integrating and strengthening the European research area". It started in September 2004 and is part of priority 8 of the policy orientated research under the Framework Programme. As one of the key aspects of the Lisbon strategy is to stimulate economic growth and employment, improvement of the knowledge based economy is one of the most important aspects. Therefore the project’s aim is to develop and improve indicators for the knowledge economy, including the analysis of aggregation issues and the use of composite indicators with special focus on data quality.

The project will finally cover nearly 100 different indicators from 30 European countries (the EU-25 plus Iceland, Norway, Switzerland, Romania and Bulgaria) and six non-European countries (the US, Japan, India, China, Australia and Canada).

The first project step consists of a review of existing concepts and definitions of the knowledge-based economy and its key components which will meet the Lisbon and Barcelona objectives for indicators classification concerning the development of main thematic areas. It follows the exploration of the necessary data and the indicator quality issues where gaps have to be identified. At the end a way forward will be mapped, identifying innovative approaches to improve the understanding and appraisal of the knowledge economy.

Compared to the analysis of single indicators, offering detailed information about special aspects of the knowledge economy, the developed composite indicators will provide the big picture of the multi-dimensional subject. They will be analysed in detail using

---

1 Münnich, Ralf, University of Trier, Universitätsring 15, Germany, 54286 Trier; Thees, Nicole, University of Trier, Universitätsring 15, Germany, 54286 Trier; Wiegert, Rolf, University of Tübingen, Mohlstraße 36, Germany, 72074 Tübingen.

2 A discussion concerning the pro and cons of composite indicators can be found under Saisana (2005).
statistical approaches, including the use of multi-criteria methods, aggregation and weighting techniques, decomposition methods, and an evaluation of analytical and presentational techniques. Simulation methods will be employed to investigate the robustness of indicators and the conclusions based on them. The quality and accuracy of indicators and the underlying data will be evaluated and the innovative use of additional information to improve indicator quality will be assessed.

In addition, the KEI project will benefit other policy objectives of the European Union and Commission Services; it contributes to a methodological framework for building effective measurements of interdisciplinary issues such as employment or social cohesion and will make recommendations for the design and use of statistical reference systems.

KEI consists of seven research work packages (WP) divided into two main groups, plus a final project report. The first group, consisting of WP1, WP2 and WP4, focuses on the theory of the knowledge based economy, policy needs and the identification of useful indicators. The output of the first group is an essential input for the second group, consisting of WP3, WP5, WP6, and WP7. These WPs will statistically analyse indicators in order to identify robust composite indicators that can meet policy requirements.

Concerning the definition of the knowledge-based economy (KBE) WP1 contains both indicators and policy. Potential indicators for measuring the drivers, characteristics and key outputs of a knowledge economy meeting policy and user needs are identified and evaluated in WP2. The objective of WP3 - statistical analysis of KBE indicators - is to analyse key aspects of the data quality of the indicators. WP4 focuses on the innovative use of indicators, either through finding solutions to missing indicators or identifying indicators that can meet future needs. After a summary of the state-of-the-art in developing composite indicators, the fifth work package deals with the construction and testing of new composite indicators as well as suggestions of improvement for the existing ones. As national indicators on R&D efforts are somewhat distorted by worldwide R&D activities of multinational companies, the aim of WP6 is to develop and test new indicators on the role of multinational companies for national indicators on R&D efforts in order to estimate the effects on national figures. In the final work package, a set of simulations will test the accuracy and reliability of the indicators in a practical environment under different realistic assumptions and data quality standards.3

The project is coordinated by R. Münnich (University of Trier - Germany) in cooperation with five partners, University of Tübingen (Germany) with E. Schaich, D. Ohly, K. Magg, R. Wiegert, the Joint Research Center in Ispra (Italy) with A. Saltelli, N. Schulze, G. Munda, S. Tarantola, M. Saisana, the Katholieke Universiteit Leuven (Belgium) with T. Van Puyenbroeck, L. Cherchye, W. Moesen, N. Rogge, the University of Maastricht (Netherlands) with A. Arundel, W. Hansen and Statistics Finland with M. Åkerblom, T. Luhtala.

The present study focuses on the analysis of the data quality. First results of an extract of the indicators will be presented. It is to emphasize, that this is a first analysis from the point of view of the end-user of indicators and that the supplier aspects of data quality have to be analysed too separately.

3 Further details can be drawn from the homepage: http://kei.publicstatistics.net.
2. Knowledge economy indicators

2.1 Indicators selection via quality information

As the project’s target is to monitor performance of the knowledge economy in different countries including the analysis of the use of composite indicators, since the indicator system itself may not be significant enough, the quality of the data cannot be neglected. There are different facts, which let arise the necessity to pay special emphasis to the data quality, not only in the sense of accuracy or reliability, but also concerning coverage, accessibility or coherence: the performance in different countries has to be measured, a variety of data sources have to be considered, e.g. survey samples or administrative data, the target group should be policy makers at the European and national levels, who would like to draw their political conclusion from the indicator result. Analysing the data quality, offers the opportunity to identify areas of indicator weakness on the one hand and to guide the selection of indicators out of the abundance of possible indicators on the other one.

But using the quality information as parameter of selection of the economic indicators intends an other problem: not only the indicators number is enormous but also the quality information - for one indicator, there exists a variety of information which differs additionally thru the countries; to handle this, the idea is, to construct a quality measure in a comparable form as an additional composite indicator. This intends obviously the aggregation of all relevant quality information.

Before presenting the results of this quality analysis, the indicator system itself has to be outlined.

2.2 Indicator system

The single indicators are discussed in terms of their relative importance for measuring the development of knowledge economy in several countries and their potential and value for composite indicators. The aspects, which have to be taken into account to ensure, that selected indicators on the knowledge economy are not limited to a ‘R&D’ view, are

- key economic sectors (e.g. low, medium or high tech sectors, private or public services),
- input and output indicators (e.g. education pipeline, R&D investment and patents),
- life and work in the knowledge economy,
- national performance for innovation (e.g. capabilities and output) and globalization (e.g. impact of globalization on work and life).

The indicators have been classified into three groups: Group A indicators concern the drivers and characteristics of a KBE and are further divided into four main sub-groups:

- A1 - production and diffusion of ICTs with 20 indicators and the subgroups:
  - economic impact of ICT sector [information & communication technology]
  - Internet use by firms
  - Internet use by individuals
  - Government ICT

4 Details will be found in Arundel (2006).
• A2 - human resources, skills and creativity with 20 indicators and the subgroups:
  - General Education indicators
  - HRST education indicators
  - Skills
  - Creativity
  - Mobility
• A3 - knowledge production and diffusion with 26 indicators and the subgroups:
  - R&D family
  - Patents (intermediate indicators of Knowledge output)
  - Bibliometrics (production of scientific knowledge)
  - Market Innovation outputs
  - Knowledge flows
  - Total investment in intangibles
• A4 - innovation, entrepreneurship and creative destruction with 20 indicators and the subgroups:
  - Entrepreneurship
  - Demand for innovative products
  - Financing of innovation
  - Market Innovation outputs
  - Organisational indicators

Group B includes seven indicators of economic outputs, concerning income, productivity and employment on the one hand and 21 indicators of social performance, concerning environment, employment and economic welfare, quality of life indicators on the other hand.

Group C consists of horizontal indicators, relevant to both the drivers and outputs, for composite indicator development.
• C1 - public sector innovation is a particular policy interest with five indicators,
• C2 - internationalization with 25 indicators and the subgroups:
  - trade
  - knowledge production and diffusion of knowledge
  - economic structure
  - human resources

At the moment, about 150 indicators are discussed. As some of the indicators are only wished indicators, it is assumed to have approximately 90 indicators in the end giving recommendations for a future evaluation of particular aspects.

The abundance of single indicators for measuring the knowledge based economy stresses the necessity of decision guidance. Due to this fact, it may be meaningful to benefit of the growing number of quality information. But which kind of quality information is accessible for the end-user? In the following part an abstract of the main existing quality structure will be given.

2.3. Existing quality information and quality composite indicator

The quality systems should be differentiated into frameworks or methodologies for assessing national statistical systems respectively data quality. One of the established frameworks is the general respectively special data dissemination system (GDDS/
SDDS\textsuperscript{5}) which promotes improved dissemination and effectiveness (Nardo, 1995; United Nations, 1995; Magg, 2005). The end-users metadata of the Eurostat NewCronos database, one of the main data sources of the KEI project, are e.g. presented in SDDS. The SDDS has five dimensions plus overlapping aspects from a methodological summary (see table 1).

### Table 1: The quality dimension of the special data dissemination system

<table>
<thead>
<tr>
<th>SDDS</th>
<th>The data: coverage, periodicity and timeliness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Access by the public</td>
</tr>
<tr>
<td></td>
<td>Integrity (transparency of practices and procedures)</td>
</tr>
<tr>
<td></td>
<td>Quality (information the user needs to assess data quality)</td>
</tr>
<tr>
<td></td>
<td>Dissemination formats</td>
</tr>
<tr>
<td></td>
<td>Concepts, definitions and classifications</td>
</tr>
<tr>
<td></td>
<td>Scope / coverage of the data</td>
</tr>
<tr>
<td></td>
<td>Accounting conventions</td>
</tr>
<tr>
<td></td>
<td>Nature of the basic data</td>
</tr>
<tr>
<td></td>
<td>Compilation practices (data processing)</td>
</tr>
<tr>
<td></td>
<td>Other aspects</td>
</tr>
</tbody>
</table>

Compared with this, the methodology for assessing data quality has much larger definition, like the Data Quality Assessment Framework (DQAF) of the IMF or the Assessment of Quality in Statistics of Eurostat (Eurostat, 2000; Pellegrino, 2004) now advanced by the Code of Practice (Grünewald, 2001; Eurostat, 2005); e.g. the seven dimensions of the Assessment of Quality in Statistics of Eurostat are relevance, accuracy, comparability, coherence, timeliness and punctuality, accessibility and clarity, completeness.

The existence of these quality definitions causes the necessity to decide first of all, which kind of quality definition has to be used. As the NewCronos is one of the main databases, most of the quality information is just presented in the SDDS standard; but also other databases have to be accounted which intend a harmonization of the quality information. In fact, it is reasonable to harmonize the given information corresponding to the Quality Assessment Framework of Eurostat respectively to the Code of Practice, as the knowledge based economy of Europe has to be analysed.

When harmonizing the quality issues, the assumption was made that the quality assessment framework describes a final target, which is not reached at the moment; therefore these assessment framework aspects, which are not provided by the dissemination standard are cancelled and for possible supplementary aspects out of the dissemination standard, a category "miscellaneous" and "extraordinary" was added.

Now, the idea of constructing a composite indicator of quality intends that each quality information of an indicator accessible has to be aggregated and to be transformed into an index which offers the opportunity of an indicator calculation. It was decided to judge the information by basic quality scores between one and zero, measuring the proportion

---

\textsuperscript{5} Definition to be found under: \url{http://dsbb.imf.org/Applications/web/dsbbhome/}.
of available issues. By doing this, the assumption was made, that each kind of information is valuable for the user himself and offers him the opportunity to draw conclusions about the indicators quality.

Similarly to the dimension reduction via composite indicators for the knowledge economy, the huge amount of quality information will have to be monitored adequately. First ideas will be depicted in the following paragraph.

3. Quality analysis

The measurement of quality started with an examination of the rate of availability of quality information and the resulted ranking of a chosen number of indicators as well as their sensitivity to weight changes of the quality concerned. This is necessary to examine the useful quality information for the discrimination of economic indicators; otherwise, the idea of constructing a composite indicator of quality seems to be not recommended.

In the first step, the rate of availability of quality information referred to one point in time only and to the general information offered by the metadata for the whole indicator, not for different countries, for only seven indicators, five from main categories A and two of category B, just presented had been analysed.

Table 2: The rate of availability of quality information

<table>
<thead>
<tr>
<th>economic indicator</th>
<th>quality indicator⁠¹</th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>TP</th>
<th>AC</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
</tr>
</thead>
<tbody>
<tr>
<td>investment in ICT / % of GDP</td>
<td></td>
<td>1.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.30</td>
<td>0.13</td>
<td>0.67</td>
<td>0.17</td>
<td>0.25</td>
<td>0.25</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>researchers / 1,000 labour force</td>
<td></td>
<td>1.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.63</td>
<td>0.29</td>
<td>0.67</td>
<td>0.17</td>
<td>0.50</td>
<td>0.50</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>BERD / % of GDP</td>
<td></td>
<td>1.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.63</td>
<td>0.29</td>
<td>0.67</td>
<td>0.17</td>
<td>0.50</td>
<td>0.50</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>enterprise churn</td>
<td></td>
<td>1.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.56</td>
<td>0.26</td>
<td>0.33</td>
<td>0.17</td>
<td>0.50</td>
<td>0.50</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>% SMEs reporting non-tech. change</td>
<td></td>
<td>1.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.75</td>
<td>0.48</td>
<td>0.67</td>
<td>0.17</td>
<td>0.75</td>
<td>0.50</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>labour productivity / hour worked</td>
<td></td>
<td>1.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.31</td>
<td>0.16</td>
<td>0.67</td>
<td>0.17</td>
<td>0.38</td>
<td>0.19</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>inequality of income distribution</td>
<td></td>
<td>1.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.31</td>
<td>0.16</td>
<td>0.67</td>
<td>0.17</td>
<td>0.38</td>
<td>0.25</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

1. legend: main quality groups: R = relevance; A = accuracy; TP = timelines and punctuality; AC = accessibility and clarity; C = comparability. The subgroups R1-R4 concern information about the “reference of specific documents where the description of more comprehensive needs could be found, if any”, “reasons for incompleteness as well as the prospects for future solutions”, the “circulation and/or readership of publications (paper of electronic) and an additional group “miscellaneous”; the subgroups A1-A3 are “general” information, information about the “sampling errors” and “non-sampling errors”; C1-C4 are “geographical comparability”, “comparability over time”, “comparability between domains” and “coherence”.

The rate of availability of the dimensions “coherence” as well as “cost & burden” and “extraordinary aspects” is zero for all economic indicators.
As the rate of availability is equal for some of the quality aspects (see table 2, previous page), those are not appropriate for an economic indicator discrimination, assuming that the proportion of availability of some quality indicators is only zero because of the small number of analysed economic indicators until now. If this isn't true, this quality information has to be examined in a separate way. Finally, there are five subdimensions which can be valuable for the indicators discrimination and for analysing its performance.

When having a look on the previous table, one may be in doubt which one of the seven indicators may be the "best" in the sense of quality availability. The median of the ranking of the quality indicators by assumption of equal weights point up, that "total researchers per thousand labour force in FTE", "Business expenditure on R&D (BERD) as percentage of GDP" and "percent SMEs reporting non technological change" are from the point of view of quality information availability recommendable.

An interesting fact is, that the methodology of two first mentioned indicators are based on the Frascati Manual, the third is result of the Community Information Survey (CIS3) which perhaps supports the hypotheses that international survey recommendations facilitate the availability of quality information.

But the assumption of equal weights for quality issues is not very reasonable, therefore the use of independent random weights with a rectangular distribution developed by a Monte Carlo simulation was analysed too. The result is very interesting: as the spread of the weights is significant between nearly zero and one: the sensitivity of the economic indicator rests quite invariant (see figure 1).

**Figure 1: Quality ranking of economic indicators**

<table>
<thead>
<tr>
<th>Economic Indicator</th>
<th>Investment in ICT / % of GDP</th>
<th>Total Researchers / 1,000 labour force (FTE)</th>
<th>BERD / % of GDP</th>
<th>Enterprise churn</th>
<th>Non-tech. change</th>
<th>% SMEs reporting non-tech. change</th>
<th>Labour productivity / hour worked</th>
<th>Inequality of income distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rectangular distribution - median</td>
<td>5.5</td>
<td>5.5</td>
<td>4</td>
<td>3.8</td>
<td>3</td>
<td>3.1</td>
<td>4</td>
<td>2.5</td>
</tr>
<tr>
<td>Random weights - min.</td>
<td>3</td>
<td>2</td>
<td>1.9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Random weights - 1st quantil</td>
<td>2</td>
<td>2</td>
<td>1.19</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Random weights - mean</td>
<td>1.19</td>
<td>1.19</td>
<td>3.8</td>
<td>3.8</td>
<td>3.8</td>
<td>3.8</td>
<td>3.8</td>
<td>3.8</td>
</tr>
<tr>
<td>Random weights - max.</td>
<td>5.5</td>
<td>5.5</td>
<td>5.5</td>
<td>5.5</td>
<td>5.5</td>
<td>5.5</td>
<td>5.5</td>
<td>5.5</td>
</tr>
</tbody>
</table>
One can recognize, that the economic indicators quality ranking is not very sensitive for the kind of quality existing. At a first glance, it seems to be important to have any quality information possible. But this assumption may be misleading due to the following question: is the information of sampling design and sampling error equal to the information about geographical comparability? To answer this question, a factor analysis with principal component which can detect the importance of single quality information and possible factors having influence on the ability of quality information was made which provides two interesting results: The quality ranking of the economic indicators can be verified once again. Only the ranking between indicators two and three is more evident: the variable "total researchers per thousand labour forces in FTE" is assessed better than the variable "BERD as a percentage of GDP" (see tab. 3).

The standard deviation underlines, that there is especially one component which has the most important influence on the quality availability. Perhaps, it is really the existence of manuals like the Frascati one, which facilitate the possible supply with quality information especially for the NSI. This is still a vague conclusion, because the quality information availability on the country level isn't analysed until now.

<table>
<thead>
<tr>
<th>Tab 3: Quality ranking as result of the principal component analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>economic indicator</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>investment in ICT / % of GDP</td>
</tr>
<tr>
<td>researchers / 1,000 labour force</td>
</tr>
<tr>
<td>BERD / % of GDP</td>
</tr>
<tr>
<td>enterprise churn</td>
</tr>
<tr>
<td>% SMEs reporting non-tech. change</td>
</tr>
<tr>
<td>labour productivity / hour worked</td>
</tr>
<tr>
<td>inequality of income distribution</td>
</tr>
</tbody>
</table>

4. Summary and Outlook

The analysis of the supply by quality information is a first adequate method to build a ranking of indicators as basis for data discrimination. In this context, methodological manuals are obviously an adequate instrument to perform and facilitate the availability of information. The data availability will be the crucial point and a severe problem to solve. A first overview for of the data availability in some countries of the seven presented indicators illustrates, that there are great vacancies for many indicators for many points in time.
Only one country has data available at one point in time for the three qualitative good indicators!
A look e.g. at NewCronos underlines, that a lot of work concerning the metadata availability for the end-user is just done; but to reach the target of the Code of Practice, the quality information offered for the end-user has to increase again. On the one hand, this is very good for the end-user, for getting a good impression of the data quality, but on the other hand, when constructing a composite indicator, the large quantity of quality information has to be handled in a way. Maybe the evaluation of a quality composite indicator will be one solution but it also has to be discussed, in which way of standardisation the additional information can be produced by the NSI and offered for the end-user by a minimum of additional burden. In order to measure performance and development of the knowledge economy correctly, e.g. in terms of significant differences, one has to get deep knowledge on data quality aspects on the surveys and indicators of interest. It is evident, that the measurement of multidimensional problems has to be made with the knowledge of what kind of data is compared. Important aspects like information about the data source, the provision of concrete standard errors and the consideration of the data availability are a must for the selection and assessment of the right indicators. Therefore, the evaluation of a quality seal when operating with an enormous number of data will be user-adequate.
In the next step of the quality analysis, the availability of quality information for all indicators especially on the country level as well as the availability of the data itself should be analysed.

5. Literature


European Conference on Quality in Survey Statistics

2006
Composite indicators & Data Envelopment Analysis

Q2006
Cardiff, April 25

Nicky Rogge
(Catholic University of Leuven)
Structure

- Composite indicators: useful but subject to controversy
- Normalization and weighting issue
- Background: why DEA?
- Incorporating (expert) opinion (sub-indicator share restrictions)
Examples

- Human Development Index (HDI)
- Technology Achievement Index (TAI)
- Environmental Sustainability Index (ESI)
- Misery Index (MI)
Composite Indicators

1. Indicators are observed (measured in different units)
2. We want an overall ‘score’ in order to compare countries…
3. …but weights for aggregation are not known with certainty

- Useful tool for policy evaluation and communication
- Useful for comparing and ranking

However, still subject to controversy
The normalization issue

Common practice: normalizing the sub-indicators before aggregating.

2 problems

- No consensus on suitable normalization method
- Results depend on normalization method

“A meaningful index is defined as an index whose underlying preference ordering is independent of admissible transformations of the variables”
  Ebert and Welsch (2004)

In practice, however, most indices and the resulting ranks depend on the normalization method → Potential point of criticism

Removing the necessity to normalize the original data
The weighting issue

Ideally, individual indicators should be weighted and combined in a manner reflecting the underlying structure of the evaluated phenomenon.

However, insufficient knowledge on the underlying structure

Often: Equal Weighting (EW)

Which weights to use? 2 issues:

1. Any predetermined common set of weights will favor some countries while harming others.
2. The choice of weights may affect the composite indicators and their ranks, undermining their credibility.

Determine endogenously the optimal set of weights
Disagreement among experts: proposed weights for Technology Achievement Index

<table>
<thead>
<tr>
<th>Expert</th>
<th>Patents</th>
<th>Royalties</th>
<th>Internet</th>
<th>Exports</th>
<th>Telephones</th>
<th>Electricity</th>
<th>Schooling</th>
<th>Enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert1</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.2</td>
<td>0.1</td>
<td>0.05</td>
<td>0.2</td>
<td>0.3</td>
</tr>
<tr>
<td>Expert2</td>
<td>0.15</td>
<td>0.15</td>
<td>0.1</td>
<td>0.15</td>
<td>0.1</td>
<td>0.05</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>Expert3</td>
<td>0.2</td>
<td>0</td>
<td>0.1</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>0.2</td>
<td>0.3</td>
</tr>
<tr>
<td>Expert4</td>
<td>0.07</td>
<td>0.07</td>
<td>0.1</td>
<td>0.15</td>
<td>0.13</td>
<td>0.06</td>
<td>0.23</td>
<td>0.19</td>
</tr>
<tr>
<td>Expert5</td>
<td>0.1</td>
<td>0.15</td>
<td>0.15</td>
<td>0.3</td>
<td>0.05</td>
<td>0.05</td>
<td>0.2</td>
<td>0</td>
</tr>
<tr>
<td>Expert6</td>
<td>0.1</td>
<td>0.05</td>
<td>0.2</td>
<td>0.1</td>
<td>0.15</td>
<td>0.05</td>
<td>0.15</td>
<td>0.2</td>
</tr>
<tr>
<td>Expert7</td>
<td>0.2</td>
<td>0.2</td>
<td>0.05</td>
<td>0.1</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
<td>0.05</td>
</tr>
<tr>
<td>Expert8</td>
<td>0.1</td>
<td>0.05</td>
<td>0.15</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>Expert9</td>
<td>0.1</td>
<td>0.1</td>
<td>0.15</td>
<td>0.15</td>
<td>0.1</td>
<td>0.05</td>
<td>0.2</td>
<td>0.15</td>
</tr>
<tr>
<td>Expert10</td>
<td>0.2</td>
<td>0.05</td>
<td>0.05</td>
<td>0.3</td>
<td>0.05</td>
<td>0</td>
<td>0.05</td>
<td>0.3</td>
</tr>
<tr>
<td>Expert11</td>
<td>0.12</td>
<td>0.15</td>
<td>0.12</td>
<td>0.15</td>
<td>0.1</td>
<td>0.12</td>
<td>0.08</td>
<td>0.16</td>
</tr>
<tr>
<td>Expert12</td>
<td>0.05</td>
<td>0.05</td>
<td>0.2</td>
<td>0.1</td>
<td>0.05</td>
<td>0.05</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>Expert13</td>
<td>0.1</td>
<td>0.05</td>
<td>0.1</td>
<td>0.25</td>
<td>0.05</td>
<td>0.15</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>Expert14</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.1</td>
<td>0.2</td>
<td>0.05</td>
<td>0.2</td>
<td>0.3</td>
</tr>
<tr>
<td>Expert15</td>
<td>0.1</td>
<td>0.15</td>
<td>0.1</td>
<td>0.23</td>
<td>0.03</td>
<td>0.02</td>
<td>0.05</td>
<td>0.15</td>
</tr>
<tr>
<td>Expert16</td>
<td>0.1</td>
<td>0.15</td>
<td>0.1</td>
<td>0.2</td>
<td>0.15</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Expert17</td>
<td>0.15</td>
<td>0.13</td>
<td>0.11</td>
<td>0.14</td>
<td>0.1</td>
<td>0.1</td>
<td>0.15</td>
<td>0.12</td>
</tr>
<tr>
<td>Expert18</td>
<td>0.1</td>
<td>0.15</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
<td>0</td>
<td>0.15</td>
<td>0.2</td>
</tr>
<tr>
<td>Expert19</td>
<td>0.12</td>
<td>0.22</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>Expert20</td>
<td>0.05</td>
<td>0.02</td>
<td>0.2</td>
<td>0.15</td>
<td>0.15</td>
<td>0.03</td>
<td>0.15</td>
<td>0.25</td>
</tr>
<tr>
<td>Expert21</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.25</td>
<td>0.15</td>
<td>0.05</td>
<td>0.05</td>
<td>0.2</td>
</tr>
<tr>
<td>Average</td>
<td>0.110</td>
<td>0.107</td>
<td>0.109</td>
<td>0.181</td>
<td>0.098</td>
<td>0.063</td>
<td>0.148</td>
<td>0.184</td>
</tr>
<tr>
<td>Max</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
<td>0.33</td>
<td>0.2</td>
<td>0.15</td>
<td>0.25</td>
<td>0.3</td>
</tr>
<tr>
<td>Min</td>
<td>0.05</td>
<td>0</td>
<td>0.02</td>
<td>0.09</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
<td>0</td>
</tr>
</tbody>
</table>
Classical DEA-problem

1. Inputs and Outputs are observed (measured in different units)
2. We want an overall expression of productivity in order to compare firms…
   - Productivity = (weighted sum of outputs) / (weighted sum of inputs)
3. …but weights for aggregation are not known with certainty
   - (production function, right prices,… unknown)

Remarkable similarity between the problem of measuring efficiency and the one of constructing composite Indicators
Towards DEA composite indicator

Step 1

• Weighted sum of indicators = ‘country total’
• Since the eventual purpose of a composite indicator is to compare (with other countries), we will express this ‘country total’ RELATIVE TO A SIMILARLY WEIGHTED SUM of BENCHMARK SUB-INDICATORS

\[
I_c = \frac{\sum_{i=1}^{m} w_{c,j} y_{c,j}}{\sum_{i=1}^{m} w_{c,j} y_{i}^B}
\]
Towards DEA composite indicator

Step 2

• Which benchmark should we choose?
• Look **WITHIN SAMPLE** of COUNTRIES for the one that yields the **HIGHEST POSSIBLE TOTAL** (given the weights)
• ‘Best Practice’ notion

\[
I_c = \frac{\sum_{i=1}^{m} w_{c,j} y_{c,j}}{\max_{y_i \in \text{\{studied countries\}}} \sum_{i=1}^{m} w_{c,j} y_{i,j}}
\]
The DEA composite indicator

**Step 3: Benefit-of-the-doubt weighting**

- Which weights should we choose?
- **CHOOSE** the weights such that the evaluated country has a **MAXIMAL COMPOSITE INDICATOR VALUE**
- ‘Benefit-of-the-doubt’-notion

\[
I_c = \max_{w_{c,j}} \frac{\sum_{i=1}^{m} w_{c,j} y_{c,j}}{\max_{y_i \in \{\text{studied countries}\}} \sum_{i=1}^{m} w_{c,j} y_{i,j}}
\]
Benefit-of-the-doubt’-notion

- **Weights?** CHOOSE them ($\geq 0$) such that the evaluated country has a maximal composite indicator value

- **Benchmark?** Look within sample for country that maximizes CI, given these weights $w_{c,i}$

$$I_c = \max_{w_{c,i}} \sum_{i=1}^{m} w_{c,i} y_{c,i}$$

$$\max_{y_{j,i} \in \{\text{studied countries}\}} \sum_{i=1}^{m} w_{c,i} y_{j,i}$$
Advantages

1. Qua model: better description of reality than (e.g.) equal weighting

2. Embedded concern for (MS’) diversity: no other weighting scheme yields higher composite indicator value *(political acceptance)*

3. Principle is easy to communicate
   - Since we are not sure about the right weights, we look for « benefit of the doubt » weights (such that your overall relative performance index is as high as possible)
• We graphically represent sub indicators as pie shares with a bigger size reflecting a higher importance.

Finland: 100.00%  
Poland: 30.03%
sub-indicator share restrictions

1. Absolute sub-indicator share restrictions

\[ \alpha \leq w_{j,i}y_{j,i} \leq \beta \]

2. Relative sub-indicator share restrictions

\[ \alpha \leq \frac{w_{j,i}y_{j,i}}{w_{j,k}y_{j,k}} \leq \beta \]

3. Proportional sub-indicator share restrictions

\[ \alpha \leq \frac{w_{i,j_i}y_{j,i}}{\sum_{i=1}^{m} w_{i,j_i}y_{j,i}} \leq \beta \]
• For the TAI, for example, the lower and upper bound could be specified by respectively the lowest and highest weight assigned over all experts to that sub-indicator.

<table>
<thead>
<tr>
<th>Country</th>
<th>Patents</th>
<th>Royalties</th>
<th>Internet</th>
<th>Exports</th>
<th>Telephones</th>
<th>Electricity</th>
<th>Schooling</th>
<th>Enrolment</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finland</td>
<td>0.0500</td>
<td>0.2068</td>
<td>0.0200</td>
<td>0.1732</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.2500</td>
<td>0.3000</td>
<td>100.00%</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.0451</td>
<td>0.2704</td>
<td>0.1712</td>
<td>0.0811</td>
<td>0.1803</td>
<td>0.0000</td>
<td>0.1534</td>
<td>0.0000</td>
<td>90.15%</td>
</tr>
<tr>
<td>Norway</td>
<td>0.0366</td>
<td>0.0000</td>
<td>0.1465</td>
<td>0.0659</td>
<td>0.1465</td>
<td>0.1099</td>
<td>0.1831</td>
<td>0.0439</td>
<td>73.25%</td>
</tr>
<tr>
<td>Poland</td>
<td>0.0150</td>
<td>0.0000</td>
<td>0.0060</td>
<td>0.0991</td>
<td>0.0000</td>
<td>0.0161</td>
<td>0.0751</td>
<td>0.0890</td>
<td>30.03%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Country</th>
<th>Percentage contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finland</td>
<td>5.00% 20.68% 2.00% 17.32% 0.00% 0.00% 25.00% 30.00%</td>
</tr>
<tr>
<td>Netherlands</td>
<td>5.00% 30.00% 18.99% 9.00% 20.00% 0.00% 17.01% 0.00%</td>
</tr>
<tr>
<td>Norway</td>
<td>5.00% 0.00% 20.00% 9.00% 20.00% 15.00% 25.00% 6.00%</td>
</tr>
<tr>
<td>Poland</td>
<td>5.00% 0.00% 2.00% 33.00% 0.00% 5.36% 25.00% 29.64%</td>
</tr>
</tbody>
</table>

Netherlands: 90.15%  Norway: 73.25%
4. Ordinal sub-indicator share restrictions

\[ w_{j,6} y_{j,6} \leq w_{j,5} y_{j,5} \leq w_{j,2} y_{j,2} \leq w_{j,3} y_{j,3} \leq w_{j,1} y_{j,1} \leq w_{j,7} y_{j,7} \leq w_{j,4} y_{j,4} \leq w_{j,8} y_{j,8} \]

- The restriction imposes an importance ranking of the sub-indicators.

5. Restrictions pertaining to category shares

**Relative**

\[ \alpha \leq \frac{\sum_{i \in Sa} w_{j,i} y_{j,i}}{\sum_{i \in Sb} w_{j,i} y_{j,i}} \leq \beta \]

**Proportional**

\[ \alpha \leq \frac{\sum_{i \in Sa} w_{j,i} y_{j,i}}{\sum_{i=1}^{m} w_{j,i} y_{j,i}} \leq \beta \]

- Focusing on the importance of key dimensions of the evaluated phenomenon may allow a more swiftly expert consensus.
Quality of knowledge economy indicators

Nicole Thees
University of Trier

European Conference on Quality in Survey Statistics
Cardiff, 25/04/2006
Aim of the KEI project

- innovative and reliable *indicator system* for a multidimensional concept
  → measuring the Lisbon goals
- analysis of aggregation issues
- composite indicators
  - significance
  - identification of trends
  - performance of KBE in different countries
Quality importance

- measuring performance
- data sources
  → survey samples, administrative
- target group
  → policy makers at the European and national levels
- indicators results for political conclusions
  → accuracy and statistical reliability of indicators and additional quality information
    - indicator weakness
    - guide the selection of indicators
Problem

- variety of economic indicators
- variety of quality information which differ thru the countries

Solution

- measuring quality by a composite indicator of quality
Data requirements

• measuring the development of KBE in several countries
• potential as composite indicator
• aspects to be considered - not limited to a ‘R&D’ view -
  → key economic sectors
  → input and output indicators
  → life and work in the KBE
  → national performance for innovation and globalization
Indicators

- four main drivers and characteristics - Group A
  → A1 - production and diffusion of ICTs (20 indicators)
  → A2 - human resources, skills and creativity (20)
  → A3 - knowledge production and diffusion (26)
  → A4 - innovation, entrepreneurship and creative destruction (20)

- two main outputs - Group B
  → B1 - economic indicators (7)
  → B2 - social performance (21)

- two horizontal indicators - Group C
  → C1 - public sector innovation (5)
  → C2 - internationalization (25)

⇒ 144 indicators
General or Special Data Dissemination System (GDDS / SDDS)

- assessing national statistical systems
- promote a improved dissemination and effectiveness
- Eurostat NewCronos: 5 dimensions + methodological summary

<table>
<thead>
<tr>
<th>SDDS</th>
<th>→ Coverage, Periodicity and Timeliness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>→ Access by the Public</td>
</tr>
<tr>
<td></td>
<td>→ Integrity (transparency of practices and procedures)</td>
</tr>
<tr>
<td></td>
<td>→ Quality (information the user needs to assess data quality)</td>
</tr>
<tr>
<td></td>
<td>→ Dissemination Formats</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>summary methodology</th>
<th>→ Concepts, definitions and classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>→ Scope / coverage of the data</td>
</tr>
<tr>
<td></td>
<td>→ Accounting conventions</td>
</tr>
<tr>
<td></td>
<td>→ Nature of the basic data</td>
</tr>
<tr>
<td></td>
<td>→ Compilation practices (data processing)</td>
</tr>
<tr>
<td></td>
<td>→ Other aspects</td>
</tr>
</tbody>
</table>
Data Quality Assessment Framework - IMF / Assessment of Quality in Statistics - Eurostat

• general methodology for assessing data quality
• international concepts and definitions in statistics
• Assessment of Quality in Statistics of Eurostat:

<table>
<thead>
<tr>
<th>main level</th>
<th>sub dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>→ relevance</td>
<td>description of more comprehensive needs</td>
</tr>
<tr>
<td>→ accuracy</td>
<td>reasons for incompleteness, future solutions</td>
</tr>
<tr>
<td>→ timeliness and punctuality</td>
<td>circulation and/or readership of publications</td>
</tr>
<tr>
<td>→ accessibility and clarity</td>
<td>general</td>
</tr>
<tr>
<td>→ comparability</td>
<td>sampling errors</td>
</tr>
<tr>
<td>→ coherence</td>
<td>non-sampling errors</td>
</tr>
<tr>
<td>→ cost and burden</td>
<td>geographical comparability</td>
</tr>
<tr>
<td></td>
<td>comparability over time</td>
</tr>
<tr>
<td></td>
<td>comparability between domains</td>
</tr>
</tbody>
</table>
Valuation

• information aggregation
  → measuring the proportion of available issues
  → quality scores

• assumption
  → each kind information is valuable for the user
  → the use of a quality information is neglected
First result - rate of availability

- 7 indicators (5 group A, 2 group B)

<table>
<thead>
<tr>
<th>group</th>
<th>main sub-group</th>
<th>indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A1 production and diffusion of ICT</td>
<td>investment in ICT / GDP</td>
</tr>
<tr>
<td>2</td>
<td>A2 skilled human resources</td>
<td>total researchers / thousand labour force in FTE</td>
</tr>
<tr>
<td>3</td>
<td>A3 knowledge production</td>
<td>BERD / GDP</td>
</tr>
<tr>
<td>4</td>
<td>A4 entrepreneurship</td>
<td>enterprise churn</td>
</tr>
<tr>
<td>5</td>
<td>A4 structural and organisational change</td>
<td>% SMEs reporting non technological change</td>
</tr>
<tr>
<td>6</td>
<td>B1 economic outputs</td>
<td>labour productivity / hour worked</td>
</tr>
<tr>
<td>7</td>
<td>B2 social performance</td>
<td>inequality of income distribution</td>
</tr>
</tbody>
</table>

- one point in time
- general information offered by the metadata
- no separation by countries
### Rate of availability

<table>
<thead>
<tr>
<th>Economic indicator</th>
<th>Relevance R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>Accuracy A1</th>
<th>A2</th>
<th>A3</th>
<th>Timeliness &amp; Punctuality TP</th>
<th>Accessibility &amp; Clarity AC</th>
<th>Comparability C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>Coherence CH</th>
<th>Cost &amp; Burden CB</th>
<th>Extra-ordinary EX</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.313</td>
<td>0.129</td>
<td>0.667</td>
<td>0.167</td>
<td>0.250</td>
<td>0.250</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.625</td>
<td>0.290</td>
<td>0.667</td>
<td>0.167</td>
<td>0.500</td>
<td>0.500</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.625</td>
<td>0.290</td>
<td>0.667</td>
<td>0.167</td>
<td>0.500</td>
<td>0.500</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.563</td>
<td>0.258</td>
<td>0.333</td>
<td>0.167</td>
<td>0.500</td>
<td>0.500</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.750</td>
<td>0.484</td>
<td>0.667</td>
<td>0.167</td>
<td>0.750</td>
<td>0.500</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.313</td>
<td>0.161</td>
<td>0.667</td>
<td>0.167</td>
<td>0.375</td>
<td>0.188</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.313</td>
<td>0.161</td>
<td>0.667</td>
<td>0.167</td>
<td>0.375</td>
<td>0.250</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
I Knowledge Economy Indicators
II Existing quality information
III Quality measurement

Rate of availability - discrimination ability
Rate of availability

→ equal weights of the quality indicators

<table>
<thead>
<tr>
<th>economic indicator</th>
<th>accuracy</th>
<th>quality indicator</th>
<th>comparability</th>
<th>quality ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A2</td>
<td>A3</td>
<td>timelines &amp; punctuality</td>
<td>accessibility and clarity</td>
</tr>
<tr>
<td>1</td>
<td>2.0</td>
<td>1.0</td>
<td>4.5</td>
<td>1.0</td>
</tr>
<tr>
<td>2</td>
<td>5.5</td>
<td>5.5</td>
<td>4.5</td>
<td>5.0</td>
</tr>
<tr>
<td>3</td>
<td>5.5</td>
<td>5.5</td>
<td>4.5</td>
<td>5.0</td>
</tr>
<tr>
<td>4</td>
<td>4.0</td>
<td>4.0</td>
<td>1.0</td>
<td>5.0</td>
</tr>
<tr>
<td>5</td>
<td>7.0</td>
<td>7.0</td>
<td>4.5</td>
<td>7.0</td>
</tr>
<tr>
<td>6</td>
<td>2.0</td>
<td>2.5</td>
<td>4.5</td>
<td>2.5</td>
</tr>
<tr>
<td>7</td>
<td>2.0</td>
<td>2.5</td>
<td>4.5</td>
<td>2.5</td>
</tr>
</tbody>
</table>
### Random weights

<table>
<thead>
<tr>
<th></th>
<th>A2 A3 sampling errors</th>
<th>A3 non sampling errors</th>
<th>timelines &amp; punctuality TP</th>
<th>comparability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>accuracy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min.</td>
<td>2.51E-07</td>
<td>3.30E-07</td>
<td>4.03E-07</td>
<td>1.28E-06</td>
</tr>
<tr>
<td>1st Qu.</td>
<td><strong>1.11E-01</strong></td>
<td>1.11E-01</td>
<td>1.11E-01</td>
<td>1.11E-01</td>
</tr>
<tr>
<td>Median</td>
<td>2.00E-01</td>
<td>2.00E-01</td>
<td>2.00E-01</td>
<td>2.00E-01</td>
</tr>
<tr>
<td>Mean</td>
<td>2.00E-01</td>
<td>2.00E-01</td>
<td>2.00E-01</td>
<td>2.00E-01</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>2.77E-01</td>
<td>2.77E-01</td>
<td>2.77E-01</td>
<td>2.77E-01</td>
</tr>
<tr>
<td><strong>comparability</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>geographica comparability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min.</td>
<td>8.30E-07</td>
<td>6.30E-07</td>
<td>1.28E-06</td>
<td>7.74E-08</td>
</tr>
<tr>
<td>1st Qu.</td>
<td><strong>1.11E-01</strong></td>
<td>1.11E-01</td>
<td>1.11E-01</td>
<td>1.11E-01</td>
</tr>
<tr>
<td>Median</td>
<td>2.00E-01</td>
<td>2.00E-01</td>
<td>2.00E-01</td>
<td>2.00E-01</td>
</tr>
<tr>
<td>Mean</td>
<td>2.00E-01</td>
<td>2.00E-01</td>
<td>2.00E-01</td>
<td>2.00E-01</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>2.77E-01</td>
<td>2.77E-01</td>
<td>2.77E-01</td>
<td>2.77E-01</td>
</tr>
<tr>
<td>C2</td>
<td>comparability over time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min.</td>
<td>8.30E-07</td>
<td>6.30E-07</td>
<td>1.28E-06</td>
<td>7.74E-08</td>
</tr>
<tr>
<td>1st Qu.</td>
<td><strong>1.11E-01</strong></td>
<td>1.11E-01</td>
<td>1.11E-01</td>
<td>1.11E-01</td>
</tr>
<tr>
<td>Median</td>
<td>2.00E-01</td>
<td>2.00E-01</td>
<td>2.00E-01</td>
<td>2.00E-01</td>
</tr>
<tr>
<td>Mean</td>
<td>2.00E-01</td>
<td>2.00E-01</td>
<td>2.00E-01</td>
<td>2.00E-01</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>2.77E-01</td>
<td>2.77E-01</td>
<td>2.77E-01</td>
<td>2.77E-01</td>
</tr>
</tbody>
</table>
Ranking of economic indicators

→ random weights

I Knowledge Economy Indicators
II Existing quality information
III Quality measurement
Conclusions

• economic indicators are not sensitive for any the kind of existing quality

  → important to have any quality information possible?
  → equivalence of information?
    - sampling design
    - sampling error
    - geographical comparability

• factor analysis with principal component

  → importance of single quality information
  → factors having influence on the ability of quality information
  → recommendation on the quality provision
Factor analysis with principal component

→ correlations between the specific quality dimensions

<table>
<thead>
<tr>
<th></th>
<th>sampling errors</th>
<th>non sampling errors</th>
<th>timeliness &amp; punctuality</th>
<th>geographical comparability</th>
<th>comparability over time</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>sampling errors</strong></td>
<td>1.0000000</td>
<td>0.93335277</td>
<td>-0.14978617</td>
<td>0.9032446</td>
<td>0.9435133</td>
</tr>
<tr>
<td>non sampling errors</td>
<td>0.9333528</td>
<td>1.00000000</td>
<td>-0.01675054</td>
<td>0.9797979</td>
<td>0.7791756</td>
</tr>
<tr>
<td>timeliness &amp; punctuality</td>
<td>-0.1497862</td>
<td>-0.01675054</td>
<td>1.00000000</td>
<td>-0.1005038</td>
<td>-0.3499482</td>
</tr>
<tr>
<td>geographical comparability</td>
<td>0.9032446</td>
<td>0.97979794</td>
<td>-0.10050378</td>
<td>1.0000000</td>
<td>0.7548264</td>
</tr>
<tr>
<td>comparability over time</td>
<td>0.9435133</td>
<td>0.77917561</td>
<td>-0.34994824</td>
<td>0.7548264</td>
<td>1.0000000</td>
</tr>
</tbody>
</table>
Factor analysis with principal component

<table>
<thead>
<tr>
<th>economic indicator</th>
<th>equal weights</th>
<th>random weighting</th>
<th>quality ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>median</td>
<td>Median</td>
<td>result of the factor analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>component number</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5 quality indicators</td>
</tr>
<tr>
<td>1</td>
<td>2.0</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>5.5</td>
<td>5.5</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>5.5</td>
<td>5.5</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>4.0</td>
<td>4.0</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>7.0</td>
<td>7.0</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>2.5</td>
<td>2.0</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>2.5</td>
<td>3.0</td>
<td>3</td>
</tr>
</tbody>
</table>
Factor analysis with principal component

- quality ranking of the economic indicators is verified
- ranking between indicator two and three is more evident
- standard deviation
  → one component contribute the main explanation
    - proportion of variance:
      74% (5 quality aspects) to 91% (5 quality aspects)
- first hypothesis
  → the existence of methodological manuals facilitate the supply with quality information
Result

- analysis of quality information supply
  - first adequate method to build a ranking of indicators
  - basis of indicators selection
- methodological manuals: adequate instrument to perform and facilitate the availability of quality information
  - may be confirmed by an analysis of quality information on the country level
- exception
  - some information (e.g. sampling errors) as well as the data availability are a must for developing composite indicator
### Data availability

<table>
<thead>
<tr>
<th>Year</th>
<th>94</th>
<th>95</th>
<th>96</th>
<th>97</th>
<th>98</th>
<th>99</th>
<th>00</th>
<th>01</th>
<th>02</th>
<th>03</th>
<th>04</th>
</tr>
</thead>
<tbody>
<tr>
<td>AU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cardiff, 25/04/2006
Outlook

• analysis of the availability of quality information
  → for all indicators
  → on the country level
• analysis of data availability
• evaluation of a composite indicator
‘Benefit of the doubt’ composite indicators

Laurens Cherchye‡
Willem Moesen‡
Nicky Rogge‡*
Tom Van Puyenbroeck‡*

(‡): Centre for Economic Studies, Catholic University of Leuven
Naamsestraat 69, 3000 Leuven, Belgium
(*): European University College, Stormstraat 2, 1000 Brussels

Octobre 2006

Abstract

Despite their increasing use, composite indicators remain controversial. The undesirable dependence of countries’ rankings on the preliminary normalization stage, and the disagreement among experts/stakeholders on the specific weighting scheme used to aggregate sub-indicators, are often invoked to undermine the credibility of composite indicators. Data Envelopment Analysis may be instrumental in overcoming these limitations. One part of its appeal in the composite indicator context stems from its invariance to measurement units, which entails that a normalization stage can be skipped. Secondly, it fills the informational gap in the ‘right’ set of weights by generating flexible ‘benefit of the doubt’-weights for each evaluated country. The ease of interpretation is a third advantage of the specific model that is the main focus of this paper. In sum, the method may help to neutralize some recurring sources of criticism on composite indicators, allowing one to shift the focus to other, and perhaps more essential stages of their construction.
1. INTRODUCTION

The mere variety of composite indicators reflects their recognition as tools for policy evaluation and communication. Yet despite their increasing prevalence, composite indicators remain the subject of controversy. The lack of a standard construction methodology, and particularly the inescapable subjectivity involved in their construction, are invoked by opponents to undermine their credibility. Subjective choices are indeed pervasive when answering the many questions bound up with a composite indicator (see Booyse, 2002): what is the overall phenomenon one purports to summarize; which sub-indicators should be included; how should they be aggregated; how to deal with missing or low quality data; to what extent can one assess how country rankings are influenced by all the foregoing questions, etc.?

We will take it here that summarizing is one of its two essential purposes, the other one being the idea of comparing several countries (or the evolution of a country over time, and the like). We will also take it that composite indicators bear, although limitedly, on public debate. Because they are so easy to use as communication tools, they inevitably do show up in media headlines and in press releases of well-respected international organizations, so at least increasing awareness of specific issues in society. In such cases, they often have an hit-parade appearance. And most probably, this feature only aggravates uneasy feelings about composite indicators in scholarly circles.

We immediately turn to the simplest form in which the composite index is formulated as a weighted average of the individual indicators:

$$CI_j = \sum_{i=1}^{m} w_{c,i} y_{c,i}$$

(1)

with $CI_j$ the composite index for country $j$, $y_{c,i}$ the (possibly normalized) value for country $j$ on indicator $i$ ($i = 1, \ldots, m$) and $w_i$ the weight assigned to indicator $i$. In general, weights are bounded in that $0 \leq w_{c,i} \leq 1$ and $\sum_{i=1}^{m} w_{c,i} = 1$. In the construction process, the lack of a standard methodology is often invoked by opponents to undermine the credibility of the composite indicators. A first typical issue of most CIs is that the sub-indicators are displayed in quite diverse measurement units. This may be problematic in that adding up apples and oranges has to be avoided. In fact, getting rid of measurement units —notably when these differ across dimensions— is one reason why CI practitioners employ normalization methods. However, this doesn’t really solve the problem. A first general remark is that normalization obscures the original purpose of the indicator: one is no longer summarizing the original data, but re-scaled scores, or distances to goalposts, or z-scores, and the like. Evidently, this also bears on the inter-country score comparisons. There is, however, an observation that is still more worrying. Keeping the weighting system fixed, the eventual rankings still depend on the particular (and so-called 'preliminary') normalization option taken. Ebert and Welsch (2004) criticize the dependency of eventual ranks on the normalization/aggregation procedure from a measurement-theoretic point of view. In a well-defined mathematical sense, a composite indicator is not meaningful when the resulting country ordering changes if the original data are transformed in such a way that there informational content is not fundamentally altered. In practice, however, most composite indicators are prone to precisely this deficiency. It is obvious that countries with lower rankings due to a specific normalization procedure may invoke this dependency to question the credibility and the use of composite indicators. Removing the requirement to normalize the data would eliminate this dependency and, thus, an important criticism.
A second issue relates to the weighting scheme used for aggregating the sub-indicators. Ideally, the sub-indicators should be weighted and combined in a manner reflecting the underlying structure of the evaluated phenomenon. Often, however, it is not at all clear what ‘paternalistic’ judgments to impute, especially since weighting information stemming from stakeholders is often characterized by strong inter-individual disagreements. Equal weighting, which is just a specific case of fixed weighting, is therefore regularly invoked as the standard in virtue of its simplicity (e.g. by Babbie, 1995). We, however, strongly dislike the use of equal weights because we believe its alleged simplicity often to be thoroughly misleading. In the absence of any specific knowledge we even very much doubt whether any fixed weighting scheme should be applied at all. The essential reason for this is actually the same as for the normalization issue, namely that country scores and rankings also depend on the specific weighting scheme. In practice, very frequently such fixed weighting schemes favor some countries while harming others inducing especially the latter ones to invoke this dependency to minimize the credibility of such rankings. Furthermore, we believe that the own specificity of each country should be taken into account as much as possible. Within this perspective, differential weighting may be desirable if not necessary to come to representative CIs.

The rest of this text discusses how Data Envelopment Analysis helps to overcome the issues just raised. This approach has already been applied to composite indicators in the context of policy performance assessment. For example, it has been used to gauge countries’ performance with regard to aggregate deprivation (Zaim, Färe and Grosskopf, 2001), to provide an alternative weighting system for the Human Development Index (Mahlberg and Obersteiner, 2001, Despotis, 2005), or as a generalized gauge for Sustainable Development (Cherchye and Kuosmanen, 2006). Especially in the European context, where tensions between the centre and member states may also bear on the precise way by which the latters’ policies are evaluated, the need for a flexible weighting system may be warranted. Indeed, besides academic contributions (e.g.: European Unemployment policy (Storrie and Bjurek, 2000), Social Inclusion policy (Cherchye, Moesen, Van Puyenbroeck, 2004), and Internal Market policy (Cherchye, Lovell, Moesen, Van Puyenbroeck, 2005)), the European Commission itself has used the technique to gauge member states’ performance with regard to the Lisbon objectives (European Commission, 2004, p. 376-378). In this paper, a miniature subset of the Technology Achievement Index (TAI) (after Desai et al. (2002), is used to provide illustrative examples. The reason for this is twofold. First, the TAI figures likewise in the Handbook on the construction of composite indicators of Nardo et al (2005) where the benefit of the doubt approach is briefly discussed. Second, the TAI acts similarly in the primer on the Benefit of the doubt by Cherchye et al. (2006) to illustrate various extensions on the methodology. The fact that we dispose of individual expert information about TAI-weights makes this application especially appealing in the current context.

Section 2 describes, for a non-specialist audience, Data Envelopment Analysis and the related Benefit of the Doubt method in more detail. Its possible elimination of the dependency of the results on preliminary normalization, and its characteristic of offering flexibility under the form of endogenous weighting, may well tone down some of the aforementioned criticisms on composite indicators. We will stress such fundamental intuitions and show some basic formulas, focusing less in this paper on technical/computational aspects of DEA. These are treated at length in various publications (see e.g. Cooper, Seiford and Zhu, 2004, or Zhu, 2003 for surveys). In section 3 we extend the basic model by appending “sub-indicator share restrictions”. Such restrictions can be interpreted as bounds for the importance of sub-indicators in the composite score. The approach allows for a straightforward pie-chart representation of composite indicators, with the total size of the pie indicating a country’s score, and the (bounded) pie shares indicating how each sub-indicator contributes to this overall value. Some different variants of
these ‘pie share’-restrictions are discussed and illustrated. Section 4 summarizes and offers some concluding remarks.

2. DATA ENVELOPMENT ANALYSIS AND “BENEFIT OF THE DOUBT”-WEIGHTING

Data Envelopment Analysis (DEA hereafter), initially developed by Charnes, Cooper and Rhodes (1978), is a (linear programming) tool for evaluating the performance of a set of peer entities that use (possibly multiple) inputs to produce (possibly multiple) outputs. The original question in the DEA literature is how one could measure each entity’s efficiency, given observations on input and output quantities in a sample of similar entities and, often, no reliable information on prices, in a setting where one has no knowledge about the ‘functional form’ of a production or cost function. However broad, one immediately appreciates the conceptual similarity between that problem and the one of constructing CIs, in which quantitative sub-indicators are available but exact knowledge of weights is not. Indeed, and unsurprisingly, the scope of DEA has broadened considerably over the last two decades, including macro-assessments of countries’ productivity performance (e.g. Kumar and Russell, 2002), and various applications to composite indicator construction (Cherchye et al., 2004, provide a list of such applications). In the latter context, the method has been labeled alternatively as the ‘Benefit-of-the-Doubt’-approach (after Melyn and Moesen (1991), who introduced it in the context of macroeconomic performance evaluation).

This label derives from one of DEA’s main conceptual starting points: (some) information on the appropriate weighting scheme for country performance benchmarking can in fact be retrieved from the country data themselves. Specifically, the core idea is that a good relative performance of a country in one particular sub-indicator dimension indicates that this country considers the policy dimension concerned as relatively important. Or, conversely, that a country attaches less importance to those dimensions on which it is demonstrably a weak performer relative to the other countries in the set. Such a data-oriented weighting method is justifiable in the typical CI-context of uncertainty about, and lack of consensus on, an appropriate weighting scheme. This perspective clearly marks a deviation from common practices in composite indicator construction. In the words of Lovell et al. (1995, p. 508): “Equality across components is unnecessarily restrictive, and equality across nations and through time is undesirably restrictive. Both penalize a country for a successful pursuit of an objective, at the acknowledged expense of another conflicting objective. What is needed is a weighting scheme which allows weights to vary across objectives, over countries and through time”.

Admittedly, some may interpret the latter quote as indicating that the cure of flexible weighting is even worse than the disease of fixed (and equal) weighting. A main objective of this and the following section is to show that this is not the case, for at least the following three reasons. First, the benefit-of-the-doubt weighting approach is inherently bound up with the idea that even under such flexible weighting a country can be outperformed by some other country in the sample (see particularly expressions (2)-(4) below). Second, it is precisely due to the flexible nature of weights, i.e. because weights can adapt to the choice of measurement units, that the normalization problem of composite indicators may be sidestepped. In DEA literature this property is commonly referred to as unit invariance. And, last but not least, in cases where additional, even rough information on appropriate weights is available, this can often easily be incorporated into the evaluation exercise (see section 3). In sum, the method may go some

We will not provide a formal proof of this statement here (see e.g. Cooper, Seiford, Tone, 2000, p. 39), but the underlying intuition should be clear: the fundamental reason for this unit invariance goes back to the feature that weights are endogenous. Endogeneity implies flexibility, and this in turn will cause weights to adapt to the units of measurement.
length in providing a practical means of implementing the idea expressed by Foster and Sen (1997, p. 206): “while the possibility of arriving at a unique set of weights is rather unlikely, that uniqueness is not really necessary to make agreed judgments in many situations.”

In what follows, we will present the benefit-of-the-doubt formula in a step-wise fashion, in order to convey its underlying intuition clearly. As stated in the introduction, the eventual purpose of composite indicators is to compare a country relative to the other countries in the set and/or to some external benchmark. The first step highlights this benchmarking objective: a country \( c \)’s composite index score is not given by a weighted sum of its sub-indicators (as is done in (1)), but rather by the ratio of this sum to a (similarly weighted) sum of the benchmark sub-indicators \( y_i^B \).

Note that one thus introduces a quite natural “degree” interpretation for the CI-value: a value of 100% implies a global performance which is similar to that of the benchmark values, a value less (more) than 1 refers to worse (better) performance.

**Step 1: the benchmarking idea**

\[
I_c = \frac{\text{actual overall performance}}{\text{benchmark overall performance}} = \frac{\sum_{i=1}^{m} W_{c,i} Y_{c,i}}{\sum_{i=1}^{m} W_{c,i} Y_i^B}
\]  

(2)

The next question relates to the identification of benchmark performance. For the time being, we concentrate on the case in which benchmarks are to be taken from the observed sample itself. This option gives a clear meaning to the notion of best practice: the eventual CI-value will be driven by comparison with other, existing observations, rather than with external (and necessarily normative) references. In particular, the benchmark observation specified in the denominator of (3) is itself obtained from an optimization problem, as indicated formally by the appearance of the max operator and its associated argument. It is in fact a country that, employing the weights \( w_{c,i} \), obtains the maximal weighted sum. Consequently, this benchmark will be endogenous too: it may well differ from one evaluated country to another.

It should be noted that this selection yields further intuition to the CI-value of 1: if, for some reason or another, a country acts as its own benchmark (that is, if no other outperforming observation is found for this country), then we have in fact retrieved the maximal composite indicator value.

**Step 2: selecting a country-specific benchmark**

\[
I_c = \frac{\sum_{i=1}^{m} W_{c,i} Y_{c,i}}{\max_{y_j,c \{\text{studied countries}\}} \sum_{i=1}^{m} W_{c,i} Y_{j,i}}
\]  

(3)

The following step pertains to the specification of the appropriate weights. Here, the benefit of the doubt-idea enters. The weighting problem is handled for each country separately, and the country-specific weights accorded to each sub-indicator are endogenously determined. The conceptual basis for this option is the data-oriented perspective mentioned above: good relative performance of a country (i.e., relative to other observed countries) on a sub-indicator dimension is considered to be revealed evidence of comparatively higher policy priority, while the reverse
position is taken for sub-indicators on which the country performs relatively poorly. Stated otherwise, since one doesn’t know a country’s true (policy) ‘weights’, one assumes that they can be inferred from looking at relative strengths and weaknesses. Specifically, this perspective entails that the analyst looks for country specific weights which make its composite indicator value as high as possible. In the absence of more verifiable information, this indeed means that each country is granted the benefit-of-the-doubt when it comes to assigning weights. To put it differently: any other weighting scheme than the one specified in (4) would worsen the position of the evaluated country vis-à-vis the other countries. Countries cannot claim that a poor relative performance is due to a harmful or unfair weighting scheme. Formally, this point is covered by the new max operator in equation (4). It also follows that this problem must be solved (separately) for each of the countries.

Step 3: selecting country-specific benefit-of-the-doubt weights

$$I_c = \max_{w_{c,i}} \left\{ \frac{\sum_{i=1}^{m} W_{c,i} Y_{c,i}}{\max_{y_{j,e}[\text{studied countries}]} \sum_{i=1}^{m} W_{c,i} Y_{j,i}} \right\}$$  \hspace{1cm} (4)

Two more features are added. One is a normalization constraint (5a), stating that no other country in the set has a resulting composite indicator greater than one when applying the optimal weights for the evaluated country. Being a scaling constraint, the precise value of this upper bound is, of course, arbitrary. Yet, once again, (5a) highlights the benchmarking idea: the most favorable weights for one country are always applied to all \(n\) observations. One is in that way effectively looking which of the countries’ sub-indicator values are such that they would lead to a worse, similar, or… better composite score, when applying the most favorable weights for the evaluated country. If there are indeed countries in the third class, a strong case can be made for the notion of ‘being outperformed’: despite the fact that one allows for country-specific benefit-of-the-doubt weights, there is then still at least one other country which, using the same weighting scheme, does even better.

Constraint (5b) limits the weights to be non-negative. Hence, the composite indicator is a non-decreasing function of the sub-indicators, and the total composite indicator value is bounded below as well. That is, \(0 \leq I_c \leq 1\) for each country, where higher values represent a better overall relative performance.

$$I_c = \max_{w_{c,i}} \left\{ \frac{\sum_{i=1}^{m} W_{c,i} Y_{c,i}}{\max_{y_{j,e}[\text{studied countries}]} \sum_{i=1}^{m} W_{c,i} Y_{j,i}} \right\}$$ \hspace{1cm} (4, repeated)

s.t.

\[
\sum_{i=1}^{m} W_{c,i} Y_{j,i} \leq 1 \hspace{1cm} \text{(5a) \(n\) constraints, one for each country \(j\))}
\]

\[
w_{c,i} \geq 0 \hspace{1cm} \text{(5b) \(m\) constraints, one for each indicator \(i\))}
\]

\(^2\)The benefit of the doubt weights can be connected to a game-theoretic set-up: they can be conceived of as Nash equilibria in an evaluation game between a regulator and an organization. See e.g. Semple (1996).
Considering the fact that, by construction, the benchmark observation attains the maximal composite indicator value of 1, the above (fractional) maximization problem can be written in a linear form, which is computationally easier to handle (e.g. by Excel-solvers, e.g. Zhu (2003)):

\[ I_c = \max_{w_{c,j}} \sum_{i=1}^{m} w_{c,j} y_{c,i}, \]

subject to constraints (5a) and (5b).

As stated above, this method is rooted in DEA. It is indeed easily verified that the model just presented is formally tantamount to the original input oriented DEA model of Charnes et al. (1978), with all sub-indicators considered as outputs and a ‘dummy input’ equal to one for all the countries. In that reading, the dummy input for each country may be interpreted as a ‘helmsman’ that pursues several policy objectives, corresponding to the different sub-indicators; see e.g. Lovell et al. (1995). Still, it should be clear from our discussion that an intuitive interpretation may also be obtained simply by regarding the model as a tool for aggregating several sub-indicators of performance, without explicit reference to the inputs that are used for achieving such performance. The problem is then indeed one in a “pure output setting” (a term coined by Cook, 2004), in which the normalization constraint (5a) is interpreted as a scaling or bounding condition (see also Cook and Kress, 1991, 1994). A valuable side-remark, which we will not pursue further in this paper, may emerge: the method just described is fully apt to deal with CI-construction in the prevailing case where input sub-indicators would appear along with achievement sub-indicators. In fact, the DEA-model of Zaim et al. (2001) exploits this characteristic.

3. SUB-INDICATOR SHARE RESTRICTIONS

Apart from the non-negativity of the weights (equation (5b)), the formal model hitherto discussed allows weights to be freely estimated in order to maximize the relative efficiency score of the evaluated country. (The weights are only restricted in that they must not make the final score exceed the upper limit of 1). The advantage of such flexibility is that it becomes hard for countries to argue that the weights themselves put them at a disadvantage. However, there are also disadvantages to this full flexibility. In some situations, it can allow a country to appear as a brilliant performer in a way that is difficult to justify. For example, if some zero weights are assigned, and if there is no prior information which backs up this possibility, some of the achievement indicators do not contribute to a country’s composite measure. One then faces the risk of basing ‘global’ performance on a small subset of all (and often meticulously selected) sub-indicators. Also, by allowing full freedom, resulting outcomes may in particular contradict prior views on weights (e.g. expert opinions). In practice, it is essential for the credibility and acceptance of composite indicators to incorporate the opinion of experts that have a wide spectrum of knowledge, to ensure that a proper weighting scheme is established. True as this may be, it is at the same time also true that, in the area of composite indicator construction, experts may (strongly) disagree about the precise value of the weights.

Fortunately, DEA models are able to incorporate such prior information by adding additional restrictions to the basic problem. This seems especially convenient in the common case where experts disagree on weights. In all probability, this is exactly the setting where the benefit of the doubt approach to CIs seems to be most powerful. When individual expert opinion is available, but when experts disagree about the right set of weights, the method is sufficiently flexible to
incorporate ‘agreed judgments’ by imposing additional (e.g., sub-indicator share) restrictions. And at the point where disagreement remains, i.e. literally where no further restrictions can be imposed, the informational gap is filled by choosing country-specific benefit-of-the doubt weights. In our opinion, and with an eye towards practical applications, the latter reasoning may as well be reversed, so as to be more in line with the remark of Foster and Sen (1997) cited in section 2.1. That is: it is easier to let experts agree a priori on restrictions than on a unique set of weights. The final result would then reflect what is actually there: limited agreement. Evidently, the nature of such restrictions can vary, and we will now briefly survey some alternatives.

Following the unit invariance inherent in DEA, one should be cautious when comparing and interpreting benefit of the doubt weights as they adapt to the units of measurement. Also, if one would impose additional restrictions on the weights (i.e., in addition to (5b)), it may well be difficult to give an instantly recognizable meaning to such restrictions. One escape route is feasible, namely to shift the focus to ‘sub-indicator shares’, which are completely independent of measurement units. Sub-indicator shares are in fact the product of the original value of the sub-indicator $y_{i,j}$ and the assigned weight $w_{i,j}$ $^3$. Referring back to equation (6), the eventual composite indicator can thus be re-interpreted as a sum of $i = 1,\ldots,m$ sub-indicator shares, one for each achievement dimension. Clearly, these $m$ terms may also be interpreted as the ‘pie shares’ that together constitute $I$: the $i$-th term represents the volume of the pie share of the $i$-th sub-indicator. The total volume of the pie accordingly captures a country’s composite indicator score, and the relative size of the shares reflects what we have earlier referred to as the relative importance/significance of the sub-indicators. In what follows, we mainly focus on restrictions on the ‘pie shares’. All such restrictions are integrated in the original benefit of the doubt framework by adding the additional constraints to the programming problem. In view of the pie share interpretation, restrictions on sub-indicator shares allow for an easy and natural representation of prior information about the importance of the CI’s components.

<table>
<thead>
<tr>
<th>Table 1: Types of pie share restrictions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Absolute Sub-indicator share restrictions</strong></td>
</tr>
<tr>
<td>$\alpha_i &lt; w_{i,j} y_{i,j} \leq \beta_i$</td>
</tr>
<tr>
<td><strong>Ordinal Sub-indicator share restrictions</strong></td>
</tr>
<tr>
<td>$w_{j,6} y_{j,6} \leq w_{j,5} y_{j,5} \leq w_{j,4} y_{j,4} \leq w_{j,3} y_{j,3} \leq w_{j,2} y_{j,2} \leq w_{j,1} y_{j,1}$</td>
</tr>
<tr>
<td><strong>Relative Sub-indicator share restrictions</strong></td>
</tr>
<tr>
<td>$\alpha_i \leq \frac{w_{j,i} y_{j,i}}{w_{j,i} y_{j,i}} \leq \beta_i$</td>
</tr>
<tr>
<td><strong>Proportional Sub-indicator share restrictions</strong></td>
</tr>
<tr>
<td>$\alpha_i \leq \frac{w_{i,j} y_{i,j}}{\sum_{i=1}^{m} w_{i,j} y_{i,j}} \leq \beta_i$</td>
</tr>
<tr>
<td><strong>Restrictions pertaining to category shares</strong></td>
</tr>
<tr>
<td>$\alpha \leq \frac{\sum_{i=1}^{m} w_{j,i} y_{j,i}}{\sum_{i=1}^{m} w_{j,i} y_{j,i}} \leq \beta$</td>
</tr>
</tbody>
</table>

$^3$ In the DEA literature, this concept is usually labelled a ‘virtual output’ (‘virtual input’). See especially Thanassoulis, Portella, and Allen (2004) for a discussion of virtual outputs (or pure weights, or exogenous benchmarks) as means to include value judgments in DEA.
In what follows, we briefly focus on the last two tabulated pie share restrictions. For a more extensive treatment of these and the other restrictions we again refer to the primer on benefit of the doubt by Cherchye et al. (2006). Wong and Beasley (1990) proposed the proportional restrictions to make it easier for the experts to quantify their opinion in terms of percentage values. These restrictions may be especially attractive in view of the fact that expert opinion is often collected by a ‘budget allocation’ approach, in which experts are asked to distribute (100) points over the different dimensions to indicate importance. The stated ‘weights’ (which actually are budget shares) are then very easy to incorporate, in the benefit-of-the-doubt model. The only remaining issue is then how to specify bounds, given the observed diversity over individual experts. In the illustrative example below, we specified the lower and upper bound by respectively the lowest and highest weight assigned over all the experts on that sub-indicator.

Figure 1 and table 2 show how all this combines into a graphical and tabular representation. The results are shown only for a subset of the countries. The figure reveals the benefit-of-the-doubt nature of the exercise: the relative importance of the pie shares/sub-indicators is different over the three countries considered. And, a fortiori, this holds for their absolute size.

**Figure 1: pie chart representation of benefit-of-the-doubt (TAI) index for selected countries**

Finland: 100.00 %  Norway: 73.24 %  Poland: 30.03 %

**Table 2: absolute values and percentage contributions to CI of sub-indicator shares**

<table>
<thead>
<tr>
<th>Country</th>
<th>Patents</th>
<th>Royalties</th>
<th>Internet</th>
<th>Tech. Export</th>
<th>Elec. Consumption</th>
<th>Schooling</th>
<th>Enrolment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finland</td>
<td>0.0500</td>
<td>0.0093</td>
<td>0.2000</td>
<td>0.2148</td>
<td>0.0000</td>
<td>0.2500</td>
<td>0.2759</td>
</tr>
<tr>
<td>Norway</td>
<td>0.0366</td>
<td>0.0000</td>
<td>0.1465</td>
<td>0.0659</td>
<td>0.1609</td>
<td>0.1831</td>
<td>0.0439</td>
</tr>
<tr>
<td>Poland</td>
<td>0.0159</td>
<td>0.0000</td>
<td>0.0060</td>
<td>0.091 0.0000</td>
<td>0.0161</td>
<td>0.0751</td>
<td>0.0890</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percentage Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finland</td>
</tr>
<tr>
<td>Norway</td>
</tr>
<tr>
<td>Poland</td>
</tr>
</tbody>
</table>

Table 2 provides more information. The upper part show the respective countries’ values of sub-indicator shares, which, as indicated, sum up to their composite score. One infers, e.g., that the absolute values of the pie shares of top-ranked Finland are not always bigger than those corresponding to the other countries that are listed. In fact, one further sees that the listed countries do not even make use of all sub-indicators to arrive at their (benefit of the doubt) score: for each country, at least one dimension is left out. The underlying ‘revealed evidence’-intuition for these observations is, again, that a country is not likely to put very much weight (and in the limit no weight at all) on dimensions in which it demonstrably has a comparative disadvantage relative to the performance of other countries in the sample. The lower part of the table shows
the percentage shares. Percentage contributions further reveal how each country is offered (some) leeway in assigning ‘importance’ to each of the components of the composite index. One notices some similarities, but some huge inter-country differences as well.

Often, composite indicators are constructed such that their sub-indicators can be classified in $p$ mutually exclusive categories $S_1, \ldots, S_p$. Each category then represents a certain orientation or focus of the evaluated phenomenon. Cherchye and Kuosmanen (2006), and Cherchye et al. (2005) show how this can be combined with weight restrictions. Here we apply this idea to restrictions on “category shares”. Imposing restrictions on these category shares involves a straightforward extension of earlier restrictions. Once more, the idea of imposing restrictions on categories arises from the common observation that it is difficult to define weights for individual sub-indicators. Again the gist of our argument holds: agreement on bounds on the level of categories is much simpler to obtain than specific weights for individual sub-indicators. Indeed, in most cases, focusing on the importance of key categories may allow one to obtain stakeholder consensus more swiftly. Imposing restrictions on categories may be taken as a first step in the quest for consensus among experts.

4. CONCLUDING REMARKS

We recall our starting point for proposing the benefit of the doubt methodology to construct composite indicators: due to insufficiently precise and probably unverifiable knowledge of the underlying structure of an evaluated composite phenomenon, uncertainty is inherent in the construction of composite indicators. The lack of a standard construction methodology, the disagreement among experts on the importance of the underlying indicators, etc., are just ways in which this uncertainty is manifested. But precisely these methodological aspects have been invoked to undermine the credibility of composite indicators. This defines a clear challenge for those who believe that composite indicators can be a useful tool for communicative purposes, as well as for those who believe that global comparisons of country performance and the closely related idea of benchmarking could eventually promote good policies. Cast against this general background, the preceding pages do certainly not offer a panacea for all problems bound up with composite indicator construction, but some aspects we touched upon may help to prevent getting bogged down in ‘merely’ methodological discussions.

Given the current stance of research in this area, the benefit-of-the-doubt approach has some virtues over other, current mainstream approaches to composite indicator construction. As we pointed out, its unit invariance allows us to transcend discussions on the undesirable impact of normalization on eventual country rankings. Its flexible approach to the weighting issue may downplay critical remarks on ‘imputed’ weighting systems. Thirdly, and importantly for practitioners, its fundamental interpretation and the concomitant country results are easy to convey (e.g. by using pie charts), a remark which also holds for the kind of information one seeks to distill from the expert community.
References


