Composite indicators: the art of mixing apples and oranges

by Stefano Tarantola and Andrea Saltelli
Econometrics and Applied Statistics Unit
Institute for the Protection and Security of the Citizen
Joint Research Centre of the European Commission
TP 361, Ispra (VA) Italy

Indicators and composite indicators

A well- known definition of *an indicator* in the literature is <<*something that provides a clue to a matter of larger significance or makes perceptible a trend or phenomenon that is not immediately detectable*>> (Hammond et al, 1995). An indicator's main defining characteristic is that it quantifies and simplifies information in a manner that promotes the understanding of the topic in question, let it be economics, environment or a sociological concept, to both decision-makers and the public.

A mathematical combination, or aggregation, of a set of indicators is most often called an *index*, or *composite indicator*. More specifically, composite indicators are based on underlying indicators that have no common meaningful unit of measurement, when there is no obvious way of weighting these underlying indicators (Saisana and Tarantola, 2002).

Composite indicators are very common in fields such as economic and business statistics, and in a variety of policy domains such as industrial competitiveness, sustainable development, globalisation and innovation (Saltelli, 2007). The proliferation of this kind of indicators is a clear symptom of their political importance and operational relevance in decision making (Nardo et al., 2005).

Challenges in composite indicators development

As reported in Saltelli et al., 2008, it is obvious that, <<*from a purely mathematical point of view, the aggregation convention used for composite indicators deal with the classical conflictual situation tackled in multi-criteria evaluation. Thus, the use of a multi-criterion framework for composite indicators in general, and for sustainability and well-being indices in particular, is relevant and desirable (Funtowicz et al., 2002; Munda, 1997 , 2005, Ülengin et al., 2001). However, the so-called "multi-criterion problem" can be solved by means of a variety of mathematical approaches, all of them plausible. This situation is due to Arrow's impossibility theorem (Arrow, 1963), which proves that it is impossible to develop a "perfect" mathematical aggregation convention. This implies that it is desirable to have mathematical algorithms that may be recommended on some theoretical and empirical grounds, or alternatively to test how robust results are with respect to different aggregation procedures. This makes sensitivity analysis a fundamental step during the development of any composite indicator (Saisana et al., 2005; Saltelli et al., 2004).*>>.

Munda (2004) insists on the importance of the quality of the aggregation convention, which is to guarantee the consistency between the assumptions made and the ranking obtained. Such quality depends crucially on the way this mathematical model is embedded in the social, political and technical structuring process. In other words, composite indicators are context-dependent and present both technical and socio-political uncertainty.

These uncertainties may not be neglected. They must be dealt with caution and above all be made as transparent as possible. Transparency must remain one of the main ingredients in developing composite indicators. This is especially true for the problem of weighting individual indicators, which remains the most important source of uncertainty and debate.

One bad example of composite indicator

Among the list of objections to the use of composite indicators one reads (Saisana and Tarantola, 2002, Nardo et al., 2005):

- *Composite indicators may send misleading, non-robust policy messages if they are poorly constructed or misinterpreted [… or] may invite politicians to draw simplistic policy conclusions […]*

An example of poorly constructed composite indicator appears in the Financial Times of September, 13th 2007. The article is about an 18-country composite indicator of "sustainability of fiscal and ecological development", developed at Germany's Allianz Insurance and Dresdner Bank, showing an unexpected result. Russia is outpacing United States, United Kingdom and Germany in securing its population's long-term economic and environmental future.

The index is a composite of five indicators, and Russia holds the top position in three of them - based on its good performance on public debt, current account and net borrowing balances, compared with the other more established economies. However, Russia performs badly on ecological development, being pushed to the bottom rank for the other two indicators because of its high carbon dioxide emissions and high energy use per unit of gross domestic product. But this is largely outweighed by the strong fiscal record.

The result came about because of Russia's huge oil and gas reserves and the sharp rise in energy prices in recent years, which have boosted significantly the country's finances. As soon as the oil price falls, then so will the country's fiscal performance. This is clearly an example of poorly constructed composite indicator as the sustainability of fiscal development of a country should not depend on the volatility of the oil price on the market. More attention should have been paid to produce a sound theoretical framework which could have guided the choice of more suitable component indicators.

In order to react to this critique, the chief economist at Allianz Insurance and Dresdner Bank declared that the findings were <<…*a thought-provoking way of looking at the* *issue, although in Russia's case it's probably more appropriate to look at each indicator separately>>*.

Notice here that the sole fact that a composite indicator has been put in place, whatever good or bad its architecture, will give the policy issue a great deal of visibility and will stimulate debate by concerned parties.

One good example of composite indicator

As stressed in Nardo et al. (2005), the strength of a composite indicator can largely depend on the availability of good quality underlying data. In July 2007, the World Bank published its sixth annual report on Worldwide Governance Indicators. The World Bank did a great effort in sifting through hundreds of indicators from 33 data sources by 30 organisations, thus providing a picture of governance across six categories over a 10-year period and more than 200 countries and territories. Given that there is no objective scale for judging a country's "rule of law", the Bank relies on up to 19 different sources for that measure.

The resulting report aimed business groups, reformists and other parts of civil society to push for better governance, which is crucial for development. The FT already expected some critics: *<<…the niceties [of CIs] are often lost when ranking countries or providing a global top or bottom 10 countries>>*. Nevertheless, a word of praise was given *<<…the inevitable talk of "hit parades" should not detract [the Bank] from its overall effort>>*.

To conclude this example, we cite here another sentence from the same FT article asserting that the main reason behind composite indicators is in their ability to quantify multidimensional and complex concepts even with all the limitations that their use implies: *<<Economists are often accused, justly, of thinking that what cannot be counted does not count. In this case, economists are trying to count what - many would say - cannot be counted. The alternatives, however, are worse. Either we ignore this fact or we make subjective guesses. For all its weaknesses, the Bank remains best-equipped to crunch the numbers and deliver the judgment, however unpalatable.>>*

Factors affecting country rankings

Although other factors may have a certain impact on the final scores and rankings (such as the imputation of missing values, the type of hierarchical structure chosen to represent the framework, or the aggregation method chosen), the issue of weighting plays a central role to the development of a composite indicator. The weighting, i.e. the need to combine in a meaningful way the underlying individual indicators, should ideally be made explicit (this is almost never the case for composite indicators that appear on the media) and agreed by an as-wide-as-possible public. This agreement is practically impossible to achieve, so specific participatory techniques (such as budget allocation) have been developed and can be used to take into account the different subjective value judgments without the need to agree on a unique set of weights.

In this paper we show a case study using the most straightforward of the participatory methods for weighting: the *budget allocation process*. Experts are given a *budget* of e.g. 100 points, to be distributed over a number of individual indicators. They can *pay* more for those indicators whose importance they want to stress, and less or nothing for those they consider less important (Jesinghaus, 1997). The budget allocation is feasible until a maximum of 10-12 indicators. If too many indicators are used, this method can give serious cognitive stress to the experts who are asked to allocate the budget. In our case study, each domain has six elements, which effectively eliminates such cognitive stress.

In such case, we strongly suggest to test the robustness of the country rankings to the weights proposed by different actors, to see whether the differences in the weights provide consistent results in terms of rankings. We will show this in our case study.

Robustness and sensitivity

Uncertainty and sensitivity analysis are key tools to investigate the robustness of the message conveyed by a composite indicator. Indeed, the building of composite indicators involves stages where judgments have to be made: the choice of a conceptual model, the selection of individual indicators, the weighting of indicators, the treatment of missing values, the aggregation rule, etc. All these sources of subjective judgment can influence the message brought by the composite indicator in a way that deserve analysis and corroboration. A combination of uncertainty and sensitivity analysis (respectively UA and SA) can help to gauge the robustness of the composite indicator, to increase its transparency and to help framing a debate around it.

Despite a synergistic use of UA and SA has proven to be more powerful (Saisana et al., 2005 and Tarantola et al., 2000), UA is more often adopted than SA (Jamison and Sandbu, 2001) and the two types of analysis are often treated separately. The types of questions for which an answer is sought via the application of UA & SA are:

• Does the use of one development strategy versus another in building the CI provide actually a partial picture of the countries' performance?

• Which constituents (e.g. countries) have large uncertainty bounds in their rank (volatile countries)?

• Which are the factors that affect the countries' rankings?

All things considered, a careful analysis of the uncertainties included in the development of a CI can render its building more robust. A plurality of methods (all with their implications) should be initially considered, because no model (CI development strategy) is a priori better than another, provided that internal coherence is always assured, as each model serves different interests. The CI is no longer a magic number corresponding to crisp data treatment, weighting set or aggregation method, but reflects uncertainty and ambiguity in a more transparent and defensible fashion.

The case study

Here we describe the results of the composite indicator of e-business readiness for European enterprises, using data from the 2006 Eurostat enterprise survey (Castaings et al., 2008). The composite is made of two core dimensions: Adoption of ICT by business and Use of ICT by business. The composite provides a valuable summary measure of the e-business readiness of enterprises in all European countries.
The e-business readiness composite indicator is one of the policy indicators selected by the Council Resolution of 28 January 2003 (5197/03) of the European Union to monitor progress in the implementation of the eEurope 2005 Action Plan and is still the benchmark for the i2010 framework. The CI is composed of two categories: 6 indicators for the group 'Adoption of ICT by business' and 6 indicators for the group 'Use of ICT by business'. The raw data for the basic indicators are expressed as percentages: 11 indicators are percentages of enterprises and one indicator (a4) is percentage of employees (see Table 1 and Table 2).

Percentage of enterprises that use Internet	***a1***
Percentage of enterprises that have web/home page	***a2***
Percentage of enterprises that use at least two 2 security facilities at the time of the survey	***a3***
Percentage of total number of persons employees using computer with their normal work routine	***a4***
Percentage of enterprises having broadband connection to internet	***a5***
Percentage of enterprises with LAN and using an Intranet and Extranet	***a6***

Table 1 e-business readiness Index: list of indicators for "Adoption of ICT"

Percentage of enterprises that have purchased products / services via the internet, Electronic Data Interchange (EDI) or any other computer mediated network where these are >1% of total purchases	***b1***
Percentage of enterprises that have received orders via the internet, EDI or any other computer mediated network where these are >1% of total turnover	***b2***
Percentage of enterprises whose IT systems for managing orders or purchases are linked automatically with other internal IT systems	***b3***
Percentage enterprises whose IT systems are linked automatically to IT systems of suppliers or customers outside their enterprise group	***b4***
Percentage of enterprises with Internet access using the internet for banking and financial services	***b5***
Percentage of enterprises that have sold products to other enterprises via a presence on specialised internet market places	***b6***

Table 2. e-business readiness Index: list of indicators for "Use of ICT"

The underlying indicators are aggregated using the budget allocation weighting scheme involving a panel of national representatives. The set of weights given in Table 3 represents the average of weights provided by twelve national representatives of the e-Business Support Network (e-BSN). However, for each indicator, twelve weights are available and the objective of the study is to assess robustness of the country scores to variability in such weights.

a1	***a2***	***a3***	***a4***	***a5***	***a6***
0.18	0.16	0.10	0.16	0.21	0.20
b1	***b2***	***b3***	***b4***	***b5***	***b6***
0.17	0.17	0.21	0.21	0.12	0.13

Table 3 Average Budget Allocation weights for the different index components

The scores and rankings for the Adoption and Use of ICT provide a relative gauge of e-business progress in 28 countries (26 European Union Member States and 2 Countries members of the European Free Trade Association).

ICT Adoption		
Countries	**Score**	**Rank**
Finland (FI)	78.06	1
Sweden (SE)	77.28	2
Iceland (IS)	76.00	3
Denmark (DK)	75.72	4
Netherlands (NL)	72.64	5
Belgium (BE)	71.72	6
Norway (NO)	71.27	7
Germany (DE)	70.10	8
France (FR)	69.11	9
United Kingdom (UK)	68.43	10
Norway (NO)	67.96	11
Luxembourg (LU)	67.92	12
Ireland (IE)	64.35	13
Slovenia (SI)	63.36	14
Spain (ES)	63.17	15
Italy (IT)	60.48	16
Czech Republic (CZ)	60.11	17
Estonia (EE)	59.91	18
Slovakia (SK)	57.44	19
Greece (EL)	55.42	20
Portugal (PT)	52.28	21
Poland (PL)	52.09	22

ICT Use		
Countries	**Score**	**Rank**
Denmark	41.42	1
Netherlands	35.22	2
Norway	34.26	3
Ireland	33.20	4
Germany	33.04	5
Iceland	32.42	6
Finland	30.85	7
Norway	30.68	8
Sweden	30.58	9
France	30.10	10
Belgium	28.35	11
United Kingdom	27.89	12
Italy	27.44	13
Luxembourg	26.99	14
Greece	26.71	15
Estonia	23.40	16
Spain	22.94	17
Portugal	22.86	18
Czech Republic	22.76	19
Slovenia	21.71	20
Lithuania	21.15	21
Cyprus	19.16	22

Lithuania (LT)	51.36	23	**Slovakia**	18.54	23
Cyprus (CY)	51.15	24	**Poland**	17.39	24
Hungary (HU)	48.75	25	**Latvia**	13.73	25
Latvia (LV)	45.35	26	**Hungary**	12.21	26
Bulgaria (BG)	43.01	27	**Romania**	11.03	27
Romania (RO)	32.42	28	**Bulgaria**	7.71	28
EU27	**63.86**		**EU27**	**26.46**	

Table 4 2006 e-Business Readiness ICT Adoption and Use – Scores and rankings according to the budget allocation weights

A graphical representation of Adoption versus Use scores for the 28 countries but also for the EU27 aggregate is proposed by Figure 1. The correspondence between the country codes with the full names of the Member States is given by Table 4. Using the EU25 aggregate, the plane is divided in 4 parts characterising the practical use of the adopted ICT infrastructures. With respect to the EU27 aggregate, the 4 zones categorize the performances of the countries with respect to the EU27 average estimated by Eurostat. Since the correlation between Adoption and Use scores is important (r = 0.905), most of the countries lie along the diagonal depicting a positive correlation. Most of the time good performances in ICT Adoption are coming along with a satisfactory level of ICT Use. With respect to the EU27 average, Greece and Italy can be distinguished for their efficiency in using ICT infrastructures given the investments made. Portugal and Spain are the only Countries from the former EU15 which are still below the European average for both adoption and use of ICT.

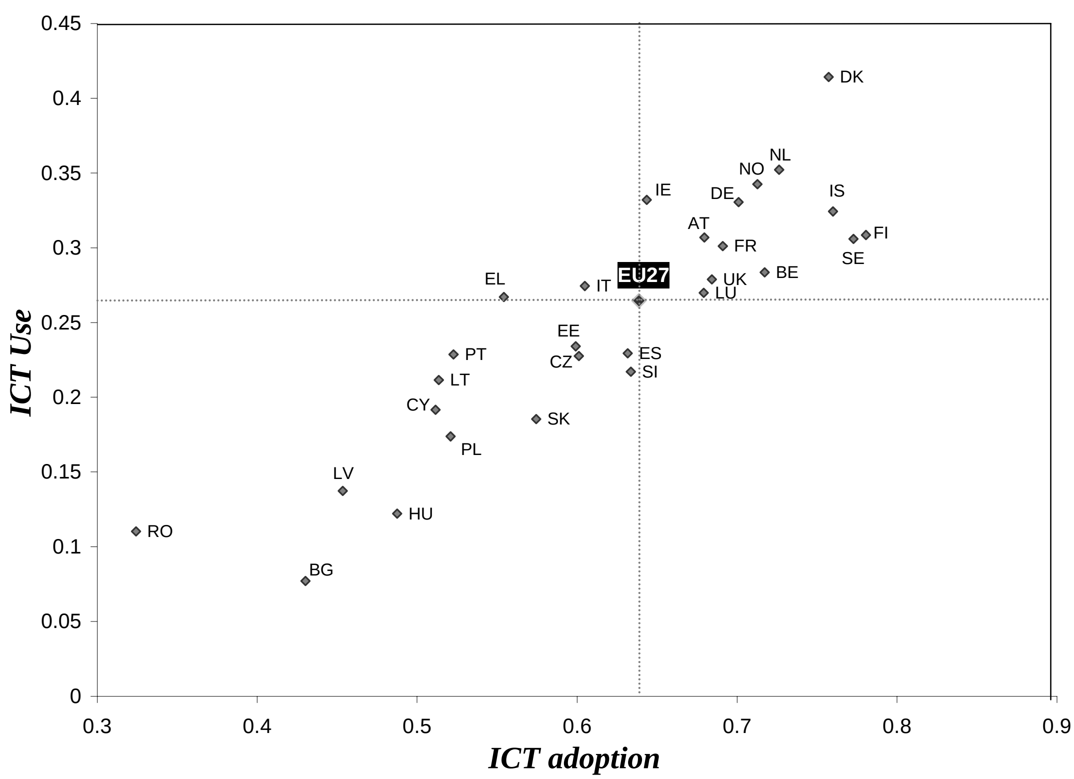

Figure 1 Adoption scores vs. Use scores employing the budget allocation weighting scheme. The black diamond indicates the EU27 aggregate

As mentioned above, the country scores are affected by uncertainty, which is due to the experts' preferences in the assignment of weights. In order to quantify this uncertainty we choose e-BSN experts at random and use averaging over the sets provided by the selected experts. Therefore, each country is characterised by a cloud of points rather than a single pair of scores (i.e those for ICT Adoption and Use). The results obtained for a cluster composed of the leading countries are displayed by Figure 2. While the leading position of Denmark cannot be contested, two groups of countries (subgroups Germany-Norway-Netherlands and Iceland-Sweden-Finland) show similar performances. However, it is also clear that the first group performs better in ICT Use and the second in ICT Adoption. Other sources of uncertainty (e.g imputation of missing values) could be taken into account to assess the robustness of the ranking. However, this point is not addressed in the current paper.

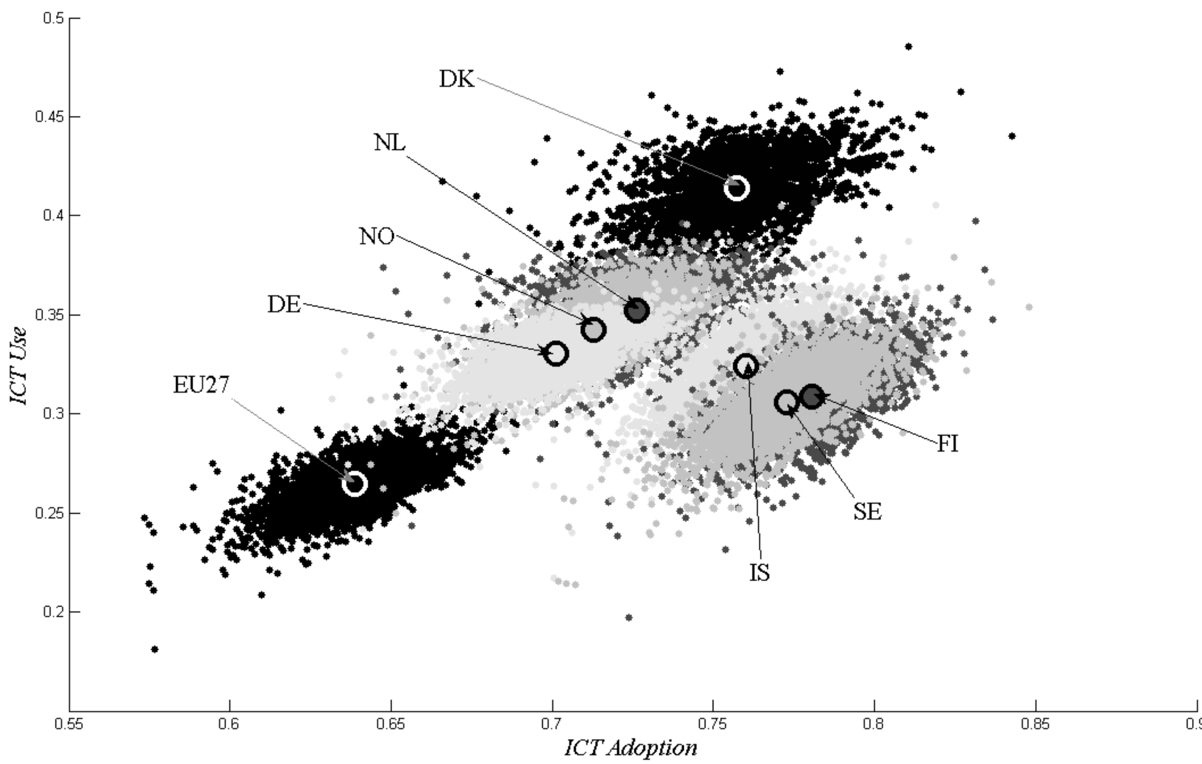

Figure 2 Uncertainty analysis for the cluster composed of leaders for e-Business Readiness – uncertainty on the weights taken into account

An appraisal of the variability of the performances achieved by the different Countries (and their evolution in time) in both Adoption and Use of ICT is provided in Figure 3 by the probability density functions obtained using the Member States scores (from 2004 to 2006). These curves are non-parametric estimates obtained using Gaussian kernels. Analysing the trend of the variance for such curves yields interesting information about the presence or absence of convergence among countries in either Adoption or Use. For ICT Adoption, the curves are characterised by a negative skew (elongated tail at the left). In other words, there is clearly a group of laggards increasing the mass of the left tail of the distribution for ICT Adoption. On the other hand, the performances are more balanced across countries for ICT Use,

In addition, the temporal shift of the curves toward larger scores is obvious for ICT Adoption yet still moderate for ICT Use. The left tails of the curves are slightly more elongated in 2004 and 2006 because Bulgaria and Romania did not participate in the 2005 survey. The improvements of countries in Adoption are characterised by an enlargement of the leading group (see the steeper right tail of the distribution). The slight progress in ICT Use comes with an increase in variance (i.e decrease of peak density and more elongated tails), meaning that the differences in countries relative performances are increasing.

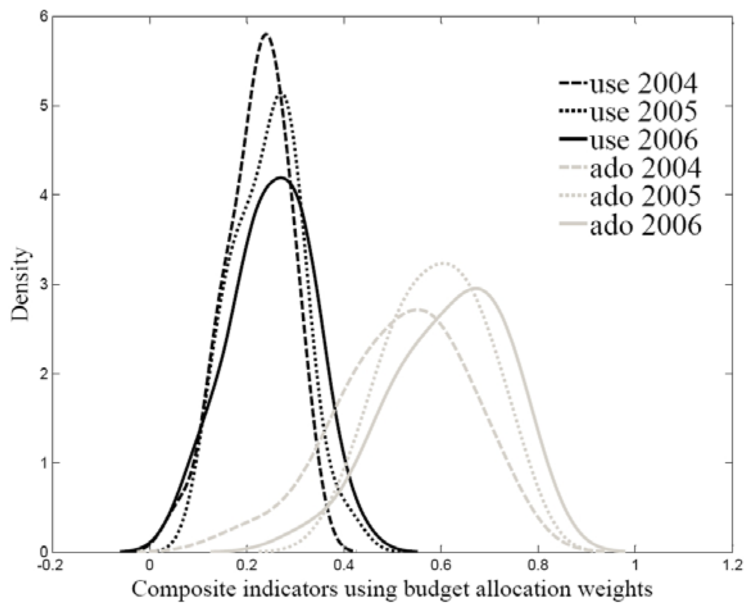

Figure 3 Probability density estimates (using normal kernel functions) for the country scores in 2004,2005 and 2006

While, the relation between the company size and the performances in ICT Adoption and Use is relatively obvious and not presented in the current paper, the picture is more complicated when analysed for different sectors of economic activity (see *Figure 4*). As expected, the sector O (Motion picture, video, radio and television activities) represents less than 1% of the population but shows outstanding performances for both Adoption and Use. It is also not surprising to observe that the F sector (construction) is lagging behind. Given the level achieved for Adoption by sector K (Real Estate, Renting and Business Activities), improvements are still needed for Use of ICT. On the contrary, sectors G (Wholesale and retail trade) and H (Provision of short stay accommodation) are characterised by a high level of Use of ICT given the investments made.

Although some economic sectors show similar performances in ICT Adoption or/and Use for the EU27 aggregate, the analysis of the distribution across countries (for the same sector) reveal very different features (see Figure 5). As an illustration, sectors I

(Transports, storage and communications) and D (Manufacturing) have achieved neighboring levels of adoption when the EU27 aggregates are compared (Figure 4). However, while the variability seems quite balanced for sector I, a group of lagging countries can be clearly distinguished for sector D (see an elongated left tail featuring another mode of the distribution). Similarly to D, the shape of the probability density function for sector O (Motion picture, video, radio and television activities) is also interesting.

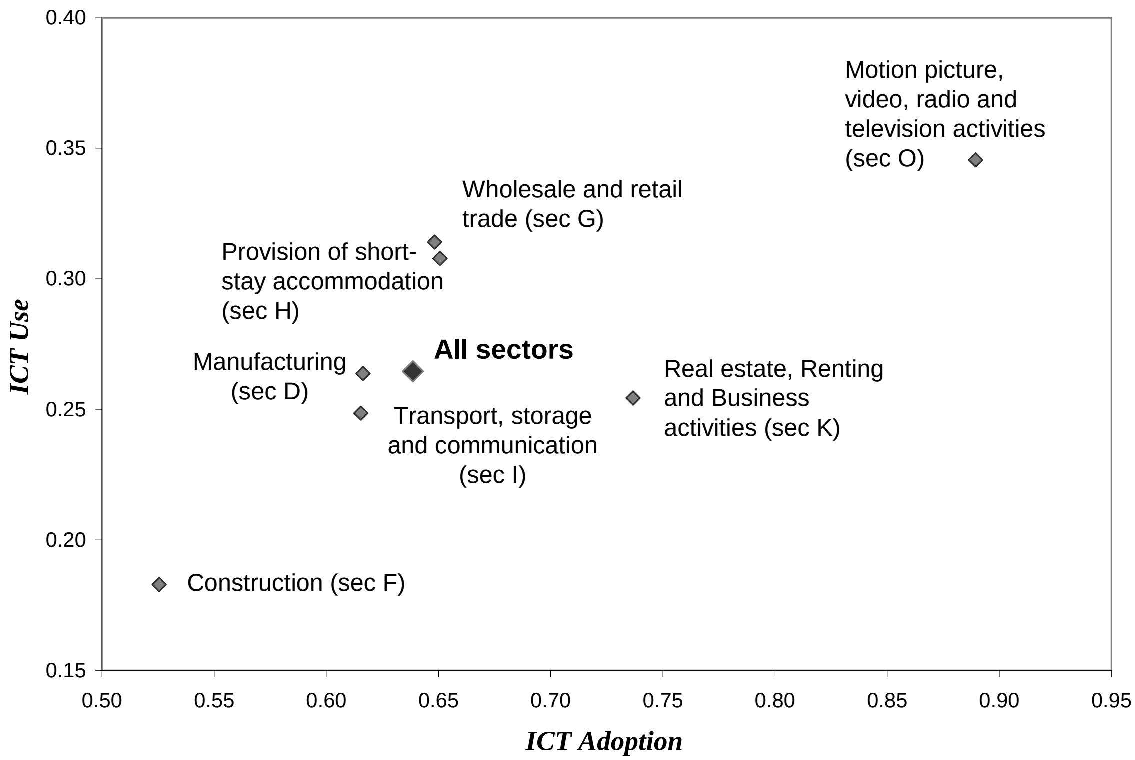

Figure 4 Adoption scores vs. ICT Use scores employing the budget allocation weighting scheme, the black diamond indicates the EU27 aggregate

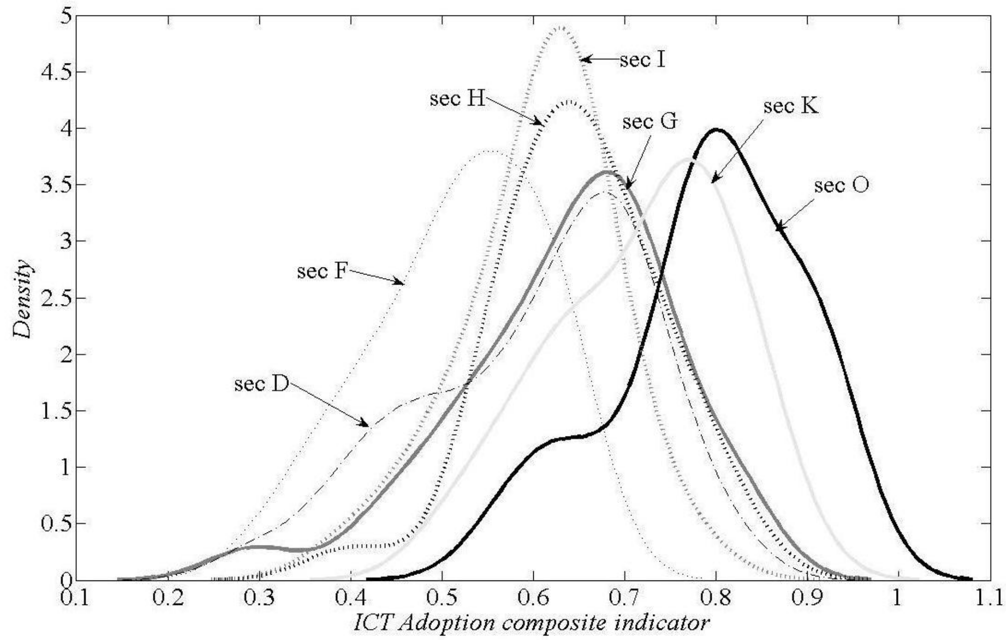

Figure 5 Probability density estimates (using normal kernel functions) for the country scores for ICT Adoption in 2006 depending on the sector of economic activity

Conclusions

We reviewed the recent debate around the use of composite indicators for policy and the suggested requirements for their development (ie the need for robustness assessment in order to improve the defensibility of the message brought by the composite indicator). We discussed a good and a bad example of composite indicator, both appeared on the Financial Times in the second semester of 2007: the *Sustainability of fiscal and ecological development* CI, developed at Germany's Allianz Insurance and Dresdner Bank and a *Governance Index* developed by the World Bank and published in the annual report on World Governance Indicators.

The case study tackled in this paper is the 2007 European e-business readiness index, a useful mechanism for comparing e-business Adoption and Use by firms in the various European countries.

Finally, we show on the case study how to make robustness tests of the country rankings, to see whether differences in weights values provide or not consistent results in terms of rankings. We propose a way to visualize the results of the robustness test.

References

Arrow, K.J. (1963) Social choice and individual values. 2d edition, Wiley, New York

Castaings, W., Tarantola, S., Marcusson, M. (2008). The 2007 e-Business Readiness Index. JRC Scientific report.

Funtowicz, S.O., Martinez-Alier, J., Munda, G., Ravetz, J. (2002) Multicriteria-based environmental policy, in H.
Abaza and A. Baranzini (eds.) Implementing sustainable development. UNEP/Edward Elgar, Cheltenham, pp. 53-77

Hammond A, Adriaanse A, Rodenburg E, Bryant D, Woodward R (1995) *Environmental Indicators: A Systematic Approach to Measuring and Reporting on Environmental Policy Performance in the Context of Sustainable Development*, Washington: World Resources Institute

Jamison, D. and Sandbu, M. (2001) WHO ranking of health system performance. Science, 293, 1595-1596.

Jesinghaus, J. (1997) in Moldan, B. And Billharz, S. (Eds), Sustainability Indicators. Report on the project on Indicators of Sustainable Development, John Wiley and Sons, Chichester.

Munda, G. (2005) “Measuring sustainability”: a multi-criterion framework. Environment, Development and Sustainability Vol 7, No. 1, pp. 117-134

Munda, G. (2004) “Social multi-criteria evaluation (SMCE)”: methodological foundations and operational consequences. European Journal of Operational Research, vol. 158/3, pp. 662-677.

Munda, G. (1997) Multicriteria evaluation as a multidimensional approach to welfare measurement, in J. van den Bergh and J. van der Straaten (eds.), Economy and ecosystems in change: analytical and historical approaches. Edward Elgar, Cheltenham, pp. 96-115

Nardo, M., M. Saisana, A. Saltelli, S. Tarantola, A. Hoffman and E. Giovannini (2005). Handbook On Constructing Composite Indicators: Methodology And User Guide, OECD Statistics Working Paper JT00188147, STD/DOC(2005)3

Saisana, M., Tarantola, S., Saltelli, A. (2005) Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. Journal of the Royal Statistical Society A, 168(2), 307-323.

Saisana M. and S. Tarantola (2002) State-of-the-Art Report on Current Methodologies and Practices for Composite Indicator Development, EUR 20408 EN

Saltelli, A., Composite indicators between analysis and advocacy, Social Indicators Research, 2007, 81, 65-77

Saltelli, A. Tarantola, S., Campolongo, F. and Ratto, M. (2004) Sensitivity Analysis in Practice. A Guide to Assessing Scientific Models. John Wiley & Sons publishers, New York.

Saltelli, A., J. Jesinghaus and G. Munda (2008) Well Being Stories, Beyond GDP Conference, Experts Workshop, 19-20 November, 2007

Tarantola, S., Jesinghaus, J. and Puolamaa, M. (2000) Global sensitivity analysis: a quality assurance tool in
environmental policy modelling. In Sensitivity Analysis (eds A. Saltelli, K. Chan, M. Scott) pp. 385-397. New York: John
Wiley & Sons.

Tarantola, S.; Nardo, M.; Saisana, M. & Gatelli, D. A new estimator for sensitivity analysis of model output: An application to the e-business readiness composite indicator Reliability Engineering and System Safety, 2006, 91(10-11), 1135-1141

Ülengin, B., Ülengin, F., Güvenç, Ü (2001) A multidimensional approach to urban quality of life: the case of
Istanbul. European Journal of Operational Research, 130, pp. 361-374