First Workshop on Indicators in the Knowledge Economy

Tübingen, 3 – 4 March 2005

2005

List of contributors:

Professor Dr. Dr. h.c. mult. Eberhard Schaich, University of Tübingen
Dr. Ian Perry, DG Research, European Commission
Dr. Ralf Münnich, University of Tübingen
August Götzfried, EUROSTAT
Anthony Arundel, University of Maastricht, MERIT
Dr. Asterios Hatziparadissis, Ministry of Development, Greece
Dr. Bart Los, University of Groningen
Mikael Åkerblom, Statistics Finland
Dr. Stefano Tarantola, Joint Research Center, ISPRA
Hans-Olof Hagén, Statistics Sweden
Dr. Beat Hulliger, Swiss Federal Statistical Office
Professor Dr. Daniela Cocchi, University of Bologna
PD Dr. Susanne Rässler, IAB Nuremberg
PD Dr. Siegfried Gabler, ZUMA Mannheim

CIS8–CT–2004–502529 KEI

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

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

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

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

`http://kei.publicstatistics.net/`

Thursday, 3 March 2005

13.30h Opening of the KEI workshop

- Welcome address by the Rector of the University of Tübingen
  Professor Dr. Dr. h.c. mult. Eberhard Schaich
- Introductory address
  Dr. Ian Perry, DG Research, European Commission
- Overview of the KEI project by the co-ordinator of the KEI project
  Dr. Ralf Münnich, University of Tübingen
- Data Quality and Indicators
  August Götzfried, EUROSTAT

Introduction by the co-ordinator of the KEI project

In the context of the Sixth Framework Programme of the European Commission the project KEI (Knowledge Economy Indicators: Development of Innovative and Reliable Indicator Systems) started in September 2004. The KEI project is part of the Policy Orientated Research section of the specific programme Integrating and Strengthening the European Research Area.

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. The project will cover 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 KEI project will review existing concepts and definitions of the knowledge-based economy and its key components. It will develop main thematic areas in relation to the Lisbon and Barcelona objectives. KEI will then use these themes to classify existing indicators and thoroughly explore data and indicator quality issues. Gaps will be identified and the way forward will be mapped, identifying innovative approaches to improve the understanding and appraisal of the knowledge economy. Composite indicators will be analysed in detail using both statistical and participatory 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 extensively to investigate the robustness of indicators and the conclusions based on them. The study will evaluate the quality and accuracy of indicators and the underlying data and assess the innovative use of additional information to improve indicator quality.

The state-of-the-art analysis, as provided by KEI for the knowledge-based economy, will benefit other policy objectives of the European Union and Commission Services. It will contribute to a methodological framework for building effective measurements of interdisciplinary issues such as sustainability, employment, social cohesion, and economic disparities. KEI will also make recommendations for the design and use of statistical reference systems.

The KEI research will be conducted by five institutions, Eberhard-Karls University of Tübingen (Ralf Münnich), Germany; the Joint Research Centre of the European Commission in Ispra (Andrea Saltelli), Italy; the Katholieke Universiteit Leuven (Tom Van Puyenbroeck), Belgium; the University of Maastricht (Anthony Arundel), The Netherlands; as well as Statistics Finland (Mikael Åkerblom), Finland. The scientific and administrative coordination will be performed by Tübingen.

KEI will organise five workshops covering specialised project topics. External experts will be invited to complement KEI research activities. The first workshop in Tübingen aimed in giving an overview of the activities of the KEI project. Within four sessions different topics on the scientific area of indicators were carried out. After a welcome address by the Rector of the University of Tübingen, Professor Dr. Dr. h.c. mult. Eberhard Schaich, the project officer, Dr. Ian Perry gave an introductory address to the audience. In the following, the co-ordinator of the KEI project, Dr. Ralf Münnich presented an overview of the KEI project and August Götzfried reported about data quality on indicators. Session

I, *Indicators for the Knowledge Economy*, chaired by Mikael Åkerblom contained a contribution by Anthony Arundel, about the work on the first two workpackages, especially, the state-of-the-art in Knowledge Economy Indicators. Dr. Asterios Hatziparadissis served as a discussant. Within session II, *Other projects in the KEI area*, chaired by Professor Dr. Giuseppe Munda, Dr. Bart Los gave an overview of the EU project EU-Klems dealing with the creation of databases on productivity by industry for EU member states. Mikael Åkerblom reviewed the NESIS project. The main scope of this project was to take stock of existing new economy indicators and classify them along a schematic representation of the Lisbon Policy process. NESIS may be regarded as an input for the KEI project. *State-of-the-Art on composite indicators* was presented in session III, chaired by Andrea Saltelli and discussed by Hans-Olof Hagén. Dr. Stefano Tarantola contributed the presentation on the session topic. The 4th and final session on *Indicator methodology* with Dr. Ralf Münnich in the chair dealt with the specific statistical methodology on indicators and their accuracy presented by Dr. Beat Hulliger, on indicators on regional level, Professor Dr. Daniela Cocchi, as well as on missing data, contributed by PD Dr. Susanne Rässler. This session was discussed by PD Dr. Siegfried Gabler. Summaries of the above mentioned contributions will be presented in the following.

Dr. Ralf Münnich, co-ordinator of the KEI project, University of Tübingen

Welcome address by Professor Schaich

As the rector of the Eberhard Karls University, I am very glad that the first workshop of the EU project KEI „Knowledge Economy Indicators“, is taking place in Tübingen. Our University, one of the oldest in Germany, combines academic heritage and tradition with highly innovative and competitive research in many disciplines. The University also has a long tradition with economics and statistics, therefore, I am especially glad that the scientific and administrative coordination of the KEI project within the scope of the 6^{th} Framework Programme will be organized by the University of Tübingen. The 6^{th} Framework Programme provides the fields of economics and statistics with increasing opportunities to participate in projects funded by the European Union. Within this context, the overall objective of the KEI project is to support the formulation and implementation of community policies by providing scientific contributions to policies which are targeted precisely on the communities' needs, and which are coherent across the various community policy areas, while at the same time sensitive to changes in policies as they take place. Thus, the project provides the EU member states and the European Commission with a set of useful and defensible indicators for evaluating policy actions, as well as providing input to national statistical agencies on feasible upgrades to their information production systems.

Within the KEI project, five participating institutions cooperate successfully: the Joint Research Center of the European Commission in Ispra, Italy; the Katholieke Universiteit Leuven, Belgium; the University of Maastricht, the Netherlands, as well as Statistics Finland, Finland and the Eberhard Karls University Tübingen, Germany. Bundling the specific skills of all project partners promises best results for the project as a whole.

The Eberhard Karls University Tübingen welcomes the opportunities which the KEI project offers and will make every effort to support it. It is with this thought in mind that I wish all partners a fruitful cooperation on this project.

Professor Dr. Dr. h.c. mult. Eberhard Schaich, Rector of the University of Tübingen

Why KEI?

The KEI research project is one of the projects on economic and environmental issues from within the Scientific Support for Policies of Sixth Framework Programme for Research and Technological Development of the European Commission. The KEI research project addresses a very important issue where Commission services need scientific research input to help improve and widen the scope and relevance of data on the Knowledge Economy. The resulting new and improved indicators should be particularly useful for policy makers at the European and national levels and to others studying the Knowledge Economy. If we accept the concept of the Knowledge (Based) Economy and manage to define what it is we must then be able to quantify it and analyse it coherently. To be able to do this a good reliable and relevant set of indicators is required. Currently though there are many problems with KE indicators, KEI must address these problems. It must also try not only to increase the knowledge base but also address issues such as the reduction of complexity of information through for example work on composite indicators. Also it should try to make indicators on the Knowledge Economy easier for policy makers to understand and use. There are already many researchers and others working hard and doing valuable work trying to improve the situation, however, if we want to understand where we are today in the Knowledge Economy, how it has developed and what developments are taking place much more work is needed. This is where the KEI project has role to play in arriving at the Lisbon objectives.

Dr. Ian Perry, DG Research, European Commission

Data quality and indicators

The presentation illustrated first of all the framework for data quality issues set at Eurostat. The data quality work within the European Statistical System focussed on the definition of statistical quality (with a number of criteria set) , the drawing up of standard data quality reports, a glossary, standard data quality indicators and specific regulations heron. All this work is done under the responsibility of the Eurostat Working Group on „Quality in Statistics“. Examples on producer oriented data quality indicators were given which are linked to the general quality criteria as such.

In the second part, the implementation of the general Eurostat quality approach was illustrated on the base three examples: Structural Business Statistics, the European Union Labour Force Survey, and the Community Innovation Survey. Based on a however harmonised all over approach, the stage of work in those three domains is different. Finally, an outlook was given on how to continue the work on data quality within the European Statistical System. More statistical domains will be involved with more and better measurement of the data quality in the years to come.

August Götzfried, EUROSTAT

Thursday, 3 March 2005

15.15h Session I: Indicators for the Knowledge Economy

Chair: Mikael Åkerblom, Statistics Finland

- State-of-the-art in Knowledge Economy Indicators
  Anthony Arundel, University of Maastricht, MERIT
- Discussant
  Dr. Asterios Hatziparadissis, Ministry of Development, Greece

Presentation on Developing indicators for a knowledge based economy

Indicators for a knowledge based economy (KBE) need to meet two basic needs: to be of relevance for policy and to be of value over the medium-term future. Past assumptions about trends may not hold in the future, requiring a careful evaluation of the usefulness and interpretation of specific indicators.

As an example of the importance of a close evaluation of indicators, a frequent misconception is that ICT and other advanced technologies such as biotechnology have driven the increase in R&D in the United States in the late 1990s. Other than a temporary blip between 1995 and 1997, this is not true, with the share of all R&D due to ICT lower in 2000 than in the early 1990s. Conversely, the fastest growth in R&D intensities has been in low technology sectors, rather than in 'high' technology sectors such as ICT. The role of ICT is not as a driver of innovation, narrowly defined, but as an enabling technology that permits productivity-enhancing innovation in many industrial sectors.

Some of the main policy concerns over the medium term future are likely to revolve around socio-political drivers of the knowledge economy, rather than in simple economic models of innovation. The main socio-political drivers are possibly demographic change, environmental challenges, and globalization. The global component creates a major challenge for KBE indicators, which is how to link indicators at the level of the region or country with firm level indicators that can span the globe. As an example, the innovative capabilities of MNEs can depend on a web of activities based in multiple countries. Another example is the link between European demographics and the rapid economic development of India and China.

To give an example, demographic change in Europe could reduce the supply of skilled human resources, due to a declining cohort of university entrants combined with declining interest in science and technology. The most common policy solution is to tap into the global market for the highly skilled, for instance by adapting European immigration policies to attract highly skilled immigrants. However, once we add a global component to this scenario, this solution appears increasingly problematic. Indicators from the UK, the US, China and India suggest that 1) the major factor driving knowledge workers to move abroad is a lack of opportunities at home and 2) relevant opportunities in the two main donor nations (China and India) have been increasing rapidly. This suggests that the window of opportunity for meeting Europe's needs for knowledge workers through immigration is likely to be limited to the next decade.

These examples show that indicators for a KBE must account for future developments, extend beyond Europe, and consider the likely impact of major socio-political changes.

Anthony Arundel, University of Maastricht, MERIT

Discussion to Session I

The general frame and the methodology developed in the Work Packages are generally well defined. The strategic objective of the project must be the creation of indicators, which will help both the policy makers and the other users.

Concerning the characteristics of the Knowledge Based Economy (KBE) more emphasis must be given on the output indicators i.e. innovation than on the input ones, the R&D. The promotion of the innovation, which is an activity between the production and the economic exploitation of knowledge it is one of the major policies of the European Union (EU). From this point of view, the indicators on „innovation“, as they are produced in the frame of the CIS (Community Innovation Survey), are a very useful instrument for the measurement of the innovation activities in a harmonized base. The role of the ICT as a technological driver is treated adequately.

The socio-political drivers for a KBE include generally the right parameters. The demographic indicators, particularly those that contribute to the increase of the innovative capabilities of the EU, are very important. The indicators on HRST developed by the OECD and EUROSTAT provide useful information on that subject. The analysis should not be limited to the absolute values of the countries but should also cover the flows. The creation of favorable conditions for the attraction of high-level scientists constitutes a basic priority in the Revised Lisbon Strategy in 2005. Therefore the study of the different policies for attracting researchers from the other countries is an important subject. The study must include not only the policies and measures of the USA but should be enlarged to other countries with rich experience on the subject as Canada, Australia, the Scandinavian Countries etc. Recent data on the mobility of the scientists should be analyzed in depth because interesting changes concerning the mobility of scientists from the Asian to the USA and the Occidental countries are observed.

The procedure for identifying indicators is well targeted because it begins with key policy questions. Concerning the scenarios of the EU they should take into consideration the Lisbon Strategy and its evolution as well. The mid-term review, the Decisions of the Ministerial Councils and other changes should be monitored very closely because they express the political will of the EU on the achievement of concrete objectives. Most of the scenarios developed by international organizations and by individual scientists as well give to China a dominant role for the next decades and a quite limited role for India. However China is concentrated mainly on the manufacturing sector which, because of the technological progress, will have a declining trend in the medium- or long-term future, as it was the case of the primary sector some decades ago. On the other hand, India is very familiar with the ICT and the services related to them and which will be the dominant sectors for the next decades. It is therefore useful to develop a scenario that gives to India a more important role.

Dr. Asterios Hatziparadissis, Ministry of Development, Greece

Thursday, 3 March 2005

16.30h Session II: Other projects in the KEI area

Chair: Professor Dr. Giuseppe Munda, University of Barcelona and Joint Research Center, ISPRA

- EU-Klems
  Dr. Bart Los, University of Groningen
- NESIS
  Mikael Åkerblom, Statistics Finland

EUKLEMS Project on Productivity in the European Union

Objectives

The purpose of the EUKLEMS project is to create a database on productivity by industry for EU member states. Next to output (measured by gross output or value added), several inputs will be considered: capital (K), labor (L), energy (E), materials (M) and service inputs (S). Besides setting up the database, the project entails analytical research „subprojects“. Examples are:

- analysis of productivity, prices, industry structures, and technology and innovation indicators;
- effects of skills formation on productivity;
- effects of innovation on productivity;
- analysis of opportunities to link industry-level productivity research to firm-level productivity research.

The project's policy relevance mainly relates to the Lisbon and Barcelona agendas. It should assist the European Commission in evaluating whether the targets will be attained or not, and in which countries or industries additional efforts will be required. To this end, data for the other two major economic forces in the world (Japan and the US) will also be constructed. Special attention wil be paid to the performance of the ten new member states and their impact on the economic performance of the EU.

From an academic point of view, the project also offers interesting opportunities. Cross-country growth regressions seem to have entered the stage of diminishing returns. Data at a lower level of aggregation are needed to test hypotheses on, for example, skill-biased technological change and the role of organizational innovation. Several project members are also expected to continue contributing to the development of standards for productivity measurement and national accounting.

Organization

Like KEI, the EUKLEMS project is a project within FP6, Priority 8. The project started on September 1, 2004 and will finish on September 1, 2007. The contribution of the European Commission will be 5.0 M€. The project is carried out by fourteen participants, of which the University of Groningen (The Netherlands) and the National Institute of Economic and Social Research (UK) are in charge of the majority of coordinating issues. Many national statistical agencies are participating, because they are expected to be the main sources of required data material.

The project contains 11 workpackages: WP1-4 deal with the construction of data on interindustry transactions, labor accounts, capital flow accounts and relative price levels,

respectively. Their outputs serves as inputs to WP5, which comprises the actual development of the database. In WP6, a „statistical roadmap“ is devised and (if necessary) adapted to facilitate collaboration with national statistical institutes. Analyses of productivity growth, relative price changes and economic structures are comprised in WP7. Such analyses will serve as inputs to WP 8and 9, which deal with the effects of skill formation and innovation, respectively. The exploration of opportunities to link industry-level data to firm-level data is done in WP10. WP11 is set up to arrange communication and dissemination of data between participants, the European Commission and, in later stages, third parties.

Dr. Bart Los, University of Groningen

Inputs from NESIS to KEI

Scope of this note is to assist the KEI project team in taking stock of the NESIS project outcomes. Only deliverables that could provide an input to KEI are considered. The NESIS project was an accompanying measure to the FP5-IST program. The main scope of the project was to take stock of existing new economy indicators and classify them along a schematic representation of the Lisbon Policy process. NESIS activities were structured along the Pillar structure of the project. The paper starts with a schematic overview taken from D5.3 *Final Report on Conceptualization and analysis of the New Information Economy*.

The project produced two kinds of deliverables. The first kind are general inventories of available quantitative information. In this respect four deliverables provide information which are relevant for the KEI project. They are presented in the paper in some detail:

D 2.2.0 Available Indicators of the New Economy;
D 5.3 Final Report on the Conceptualisation and Analysis of the New Economy;
D 2.2.2 The EU-15's New Economy A statistical Portrait;
D 2.2.1.3 Improving the measurement of the New economy, recommendations to the ESS.

The results of pilot research deliverables relevant to KEI are presented next:

D 5.1.2 Final report, Conclusions about Knowledge-based indicators
D 5.6.3 The measurement of Knowledge Stocks and Flows in the New Economy
D 3.1.2 European policy indicators survey

In this note we tried to provide the KEI team with an overview of the results from NESIS project that could be of use for KEI. For a more complete overview one should refer to the original documents. The table below gives the contact persons list. It is also possible to refer to the project site `http://nesis.jrc.cec.eu.int/` and access the deliverables from the left hand side menu. Many of the topics treated in NESIS are not particularly relevant for KEI, but one could run a text search on the over 1,000 documents produced by the project.

Deliverable	Contact person
D2.2.0 Available Indicators of the New Economy	Teun Wolters TWLS@cbs.nl
D5.3 Final Report on the Conceptualisation and Analysis of the New Economy	Graham Room hssgjr@bath.ac.uk
D2.2.2 The EU-15's New Economy A statistical Portrait	Teun Wolters TWLS@cbs.nl
D2.2.1.3 Improving the measurement of the New economy, recommendations to the ESS	Mikael Åkerblom Mikael.Akerblom@stat.fi
D5.1.2 Final report, Conclusions about Knowledge-based indicators	Teun Wolters TWLS@cbs.nl
D5.6.3 The measurement of Knowledge Flows in the New Economy	Mikael Åkerblom Mikael.Akerblom@stat.fi
D3.1.2 European policy indicators survey	Riccardo Girardi Riccardo.girardi@jrc.it

Mikael Åkerblom, Statistics Finland

Friday, 4 March 2005

9.00h Session III: State-of-the-art on composite indicators

Chair: Andrea Saltelli, Joint Research Center, ISPRA

- State-of-the-art on composite indicators
  Dr. Stefano Tarantola, Joint Research Center, ISPRA
- Discussant
  Hans-Olof Hagén, Statistics Sweden

Composite Indicators - State of the Art

1. Pros and Cons of Composite Indicators

Composite indicators can be used to summarize complex or multi-dimensional issues in view of supporting decision makers. They help providing the big picture, can attract public interest and add a layer of information to the underlying list of indicators. On the other hand, they may send misleading, non-robust policy messages if they are poorly constructed or misinterpreted. Furthermore, the development of composite indicators involves stages where judgment has to be made which can lead more scope for disagreement among decision makers.

2. Examples

The list of existing composite indicators related to the knowledge economy include among others the European Innovation Scoreboard (EIS), Technology Achievement Index (TAI), Summary Innovation Index, e-Business readiness indicator, and Welfare of Nations Index.

3. Construction of Composite Indicators

The development of composite indicators comprises several important decisions to be made:

Selection: Statistical methods like principal component analysis and factor analysis can be used to derive the optimal composition of the final indicator.

Imputation: Missing data can comprise different patterns (missing completely at random, missing at random, not missing at random) that can be best dealt with different single and multiple imputation methods.

Normalization: Several methods are available to normalize the sub-indicators. The approaches most frequently adopted in the literature are based on standardized, re-scaled or raw values of the variables.

Weighting: Possible alternatives to the simplest approach of equal weighting are e.g. budget allocation (BAL), analytic hierarchy process (AHP) and benefit of the doubt (BOD).

Aggregation: Different aggregation methods can be applied. The most common are linear aggregation (LIN), geometric aggregation (GME) and multi-criteria analysis (MCA).

4. Uncertainty Analysis

Uncertainty Analysis (UA) focuses on how uncertainty in the input factors propagates through the structure of the composite indicator. The possible sources of uncertainty (e.g. selection, imputation, normalization, weighting, aggregation) can be assessed by applying Monte Carlo simulations.

5. Sensitivity Analysis

Sensitivity Analysis (SA) studies how much each individual source of uncertainty contributes to the output variance of the composite indicator.

Reference: **M. Saisana, S. Tarantola and A. Saltelli (2005)**, Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators, *Journal of the Royal Statistical Society A*, **168(2)**, 1-17.

Dr. Stefano Tarantola, Joint Research Center, Ispra

Discussion to Session III - Composite indicators controversy

1. It is on one hand quite understandable but on the other life is full of choices between quite different alternatives for example should I pick an interesting low income job or a dull high income job. And politics is definitive about choices of incomparable things as security, freedom and consumption. Why should we not produce figures and facts as a base for these decisions?

2. I do like that you have written on the subject, and I agree with you but I think that you have wanted to have more arguments for pros so you have made four out just one argument. And unfortunately all their five argument against CI is however valid.

3. I believe that the introducing of the model thinking give a new dimension to the CI-concept. And I really liked the statement by Box *all models are wrong, some are useful.*

4. The stress on negotiation is indeed very important.

5. The criteria for CI:s relevance, accuracy, credibility, timeliness, accessibility, interpretability and coherence is quite fine.

6. In the normalization discussion I miss the role of extreme values and the impact of these. If you have these it is necessary to test the robustness then you dismiss these.

7. In the correlation discussion you frequently use the example of that you are not prepared to trade speed and beauty of a car this. It is a quite illustrative example, but it is a very Italy too. A more Swedish example would be the choice between safety and reliability two factors that are essential to us. In your example you clearly show that two rather correlated indicator can differ substantially in relation to other indicators.

8. The discussion of compensability end out in the implicit recommendation of MCA multi-criteria procedure. I believe that is not any better to take away the possibility of that, since as I already have pointed out politics is about choices, and than it do matter how big the difference are, much more than the ranking as in some foot-ball league.

9. Analyses of the Robustness can't be stressed enough.

10. In the conclusion it is stressed the very import fact that the CI is the starting point.

Hans-Olof Hagén, Statistics Sweden

Friday, 4 March 2005

10.30h Session IV: Indicator methodology

Chair: Dr. Ralf Münnich, University of Tübingen

- On displaying indicators and their accuracy
  Dr. Beat Hulliger, Swiss Federal Statistical Office
- Indicators on regional levels
  Professor Dr. Daniela Cocchi, University of Bologna
- Handling missing data for indicators
  PD Dr. Susanne Rässler, IAB Nuremberg
- Discussant
  PD Dr. Siegfried Gabler, ZUMA Mannheim

On displaying indicators and their accuracy

Displays of indicators are addressed to politicians and managers with limited statistical knowledge and less time. The purpose of an indicator display is not to explain a problem but to give a quick overview and highlight salient features. The time devoted to the display may be 10 seconds on TV or 1 minute in a report. Therefore much care is needed to create easily readable graphics and to display at the same time as the indicator any information relevant to its interpretation, in particular its variance. Unfortunately most indicators are published without a variance estimate. E.g. the coefficient of variation of the expenditure of Swiss industry for Research and Development in 2000 was estimated at 5% but was not published!

Bar charts for categories and line plots for time series are the most useful and most frequent displays for indicators. The pie chart fortunately is now rarely used because of its sever perceptual problems. However the good composition of displays is still a challenge. For example there are still many time series displayed as bar charts.

The working horse of depicting variance is the confidence interval. However, the main question of interest usually is a test, for example when comparing a national indicator with the EU-indicator. When comparing two indicators the so-called overlap test (checking whether the individual confidence intervals overlap) is conservative for independent indicators and may often be used safely. Multiple comparisons are, of course, an issue but no simple overall solution to prevent over-interpretation seems at hand.

For bar charts the confidence interval may be depicted by a (white) rhombus at the top of the (coloured) bar. We call this plot „candle plot“. The rhombus makes the confidence interval more prominent than whiskers. Its surface is half of a corresponding confidence bar. For stacked bar charts the problem of dependence of the proportions is an additional problem. The confidence interval for the cumulative proportions may be still be depicted and seems to be useful.

When comparing two or more time series at specific time points the individual confidence intervals at the time points are appropriate. However their endpoints should not be connected since this would create a false impression of a joint confidence band around the line of a time series.

To test the evolution of a time series funnels may be plotted which start at the indicator at time t and open up to the confidence interval for the difference between time $t+1$ and time t centred at the indicator of time $t+1$. The confidence interval of the difference should take into account the correlation. If the funnel does not cover a horizontal line from the indicator at time t then the change from t to $t+1$ is significant. Of course this is just one possibly interesting comparison among many. It is not possible to plot funnels for all possible tests.

Finally note that a ranking in the form of a league table is a poor statistical summary in spite of its wide acceptance by the media and the public. Displaying the corresponding performance indicators as a series of bars in their order of magnitude gives, in addition to the ranking, the visual information on the quantitative differences between the indicators.

Adding a confidence interval, e.g. with a candle plot for each bar, it may become clear that much of the ado about a particular ranking is, in fact, about nothing.

Dr. Beat Hulliger, Swiss Federal Statistical Office

Indicators on Regional Levels

The demand for statistics helping in decision at local level is increasing and reliable regional level indicators are often requested. Important differences occur according to whether these indicators are conceived according to a top-down or a bottom-up strategy. Direct estimators (ratio, post-stratified ratio, forms of weighting) can be computed using only the sample information coming from the domain under study, but the variance of such estimators may become too high.

Small area estimation tools may be useful for constructing indicators at regional level. These techniques are proposed for estimation in geographical areas or population subgroups when the sample size is insufficient for obtaining stable results. Improvements in estimation are due to using (borrowing) information out of the domain, by means of the so-called synthetic estimators. Synthetic estimators can be computed using information external to the domain, exploiting the idea of similarity (partial or total exchangeability) between domains. The construction of the estimator starts from a direct estimator for each domain and proposes a link among them. The nature of the link depends on the available information. For instance, information can be borrowed across space (cross-section), about aggregations of domains containing the small area or about the whole population (borrowing information, or smoothing, or shrinking across space). Alternatively, it can be borrowed from time series information within the small area, or time and space information are considered together. Additivity of estimates obtained for sub-domains of a greater domain ought to be possible. Auxiliary information can be enclosed in the estimator. Possible negative characteristics of small area estimators are the reduction of variance at the price of bias or the lack of robustness if the implicit or explicit model does not hold.

Most small area estimators suggested in the literature may be expressed as „composite estimators“. Design based small area estimators can be proposed, namely the optimal composite estimators, where weights are obtained minimizing the MSE of the composite estimator, or the generalized regression estimator (GREG).

As a different alternative, model-based estimators are proposed, where models include specific small area effects. Hierarchical modeling is a tool which integrates well with the solution to other problems dealt in this session: evaluation of accuracy and treatment of missing values. Two possible versions can be proposed: genuine hierarchical models or Linear Mixed Models. Small area estimators are obtained from these models by applying one of the following methods: empirical best-linear unbiased prediction (EBLUP), empirical Bayes (EB) and hierarchical Bayes (HB). We point out the flexibility of hierarchical model based inference under the Bayesian paradigm. Two different examples are illustrated: the first one at the area level and the second one at the unit level.

For area level hierarchical models, Bayesian solutions can be proposed, where the posterior means of second level model parameters are seen as composite estimators. A hierarchical model for local census undercount in Italy is illustrated, where evidence is found that that the municipality population size influences census undercount. The guess that covariates more useful than the geographical area could be used for the design of a new PES and its subsequent analysis and that physical contiguity may be unimportant is confirmed by the

analysis. A main difficulty in model construction has been how to deal with heterogeneity, which had individual and municipal components.

As an example of LMM we illustrate the estimation for the Average Equivalent Household Income at the level of NUTS2 and NUTS3 regions, using data from the European Community Household Panel (ECHP) survey. This variable is the amount of income that an individual, living alone, should dispose of to reach the same level of economic welfare he enjoys in his household.

A set of covariates on household characteristics whose area-level totals can be obtained from Census results is available, but their predictive power is moderate. The check is made by comparing EBLUP estimators associated to unit level Linear Mixed Models and robust design-based alternatives like the optimal composite estimator and the generalized regression estimator (GREG).

Professor Dr. Daniela Cocchi, University of Bologna

Missing data

Missing data is a problem nearly everybody dealing with survey statistics has to deal with. A common procedure in dealing with missing data is deleting all observations containing missing items but this can lead to a remarkable loss of the available data and biased estimates. Other possibilities include weighting procedures, model-based correction of parameter estimates, or single imputation procedures combined with correction of the variance estimates to allow for the uncertainty that comes with imputation. However, the most flexible approach for multipurpose complex surveys as analyzed in KEI seems to be multiple imputation (MI), since this method is designed to account for the uncertainty about which value to impute and allow standard complete-data analyses of the imputed data.

When using MI as approach for dealing with missing data, m data sets are independently randomly drawn to round off the incomplete data. Based on each set, values for the parameter of interest and its variance ($\widehat{\theta}^{(m)}$ and $\widehat{\text{var}}(\widehat{\theta}^{(m)})$) are estimated. The parameter we are interested in is then calculated as the mean of the m estimators $\widehat{\theta}^{(m)}$, its variance can be calculated via two components of variances, one defined by the estimators $\widehat{\text{var}}(\widehat{\theta}^{(m)})$, the other caused by the uncertainty implemented in the MI method.

The project KEI deals with numerous indicators from EU-countries, the USA, and Japan for a period of several years beginning in 1995. So missing indicators are to be expected for some combinations of countries and years but first of all at the most recent point in time since only estimates of the indicators are available for the last year under consideration.

The model proposed to be used in KEI is a multivariate linear mixed-effects model. The general approach applied for KEI has already been applied by Rubin (2003) in the National Medical Expenditure Survey (NMES). Its iterative procedure can be described as follows:

- Facing a data set with - for the sake of better illustration only - three variables (A, B and C) each with missing data, we begin with arbitrarily filling in all missing values for two of the variables, B and C.
- In a next step we fit a model for A given B and C where A is observed and then impute the missing values of A.
- Now the imputed values for B and C are replaced by fitting models for these variables like for A in the step before and imputing the missing values regarding the conditional models.
- These steps are iteratively repeated until a satisfying solution results, guaranteing great flexibility due to the conditional operations.

Considering the characteristics of KEI missing indicators can be assumed being missing at random (MAR). The mixed-effects models are fitted for each indicator separately:

$$y_i = X_i\beta + Z_i b_i + \epsilon_i \quad ,$$

where y_i is the KEI indicator number i and the right hand side explains the various fixed and random effects including the random error ϵ_i.

First results with the KEI data based on $m = 10$ imputations are already available. Important in the context of this project is the provision for correlation between the indicators and heteroscedasticity. Both problems can be dealt with using the approach above and tools already available. Further research is to be done to allow for more flexible serial autocorrelation and for spacial autocorrelation since missing values can be expected to be linked to their neighbors in previous or following years as well as in similar, bordering countries.

Reference: **D.B. Rubin (2003)**, Nested multiple imputation of NMES via partially incompatible MCMC , *Statistica Neerlandica* , **57(1)**, 3-18.

PD Dr. Susanne Rässler, IAB Nuremberg

Discussion to Session IV

First of all I would like to thank all three speakers for their interesting presentations which may help to investigate indicators from a statistical viewpoint.

Graphical representation of indicators is usually provided only in form of bar charts if at all. Beat Hulliger mentioned that the purpose of displays of indicators is to help policy makers with limited statistical knowledge and less time to understand easily the representation, to give a quick overview of the indicator and to highlight salient features. In principle, I agree with Beat that at least variability should be displayed. But the computation of variances for a composite indicator is not an easy task because of considering different time points or different sample designs. The proposed candle and funnel plots are fancy and it is time to ask the user of their usefulness.

There is an increasing demand for statistics helping for decision at the local level. Small area estimation (SAE) is a good method for estimating indicators in small areas if only a few sample data in those areas are available. The question is the type of estimator (direct, synthetic or composite), design-based or model-based and the method someone should use, e.g. empirical best-linear unbiased prediction, empirical Bayes or hierarchical Bayes maybe with Linear Mixed Models. If one uses Bayesian methods one needs an a-priori distribution and the question arises which one the most reasonable is considering indicators. The two examples presented by Daniela Cocchi showed the possible application of SAE.

Multiple imputation (MI) is a method for filling in missing data with plausible values which is more and more applied nowadays. Susanne Rässler is an expert on MI. One difficulty may be to convince a user that he needs not only one full dataset but maybe 15. Critics (Fay 1992) have demonstrated that variance estimates based on multiple imputation may be inconsistent or require the complex assumption that the imputation is proper which is sometimes hard to ascertain.

Reference: **R.E. Fay (1992)**, When are inferences from multiple imputation valid?, *Proceedings of the Survey Research Methods Section. American Statistical Association, Alexandria*, 227-232.

PD Dr. Siegfried Gabler, ZUMA Mannheim

Acknowledgements

Ladies and Gentleman, dear participants of this first workshop,

Please allow me some final remarks now that our first KEI workshop is close to its end.

Of course, I can only speak for myself but I think that this first official workshop of our project has been very productive and informative. I hope all those of you not involved in KEI as partners got some idea of what may have driven the KEI partners to dedicate a considerable part of their time for the next months to the analysis of composite indicators.

I would like to thank all of you for being here and for your interest in the work of KEI. Especially I want to thank all speakers and discussants for their interesting contributions to this workshop.

And last but not least I want to express my gratitude to all those people who made this workshop possible working in the background, who helped in organizing and carrying out this meeting.

For now, I wish you all the best - first of all a good and save trip home. Especially for those having to take the plane I hope they do not face the same problems they had in getting here.

I hope you all keep these two days in pleasant mind.

Good bye and I hope we will meet again soon on due occasion!

Dr. Ralf Münnich, co-ordinator of the KEI project, University of Tübingen

Outlook to the next workshop by the co-ordinator

The next KEI workshop will take place in Maastricht, 6 - 7 October 2005. The purpose of this workshop is to provide a forum for critical evaluations of the future challenges for developing indicators to policy, given the challenges facing local and globally-integrated *knowledge economies*. Presentations within six sessions are foreseen:

1. Sectoral and technology based approaches to indicators for the knowledge economy
2. The challenge for human resources and knowledge production
3. What do we want from a knowledge economy?
4. Integrating the local with the global
5. Composite indicator approaches to measuring the knowledge economy
6. Policy challenges

More details on the Maastricht workshop including the leaflet are available at the KEI home page (`http://kei.publicstatistics.net`).