Open access • Journal Article • DOI:10.1057/PALGRAVE.JORS.2602445

Creating composite indicators with DEA and robustness analysis: the case of the Technology Achievement Index — Source link

Laurens Cherchye, Wim Moesen, Nicky Rogge, T Van Puyenbroeck ...+4 more authors

Institutions: Catholic University of Leuven

Published on: 01 Feb 2008 - Journal of the Operational Research Society (Palgrave Macmillan UK)

Topics: Data envelopment analysis, Benchmarking and Robustness (economics)

KATHOLIEKE UNIVERSITEIT LEUVEN

Faculty of Economics and Applied Economics

Department of Economics

Creating composite indicators with DEA and robustness analysis: the case of the technology achievement index

by

L. CHERCHYE
W. MOESEN
N. ROGGE
T. VAN PUYENBROECK & M. SAISANA
A. SALTELLI
R. LISKA
S. TARANTOLA

Public Economics

Center for Economic Studies
Discussions Paper Series (DPS) 06.03
http://www.econ.kuleuven.be/ces/discussionpapers/default.htm

January 2006

DISCUSSION PAPER

Creating Composite Indicators with DEA and Robustness Analysis: the case of the Technology Achievement Index*

L. Cherchye‡, W. Moesen‡, N. Rogge‡*, T. Van Puyenbroeck‡*

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

M. Saisana, A. Saltelli, R. Liska, S. Tarantola

(Joint Research Centre, European Commission; Italy)

January 2006

Abstract

Composite indicators are regularly used for benchmarking countries' performance, but equally often stir controversies about the unavoidable subjectivity that is connected with their construction. Data Envelopment Analysis helps to overcome some key limitations, viz., the undesirable dependence of final results from the preliminary normalization of sub-indicators, and, more cogently, from the subjective nature of the weights used for aggregating. Still, subjective decisions remain, and such modelling uncertainty propagates onto countries' composite indicator values and relative rankings. Uncertainty and sensitivity analysis are therefore needed to assess robustness of final results and to analyze how much each individual source of uncertainty contributes to the output variance. The current paper reports on these issues, using the Technology Achievement Index as an illustration.

1. Introduction

Organisations such as the United Nations, the European Commission, and others have developed and used "composite indicators" in which single indicators are aggregated into one index. These composite indicators provide comparisons of countries in complex and sometimes elusive policy issues. These measures are increasingly recognised as a tool for policy making and, especially, public communications on countries' relative performance in wide ranging fields such as the environment, the economy, or technological development.[1]

Composite indicators (CIs) are much like mathematical or computational models. Just as for models, the justification for a CI lays in its fitness to the intended purpose and peer acceptance. Also, their construction owes more to craftsmanship than to universally accepted scientific rules for encoding. The construction of CIs involves stages where subjective judgement has to be made: the selection of indicators, the treatment of missing values, the choice of aggregation model, the weights of the indicators, and so on. These choices can even be used to manipulate the results. It is, thus, important to identify the sources of subjective assessment and data errors and use uncertainty and sensitivity analysis to gain useful insights during the process of CI building, including an appraisal of the reliability of countries' ranking. These considerations are a central theme of the current paper.

* This paper is an offshoot of the KEI-project (contract n° 502529) that is part of priority 8 of the policy orientated research under the European Commission's Sixth Framework Programme (see http://kei.publicstatistics.net/). Laurens Cherchye thanks the Fund for Scientific Research-Flanders (FWO-Vlaanderen) for his postdoctoral fellowship.

[1] For an overview, see the JRC information server on composite indicators: http://farmweb.jrc.cec.eu.int/ci/.

The construction methodology that is used in the present paper is rooted in Data Envelopment Analysis (DEA). The original question in the DEA-literature is how one could measure each decision making unit's (e.g., a firm's) relative efficiency, given observations on input and output quantities in a sample of peers and, often, no reliable information on prices (e.g., Charnes and Cooper, 1985). One immediately appreciates the conceptual similarity between that original problem and the one of constructing CIs. In the latter case, quantitative sub-indicators for overall benchmarking are available, but as a rule there is only disparate expert opinion available about the appropriate weights to be used in an aggregator function. Yet there are differences between the two settings as well, the most notable one perhaps being that CIs typically look at 'achievements' without taking into account the input-side. Though there are some interesting exceptions (see the work of the European Commission on the Summary Innovation Index in 2005)

A known remarkable feature of the DEA-methodology is that it looks for endogenous (possibly constrained) weights/shadow prices, yielding an overall score that depicts the analyzed decision making unit in its best possible light relative to the other observations. This quality explains a major part of the appeal of DEA-based CIs in real settings. For example, several European policy issues entail an intricate balancing act between supra-national concerns of the centre and the country-specific policy priorities of member states. If one opts to compare composite performance of member states by subjecting them to a similar weighting scheme, this may prevent acceptance of the entire exercise. To take an example: with reference to European social inclusion policy, Atkinson *et al.* (2002) remark that "in the context of the EU, there are evident difficulties in reaching agreement on such weights, given that each member state has its own national specificity." As the essence of DEA is that it yields most favourable, country-specific weights, it may help to counteract such problems. However, the typical DEA set-up, which only requires the endogenous weights to be non-negative, is insufficient to guarantee peer acceptance. Usually some expert information about the most appropriate weights to be used for aggregating the individual sub-indicators is available, and such opinions should ideally be incorporated to make the weights acceptable. We will provide a typical example below.

DEA-based CIs have *inter alia* been used to assess European labour market policy (Storrie and Bjurek, 2000), European social inclusion policy (Cherchye, Moesen and Van Puyenbroeck, 2004), and internal market policy (Cherchye, Lovell, Moesen and Van Puyenbroeck, 2005). A similar model has been tested to assess progress towards achieving the so-called Lisbon objectives (European Commission, 2004, p. 376-378). Similarly, some authors have proposed a DEA-approach for the well-known Human Development Index (Mahlberg and Obersteiner, 2001; Despotis, 2005). In this paper, we will use the Technology Achievement Index (TAI) to illustrate our approach. Together with the Human Development Index, the TAI is developed by the United Nations for the Human Development Reports. The main reason for using it here as an illustrative example is that it figures likewise, and in an extensive fashion, in the JRC-OECD *Handbook on Constructing Composite Indicators* (see Nardo *et al.*, 2005a; 2005b).[2] We will complement the handbook's results by providing a more in-depth analysis of the DEA approach.

We will start in section 2 by briefly discussing the TAI as well as the available information on possible weighting schemes, obtained by a panel of experts. Section 3 presents the basic model and indicates its relationship with more conventional DEA-models. We then

[2] Regarding possible methodologies for composite indicator construction, both references have a considerably broader scope than the current paper, which only focuses on DEA-based indices. For instance, the interested reader may find there sensitivity and uncertainty analyses, using the TAI-data, that compare DEA-based results with those stemming from otherwise obtained indices (e.g. via exogenous weighting, or via a non-compensatory multicriteria approach).

address uncertainty and sensitivity analysis in section 4. The current mainstream literature on sensitivity analysis for DEA-models is primarily concerned with the sensitivity of (in)efficiency scores following data perturbations in a *given* set of inputs and outputs (see e.g. Cooper *et al.*, 2004). In the case of CI's, however, one is typically also concerned with the robustness of results if performance dimensions are added or deleted, if the expert information would have been different, and so on. Such choice-of-model concerns have been addressed rather infrequently in the DEA-literature (e.g. Valdmanis, 1992; Wilson, 1995; Banker *et al.*, 1996; Simar and Wilson, 1998; Simar, 2003). Even here the parallel between composite indicators and mathematical models is useful. In mathematical models of natural or man-made systems uncertainty and sensitivity analysis relative to modelling assumptions or scenarios has been studied (see Saltelli *et al.*, 2004, for a review). The methodology that we present in section 4 may therefore be valuable for a broader DEA-audience as well. Section 5 concludes and offers some final remarks.

2. The Technology Achievement Index and expert opinions

The United Nations introduced the TAI to capture how well a country is creating and diffusing new as well as existent technologies and building a human skill base for technology creation, with the intention of helping policy-makers to define technology strategies (UN, 2001). As explained by Desai *et al*. (2002), these dimensions are captured by eight *achievement* indicators: (*I*) the number of *patents* granted per 1,000,000 people, (*II*) the receipt of *royalties* (in US$, per 1000 inhabitants), (*III*) the number of *Internet* hosts per 1,000 people, (*IV*) *exports* of high and medium technology products (as a share of total goods exports), (*V*) the number of *telephone* lines per 1,000 people (in logs), (*VI*) *Electricity* consumption per capita (in logged kWh), (*VII*) the mean years of *schooling*, and (*VIII*) the gross *enrollment* ratio of tertiary students in science, mathematics and engineering. We refer to the actual figures, used in the current paper, and extensive explanations of each sub-indicator in Desai *et al*. (2002). This list exhibits a typical feature of most CIs, i.e. that the sub-indicators are displayed in quite diverse measurement units. According to current practice, the TAI's authors deal with this problem by *normalising* the original data, a feature to which we will comment further on.[3]

The normalised sub-indicators are next *weighted* and *added*. Specifically, the UN uses equal weights for each of the components. We will now depart from that approach. One reason for doing so has already been mentioned in the introduction: applying "benefit-of-the-doubt" weights may help to foster acceptance of the eventual results by the national stakeholders considered.[4] A second one is that we have information on the subject, stemming from an internal JRC survey conducted on 21 individuals, on the set of weights which each individual would consider as appropriate. These weights were obtained using 'Budget Allocation', a participatory method in which experts are given a budget of *N* points, to be distributed over a number of sub-indicators. In budget allocation each expert can "pay" more

[3] The transformation method used for the TAI re-expresses the original value for each sub-indicator on a (unit free) scale from 0 to 1, using the formula (*original value – observed minimum value) / (observed maximum value - observed minimum value)*. For the *telephone* and *electricity* sub-indicators, logarithms rather than original values are taken.

[4] A large majority of composite indicators are of the equal weighting type. We here instead allow for (constrained) country-specific weights, which may be justified by considerations such as those figuring in the introduction, but also on pure modelling grounds: forcing weights to be equal neglects the reality that there are, often, a compilation of possibly conflicting opinions available. Hence, equal weighting is in general not even an adequate *description* of a core issue in composite indicator construction. Finally, "simple" equal weighting implies fixed weighting, which in turn implies that country rankings may change merely because another normalisation method has been used (see further).

for those indicators whose importance he/she want to stress. Summary information about the distributions of the points so-obtained, as applied in the TAI setting, is provided in Table 1.

Table 1: summary statistics on TAI-weights, retrieved from expert panel

Weights	Patents	Royalties	Internet	Exports	Telephone	Electricity	Schooling	Enrollment
Mode	0,10	0,05	0,10	0,20	0,10	0,05	0,20	0,20
Average	0,11	0,11	0,11	0,18	0,10	0,06	0,15	0,18
St. dev.	0,05	0,07	0,05	0,07	0,05	0,04	0,06	0,08
Min	0,05	0,00	0,02	0,09	0,00	0,00	0,05	0,00
10th percentile	0,05	0,05	0,05	0,10	0,05	0,00	0,05	0,10
90th percentile	0,20	0,20	0,20	0,30	0,15	0,12	0,20	0,30
Max	0,20	0,30	0,20	0,33	0,20	0,15	0,25	0,30
# ranked (*)								
- on top	2	2	1	8	0	0	5	9
- at bottom	4	8	6	1	5	15	3	3

(*): entries provide the number of times a sub-indicator figures on top (resp. at the bottom) of experts' rankings. The horizontal sum exceeds the number of experts, which is due to tied rankings for first (resp. last) places.

As one notices, there are considerable inter-individual differences in the proposed weighting schemes, with not a single pair of experts sharing a similar proposal. This holds for the magnitudes as well as for the relative importance of the different sub-indicators. One can infer from Table 1 that a limited consensus emerges from the panel on the relative importance of the variables, and that unanimity is only achieved in judging that the telephone and electricity indicators are less important than all other indicators. No additional consensus about the dimensions' ranking emerges from the panel. Also, although equal weights (of 1/8) fall within the upper and lower bounds over the sample of experts, nobody in the panel proposed to weigh all sub-indicators equally, in contrast with the actual TAI. This clearly illustrates one stage in the TAI's construction where subjective judgement has been made. For example, if alternatively a dimension-wise plurality vote (among *this particular* expert-panel) had been used, the eventual weighting scheme would have been the one figuring in the first line of Table 1. The questions to be taken up in the following sections are how such information can be incorporated when calculating an overall index, and to what extent perturbations in this setting have an impact on eventual country rankings.

Before doing so, one more remark is in order. It is well-known that weights in a linear aggregate $\Sigma y_{ij} w_i$ have the meaning of trade-offs. Hence, what matters in the linear composite are the relative weights (which directly refer to the substitutability of the different dimensions) rather than the absolute weights. It has however been observed (e.g. by Munda and Nardo, 2003), that experts usually interpret weights, such as those stemming from a Budget Allocation method, as 'importance coefficients' (cf. Freudenberg, 2003, p. 10: "Greater weight should be given to components which are considered to be more significant in the context of the particular composite indicator"). In fact, the 21 experts were literally asked to assign more points to a sub-indicator "the more important this indicator is". We will consequently adhere to such an interpretation of the above weights in what follows.[5]

[5] In fact, in the aforementioned *handbook* one finds that two types of weighting information were gathered from this panel. That is, weights were *also* obtained using the Analytical Hierarchy Process (see Nardo *et al.* (2005a,b) for results based on the AHP-information), and such weights typically have a relative interpretation. Due to space limits, we focus on the budget allocation results. However, as demonstrated in Cooper, Seiford and

3. A basic DEA-model

To introduce DEA as a tool for constructing "benefit-of-the-doubt" CIs, we consider a cross-section of m sub-indicators and n countries, with y_{ij} the value of sub-indicator i in country j. In the following, and in line with the more common DEA terminology, we will often refer to sub-indicators as "outputs". In the TAI case, each sub-indicator/output i has the following interpretation: if $y_{ij} > y_{ik}$ then country j performs better than country k.

Our objective is to merge these individual sub-indicators/outputs into a single-valued composite indicator, defined as the weighted average of the m sub-indicators; we use w_i to represent the weight of the i-th sub-indicator. As discussed above, the available expert information does not allow us to specify *a priori* a unique vector of generally acceptable weights. Therefore, we endogenously select those weights that maximize the CI value for the country under consideration. This gives the following linear programming problem for each country j:

$$CI_j = \max_{wi} \sum_{i=1}^{m} y_{ij} w_i$$

Subject to

$$\sum_{i=1}^{m} y_{ik} w_i \le 1 \qquad \forall k = 1,\ldots,n \qquad (\textit{normalisation constraint})$$

$$w_i \ge 0 \qquad \forall i = 1,\ldots,m \qquad (\textit{non-negativity constraint})$$

The objective function reveals the benefit-of-the-doubt interpretation of the methodology: the problem chooses those weights w_i that maximize the resulting indicator value CI_j. As a result, the highest relative weights are accorded to those dimensions for which the country j achieves the best performance (in relative terms) when compared to the other countries in the sample. The weights are not fixed *a priori*; the only restriction in the formulation above is that they should be non-negative, which implies that the CI is a non-decreasing function of the sub-indicators (see the *non-negativity constraint*; we discuss the inclusion of additional weight restrictions below). To guarantee an intuitive interpretation of the CI, we impose that no country in the sample can achieve a value that is greater than unity under these weights (see the *normalisation constraint*). We obtain $0 \le CI_j \le 1$ for each country j, with higher values indicating a better relative performance.

As pointed out by Despotis (2005), this model is formally equivalent to the original input oriented, constant-returns-to-scale DEA model presented by Charnes *et al.* (1978), when using the sub-indicators to represent the different outputs and allocating a single 'dummy input' with value unity to each country. In that interpretation, the dummy input for each country may be interpreted in terms of 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 above discussion that an intuitive interpretation may also be obtained by simply 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.[6]

Tone (2000, p. 169-174), AHP-based information can also be appended to DEA-based indicators by creating assurance regions for the weights (rather than by constraining the virtual outputs, as we do).

[6] Conceptually, the dummy input/helmsman approach may be difficult to reconcile with the fact that one is actually using an *input-oriented* DEA model (which looks for feasible downward adjustments of inputs, holding outputs fixed). Moreover, the argument that an input- and an output-orientation are fully equivalent for the class of DEA-models introduced by Charnes *et al.* (1978), only holds for models without (or with specific kinds of) weight restrictions. Therefore we prefer to think of the problem as one in a "pure output setting" (a term coined by Cook, 2004), in which the normalization constraint is interpreted as a scaling or bounding condition (see also

Interestingly, the CI values are independent of the units in which the constituent sub-indicators are measured, i.e. the CI meets the important property of 'units invariance'. Indeed, units invariance is a well-known property of the original DEA model introduced by Charnes *et al*. (1978). At this point, it is worth stressing that the composite indicators that use the most common practice of fixed weighting (with equal weighting as a special case) do *not* meet the units invariance property. In fact, this units invariance property of the 'benefit-of-the-doubt' alternative makes the normalisation stage (see section 2) redundant. This is particularly convenient from a practical point of view; see, e.g., the discussion in Freudenberg (2003) on the sensitivity of CI results with respect to the specific normalisation scheme that is used.

The above model implies a most generous CI; the only restriction on the weights is that they should be non-negative. Somewhat inconveniently, this does not exclude extreme scenarios. For example, all the relative weight can be assigned to a single sub-indicator, which would then completely determine the overall CI performance; the other sub-indicators would 'not matter' as their relative weight would equal zero. Moreover, we *do* have expert opinion, and to neglect their contribution means running the risk that the eventual composite indicator is rejected. This indicates a need for further restricting the endogenously selected CI weights.

In fact, the issue of imposing additional *a priori* weights has attracted considerable attention in the DEA literature; see, e.g., Thanassoulis *et al*. (2004) for a survey. In the present context, restrictions regarding the so-called 'virtual outputs' are particularly interesting as these (i) do not depend on measurement units and (ii) directly reveal how the respective outputs contribute to a composite indicator value. In DEA terminology, virtual outputs refer to the product of each separate sub-indicator/output and the associated weight: formally, the l-th virtual output for country j is given as the product $y_{lj}w_l$. Clearly, these virtual outputs may also be interpreted as the 'pie shares' that together constitute the CI_j ($=\sum_{i=1}^{m} y_{ij}w_i$): the i-th virtual output represents the (volume of the) pie share of the i-th sub-indicator, thus revealing the importance of that sub-indicator in the computation of CI_j. As explained in greater detail above, the available (budget allocation) expert information in the specific TAI case is consistent with formulating upper and lower bounds regarding the virtual outputs. Specifically, we will be concerned with a type of constraints known in the DEA-literature as 'proportional virtual weight restrictions' (Wong and Beasley, 1990), which for the reasons just indicated can alternatively be labelled *pie share constraints* for each sub-indicator/output l :

$$L_l \leq \frac{y_{lj}w_l}{\sum_{i=1}^{m} y_{ij}w_i} \leq U_l \qquad \text{(\textit{pie share constraint})}$$

with L_l and U_l the respective (pre-specified) lower and upper bounds. Such a restriction is equivalently expressed as

$$L_l\left(\sum_{i=1}^{m} y_{ij}w_i\right) \leq y_{lj}w_l \leq U_l\left(\sum_{i=1}^{m} y_{ij}w_i\right)$$

Cook and Kress, 1994). See Cherchye *et al*. (2004) for the 'pure output' fractional program formulation of the linear program stated in the main text.

Obviously, in this presentation the pie share constraints do not interfere with the linear nature of the programming problem. In fact, and importantly, the resulting construction of CI_j remains invariant to the units of measurement.[7]

One can interpret the CI_j (possibly calculated under pie share constraints) from a benchmarking perspective. In that respect, a value below unity means that there is some other country in the sample that demonstrably outperforms the evaluated country *even when using the latter's most favourable weighting scheme*. If this is the case, such an outperforming country may be conceived as a suitable benchmark for the evaluated country. More generally, the value of CI_j reveals the degree of superior performance. This interpretation is intuitive and straightforward to convey to the target audience: "Combine the sub-indicator values of another country with *your* most favourable, possibly constrained, weights; this weighted sum may in fact be higher than the one based on your own sub-indicator values. Look specifically for the country that maximizes this similarly weighted average; the ratio of 'your' weighted sum and the similarly weighted sum of this benchmark country yields your CI-value."

We end this section by presenting results for the TAI as obtained by the above methodology. In this baseline scenario, 23 countries and all 8 sub-indicators are included. Contrary to the original TAI, the benefit-of-the-doubt aggregation is performed on the *original* rather than on transformed data, since normalization is redundant in our approach (this also means that we have taken the original rather than the logged values for the *telephone* and *electricity* values). For the baseline scenario, we additionally appended "pie share" constraints that are directly inspired by the experts' stated weight sets. Specifically, we required that the relative pie share of each indicator should not lie outside the minimum and maximum bounds as tabulated in Table 1 (i.e., the pie share of *patents* is between 5% and 20%, the pie share of *royalties* between 0 and 30%, etc.).

Figure 1 is used to demonstrate how the results of our approach can be presented graphically (we refer to Nardo *et al.* 2005b, for comments on the issue of presenting CIs and for a list of possible alternatives). We have taken two examples, to wit, top ranked Finland and Singapore, for which the baseline scenario in fact entails a significant drop as compared with the actual TAI-figures. The difference in the total composite indicator value is indicated by the size of the pies, the importance of the sub-indicators by the pie-shares. Directly below the figures, in Table 2, one finds the values of these and some other countries' "pie shares" (measured in *absolute* numbers), so that the sum of these shares yields their composite indicator value. Recalling Table 1, one can readily inspect that all tabulated pie shares are in accordance with our starting point of granting leeway to each country when assigning shares, without however violating the upper and lower bounds on the relative shares as retrieved from the expert group. One further infers that the so-obtained pie shares can in fact be quite diverse in terms of their *relative importance.* Compare e.g. Finland with Singapore, with Finland assigning 1/4 of its total to schooling, and Singapore less than 1/14; or with Finland assigning 16/100 to *royalties* whereas Singapore actually maximizes its (duly constrained) score by completely neglecting that sub-indicator, etc. This is even the case for countries having a similar composite indicator value (e.g. Belgium and New Zealand). Note that assigning zero weights is consistent with the idea of respecting the lower bounds as provided by the panel (for example, three experts recommended to discard the electricity indicator). Using superscripts, we have also indicated in Table 2 whether the pie share-constraints are

[7] Essentially, the pie share constraint limit weight flexibility across performance dimensions. In fact, it is also possible to limit weight flexibility across countries, i.e. weights cannot vary (too much) over different country observations. See Cherchye and Kuosmanen (2004) for a detailed discussion. We will refrain from pursuing this further in this paper, and instead build explicitly on the information as provided by our experts. In doing so, we bear in mind the remark of Foster and Sen (1997, p. 206) that 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".

binding at the lower or upper bound (or not). Again, one notices considerable differences in this respect among the different countries.

Figure 1: TAI for Finland (100%) and Singapore (14.3%) in baseline DEA scenario

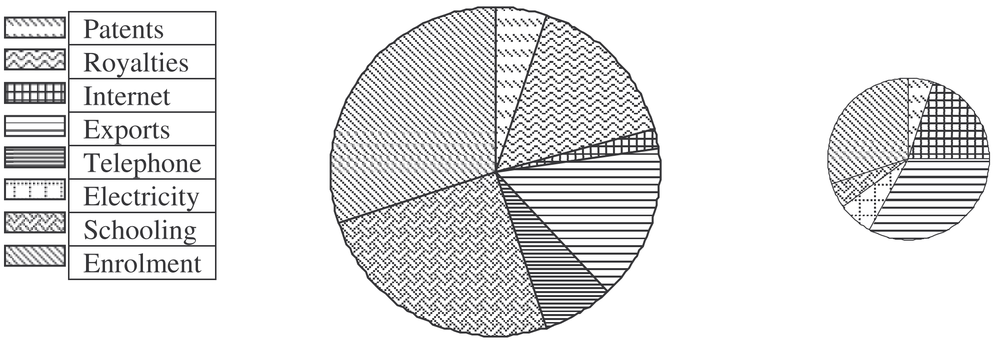


Table 2: Pie shares (in absolute terms) and their ('composite') sum for selected countries

Pie shares	Patents	Royalties	Internet	Exports	Telephone	Electricity	Schooling	Enrolment	Σ
Finland	$0,05^{L}$	0,16	$0,02^{L}$	0,15	0,07	$0,00^{L}$	$0,25^{U}$	$0,30^{U}$	**1.000**
Japan	$0,20^{U}$	$0,00^{L}$	$0,02^{L}$	$0,33^{U}$	$0,20^{U}$	0,12	0,09	0,04	**1.000**
Belgium	$0,03^{L}$	$0,18^{U}$	$0,01^{L}$	0,07	$0,01^{L}$	$0,00^{L}$	$0,15^{U}$	0,15	**0.616**
NZ	$0,03^{L}$	$0,00^{L}$	$0,12^{U}$	$0,06^{L}$	0,07	$0,09^{U}$	$0,15^{U}$	0,09	**0.614**
Italy	$0,01^{L}$	$0,00^{L}$	$0,00^{L}$	0,04	$0,04^{U}$	$0,00^{L}$	$0,05^{U}$	$0,06^{U}$	**0.204**
Singapore	$0,01^{L}$	$0,00^{L}$	$0,03^{U}$	$0,05^{U}$	$0,00^{L}$	0,01	$0,01^{L}$	$0,04^{U}$	**0.143**

Superscript 'L' (resp. 'U') indicates that this value equals the lower (resp. upper) bound of the relative pie share constraint associated with this indicator.

In Table 3, the composite indicator values of this baseline scenario (in bold) are compared with two other cases. On the left one finds the actual TAI values as calculated by the UN, and countries have been ranked in the table accordingly. The second column provides DEA-based CI-values for a benefit-of-the-doubt model that only uses the non-negativity constraints on weights. All countries get a higher CI value than they have on the basis of the UN's fixed weighting scheme, and several of them even get the maximum score of 100%. This is what one would expect from 'benefit-of-the-doubt' weights. As we emphasized before, the unrestricted DEA model allows for extreme weight scenarios; e.g., in our application, we get a zero weight in 63.5% of all 184 cases (= 23 countries x 8 dimensions). Still, differences are also partially due to the 'artificial' normalisation stage of the actual TAI. Norway, for example, considerably increases its score due to the fact that its relatively high figures for *telephone* and *electricity* are no longer smoothed out by taking logarithms. Similarly, the sharp drop for Singapore as one moves from the first to the third column is largely driven by this phenomenon, working in the opposite direction.[8] Once again of course, this highlights the need for uncertainty analysis.

[8] If one calculates the equally weighted TAI with normalized data, but without using logs for these 2 indicators, Norway would move up to the 4th place. Similarly, Singapore would be positioned on the 13th rather than on the 8th place. The maximal score of Singapore in the second column is a direct consequence of the absence of any

Table 3: TAI-values

Countries	Original (UN)	Unconstrained BotD	**Baseline**
Finland	0.744	1.000	**1.000**
US	0.733	1.000	**1.000**
Sweden	0.703	1.000	**1.000**
Japan	0.698	1.000	**1.000**
Rep. of Korea	0.666	1.000	**0.625**
Netherlands	0.630	0.994	**0.901**
UK	0.606	0.976	**0.750**
Canada	0.589	0.982	**0.435**
Australia	0.587	1.000	**0.618**
Singapore	0.585	1.000	**0.143**
Germany	0.583	0.921	**0.818**
Norway	0.579	1.000	**0.732**
Ireland	0.566	0.831	**0.735**
Belgium	0.553	0.802	**0.616**
New Zealand	0.548	0.975	**0.614**
Austria	0.544	0.820	**0.729**
France	0.535	0.849	**0.736**
Israel	0.514	0.813	**0.565**
Spain	0.481	0.756	**0.436**
Italy	0.471	0.822	**0.204**
Czech Republic	0.465	0.792	**0.331**
Hungary	0.464	0.856	**0.320**
Slovenia	0.458	0.684	**0553**

4. Uncertainty and sensitivity analysis

4.1. Uncertainty analysis

In the general case, uncertainties in the development of a composite indicator would be linked to a number of factors, including (Nardo *et al.*, 2005b):

a) The model chosen for estimating the measurement error in the data, e.g. based on available information on variance estimation.
b) The mechanism for including or excluding sub-indicators in the composite.
c) The transformation and/or trimming of sub-indicators, e.g. removing outliers.
d) The type of normalisation scheme, e.g. re-scaling or standardisation, applied to remove scale effects from the sub-indicators.
e) The amount of missing data and the choice of imputation algorithm used to replace missing data.
f) The choice of the weights, e.g., equal weights or weights derived from a DEA-based approach.
g) The level of aggregation, if more than one levels are used, e.g., at the indicator or at the sub-indices level.
h) The choice of aggregation system, e.g., additive, multiplicative, or multi-criteria analysis.

pie-share constraints: it then assigns no less than 85% of its pie to the exports indicator, and effectively neglects five other dimensions (viz. those for which it reaches the lower bound in table 2).

All these assumptions can heavily influence countries scores in a composite indicator and should be taken into account before attempting any interpretation of the results. Saisana *et al.* (2005) studied the uncertainties in the Technology Achievement Index focusing on the type of normalisation for the sub-indicators, the weighting scheme, and the sub-indicators' weights.

Even when restricting ourselves in this work to a DEA-model that incorporates expert opinion, our baseline scenario still is characterized by specific modelling choices. We focus on two points of the chain of composite indicator building, which can introduce uncertainty in the countries scores: point (c) on the consideration of logarithms for "Telephones" and "Electricity", as applied in the original version of the Index by the UN and point (f) on the weights provided by experts and the weight bound scenarios for the DEA-model. We remind the reader that in the DEA-model, normalisation is not required.

The uncertain input factors in our analysis are described in Table 4. The triggers X_1 to X_{21} decide whether to consider an Expert's set of weights. The experts are sampled independently of one another. Next, factor X_{22} determines the type of the weight bound scenario for the DEA, be it either the min-max values of the set of weights of the selected experts, or the 5^{th}-95^{th} percentiles, or the 10^{th}-90^{th} percentiles. Finally, trigger X_{23} determines whether to use logarithms for "Telephones" and "Electricity". Note that in the $K = 23$ dimensional space of uncertainties there are $2^{21}\times3\times2$ = 12,582,912 possible combinations of the input factors values. Given that we cannot afford a full design with so many simulations, we need a representative sampling of the space of uncertainties. We anticipate here that we use an **LP-τ** sampling scheme (Sobol', 1967) of size $N = 24{,}576$ for the purposes of the sensitivity analysis to be discussed in detail in Section 4.2. With **LP-τ** we guarantee that all combinations consider more than three experts.

Table 4: The 23 uncertain input factors for the analysis

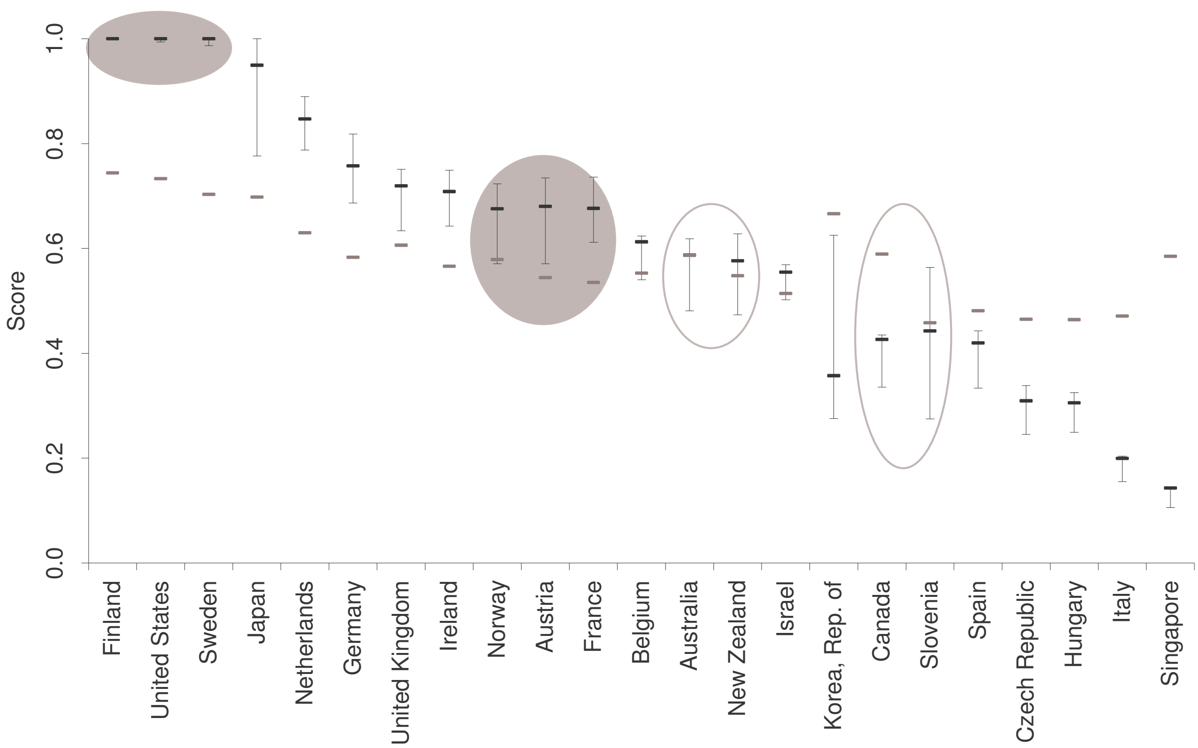
Input factor	*Definition*	*Alternatives*
X_1	Consideration of Expert 1	▪ Included ▪ Excluded
X_2	Consideration of Expert 2	▪ Included ▪ Excluded
…	…	…
X_{21}	Consideration of Expert 21	▪ Included ▪ Excluded
X_{22}	Weight bound scenario	▪ Min-max ▪ 5th -95th percentile ▪ 10th -90^{th} percentile
X_{23}	Data transformation	▪ Raw data ▪ Logarithms for Telephones & Electricity

All these uncertainties are translated into a set of N combinations of scalar input factors, which are sampled from their discrete distributions (2 levels for X_1, X_{21} and X_{23}, and 3 levels for X_{22}) in a Monte Carlo simulation framework. The composite indicator is then evaluated N times and the values obtained are associated to the corresponding draws of uncertain factors to appraise their influence. As a result, all composite indicator values are non-linear functions of the uncertain input factors, and the estimation of their probability distribution functions (pdf) is the purpose of the uncertainty analysis (UA).

Figure 2 presents the results of the uncertainty analysis for the Technology Achievement Index, which are summarised by country scores statistics (median, 5^{th} and 95^{th} percentiles).

The graph should be read "horizontally": sets of whisker plots partially overlapping indicate situations when the ranking of the corresponding countries can interchange, so showing similar degree of performance. If two countries have non-overlapping bounds, the policy inference is robust, independently of the level of uncertainty in the data. Finland, USA, and Sweden are unarguably the best performing countries, both in the original UN version of TAI and in the present case in which we acknowledge uncertainties related to the DEA-model. There are, however, several countries whose relative performance is strongly influenced by the assumptions in the evaluation model.

Figure 2: Results of Uncertainty analysis - Countries scores



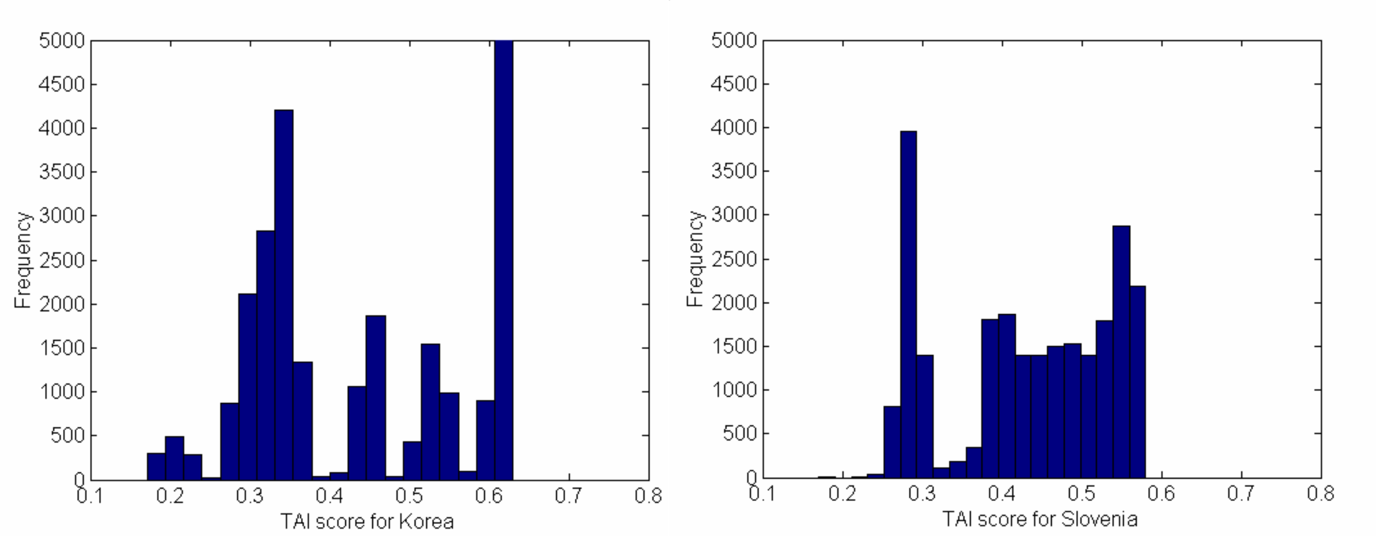
Note: Original TAI scores in the UN version (grey marks), median TAI scores (black mark), 5th and 95th percentiles (bounds). Countries are ordered according to the median of ranks.

Table 4 gives the countries ranks based on the original TAI and the median of ranks from the robustness analysis. For about 13 countries the difference between the TAI rank and the median rank when considering the DEA-related assumptions is less than 2 positions. Singapore and Korea decline the most (more than 10 positions). Conversely, Austria, France and Slovenia improve their rank between 6 and 7 positions. Looking at the range of the uncertainty bounds, Korea and Slovenia are two of the most volatile countries. Their distributions are plotted in Figure 3. Korea's score, considering the 5^{th} and 95^{th} percentiles of the distribution, can range between 0.276 and 0.625, while for Slovenia the performance is estimated between 0.275 and 0.564.

Table 4: Countries ranks and median scores

	Countries	Rank based on original score	Median of Ranks	Difference in rank (original-median)	Median of scores
FIN	Finland	1	2	-1	1.000
USA	United States	2	2	0	1.000
SWE	Sweden	3	2	+1	1.000
JPN	Japan	4	4	0	0.950
KOR	Korea, Rep. of	5	16	-11	0.357
NLD	Netherlands	6	5	1	0.847
GBR	United Kingdom	7	7	0	0.719
CAN	Canada	8	17	-9	0.426
AUS	Australia	9	14	-5	0.587
SIN	Singapore	10	23	-13	0.143
DEU	Germany	11	6	5	0.758
NOR	Norway	12	10	2	0.675
IRL	Ireland	13	8	5	0.708
BEL	Belgium	14	12	2	0.613
NZL	New Zealand	15	14	1	0.576
AUT	Austria	16	10	6	0.680
FRA	France	17	10	7	0.676
ISR	Israel	18	15	3	0.555
ESP	Spain	19	18	1	0.420
ITA	Italy	20	22	-2	0.200
CZE	Czech Republic	21	20	1	0.309
HUN	Hungary	22	21	1	0.305
SVN	Slovenia	23	17	6	0.443

Figure 3: Uncertainty analysis of the composite indicator TAI for Korea and Slovenia who present very large uncertainty bounds (most volatile countries)



An evident question related to the overlapping in the countries scores, as shown in Figure 3, is: which countries have significantly different performance in the technological development? Can we argue that France (median score = .676) performs significantly better than Norway (median score = .675), or that Canada's level of technological achievement (median score = .426) is superior to that of Spain (median score = .420)? A hypothesis test could provide such an answer. One of the advantages of uncertainty analysis, which has not been exploited so far in the literature of composite indicators development, is that it allows for an estimation of the pdf for a country's score and the respective pdf of its rank. We applied the Wilcoxon signed rank test, also known as the Wilcoxon matched pairs test, to test the median difference in paired TAI scores (Conover, 1980). This test is the non-parametric equivalent of the paired t-test, and it does not require the distributional assumption that the differences follow a normal distribution, which was not confirmed in our case. On the contrary, the only assumption required for this test, that the distribution of the differences is symmetric, was confirmed. Applying this test we identify four groups of countries for which no distinction should be made on their technological achievement level. The groups are shaded in grey in Figure 3. The first group contains the top three performing countries Finland (1.00), USA (1.00) and Sweden (1.00). The second group contains Austria (.680), France (.676) and Norway (.675). Australia (.587) and New Zealand (.576) belong to the third group. Finally, Slovenia (.443) and Canada (.426) belong to the forth group. Note that, although the median score for Spain is .420, which is very close to that of Canada, the performance of the two countries can be clearly distinguished.

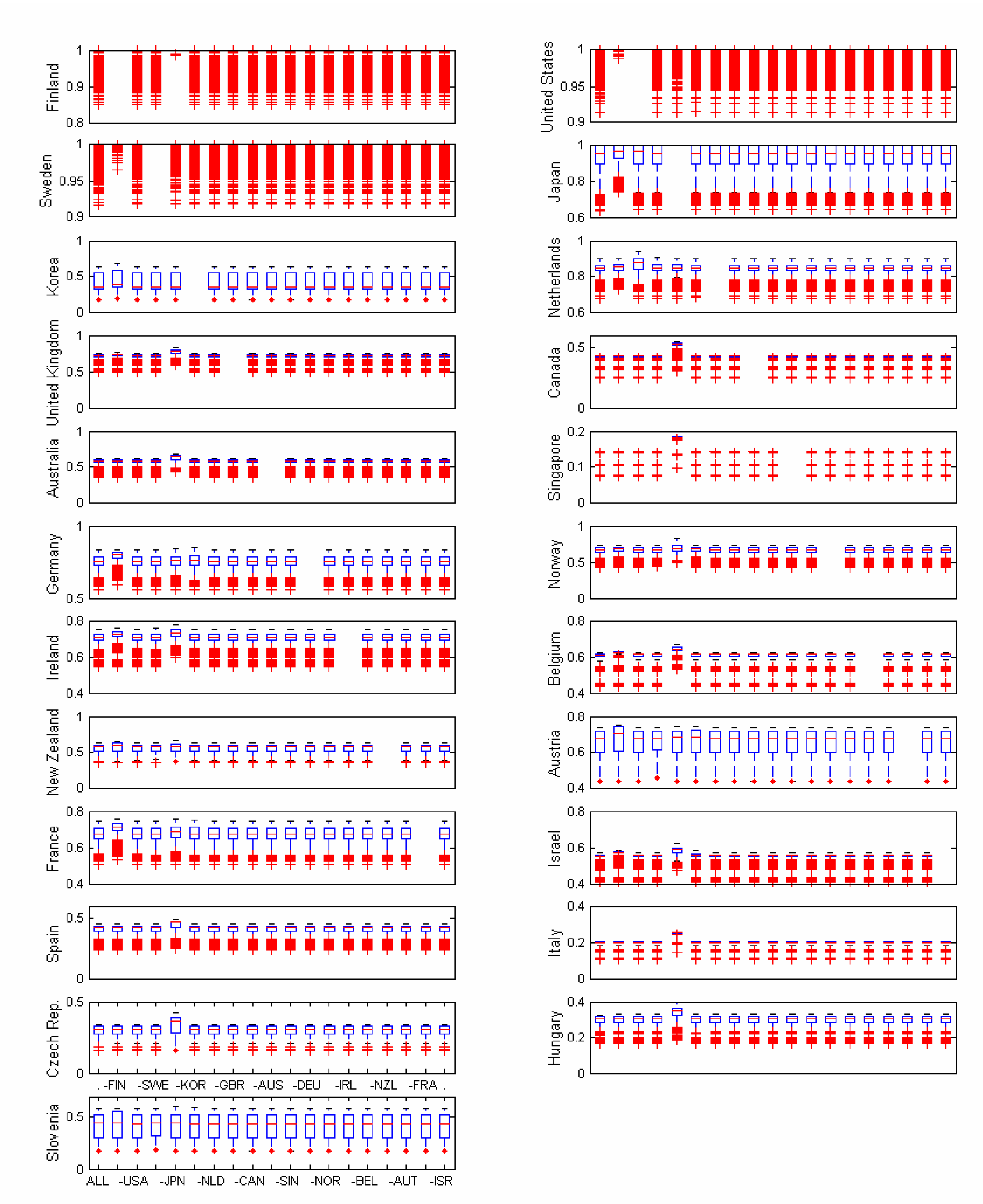
4.2 Sensitivity analysis

We next complement our uncertainty analysis with sensitivity analysis. We first investigate sensitivity of the above uncertainty results with respect to outlier countries. Subsequently, we use variance-based techniques to apportion the calculated (aggregate) variance/uncertainty in the country scores to the uncertain input factors in our analysis; this provides insight into the sensitivity of the countries scores with respect to each individual source of uncertainty.

Sensitivity analysis for outlier countries

Procedures that randomly omit some observations (in our case countries; e.g. one randomly excludes one country at a time) have been suggested in the DEA literature as a way to correct for the impact of outlier observations (e.g., Wilson, 1995; Cazals *et al.*, 2002; Simar, 2003). With a view to assess such an impact on the countries scores we have repeated the Monte Carlo approach described above eliminating one country at a time from the set of 23 countries. The $N = 24{,}576$ composite indicator values are estimated for each group of 22 countries. The boxplots are presented in Figure 4, for the entire set of 23 countries (ALL) and for each country's elimination starting from Finland to Israel. The two countries that have the greatest impact in the countries scores are Japan and Finland. When Japan is eliminated from the set, the countries that improve their score are: Finland, United Kingdom, Australia, Ireland, Spain, Czech. Rep., Canada, Singapore, Norway, Belgium, Israel, Italy and Hungary. When Finland is eliminated from the set, the countries that improve their score are: Sweden, United States, Korea, Germany, France, and to a lesser degree Japan. The elimination of the remaining countries does not have any notable impact on the countries scores. This result confirms a quite robust DEA-model to outlier observations. In our opinion, this justifies that we do not explicitly account for outlier countries in our remaining analysis.

Figure 4: Boxplots of countries TAI scores when eliminating one country at a time



Note: The box has lines at the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the box to show the extent of the rest of the data. Outliers (+) are data with values beyond the ends of the whiskers. If there are no data outside the whisker, a dot is placed at the bottom whisker.

Sensitivity analysis using variance-based techniques

At this step it is useful to use sensitivity analysis to apportion the variance (uncertainty) in the countries scores to the different *K* uncertain input factors (in our case *K* = 23; see Table 4). The starting point of the variance-based methods is the variance decomposition $V(Y) = V(E(Y \mid X_k)) + E(V(Y \mid X_k))$, where X_k is an uncertain input factor, $k = 1,...,K$. Note that in both expressions V(E(.)) and E(V(.)) the outer operator is taken over the conditioning argument while the inner operator is taken over its complementary set, i.e.

$V(E(Y \mid X_k)) \equiv V_{X_k}\left(E_{\mathbf{X}_{-k}}\left(Y \middle| X_k\right)\right)$ where $\mathbf{X}_{-k}$ is the vector of all-but-k factors.

The first-order sensitivity measures can be calculated as $S_k = V(E(Y \mid X_k))/V(Y)$ for each uncertain factor. The higher S_k, the higher the importance of X_k, as the larger the average drop in variance $V(Y)$ obtained when fixing X_k within its range.

In the case of an additive (and hence linear) model where no interactions between its uncertain factors occur, we have $\sum_{k=1}^{K} S_k = 1$. For non-additive models, higher order sensitivity measures that capture interaction effects among sets of input factors have to be computed, to help us improve our understanding of the model structure. However, higher order measures are usually not estimated, as in a model with *K* factors the total number of sensitivity measures (including the first-order) that should be estimated is as high as 2^K-1. For this reason, a more compact sensitivity measure is used. This is the total-effect measure that concentrates in one single term all the interactions involving a given factor X_k (Homma and Saltelli, 1996). We indicate with S_{Tk} the average of the four estimates of total-effect measures.

When several layers of uncertainty are simultaneously activated, composite indicators turn out to be non-linear, possibly non-additive models due to interactions between the uncertain input factors (Saisana *et al.*, 2005). As a result, all TAI scores and ranks are non-linear functions of the uncertain input factors. As argued by practitioners (Saltelli *et al.*, 2000, EPA 2004), robust, "model-free" techniques for sensitivity analysis should be used for non-linear models. Variance-based techniques have been shown to yield useful results for sensitivity analysis. The discussion of their methodological formulation to compute sensitivity measures that account for the interaction between the input factors goes beyond the scope of this report and the reader is referred to Saltelli *et al.* (2000). Here we only display those additional properties of model-free variance-based techniques that are convenient for the present analysis:

- they allow an exploration of the whole range of variation of the input factors, instead of just sampling factors over a limited number of values, as done e.g. in fractional factorial design (Box *et al.* 1978);
- they are quantitative, and can distinguish main effects (first order) from interaction effects (second and higher order);
- they are easy to interpret and to explain;
- they allow for a sensitivity analysis whereby uncertain input factors are treated in groups instead of individually.

The extended variance-based methods, including the version we used here based on the work of Saltelli (2002), are implemented in the freely distributed software SIMLAB (Saltelli *et al.*, 2004).

The pair (S_k, S_{Tk}) gives a fairly good description of the DEA-model sensitivities at a reasonable cost, which for the improved method is of $2N(K+1)$ model evaluations. In our analysis, the base sample is of size N = 512 and the composite indicator value for each country is evaluated performing $2 \cdot 512 \cdot (23+1) = 24{,}576$ DEA-model runs.

The sensitivity measures S_k and S_{Tk} are given in Table 5. When we use S_k for sensitivity analysis, we are looking for important input factors that - if fixed singularly - would reduce the most the variance in the output variable. "Importance" in sensitivity analysis, though, is a relative notion and there is no established threshold: one usually looks at the S_k values and the distances between them and considers the first few factors as important. In this work, an input factor will be considered as important if $S_k > 0.10$ (i.e. if the input factor explains more than 10% of the variance in a country's score). The greater the value of the measure $S_{Tk} - S_k$, the more that factor is involved in interactions with other factors.

The countries with the largest uncertainty bounds are Slovenia and Korea. Some 70% of Slovenia's variance is mainly explained by consideration of Expert 3. A total of 82.6% of the country's variance in technological achievement is explained by considering the input factors singularly. The remaining 17.4% of the variance is due to interactions among the factors. Korea's variance is due to the consideration of Expert 15 (73% variance explained) and to interactions of this factor with the consideration of Expert 3.

For the entire set of countries, Experts 1, 3, 12, 14, and 20 are those driving most of the variance in the countries scores. The weight bound scenario (X_{22}) is influential only to a few countries, i.e. Japan, Netherlands, Germany, Norway, Ireland and France. Finally, the data transformation which consists in considering the logarithms of 'Telephones' and 'Electricity' is not influential to any country's variance, when total-effects measures are analysed. This means that it would be meaningless to discuss on the use of scale transformations for those two indicators (*in the class of DEA-models*). In principle, one could omit the log-transformation as the results, in terms of country scores, are not affected. This result may seem to contradict the conclusion in Section 3 on Norway's and Singapore's score being affected by the logarithmic transformation. However, the right interpretation is that our discussion in Section 3 pertained to a comparison of the original TAI model (which includes logarithmic transformation and fixed weighting) with the unconstrained DEA model (which includes flexible weighting without a prior logarithmic transformation), whereas the current discussion relates to comparing alternative DEA-models (i.e., with or without the logarithmic transformation). In fact, the result in Table 5 suggests that flexible weighting DEA-models (to an important extent) effectively accommodate for sensitivity with respect to prior transformation schemes.

Table 5. Sensitivity measures of first order and total effect for the composite indicators scores.

First-order sensitivity measures, S_k

	X_1	X_2	X_3	X_4	X_5	X_7	X_{10}	X_{12}	X_{14}	X_{15}	X_{16}	X_{17}	X_{19}	X_{20}	X_{22}	X_{23}	Sum
Finland	.01	.00	.00	.00	.00	.00	.00	.04	.02	.00	.00	.00	.00	.01	.00	.00	.091
United States	.00	.00	.00	.00	.09	.06	.00	.00	.00	.00	.14	.02	.00	.00	.00	.00	.317
Sweden	.01	.00	.01	.00	.03	.11	.03	.00	.00	.02	.10	.00	.01	.01	.05	.01	.460
Japan	.00	.00	.04	.00	.06	.12	.14	.00	.00	.16	.00	.00	.00	.00	.13	.01	.685
Korea	.00	.00	.10	.00	.00	.00	.00	.00	.00	.73	.00	.00	.00	.00	.00	.00	.855
Netherlands	.00	.00	.00	.00	.13	.12	.01	.00	.01	.14	.00	.00	.02	.00	.17	.01	.653

U. Kingdom	.07	.00	.00	.00	.00	.00	.00	.03	.05	.09	.00	.00	.01	.03	.03	.00	.323
Canada	.06	.00	.01	.00	.00	.00	.00	.08	.06	.00	.00	.00	.00	.09	.00	.00	.316
Australia	.05	.01	.17	.01	.00	.00	.01	.08	.05	.00	.00	.00	.03	.15	.02	.00	.615
Singapore	.07	.00	.00	.00	.00	.00	.00	.07	.06	.03	.00	.00	.00	.07	.00	.00	.332
Germany	.00	.00	.05	.00	.00	.00	.04	.00	.01	.31	.00	.00	.00	.01	.13	.07	.627
Norway	.02	.00	.12	.01	.00	.01	.00	.09	.03	.00	.00	.00	.01	.19	.09	.02	.621
Ireland	.05	.01	.00	.00	.00	.00	.00	.00	.05	.19	.00	.00	.01	.01	.06	.03	.401
Belgium	.07	.00	.00	.00	.00	.00	.00	.05	.07	.02	.00	.00	.00	.05	.00	.04	.315
New Zealand	.01	.01	.48	.01	.00	.00	.00	.03	.00	.00	.00	.00	.01	.18	.02	.05	.808
Austria	.00	.00	.58	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.13	.02	.01	.748
France	.00	.00	.06	.00	.00	.00	.03	.00	.02	.39	.00	.00	.00	.00	.12	.02	.646
Israel	.05	.00	.00	.02	.00	.00	.00	.05	.10	.00	.00	.00	.00	.03	.01	.01	.287
Spain	.05	.01	.28	.01	.00	.00	.00	.02	.06	.01	.00	.00	.00	.14	.00	.02	.583
Italy	.06	.00	.04	.00	.00	.00	.00	.06	.07	.00	.00	.00	.00	.10	.00	.00	.347
Czech Rep.	.02	.01	.43	.00	.00	.00	.00	.02	.02	.00	.00	.00	.00	.17	.00	.01	.698
Hungary	.04	.00	.25	.00	.00	.00	.00	.03	.04	.00	.00	.00	.00	.14	.00	.02	.545
Slovenia	.00	.00	.70	.00	.00	.00	.00	.00	.00	.03	.00	.00	.00	.08	.00	.01	.826

Total-effect sensitivity measures, S_{Tk}

	X_1	X_2	X_3	X_4	X_5	X_7	X_{10}	X_{12}	X_{14}	X_{15}	X_{16}	X_{17}	X_{19}	X_{20}	X_{22}	X_{23}	
Finland	.65	.00	.00	.43	.00	.01	.00	.66	.57	.01	.01	.00	.10	.69	.00	.00	
United States	.02	.48	.08	.02	.56	.56	.13	.02	.02	.16	.51	.14	.00	.02	.09	.00	
Sweden	.16	.22	.02	.04	.35	.56	.11	.04	.14	.29	.26	.08	.09	.05	.10	.07	
Japan	.04	.04	.12	.05	.18	.24	.31	.03	.03	.28	.03	.04	.03	.04	.33	.01	
Korea	.03	.01	.15	.02	.01	.03	.03	.01	.02	.78	.02	.01	.00	.04	.03	.00	
Netherlands	.08	.07	.05	.06	.30	.28	.04	.08	.11	.21	.07	.03	.05	.09	.27	.01	
U. Kingdom	.46	.01	.01	.10	.01	.02	.02	.42	.43	.12	.01	.01	.02	.42	.07	.00	
Canada	.50	.01	.01	.09	.01	.00	.01	.54	.52	.00	.01	.00	.00	.53	.01	.00	
Australia	.26	.02	.21	.06	.02	.04	.02	.36	.33	.00	.02	.01	.06	.39	.06	.00	
Singapore	.54	.01	.00	.08	.01	.00	.01	.55	.54	.00	.01	.00	.00	.55	.00	.00	
Germany	.06	.03	.15	.04	.07	.06	.14	.04	.06	.43	.02	.02	.02	.08	.25	.07	
Norway	.15	.02	.17	.05	.03	.07	.01	.30	.22	.01	.03	.02	.05	.38	.19	.02	
Ireland	.34	.02	.03	.10	.02	.06	.05	.27	.31	.24	.02	.01	.03	.26	.12	.03	
Belgium	.50	.01	.00	.13	.00	.01	.01	.50	.50	.02	.01	.00	.01	.47	.00	.05	
New Zealand	.05	.01	.59	.03	.01	.03	.01	.11	.08	.01	.01	.01	.02	.31	.06	.05	
Austria	.01	.01	.75	.01	.02	.01	.02	.01	.02	.02	.01	.01	.01	.29	.07	.01	
France	.06	.02	.13	.03	.06	.06	.12	.02	.07	.49	.02	.02	.02	.07	.22	.03	
Israel	.44	.01	.01	.14	.02	.02	.02	.46	.50	.02	.01	.00	.00	.41	.06	.01	
Spain	.27	.01	.35	.07	.02	.01	.02	.25	.28	.03	.02	.00	.00	.40	.02	.02	
Italy	.49	.01	.05	.08	.01	.00	.01	.49	.50	.00	.01	.00	.00	.54	.00	.00	
Czech Rep.	.17	.01	.52	.05	.01	.01	.00	.17	.18	.01	.01	.00	.00	.39	.02	.01	
Hungary	.30	.01	.31	.07	.01	.01	.01	.30	.30	.01	.01	.00	.00	.43	.02	.02	
Slovenia	.01	.00	.81	.01	.01	.01	.01	.01	.01	.06	.01	.01	.00	.18	.03	.01	

Marked values are in yellow (>0.10), grey (>0.30) and black (>0.50). S_k is the average of the eight estimates of first-order measures for factor X_k, and S_{Tk} is the average of the four estimates of total effect for factor X_k.

We are further interested in the difference in the technological achievement levels between Slovenia (median score = .4425) and Korea (median score = .3571), which present significant overlapping in their scores. In such cases, where partial overlapping between two countries occurs, the difference in the countries scores can be further analyzed in a sensitivity framework to identify the most influential factors and provide insight into the situation. Figure 5 provides the relative performance of the two countries acknowledging the uncertainties in the DEA-approach. Note that, although Slovenia has a higher median score of technological performance than Korea, 61% of the score differences fall in the left-hand region, where Korea performs better than Slovenia. In fact, this was the message, i.e. the better performance of Korea with respect to Slovenia, which was conveyed when examining the median of the ranks. We recall that in the original version of TAI, Korea is ranked 5th, whilst Slovenia is situated on the 23rd position. The next issue that comes into question is: which factors are mostly responsible for that uncertainty? The results of the sensitivity analysis are given in Table 6. Taken singularly, the factors account for 91.3% of the variance in the difference between the two countries. Most of the variance is due to the consideration of the weights provided by Expert 15 (70.5%) and by Expert 3 (16.3%). The remaining small portion of the output variance, i.e. 8.7%, is explained by the interactions among the factors themselves. Previously, in the study of the effect of expert inclusion/exclusion on the individual country scores, Expert 15 was mainly responsible for Korea and Expert 3 for Slovenia. Yet, one could not foresee the degree to which these two Experts determine the difference in scores between the two countries. This is now feasible thanks to the results presented in Table 6, from which we can see that Expert 15 drives the preference between Korea and Slovenia. As expected, the importance of factors depends from the output being considered. Here it makes a difference whether we look at a country score or at the difference between two countries. In general, prior to applying sensitivity analysis, the questions to be answered need to be clearly specified and the output variables clearly identified.

Figure 5: Uncertainty analysis for the Technology Achievement Index for Slovenia vs. Korea

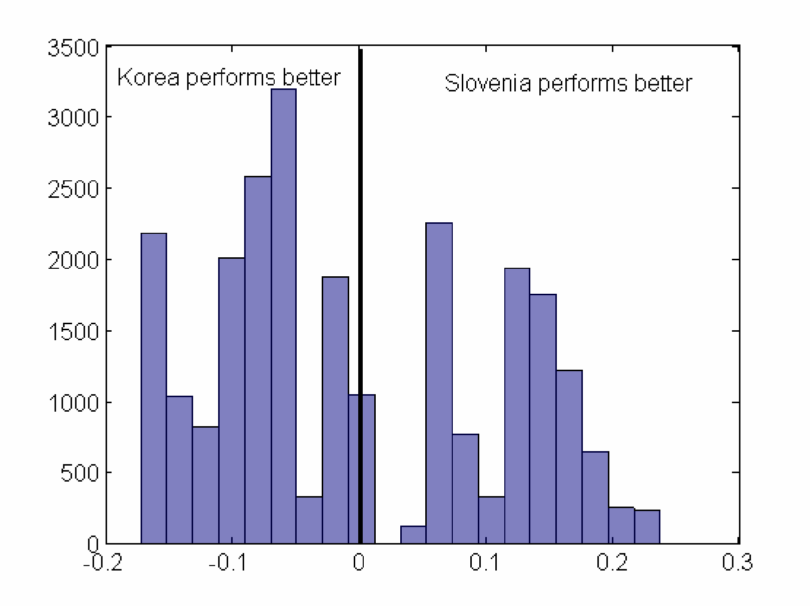


Table 6: Sensitivity measures of first-order and total effect for the difference between the composite indicator score for Slovenia and Korea

	S_k	S_{Tk}
X_1	0.000	0.018
X_2	0.002	0.006
X_3	**0.163**	**0.209**
X_4	0.000	0.007
X_5	0.001	0.004
X_6	0.003	0.002
X_7	0.000	0.015
X_8	0.001	0.003
X_9	0.000	0.006
X_{10}	0.002	0.015
X_{11}	0.002	0.004
X_{12}	0.000	0.008
X_{13}	0.003	0.001
X_{14}	0.000	0.021
X_{15}	**0.705**	**0.731**
X_{16}	0.003	0.007
X_{17}	0.000	0.001
X_{18}	0.000	0.005
X_{19}	0.000	0.000
X_{20}	0.019	0.047
X_{21}	0.000	0.001
X_{22}	0.000	0.015
X_{23}	0.008	0.008
Sum	*0.913*	

Conclusion

Media and policy-makers look with increasing interest at composite indicators as appealing tools to attract the attention of the community, build narratives and help focusing policy debates. Methodological gaps or fragilities in their design and construction may invite politicians to draw simplistic conclusions or the press to communicate misleading information. That is why national and international organisations believe that it is important to focus on methodological issues in the design of composite indicators. Here, we have illustrated a generalisation of the DEA-model for the selection of weights combined with a variance-based sensitivity analysis method. In addition, we have tested it on a practical case study related to the design stage of composite indicators, where rarely robustness and sensitivity analysis are applied.

References

Allen, R., A. Athanassopoulos, R. G. Dyson and E. Thanassoulis (1997), Weights restrictions and value judgements in data envelopment analysis: evolution, development and future directions, *Annals of Operations Research* 73, 13 -34.

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

Banker, R.D., H. Chang and W.W. Cooper (1996), Simulation Studies of effficiency, returns to scale and misspecification with nonlinear functions in DEA, *Annals of Operations Research* 66, 233-253.

Box, G., Hunter, W. and Hunter, J. (1978), Statistics for experimenters, New York: John Wiley and Sons.

Charnes A. and W.W. Cooper (1985), Preface to Topics in Data Envelopment Analysis, *Annals of Operations Research* 2, 59-94.

Charnes, A., W.W. Cooper and E. Rhodes (1978), Measuring the efficiency of decision making units, *European Journal of Operations Research* 2, 429-444.

Cherchye, L. and T. Kuosmanen (2004), Benchmarking Sustainable Development: A Synthetic Meta-Index Approach, in M. McGillivray and M. Clarke (eds.), *Perspectives on Human Development*, United Nations University Press, to appear.

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

Cherchye, L. , W. Moesen , C.A.K. Lovell and T. Van Puyenbroeck (2005), One Market, One Number: A Composite Indicator Assessment of EU Internal Market Dynamics, CES DP Series 05.16, KU Leuven.

Cazals, C. Florens, J.P. and L. Simar (2002), Nonparametric Frontier Estimation: a Robust Approach , *Journal of Econometrics,* 106, 1–25.

Conover, W. J. (1980), *Practical Nonparametric Statistics,* New York, Wiley.

Cook, W.D. (2004), Qualitative data in DEA, in W.W. Cooper, L.M. Seiford and J. Zhu (eds.), *Handbook on Data Envelopment Analysis*, Kluwer Academic Publishers, Boston.

Cook, W.D., M. Kress (1994), A multiple criteria composite index model for quantitative and qualitative data, *European Journal of Operations Research* 78, 367-379.

Cooper, W.W., L. Seiford and K. Tone (2000), *Data Envelopment Analysis*, Kluwer Academic Publishers, Boston.

Cooper, W.W., Li, Shanling, Seiford, L.M. and Zhu, Joe (2004), Sensitivity Analysis in DEA, in W.W. Cooper, L.M. Seiford and J. Zhu (eds.), *Handbook on Data Envelopment Analysis*, Kluwer Academic Publishers, Boston.

Desai M., Fukuda-Parr S., Johansson C. and Sagasti F. (2002), Measuring the Technology Achievement of Nations and the Capacity to Participate in the Networking Age*, Journal of Human Development* 3 (1), 301-311.

Despotis, D.K. (2005), A reassessment of the human development index via data envelopment analysis, *Journal of the Operational Research Society* 56, 969-980.

EPA (2004) Council for Regulatory Environmental Modeling, "Draft Guidance on the Development, Evaluation, and Application of Regulatory Environmental Models". (http://www.epa.gov/osp/crem/library/CREM%20Guidance%20Draft%2012_03.pdf.)

Foster, J., A. Sen (1997), *On Economic Inequality (2nd , expanded edition)*, Clarendon Press.

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

Homma, T. and Saltelli, A. (1996), Importance measures in global sensitivity analysis of model output, *Reliability Engineering and System Safety* 52 (1), 1-17.

Kao, C. And H.T. Tung (2005), Data envelopment analysis with common weights: the compromise solution approach, *Journal of the Operational Research Society*, 56, 1196-1203.

Lovell, C. A. K., Pastor, J. T. and Turner, J. A. (1995), Measuring Macroeconomic performance in the OECD: a comparison of European and non-European countries, *European Journal of Operational Research* 87, 507-518.

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

Nardo, M., M. Saisana, A. Saltelli and S. Tarantola (EC/JRC), A. Hoffman and E. Giovannini (OECD) (2005a), *Handbook on Constructing Composite Indicators: Methodology and User's Guide*, OECD Statistics Working Paper, JT00188147.

Nardo, M., M. Saisana, A. Saltelli, S. Tarantola (2005b), Tools for Composite Indicators Building, European Commission-Joint Research Centre, EUR 21682 EN.

Pedraja-Chaporro, F., J. Salinas-Jimenez, P.Smith (1997), On the Role of Weight Restrictions in Data Envelopment Analysis, *Journal of Productivity Analysis* 8, 215-230.

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

Saisana M., Saltelli A., Tarantola S. (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.

Saltelli A. (2002), Making best use of model valuations to compute sensitivity indices, *Comput. Phys. Commu.* 145, 280–97.

Saltelli, A., Chan, K. and Scott, M. (2000), *Sensitivity analysi*s, Probability and Statistics series, New York: John Wiley & Sons.

Saltelli A, Tarantola S, Campolongo F, Ratto M. (2004), *Sensitivity analysis in practice. A guide to assessing scientific models*, England: Wiley.

Sarrico, C.S., R.G. Dyson (2004), Restricting virtual weights in data envelopment analysis, *European Journal of Operational Research* 159, 17-34.

Simar, L. (2003), Detecting Outliers in Frontiers Models: a Simple Approach, *Journal of Productivity Analysis* 20, 391–424.

Simar, L. and P. Wilson (1998), Sensitivity of efficiency scores : How to bootstrap in Nonparametric frontier models, *Management Science* 44, 1, 49-61.

Sobol' IM (1967), On the distribution of points in a cube and the approximate evaluation of integrals, *USSR Comput. Math. Phys.* 7, 86–112.

Storrie, D., H. Bjurek (2000), Benchmarking European Labour Market Performance with Efficiency Frontier Techniques, CELMS Discussion paper, Göteborg University.

Thanassoulis, E., M.C. Portela and R. Allen (2004), Incorporating value judgements in DEA, in W.W. Cooper, L.M. Seiford and J. Zhu (eds.), *Handbook on Data Envelopment Analysis*, Kluwer Academic Publishers, Boston.

United Nations (2001), Annex 2.1: The Technology Achievement Index: A New Measure of Countries' Ability to Participate in the Network Age, in Human Development Report, United Kingdom: Oxford University Press. (http://www.undp.org)

Valdmanis, V. (1992), Sensitivity analysis for DEA models : An empirical example using public vs. NFP hospitals, *Journal of Public Economics* 48, 185-205.

Wilson, P.W. (1995), Detecting influential observations in data envelopment analysis, *Journal of Productivity Analysis* 6, 27-46.

Wong, Y-H B. and J.E. Beasley (1990), Restricting weight flexibility in data envelopment analysis, *Journal of the Operational Research Society* 47, 136-150.