KATHOLIEKE UNIVERSITEIT
LEUVEN

Faculty of Economics and
Applied Economics

Department of Economics

Constructing a knowledge economy composite indicator with imprecise data

by

Laurens CHERCHYE
Willem MOESEN
Nicky ROGGE
Tom VAN PUYENBROECK

Public Economics

Center for Economic Studies
Discussions Paper Series (DPS) 09.15
http://www.econ.kuleuven.be/ces/discussionpapers/default.htm

July 2009

DISCUSSION
PAPER

Constructing a Knowledge Economy Composite Indicator with Imprecise Data

Laurens Cherchye‡
Willem Moesen‡
Nicky Rogge‡*[1]
Tom Van Puyenbroeck‡*

(‡): Katholieke Universiteit Leuven (KULeuven)
Faculty of Business and Economics
Naamsestraat 69, 3000 Leuven (Belgium);
and
(*) : Hogeschool-Universiteit Brussel (HUBrussel)
Centre for Economics & Management (CEM)
Stormstraat 2, 1000 Brussels (Belgium)

July 2009

Abstract

This paper focuses on the construction of a composite indicator for the knowledge based economy using imprecise data. Specifically, for some indicators we only have information on the bounds of the interval within which the true value is believed to lie. The proposed approach is based on a recent offspring in the Data Envelopment Analysis literature. Given the setting of evaluating countries, this paper discerns a 'strong country in weak environment' and 'weak country in strong environment' scenario resulting in respectively an upper and lower bound on countries' performance. Accordingly, we derive a classification of 'benchmark countries', 'potential benchmark countries', and 'countries open to improvement'.

Key words: knowledge economy indicators; composite indicators; Multiple Imputation; Benefit of the Doubt; weight restrictions; Data Envelopment Analysis; data impreciseness
JEL-classification: C14, C61, C82

[1] Corresponding Author: Nicky Rogge, Centrum voor Economie en Management (CEM), Hogeschool Universiteit Brussel, Stormstraat 2, 1000 Brussels, e-mail: Nicky.Rogge@hubrussel.be

1. Introduction

In both the definition and the monitoring of reforming policies, composite indicators (CIs, hereafter) often play an explicit role in shaping the frame of thinking of policy actors by drawing attention to particular issues, identifying benchmark performances and trends, and helping to set policy priorities. Furthermore, compiling a multitude on individual indicators on complex issues into a single index, CIs often seem easier to interpret by policy makers, the media and the general public. Consequently, CIs are increasingly recognized as a tool for policy making and, especially, public communication on countries' relative performances in wide ranging fields such as the environment (e.g., the Environmental Performance Index), the economy (e.g., the Internal Market Index), human development (e.g., the Human Development Index), technological development (e.g., the Technology Achievement Index), etc.

Nevertheless, CIs remain the subject of controversy. Opponents question their credibility mainly by pointing at the subjectivity involved in their construction. Subjective choices are indeed pervasive when answering the many questions bound up with a CI (see Booysen, 2002): what is the overall phenomenon one purports to summarize; which indicators should be included; how should they be aggregated; how to deal with low quality or imprecise data; to what extent can one assess how country rankings are influenced by all the foregoing questions, etc.? In its request to come up with a CI for assessing countries' performance in the Knowledge-Based Economy, the European Commission specifically required to address the problems just indicated. The current paper is an offshoot of this project. Here, the focus is solely on how to deal with missing data in the construction of CIs. If timely data are required, the knowledge-based economy being a case in point, missing data are far from uncommon. Neglecting this fact, as is sometimes done (e.g. by substituting the last data available for each country, or by plainly neglecting the specific indicator for the concerned countries), impedes a genuinely global and up-to-date benchmarking exercise in general, and the construction of a reliable CI in particular.

Note that we do not address the issue of handling missing data as such, but the way in which such handling may be coupled with the subsequent stage of constructing a CI. During the collection stage of data that are related to the Knowledge-Based Economy, multiple imputation was proposed as a technique to fill in the wholes in the dataset. As its label suggests, this methodology generates multiple values, without providing any information whatsoever on which of these values should be perceived as the right one. Against the background of the critique that selecting any *specific* CI model is hard to justify substantially, and taking the multiple imputation technique as given, we opt to incorporate uncertainty in the creation of CI-values. Loosely speaking, we generate CI-confidence intervals based on the several imputations. Specifically, we amend the so-called 'Benefit of the Doubt' construction method (after Melyn and Moesen, 1991 and Cherchye et al., 2007b), to handle mixtures of exact and imprecise data. In doing so, we follow a recent extension discussed in the Data Envelopment Analysis (DEA) literature. We also demonstrate that this proposition is flexible to incorporate a priori information or expert opinion by imposing additional weight restrictions. Especially with an eye towards enhancing credibility and acceptance of CIs in practical applications, this ability of adding extra information seems very convenient.

The rest of the paper is organized as follows. Section 2 presents the data used in the 'KEI'-project (an acronym for 'Knowledge Economy Indicators: Development of Innovative and Reliable Indicator Systems'), whereas in section 3 we shortly discuss the issue of handling missing data with Multiple Imputation. Section 4 briefly reports on the essential characteristics of a Benefit-of-the-Doubt CI. In section 5 we brings both components together and discuss how to introduce mixtures of precise and imprecise (i.e. multiple imputed) data in a 'Benefit of the Doubt'-CI . We apply the approach of Entani et al. (2002), Despotis et al. (2002), and Park (2006) to the final (shortened) KEI dataset for the year 2004 and discuss the most striking results in detail in Section 5. In the final section, we offer some concluding remarks and some avenues for further research.

2. The knowledge economy indicator (KEI) data

In the literature, a considerable number of potential indicators has been proposed for measuring the many constituent drivers, characteristics, and key outputs of a knowledge economy. The data set upon which the subsequent analysis builds, contains not less than 115 of such indicators for the year 2004 for 27 countries, and both the EU-15 and EU-25.[2,3] From this larger set, a subset of 23 indicators was selected, describing 97.4% of the variation in the overall set.[4] Specifically, these 23 indicators are (i) *ICT value added* (% of total business sector value added), (ii) *SMEs ordering over the internet* (% of total SMEs), (iii*) Individuals using the internet for banking* (% total), (iv) *Pisa reading literacy of 15y* (average score), (v) *Total researchers* (per 1000 labour force in FTE), (vi) *Participation in lifelong learning* (% of working 25-64y), (vii) *Employed in creative occupations* (% total), (viii) *BERD performed in service industries* (%), (ix) *EPO high tech patent applications* (per million population), (x) *Triadic patent families* (per million population), (xi) *Firm entries* (birth rate), (xii) *GDP* (per capita), (xiii) *Early-stage venture capital* (% GDP), (xiv) *SMEs reporting non technological change* (%), (xv) *GDP per capita* (in PPS), (xvi) *Real GDP growth rate* (% change on previous year), (xvii) *Total employment growth* (% change on previous year), (xviii) *Long term unemployment rate* (% change on previous year),

[2] Of course, all of the 115 sub-indicators already passed an initial phase of judgment as they were included in the broad data set. They were categorized into seven main areas (i.e., (I) production and diffusion of ICT, (II)human resources, skills and creativity, (III) knowledge production and diffusion, (IV) innovation, entrepreneurship and creative production, (V) economic outputs, (VI) social performance, and (VII) Internationalisation) .

[3] 15 old EU countries: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden, U.K. 10 new EU countries: Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Slovakia, Slovenia + US and Japan.

[4] The reduction of the larger set to 23 indicators was undertaken by a team at the Econometrics and Applied Statistics Unit of the Joint Research Centre of the European Commission (Ispra). Although it is not the focus of the current paper, we point out that this reduction resulted from using forward and backward stepwise regression techniques. Conversely, the reliability of the resulting subset of 23 indicators as a representation of the seven KEI-dimensions was checked with canonical analysis (see Saisana et al., 2008).

(xix) *Hampered in daily activities because of chronic conditions* (% of population 15+), (xx) *Rooms per person by tenure status and type of housing* (average number of rooms), (xxi) *Technology balance of payments* (% GERD), (xxii) *Co authorship share on international S&E articles* (% of international articles), and (xxiii) *Foreign PhD students* (% of total PhD enrolment).

As is often the case with CI's, many of the CI constituent indicators are measured in different units. With higher values reflecting better performances, most indicators are considered as 'goods'. The exceptions are 'Long term unemployment rate' and 'Hampered in daily activities because of chronic conditions', both perceived to be "bads" as higher values represent a worse performance. To put all indicators on a common basis, we normalize the data using a re-scaling technique.[5] Depending on whether an indicator is considered as 'good' or 'bad', we apply respectively (1a) and (1b).

$$y_{j,i}^{n} = \frac{y_{j,i} - y_i^{\min}}{y_i^{\max} - y_i^{\min}} \qquad (1a) \quad \text{and} \quad y_{j,i}^{n} = \frac{y_i^{\max} - y_{j,i}}{y_i^{\max} - y_i^{\min}} \qquad (1b)$$

As a result, all normalized indicators $y_{j,i}^{n}$ have identical ranges (0,1), with a normalized value of 0 and 1 corresponding respectively to the poorest and strongest performance on indicator *i*.

We point out that the 23 indicators relate to seven thematic dimensions, as developed in relation to the Barcelona and Lisbon objectives;

I. 'Production and diffusion of ICT': indicators i to iii,
II. 'Human resources, skills and creativity': indicators iv to vii,
III. 'Knowledge production and diffusion': indicators viii to x ,
IV. 'Innovation, entrepreneurship and creative production': indicators xi to xiv,
V. 'Economic outputs': indicators xv to xvii,
VI. 'Social performance': indicators xviii to xx.
VII. 'Internationalisation': indicators xxi to xiii.

3. Handling missing data: Multiple Imputation

As in many practical applications, the selection of thematically ideal indicators ran into an operational problem, viz, missing data values. Evidently, this is troublesome for the development of robust, credible, and up-to-date CIs. There is an extensive literature on how to handle missing data. In essence, there are three generic approaches: Case Deletion, Single Imputation, and

[5] A similar method is, for example, used in the construction of the Human Development Index. For an overview of other normalization procedures, see Nardo et al. (2008) with the 'Handbook on Constructing Composite Indicators: Methodology and User Guide', OECD, Joint Research Centre.

Multiple Imputation.[6,7] While the first one simply omits the missing records from the dataset, the latter two ones perceive the missing data as part of the analysis and seek to provide imputation value(s). Due to its ability to better reflect the uncertainty inherent in the imputation, often, Multiple Imputation is preferred.

Essentially, Multiple Imputation requires three steps to come to the final result: imputation, analysis and pooling. The most challenging step, is the 'imputation' step, in which N imputations are generated for the missing values. To create such imputations, one identifies a random process that reflects the uncertainty in the complete data set Y (consisting of observed values Y_{obs} and missing values Y_{mis}). In particular, uncertainty relates to the conditional distribution of Y_{mis}, $f\left(Y_{mis}\middle|Y_{obs},\theta\right)$, with parameter vector θ depending on Y. If this distribution was known, it would be straightforward to reconstruct Y_{mis}. However, since θ depends on the incomplete Y, $f\left(Y_{mis}\middle|Y_{obs},\theta\right)$ is unkown. Thus, a first step is to obtain the posterior distribution of θ, $f\left(\theta\middle|Z,Y_{obs}\right)$ from the data.[8] Using this posterior distribution, an imputation algorithm is applied N times. In each of these N steps, one draws θ^* from $f\left(\theta\middle|Z,Y_{obs}\right)$ and, next, defines $f\left(Y_{mis}\middle|Z,Y_{obs},\theta^*\right)$ of which we draw Y_{mis}. This yields N 'completed' datasets that consist of observed data Y_{obs} and imputed data $Y_{mis,k}\ \left(k=1,...,N\right)$ for the original missing values. Evidently, by its very nature Multiple Imputation does not seek to provide any information whatsoever on which of these N imputed values should be considered as the 'true value'. As a matter of fact, there is no basis for the evaluator to judge the correctness of each of these estimations. Therefore, in the second and third step, all produced N datasets are further examined and pooled. The second 'analysis' step mainly involves computing the within-imputation variances (2) for each of the N datasets.

$$\bar{V}=\frac{1}{N}\sum_{k=1}^{N}V\left(Y_{mis,k}\right) \qquad (2)$$

The third 'pooling' step entails computing the average imputed value $(3a)$ and the between-imputations variance $(3b)$.

$$\overline{Y^j_{mis}}=\frac{1}{N}\sum_{k=1}^{N}Y^j_{mis,k} \quad (3a) \quad \text{and} \quad B=\frac{1}{N-1}\sum_{k=1}^{N}\left(Y^j_{mis,k}-\overline{Y^j_{mis}}\right)\left(Y^j_{mis,k}-\overline{Y^j_{mis}}\right)' \quad (3b)$$

---

[6] For a brief survey of these three methods, see Nardo et al. (2008, pp. 55-58). More comprehensive surveys can be found in Little and Rubin (2002), Little (1997) and Little and Schenker (1994).

[7] Recommended readings on Multiple Imputation are Rubin (1987) and Schafer (1999).

[8] In the posterior distribution, $f\left(\theta\middle|Z,Y_{obs}\right)$, Z denotes exogenous variables which may have an influence on θ.

Based on the outcomes from those two steps, confidence intervals can be obtained by taking the overall estimate plus or minus a multiple of standard error, where that number is a quantile of Student's t-distribution with degrees of freedom:

$$df = (N-1)\left(1+\frac{1}{r}\right)^2 \qquad (4)$$

where r is the between-to-within ratio:

$$r = \left(1+\frac{1}{N}\right)\frac{B}{\overline{\overline{V}}} \qquad (4a)$$

In the way outlined above, several completed datasets can be generated.[9] In turn, such datasets can be used to create a CI. The question at hand is, hence, how to deal with the fact that *multiple* datasets eventually translate into a CI.

One way to proceed is to use a set that comprises both observed data and *mean* imputed values (i.e. as found by (3)). When using this dataset, one implicitly assumes that average imputations are most likely to correspond with real values. These (pseudo-) precise data constitute one of the datasets we will use in the sequel.

Clearly, however, the Multiple Imputation technique allows for alternative, 'imprecise' datasets, notably by appending the precise (observed) data with (imputed) data *intervals*. Specifically, we construct such intervals by applying expression (4a) and its constituent components such that we get, for each missing value, a range of values that constitutes a 50 % confidence interval around its imputed mean.. Here, we assume that the imputation methodology provides us with sufficient information to mark out the range (i.e., confidence intervals) in which the true value $y_{j,i}$ is believed to lie.

As an example, Table 1 illustrates the steps to come to a complete, imprecise, and normalized dataset of 23 indicators. The upper part of this table displays the data as originally observed. As indicated by 'Na', several data values are missing. The middle part of Table 1 displays the mixture of observed data and imprecise interval data. Finally, the lower part of Table 1 lists normalized indicator data resulting from re-scaling the completed dataset as in (1a) and (1b).

To recall, all these steps merely provide the groundwork for the issue that is the focus of our concern. In section 5, we hence discuss proposals to incorporate such data mixtures in a so-called 'Benefit of the Doubt' composite indicator. Before doing so, we first shortly explain what is meant with the latter construct.

[9] For the application we study in this paper, this work was done by a team at the Universities of Tübingen and Trier (i.e., Ralf Münnich, Nicole Thees, and Luis Huergo) who participate in the KEI project.

Table 1: Original, imputed and re-scaled data for 23 knowledge-based economy indicators

Country	Ind.i	Ind.ii	Ind.iii	Ind.iv	…	Ind. xx	Ind. xxi	Ind. xxii	Ind. xxiii
Austria	0.0424	0.2100	0.1800	491.00	…	2.10	Na	0.0820	0.2125
Belgium	0.0444	0.0900	Na	507.00	…	2.10	0.7965	0.1080	0.3131
Germany	0.0395	0.4600	0.2600	491.00	…	1.90	0.3765	0.1050	Na
Denmark	0.0420	0.2800	0.4500	492.00	…	2.00	Na	0.0990	0.2036
…	…	…	…	…	…	…	…	…	…
Malta	0.0463	Na	Na	Na	…	2.29	Na	Na	0.1176
Poland	0.0393	0.0900	0.0400	497.00	…	1.86	Na	0.0660	Na
Slovenia	0.0413	0.1600	0.0900	Na	…	1.77	Na	0.0660	Na
Slovakia	0.0384	0.0300	0.1000	469.00	…	1.84	Na	0.0610	0.0121

Country	Ind.i	Ind.ii	Ind.iii	Ind.iv	…	Ind. xx	Ind. xxi	Ind. xxii	Ind. xxiii
Austria	0.0424	0.2100	0.1800	491.00	…	2.10	[0.5858,0.9245]	0.0820	0.2125
Belgium	0.0444	0.0900	[0.0284,0.1700]	507.00	…	2.10	0.7965	0.1080	0.3131
Germany	0.0395	0.4600	0.2600	491.00	…	1.90	0.3765	0.1050	[0.1308,0.2351]
Denmark	0.0420	0.2800	0.4500	492.00	…	2.00	[0.2205,0.6172]	0.0990	0.2036
…	…	…	…	…	…	…	…	…	…
Malta	0.0463	[0.2306,0.6582]	[0.0296,0.1959]	[497.69,511.71]	…	2.29	[0.3044,0.6749]	[0.0969,0.1644]	0.1176
Poland	0.0393	0.0900	0.0400	497.00	…	1.86	[0.7527,0.9681]	0.0660	[0.0049,0.0103]
Slovenia	0.0413	0.1600	0.0900	[493.66,502.23]	…	1.77	[0.3524,0.6531]	0.0660	[0.0142,0.0294]
Slovakia	0.0384	0.0300	0.1000	469.00	…	1.84	[0.1920,0.6282]	0.0610	0.0121

Country	Ind.i	Ind.ii	Ind.iii	Ind.iv	…	Ind. xx	Ind. xxi	Ind. xxii	Ind. xxiii
Austria	0.1007	0.3600	0.3469	0.2973	…	0.5833	[0.3414,0.5518]	0.1342	0.1533
Belgium	0.1350	0.1200	[0.1318,0.2322]	0.5135	…	0.5833	0.4723	0.2084	0.2264
Germany	0.0534	0.8600	0.5102	0.2973	…	0.4167	0.2115	0.1998	[0.0939,0.1697]
Denmark	0.0953	0.5000	0.8980	0.3108	…	0.5000	[0.1950,0.2806]	0.1827	0.1468
…	…	…	…	…	…	…	…	…	…
Malta	0.1674	[0.6803,0.9774]	[0.1508,0.2686]	[0.4495,0.5153]	…	0.7492	[0.2781,0.3968]	[0.2396,0.3065]	0.0844
Poland	0.0491	0.1200	0.0612	0.3784	…	0.3808	[0.4887,0.5352]	0.0885	[0.0024,0.0064]
Slovenia	0.0822	0.2600	0.1633	[0.3710,0.4113]	…	0.3114	[0.1965,0.3833]	0.0885	[0.0092,0.0202]
Slovakia	0.0343	0.0001	0.1837	0.0001	…	0.3645	[0.1853,0.2794]	0.0742	0.0077

4. ‘Benefit of the Doubt’ composite indicators

Summarizing several performance indicators into one global figure entails making judgments about the relative ‘worth’ of each of the indicators. The core idea in this respect is that greater weight should be given to components which are considered to be more important in the context of the particular composite indicator (Freudenberg, 2003, p.10). However, as verifiable information regarding countries’ true policy weights for the multiple indicators is usually lacking, it is not at all clear what judgments to impute. In addition, this ambiguity is usually not overcome by resorting to expert opinion: strong inter-individual disagreement is often present in the weighting information stemming from stakeholders.

Melyn and Moesen (1991) were the first to recognize the conceptual similarity between this problem and the one conventionally discussed in the Data Envelopment Analysis literature. The latter addresses the problem of how to measure firms’ relative performance, given observations on (possibly multiple) input and output quantities and, often, no reliable information on prices and no exact knowledge on the ‘functional form’ of a production or cost function. In the context of macroeconomic performance evaluation, they alternatively labeled DEA as the ‘Benefit of the Doubt” approach.

The label ‘benefit of the doubt’ highlights one of DEA’s main conceptual starting points: information on the appropriate weighting scheme for country performance benchmarking can actually be retrieved from the country data themselves. The intuition behind this idea is that a good relative performance of a country in one particular indicator dimension indicates that the country concerned considers this policy dimension as relatively important. Or, conversely, that a country attaches less importance to those dimensions on which it is demonstrably a weak performer relative to the other countries in the set. Thus, with relative strengths being interpreted as a ‘revealed preference’/higher importance and the influence of relative weaknesses conversely being curtailed, each country is in fact entitled to its own ‘optimal’ weighting scheme. ‘Benefit of the doubt’-weighting thus implies that, absent any information about the correct weights, each country is put in the best possible light relative to the other countries in the sample when its aggregate performance is gauged. Formally, the endogenous specification of the country-specific optimal weights is highlighted by the max operator in the following linear programming problem.

$$CI_0 = \max_{w_{c,i}} \sum_{i=1}^{m} w_{c,i} y_{c,i}^{n} \qquad (5)$$

$$s.t.$$

$$\sum_{i=1}^{m} w_{c,i} y_{j,i}^{n} \leq 1 \qquad \forall \; j = 1,\ldots,n \qquad (5a)$$

$$w_{c,i} \geq 0 \qquad (5b)$$

$$w_{c,i} \in W \qquad (5c) \quad \textit{optional}$$

This weighting problem is handled for each country separately, hence n times. The $y_{j,i}^n = \left(y_{j,1}^n, y_{j,2}^n, ..., y_{j,m}^n \right)^T$ represents the column vector of m $(i = 1,2,...,m)$ re-scaled indicators $y_{c,i}^n$ for country j. In our application, there are 23 indicators which compromise the knowledge-based economy CI, so m=23. Similarly, $y_{c,i}^n$ are the re-scaled indicator values for the country c under evaluation. The $w_{c,i}$ are the weights assigned to the indicators by the evaluated country.

As the objective indicates, these weights are chosen such that the CI-value for each country is maximal. Observe that the full model has some additional and important features. A first one is the presence of a normalization constraint $(5a)$, that ensures that a country's composite indicator value can at most be 100%. A constraint such as $(5a)$ is necessary insofar as one genuinely aims for assessing the *relative* performance of companies, i.e. when the concern is with best *practices* rather than with a country's positioning relative to some exogenously imposed (normative) aggregate target. Indeed, the relative, best practice nature of gauging performance is, secondly, underscored by the fact that $(5a)$ is in fact applied *n* times: the optimal weights of country *c* are thus also used for assessing the data of each other country. In this sense, benefit-of-the-doubt weighting may well entail that even with country-specific favorable weights, there is another country that gets a higher aggregate score in one of the *n* comparison constraints $(5a)$. Put differently, even when one grants the benefit of the doubt to a country, it is still possible that its final score is only a fraction of that obtained by another country, since its most favorable weighting scheme is even more favorable to that other country's underlying data. Thirdly, the 'non-negativity' constraint $(5b)$ limits weights to be non-negative. This restricts the CI to be a non-decreasing function of its composing indicators.

It is easy to verify that this Benefit of the doubt-approach (BoD) is formally equivalent to the original input-oriented CCR-DEA model (after Charnes, Cooper and Rhodes, 1978), with all indicators considered as outputs and a 'dummy input' equal to one for all countries.[10] Intuitively, the model can be simply regarded as a tool for aggregating several indicators into an overall CI, without explicit reference to the inputs that are used.[11]

The appeal of using BoD to construct CIs is mainly due to the method's flexibility in weight choice. Any other weighting scheme than the one specified by BoD would worsen the position of the evaluated country vis-à-vis the other countries. Hence, countries cannot claim that a poor relative performance is due to a harmful or unfair weighting scheme. However, there are also disadvantages to this nearly unlimited flexibility in weight choice. In some situations, it can

[10] See Cooper et al. (2004), Chapter 1, pp.8-19 for an extensive treatment and illustration of traditional CCR-DEA model.
[11] The problem is then one in a pure output setting in which all countries are regarded to be equally able at achieving performances. The dummy input may be interpreted as a 'helmsman' that pursues several policy objectives, corresponding to the several policy objectives, see e.g. Lovell et al. (1995).

allow a country to appear as a brilliant performer in a way that is difficult to justify. One inconvenience is that, without violating constraints $(5a)$ and $(5b)$, countries may ignore some indicators on which they perform relatively poor by assigning (quasi) zero-weights. One then faces the risk of obtaining CIs reflecting only the performance on a small subset of all (often meticulously selected) indicators. Another concern is that weights may too much deviate from what experts believe is realistic, even if these experts fail to agree about a unique common weighting scheme. In both cases, 'full flexibility' weights contradict prior views (e.g. expert opinion) on weights, which cause CIs to be unreliable evaluations of policy performance. To avoid inappropriate weights, additional restrictions can be added to curb the freedom in weight choice (see the optional constraint $(5c)$). In practice, the involvement of experts in the development of CIs clearly benefits the credibility and acceptance of the results. On the other hand, if their opinions are available but ignored, eventual CIs risk being rejected.

5. Data impreciseness & Benefit of the Doubt

Usually, BoD-CIs build on precise data only.[12] Recalling that the timeliness of our knowledge economy CI requires using imputed values for missing data, this would mean using a (pseudo-)specific dataset that consists solely of observed and (average) imputed data values. However, this is a strong assumption. For that reason, we here explore the avenue of replacing missing values by unknown, yet, bounded decision variables: $\underline{y}^n_{j,i} \leq y^n_{j,i} \leq \overline{y}^n_{j,i}$. In our present case study, this boils down to using the interval datasets where $\overline{y}^n_{j,i}$ and $\underline{y}^n_{j,i}$ respectively correspond to the upper and lower bound re-scaled values of the confidence intervals. The standard BoD-approach is not readily equipped to handle such imprecise data, and we accordingly need to alter problem (5)-(5c)

Unsurprisingly, we can draw some inspiration from recent developments in the DEA literature. Specifically, our 'imprecise BoD'-model builds on the contributions of Entani et al. (2002), Despotis et al. (2002), and Park, (2006). Basically, these authors proposed to simply convert the imprecise data into exact counterparts by specifying several evaluation *scenarios* and just apply regular BoD to each of these scenarios.

Applied to the setting at hand, we will specifically focus on two extreme case scenarios in which countries can be evaluated: a '*strong country in a weak environment*' scenario and a '*weak country in a strong environment*' scenario. In the '*strong country in a weak environment*' scenario (6a), the evaluated country receives his highest possible normalized indicator values

[12] Numerous CIs of this type are extensively reported in the literature (e.g. European Unemployment policy (Storrie et al., 2000), Internal Market Index (Cherchye et al., 2007a), Sustainable Development (Cherchye et al., 2004a), Social Inclusion Index (Cherchye et al., 2004b), and Human Development Index (Despotis et al., 2005), Technology Achievement Index (Cherchye et al., 2008), etc.)

$\overline{y}^{n}_{c,i}$, while we select for the other countries their lowest possible ones $\underline{y}^{n}_{j,i}$. Thus, the evaluated country is positioned in his strongest position, while the opposite holds for the other countries. Obviously, this setting yields the evaluated country's maximum CI-value (i.e., CI_c^{upper}). The converse happens in the '*weak country in a strong environment*' case (6b), where the evaluated country is assigned its lowest possible re-scaled indicator values $\underline{y}^{n}_{c,i}$ and the other countries get highest possible ones $\overline{y}^{n}_{j,i}$. As a result, the evaluated country receives the lower bound CI-value (i.e., CI_c^{lower}).

$$
\begin{array}{ll}
CI_c^{upper} = \max\limits_{w_{c,i}} \sum_{i=1}^{m} w_{c,i}\, \overline{y}^{n}_{c,i} \quad (6a) & \qquad CI_c^{lower} = \max\limits_{w_{c,i}} \sum_{i=1}^{m} w_{c,i}\, \underline{y}^{n}_{c,i} \quad (6b) \\
s.t. & \qquad s.t. \\
\sum_{i=1}^{m} w_{c,i}\, \overline{y}^{n}_{c,i} \leq 1 \quad \forall\ j = 1,\ldots,n & \qquad \sum_{i=1}^{m} w_{c,i}\, \underline{y}^{n}_{c,i} \leq 1 \quad \forall\ j = 1,\ldots,n \\
\sum_{i=1}^{m} w_{c,i}\, \underline{y}^{n}_{j,i} \leq 1 \quad \forall j \neq c & \qquad \sum_{i=1}^{m} w_{c,i}\, \overline{y}^{n}_{j,i} \leq 1 \quad \forall j \neq c \\
w_{c,i} \geq 0 \quad \forall i & \qquad w_{c,i} \geq 0 \quad \forall i \\
w_{c,i} \in W \quad \forall i & \qquad w_{c,i} \in W \quad \forall i
\end{array}
$$

The general difference between the original formulation (5) and both models above is that the normalization/comparison constraints are split into two sets: one (singleton) set pertains to the evaluated country, the other set to the other countries. The CI_c^{upper} and CI_c^{lower} mark out the range of CI-values in which the exact one is believed to lie, given the pre-specified confidence level and the estimates as obtained by multiple imputation of the missing values. By construction, all possible CI values resulting from alternative scenarios are within this 90%-confidence interval range. The length of the obtained interval reflects the uncertainty about the real CI-value. This ambiguity may vary from country to country. Clearly, $CI_c^{lower} \leq CI_c^{upper}$ where equality holds in the case of only exact data or when a country is evaluated as benchmark in both extreme scenarios.

It follows from the last observation that, even when there is some ambiguity in a CI's constituent components, there may be instances in which a fairly convincing case can be made about the (benchmark) nature of a country's performance. In fact, this line of thinking can be extended. Park (2006) proposed a three group-classification system in which he labelled evaluated observations as either perfectly efficient, potentially efficient, or inefficient, depending on the efficiency values such observation receive in the two 'extreme' scenario's as captured by programs (6a) and (6b). A similar classification system can be applied in the current setting of evaluating countries. Below, we will refer to a “benchmark country” whenever a country is designated as such in *both* extreme scenarios. Conversely, there is a group of “countries open to improvement” that can be unambiguously labelled as such, since in neither of the two scenarios they emerge as benchmark performers. Ambiguity is only associated with “potential benchmark countries”, for which the actual classification as a strong performer or not is contingent on the actual imputation of specific data values. Thus, while uncertainty about some data values still allows for a crisp classification of some countries, the last label explicitly recognizes that a CI cannot always deliver crisp answers in a context of data uncertainty. In terms of our specific set-up, it is easy to see that this category contains countries that fare well in a ‘strong country in weak environment’ scenario but not in the opposite extreme ‘weak country in strong environment’ scenario.

When a timely follow-up of global evolutions is needed, and when it is decided that CIs can be instrumental for such an aggregate follow-up, this interval and classification approach seems valuable. First, it highlights rather than conceals the underlying imperfect nature of the exercise. Second, the results so generated may be more swiftly acceptable, even for poorer performing countries. To illustrate this, consider Table 2 which displays precise and interval CI estimates for countries A, B and C. Their CI-point estimates of respectively 0.92, 0.76 and 1, result from conventional BoD as in (5), and hence build on the (pseudo-specific) assumption that mean imputed values are most likely to substitute for originally missing data. Based on these results, country C is perceived to be better performing and, therefore, a benchmark for country A and B. With a CI-value of 0.76, Country B is believed to be the worst performer.

Table 2: Illustrative example

	CI	CI-interval	Classification
Country A	0.92	[0.83,1.00]	Potential benchmark country
Country B	0.76	[0.65,0.87]	Country open to improvement
Country C	1.00	[1.00,1.00]	Benchmark country

When the alternative approach is followed, this crisp verdict is somewhat nuanced. Country C is ranked on top in both extreme evaluation scenarios, so countries A and B will not deny that this is a benchmark performer. Country A evidently can claim that its relative performance assessment is at least partially driven by data problems. Country B can make a similar claim, but only to a limited extent.

6. An imprecise knowledge-based economy CI

In this section, we discuss and interpret the Knowledge-based Economy CI for the year 2004. More specifically, we compute CI-estimates applying an unrestricted and a restricted version of BoD to a precise and an imprecise dataset where mean imputed values and 50% confidence interval estimates is substituted for originally missing data. Based on the precise dataset, we compute data point estimates for the CIs by conventional BoD as in (5). Results are displayed in the left part of Table 3 under the heading 'Average Dataset". The estimation of CI intervals based on imprecise data requires one to use the adjusted version of BoD as formulated in (6a) and (6b). Resulting imprecise CIs are shown in the middle part and the right part of Table 3. The official KEI-rank as computed by the European Commission is included between brackets.

In case of no extra weight restrictions, we omit the optional constraint $w_{c,i} \in W$. In this situation, countries are granted the most leeway in defining their optimal weights. By consequence, CI-values are highest. In Table 3, this can be quite readily denoted by comparing full flexibility results with restricted results. With most countries receiving scores of 100%, clearly, unrestricted BoD doesn't discriminate adequately between country performances. For the imprecise CIs, for instance, respectively 18 and 7 countries are evaluated as benchmark or potential benchmark countries. While being a 'good news' show for countries themselves, there is no doubt that the added value of such evaluations from point of view of improving policies is rather limited. The problem comes from the fact that too much flexibility can allow countries to appear as brilliant performers in a way that is difficult to justify. As pointed above, one particular concern is that countries can assign zero weights to indicators on which they perform relatively poor. A deeper analysis of the optimal weights reveals that this is exactly what happens in the unrestricted BoD evaluation. More specifically, CIs resulting from full flexibility BoD comprise, on average, only 5.5 indicators. Stated differently, of the 23 meticulously selected KEI-indicators, countries ignore, on average, 17.5 indicators. An illustrative example is Greece that is somewhat surprisingly evaluated to be a benchmark country. An analysis of the assigned weights reveals that this result is, to say the least, unrepresentative for the 'global' performance of Greece in knowledge-based economy. More in particular, Greece respectively uses three and ten indicators to construct its CI in the 'weak country in strong environment' and 'strong country in weak environment' scenarios (see respectively Appendix A and B). All other performance indicators are ignored. Of all countries in the sample, Finland uses the most indicators with 10 non-zero weights for both scenarios. This in mind, it seems reasonable to say that all full flexibility results are only partially reflecting knowledge-based economy performances and, hence, unreliable for global policy performance evaluation.

Table 3: CIs and evaluations for unrestricted and restricted model

	Average Dataset		**90% Conf. Int. Dataset (Full Flexibility)**			**90% Conf. Int. Dataset (Restricted)**		
Country	**Full flexibility**	**Restricted**	**Scenario 1**	**Scenario 2**	**Evaluation**	**Scenario 1**	**Scenario 2**	**Evaluation**
Austria (10)	100.00%	89.93%	100.00%	100.00%	*Benchmark Country*	82.51%	100.00%	*Potential Benchmark Country*
Belgium (11)	100.00%	100.00%	100.00%	100.00%	*Benchmark Country*	90.69%	100.00%	*Potential Benchmark Country*
Germany (15)	100.00%	100.00%	100.00%	100.00%	*Benchmark Country*	90.57%	100.00%	*Potential Benchmark Country*
Denmark (2)	84.93%	100.00%	100.00%	100.00%	*Benchmark Country*	100.00%	100.00%	*Benchmark Country*
Spain (20)	100.00%	92.48%	100.00%	100.00%	*Benchmark Country*	87.92%	100.00%	*Potential Benchmark Country*
Finland (4)	100.00%	100.00%	100.00%	100.00%	*Benchmark Country*	100.00%	100.00%	*Benchmark Country*
France (12)	100.00%	87.44%	80.92%	100.00%	*Potential Benchmark Country*	70.76%	100.00%	*Potential Benchmark Country*
Greece (24)	100.00%	81.15%	100.00%	100.00%	*Benchmark Country*	76.56%	100.00%	*Potential Benchmark Country*
Ireland (9)	100.00%	100.00%	100.00%	100.00%	*Benchmark Country*	99.73%	100.00%	*Potential Benchmark Country*
Italy (23)	92.96%	66.95%	99.73%	100.00%	*Potential Benchmark Country*	67.58%	76.68%	*Country Open to Improvement*
Luxembourg (3)	100.00%	100.00%	100.00%	100.00%	*Benchmark Country*	99.30%	100.00%	*Potential Benchmark Country*
Netherlands (8)	95.91%	100.00%	100.00%	100.00%	*Benchmark Country*	99.06%	100.00%	*Potential Benchmark Country*
Portugal (27)	100.00%	54.43%	81.85%	83.02%	*Country Open to Improvement*	52.88%	57.05%	*Country Open to Improvement*
Sweden (1)	100.00%	100.00%	100.00%	100.00%	*Benchmark Country*	100.00%	100.00%	*Benchmark Country*
United Kingdom (7)	100.00%	100.00%	100.00%	100.00%	*Benchmark Country*	99.57%	100.00%	*Potential Benchmark Country*

Table 3: CIs and evaluations for unrestricted and restricted model (continued)

	Average Dataset		90% Conf. Int. Dataset (Full Flexibility)			90% Conf. Int. Dataset (Restricted)		
Country	**Full flexibility**	**Restricted**	**Scenario 1**	**Scenario 2**	**Evaluation**	**Scenario 1**	**Scenario 2**	**Evaluation**
Cyprus (19)	99.89%	100.00%	100.00%	100.00%	*Benchmark Country*	93.94%	100.00%	*Potential Benchmark Country*
Czech Republic (21)	100.00%	73.76%	81.62%	88.57%	*Country Open to Improvement*	71.76%	79.76%	*Country Open to Improvement*
Hungary (26)	100.00%	89.93%	100.00%	100.00%	*Benchmark Country*	86.96%	95.29%	*Country Open to Improvement*
Estonia (17)	100.00%	100.00%	100.00%	100.00%	*Benchmark Country*	98.03%	100.00%	*Potential Benchmark Country*
Lithuania (25)	100.00%	93.39%	91.95%	100.00%	*Potential Benchmark Country*	77.61%	100.00%	*Potential Benchmark Country*
Latvia (22)	100.00%	97.31%	100.00%	100.00%	*Benchmark Country*	91.96%	100.00%	*Potential Benchmark Country*
Malta (18)	100.00%	96.44%	92.39%	100.00%	*Potential Benchmark Country*	68.62%	100.00%	*Potential Benchmark Country*
Poland (29)	97.34%	79.88%	83.57%	100.00%	*Potential Benchmark Country*	73.09%	88.19%	*Country Open to Improvement*
Slovenia (16)	82.19%	74.31%	92.47%	100.00%	*Potential Benchmark Country*	68.92%	92.76%	*Country Open to Improvement*
Slovakia (28)	100.00%	67.13%	87.25%	98.43%	*Country Open to Improvement*	65.48%	76.39%	*Country Open to Improvement*
EU15 (13)	92.47%	91.19%	85.32%	100.00%	*Potential Benchmark Country*	74.58%	100.00%	*Potential Benchmark Country*
EU25 (14)	90.28%	85.43%	79.36%	100.00%	*Potential Benchmark Country*	70.02%	100.00%	*Potential Benchmark Country*
USA (5)	100.00%	100.00%	100.00%	100.00%	*Benchmark Country*	91.15%	100.00%	*Potential Benchmark Country*
Japan (6)	100.00%	100.00%	100.00%	100.00%	*Benchmark Country*	94.93%	100.00%	*Potential Benchmark Country*

To ensure that a proper weighting scheme is established, we add proportional weight restrictions to limit the freedom in weight choice. We previously denoted that restrictions are preferably defined based on expert opinions. In our current case study, suppose that experts agree on more or less equal importance of dimensions. In line with this consensus, we impose importance of each of the seven dimensions to vary around the equal importance of 1/7. More in particular, we allow for a certain amount of variation around these 'point values', viz. minus 25 % (lower bound) and plus 25 % (upper bound) of this 1/7. Moreover, experts agree on the relevance of each of the 23 performance indicators in the construction of CIs. This in mind, to prevent too little emphasis on particular indicators, we force importance of individual indicators within the dimensions concerned to be at least 5%. Formally, optional restrictions $w_{c,i} \in W$ in (3a) and (3b) respectively entail the following:

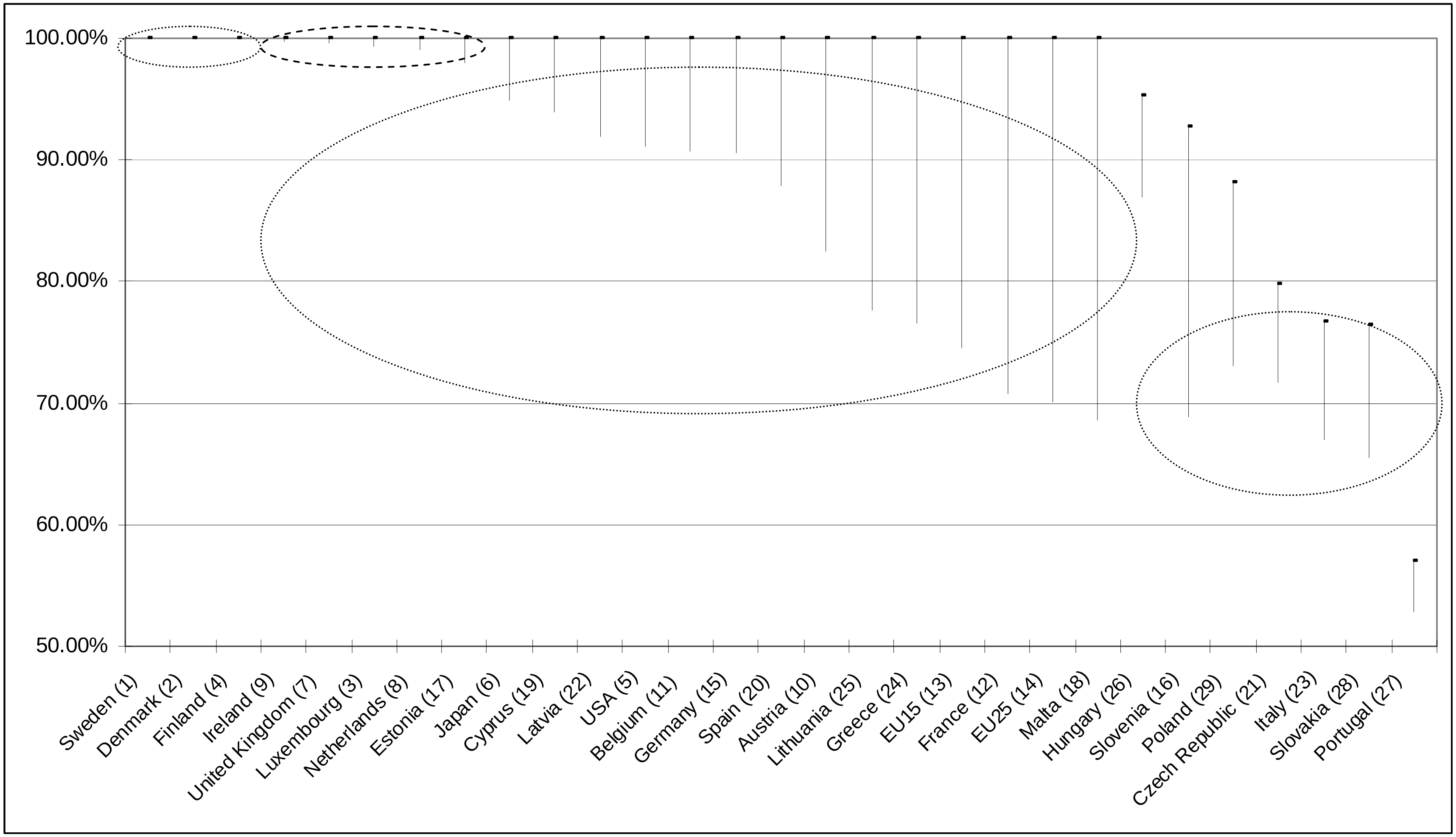
$$\left(\frac{1}{7}\right) \times 0.25 \leq \frac{\sum_{i \in Dim\,k} w_{c,i}}{\sum_{i=1}^{m} w_{c,i}} \leq \left(\frac{1}{7}\right) \times 1.25 \qquad (7a)$$

$$\frac{w_{c,i}}{\sum_{i \in Dim\,k} w_{c,i}} \geq 0.05 \qquad (7b)$$

Countries can then still choose their optimal weights freely but now with one proviso: they have to respect the imposed weight restrictions. Results are displayed in the third column and the complete right part from Table 3. For the remainder of this section, the focus is completely on these restricted BoD results.

Let us start with denoting some interesting features. First of all, the precise estimates of the CIs are always situated in the imprecise CI-intervals, $\left[CI_c^{upper}, CI_c^{lower}\right]$. Second, CI-values in the 'strong country in weak environment' set-up (scenario 2) are always higher or equal than the ones in the 'weak country in strong environment' setting (scenario 1). Only in the situation where both resulting values coincide, does the interval reduce to one exact value. In our application, this only appears to happen when countries are evaluated to be benchmark countries with CI-values of 100.00% in both scenarios. Third, reducing the flexibility in weight choice, (extra) weight restrictions always lower resulting CIs. Reasoning by analogy, countries can never move up to a higher group after adding weight restrictions. With all countries assigned to equally good or better classification group in unrestricted BoD, this feature is quite easy to denote in Table 3. Straightforwardly, countries evaluated to be benchmark countries in restricted BoD (i.e., Denmark, Finland, and Sweden) are also benchmarks in the case where no restrictions have to be satisfied. Finally, CIs resulting from both boundary scenarios mark out the range of CI-values in which the exact one is believed to lie. As discussed before, the length of this interval reflects the ambiguity about the real CI-value. As is shown in Figure 1, interval lengths vary from country to country reflecting the uncertainty on the underlying data. Portugal, for example, has lower and

Figure 1: Imprecise country evaluations: mean, 'strong country in weak environment' and 'weak country in strong environment' CIs



upper bound values (respectively 53.85 % and 55.14 %) which almost coincide. By result, such an interval gives a very good indication of what the exact CI-value might be. A quite opposite result is obtained for France for whom, with bounds of 80.03% and 96.51%, much obscurity remains about its real CI-value.

Another interesting finding is that our resulting ranking is very similar to the one computed by the JRC team using a multi-modeling approach (see Saisana et al., 2008).[13] Unsurprisingly, the Scandinavian countries top the list of countries with Sweden, Denmark and Finland being evaluated as benchmark countries (see Figure 1). Twelve countries emerge in the group of potential benchmark countries (middle group in Figure 1): 6 old EU-15 countries (i.e., Belgium, Germany, Ireland, Luxembourg, the Netherlands, and the United Kingdom), 3 new EU Member States (i.e., Cyprus, Estonia, and Malta), and USA and Japan. Note that some of these countries (i.e., Ireland, United Kingdom, Luxembourg and the Netherlands) achieve CIs near to unity under the 'weak country in strong environment' scenario. This implies that their performance is near to be perfect as indicated in Figure 1 by the dashed subgroup. In correspondence with the official KEI ranking, Southern European countries (i.e., Spain, France, Greece, Italy, and Portugal) are considered to be 'Countries Open to Improvement' with respect to a knowledge based economy. Particularly striking is the very bad performance of Italy and Portugal which appear in the group of 'Countries Open to Improvement' together with new EU Member States (cfr. Figure 1). Another interesting result is also that overall both USA and Japan have a better performance than EU 15 that on its turn is better performing than the EU 25. This result should not be a surprise as both the USA and Japan are known to be moving more rapidly towards a knowledge-based economy than the EU. One explanation could be the higher investment in knowledge to GDP ratio as illustrated in Table 4.

Table 4: Average of annual investment in knowledge as a percentage of GDP (1997-2003)

Country	Annual investment	Country	Annual investment
Sweden	6.28 %	EU 15	3.48 %
United States	6.19 %	Belgium	3.40 %
Finland	5.61 %	Austria	3.27 %
Denmark	4.74 %	Ireland	2.49 %
Japan	4.63 %	Spain	2.47 %
France	4.09 %	Italy	2.24 %
Germany	3.73 %	Greece	1.80 %
Netherlands	3.69 %	Portugal	1.69 %
United Kingdom	3.50 %		

Source: OECD Factbook 2008

[13] This multi-modeling approach consists of about 2,000 simulations based on combinations of the imputation method, number of dimensions, number of performance indicators, applied normalization procedure, weighting and aggregation method, etc.

Clearly, the USA and Japan outperform most EU countries when it comes to investing in knowledge. Notable exceptions are the Nordic countries. In fact, the difference between top investors Sweden, USA and Finland on the one hand, and Germany, the largest European economy, on the other hand, is substantial. What is even worse is that, with the exception of the Nordic countries, most EU countries failed to increase knowledge investment between 1997 and 2003. Southern European countries appear at the bottom of the table and considerable room for improvement in the knowledge based economy-related policies. Pushing through reforms aimed at raising investment rates as quickly as possible, the situation for most new EU Member States is one of trying to catch up with the old EU Member States.

7. Concluding remarks

Despite their increasing prevalence, CIs remain the subject of controversy. Particularly the traditional inability of their construction tools to deal with imprecise data is problematic in the evaluation of countries' policy performance where, on regular basis, indicator values are vague or even missing. Given that data are frequently gathered through the countries' own statistical offices, each following its own collection procedures, it seems even inconceivable to think that this impreciseness will ever be fully eliminated. Therefore, to improve the usefulness and credibility of CIs, construction tools should be adjusted so as to be able to handle missing or imprecise data.

Throughout this paper, we discussed an extension of the 'Benefit of the Doubt' methodology (after Melyn and Moesen, 1991) to deal with mixtures of exact and imprecise data. The extension to allow for imprecise data originates in the DEA-literature (i.e., Entani et al., 2002 and Despotis et al., 2002) and, essentially, boils down to focus on the boundary values of the imprecise (interval) data. More specifically, inserting these boundary values in certain combinations allows one to distinct between two scenarios: a 'strong country in weak environment' scenario and a 'weak country in strong environment' scenario. Based on this distinction, BoD calculates the upper and lower bound values of the CI-interval in which the exact CI-value is supposed to lie. By result, our tool is in accordance with the intuition that CIs based on imprecise data should be imprecise as well. The provision of CI-intervals not only keeps in mind the impreciseness of the data, it also renders the message of the evaluation somewhat less harsh. In accordance with the ideas discussed in the DEA-literature, we extend BoD with a three-group classification system in which countries are subdivided based on resulting CI-intervals (i.e., benchmark countries, potential benchmark countries, and countries open to improvement). In our opinion, such a classification renders evaluation results more easy to communicate and, more importantly, more swiftly acceptable to countries and policy performers.

The aforementioned extensions to handle imprecise data also can be used when BoD is used to evaluate country policy performances in a dynamic setting, i.e. one assesses the performance of a group of countries over time. The decomposition of the overall change into a catching-up and environmental change component allows one to detect whether this change is attributable to country-specific improvement or just because of an overall improvement in the policy environment. For both the ‘strong country in weak environment’ and ‘weak country in strong environment’ scenario, decompositions can be performed and interpreted. Other topics in the literature which offer plenty of scope for further research for handling imprecise data in BoD, are the uncertainty and sensitivity analysis of CIs (e.g. Saisana et al., 2005 and Cherchey et al. 2008) and the statistical inference based on CIs (see e.g. Simar et al, 2000 and 2006).

Acknowledgments

This paper is an offshoot of the KEI project (contract no.502529) that is part of priority 8 of the policy-oriented research under the European Commission’s Sixth Framework Programme. Laurens Cherchye gratefully acknowledges financial support from the Research Fund K.U.Leuven through the grant STRT1/08/004.

References

Booysen, F., 2002. An Overview and Evaluation of Composite Indices of Development. Social Indicators Research 59, 115-151.

Charnes, A., Cooper, W.W., Rhodes, E., 1978. Measuring the Efficiency of Decision Making Units. European Journal of Operational Research 2, 429-444.

Cherchye, L., Kuosmanen, T., 2004a. Benchmarking Sustainable Development: A Synthetic Meta-Index Approach. Chapter 7 in M. McGillivray and M. Clarke (eds.), Perspectives on Human Development, United Nations University Press.

Cherchye, L., Moesen, W., Van Puyenbroeck, T., 2004b. Legitimately Diverse, yet Comparable: On Synthesizing Social Inclusion Performance in the EU. Journal of Common Market Studies 42, 919-955.

Cherchye, L., Lovell, C.A.K., Moesen, W. and Van Puyenbroeck T., 2007a. “One Market, One number? A Composite Indicator Assessment of EU Internal Market Dynamics. European Economic Review 51, 749-779.

Cherchye, L., Moesen, W., Rogge N., Van Puyenbroeck, T., 2007b. An introduction to ‘benefit of the doubt’ composite indicators. Social Indicators Research 82, 111-145.

Cherchye, L., Moesen, W., Rogge N., Van Puyenbroeck, T., Saisana, M., Saltelli, A., Liska, R., Tarantola, S., 2008. Creating Composite Indicators with DEA and Robustness Analysis: the case of the Technology Achievement Index. Journal of the Operational Research Society 59, 239-251.

Cooper, W.W., Seiford, L.M., Zhu, J., 2004. Data Envelopment Analysis: History, Models and Interpretations. in Handbook on Data Envelopment Analysis, W.W. Cooper, L.M. Seiford, J. Zhu (eds), Chapter 1, 1-39, Kluwer Academic Publishers, Boston. 2004.

Despotis, D.K., Smirlis, Y.G., 2002. Data Envelopment Analysis with Imprecise Data. European Journal of Operational Research 140, 24-36.

Despotis, DK., 2005. A reassessment of the human development index via Data Envelopment Analysis. Journal of the Operational Research Society 56(8), 969-980.

Entani, T., Maeda, Y., Tanaka, H., 2002. Dual models of interval DEA and its extensions to interval data. European Journal of Operational Research 136, 32-45.

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

Little, R.J.A., Rubin, D.B., 2002. Statistical Analysis with Missing Data. Wiley Interscience, J. Wiley & Sons, Hoboken, New Jersey.

Little, R.J.A., 1997. Biostatistical Analysis with Missing Data. in Encyclopedia of Biostatistics, P. Armitage and T. Colton (eds.), London: Wiley.

Little, R.J.A., Schenker, N., 1994. Missing Data. in Handbook for Statistical Modeling in the Social and Behavioral Sciences, G. Arminger, C.C Clogg, and M.E. Sobel (eds.), pp.39-75, New York: Plenum.

Lovell, C.A.K, Pastor, J.T., Turner, J.A., 1995. Measuring Macroeconomic Performance in the OECD: A Comparison of European and Non-European Countries. European Journal of Operational Research 87 (1995), 507-518.

Melyn, W., Moesen, W., 1991. Towards a Synthetic Indicator of Macroeconomic Performance: Unequal Weighting when Limited Information is Available. Public Economics Research Paper 17, CES KULeuven.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S. (European Commission, JRC) and Hoffman, A., Giovannini, E., (OECD) 2008. Handbook on Constructing Composite Indicators: Methodology and User Guide.

OECD Factbook 2008: Economic, Environmental and Social Statistics.

Park, K.S., 2006. Efficiency bounds and efficiency classifications in DEA with imprecise data. Journal of the Operational Research Society, 1-8.

Rubin, D.B., 1987. Multiple Imputation for Nonresponse in Surveys, New York: John Wiley and Sons.

Saisana, M., Munda, G., 2008. Knowledge Economy: Measures and drivers. EUR Reports 23486 EN, European Commission, JRC-IPSC

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

Schafer, J., 1999. Multiple Imputation: A Primer. Statistical Methods in Medical Research 8, 3-15.

Simar, L., Wilson, P.W., 2000. Statistical Inference in Nonparametric Frontier Models: The State of the Art. The Journal of Productivity Analysis 13, 49-78.

Simar, L., Wilson, P.W., 2006. Statistical Inference in Nonparametric Frontier Models: recent Developments and Perspectives. Chapter in The Measurement of Productive Efficiency, 2nd Edition, Harold Fried, C.A. Knox Lovell and Shelton Schmidt (eds), Oxford University Press, 2006.

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

Appendix A: Absolute contributions of the 23 performance indicators: “weak country in strong environment” scenario as computed by full flexibility BoD

Country	CI_c^{lower}	Ind i	Ind ii	Ind iii	Ind iv	Ind v	Ind vi	Ind vii	Ind viii	Ind ix	Ind x	Ind xi	Ind xii	Ind xiii	Ind xiv	Ind xv	Ind xvi	Ind xvii	Ind xviii	Ind xix	Ind xx	Ind xxi	Ind xxii	Ind xxiii
Austria (10)	100.00%	0	0	0	0	0	0	0	0.1634	0	0	0	0	0	0	0.2093	0	0	0	0.6273	0	0	0	0
Belgium (11)	100.00%	0	0	0	0.0381	0.008	0	0.0283	0.1	0	0.023	0	0	0.0055	0	0	0	0	0	0.5863	0.1662	0.0056	0.039	0
Germany (15)	100.00%	0.0047	0.3683	0	0	0	0	0	0	0.0777	0.2901	0	0	0	0.2007	0	0	0	0	0	0	0.0584	0	0
Denmark (2)	100.00%	0	0	0.1832	0	0	0	0	0	0.0384	0.1041	0	0	0.1094	0.5585	0	0	0	0	0	0	0	0.0063	0
Spain (20)	100.00%	0	0	0.1188	0	0	0	0	0	0.003	0	0	0	0	0	0	0.0572	0.4285	0	0.373	0	0	0.0194	0
Finland (4)	100.00%	0.0857	0.02	0	0	0.4247	0	0	0.0808	0	0	0.037	0	0	0	0.0023	0.0367	0.0035	0	0.2381	0.0712	0	0	0
France (12)	80.92%	0	0.0359	0	0.0168	0.307	0	0	0.0118	0	0	0	0	0	0	0	0	0	0	0.1932	0.2107	0	0.0337	0
Greece (24)	100.01%	0	0	0	0	0	0	0.0469	0	0	0	0	0	0	0	0	0.356	0.5972	0	0	0	0	0	0
Ireland (9)	100.00%	0	0	0	0.0421	0	0.0084	0	0	0	0.0152	0	0	0	0	0	0.2216	0.0964	0.4496	0.0176	0.1491	0	0	0
Italy (23)	99.73%	0	0	0	0	0	0.0239	0.0275	0	0	0	0	0	0	0	0	0	0	0	0.8841	0	0	0.0619	0
Luxembourg (3)	100.00%	0	0	0.0301	0	0	0	0	0.1065	0	0	0	0.2716	0	0	0	0.264	0	0	0.1672	0.1605	0	0	0
Netherlands (8)	100.00%	0.008	0	0	0	0	0	0.0955	0.0916	0	0	0	0	0.003	0	0	0	0	0	0.3777	0.4086	0	0.0157	0
Portugal (27)	81.85%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.7843	0.0179	0	0.0163	0	0
Sweden (1)	100.00%	0	0	0	0.0312	0	0	0	0	0	0	0	0	0.068	0	0	0	0	0.8981	0.0027	0	0	0	0
United Kingdom (7)	100.00%	0.1909	0.3636	0	0	0	0	0	0	0	0	0.1163	0.1029	0	0	0	0.1903	0.036	0	0	0	0	0	0
Cyprus (19)	100.00%	0	0	0	0	0	0.031	0	0	0.0149	0	0	0	0	0	0	0	0.0939	0	0.7802	0	0	0.08	0
Czech Republic (21)	81.62%	0	0	0	0	0	0	0.1311	0	0	0	0.1284	0	0	0	0	0.1945	0	0.1544	0.1137	0.0692	0.0249	0	0
Hungary (26)	100.00%	0	0	0	0	0	0	0	0	0	0.0023	0.5555	0.0065	0	0	0	0.2329	0	0.2029	0	0	0	0	0
Estonia (17)	100.00%	0	0	0.1916	0	0	0	0	0	0.0084	0	0	0	0	0.0948	0	0.2801	0	0	0.0792	0.3461	0	0	0
Lithuania (25)	91.95%	0	0	0	0	0	0	0	0.0736	0	0	0	0	0	0	0	0.5016	0	0	0.1608	0.1835	0	0	0
Latvia (22)	100.00%	0	0	0	0.1291	0	0.029	0	0	0	0	0	0	0	0	0	0.6556	0	0	0.0988	0.0741	0	0.0134	0
Malta (18)	92.39%	0.011	0	0	0	0	0	0	0.1169	0	0	0	0	0	0	0	0	0	0	0.4161	0.3799	0	0	0
Poland (29)	83.57%	0	0	0	0.112	0	0	0	0	0	0	0.0797	0	0	0	0	0.2345	0	0	0.1299	0.0813	0.1982	0	0
Slovenia (16)	92.47%	0	0	0	0	0	0.1275	0	0	0	0	0	0	0	0	0	0.337	0	0	0.4602	0	0	0	0
Slovakia (28)	87.25%	0	0	0	0	0	0	0	0	0	0	0.1969	0	0	0	0	0.3563	0	0	0.3193	0	0	0	0
EU15 (13)	85.32%	0	0	0	0.1357	0	0.0531	0	0	0.0255	0	0	0	0	0	0	0	0	0	0.323	0.113	0	0.2029	0
EU25 (14)	79.36%	0	0.035	0	0	0.1205	0	0	0.0387	0	0	0.03	0	0	0	0	0.0366	0.0031	0	0.2779	0.2517	0	0	0
USA (5)	100.00%	0.0159	0	0	0.0447	0	0	0	0	0	0	0	0	0	0	0	0.1869	0.0018	0.7141	0	0.0065	0.0007	0.0294	0
Japan (6)	100.00%	0	0	0	0.0261	0	0	0	0.1493	0.0937	0.1176	0	0	0	0	0.0915	0	0	0	0.4581	0.0507	0	0.013	0

Appendix B: Absolute contributions of the 23 performance indicators: “strong country in weak environment” scenario as computed by full flexibility BoD

Country	CI_c^{upper}	Ind i	Ind ii	Ind iii	Ind iv	Ind v	Ind vi	Ind vii	Ind viii	Ind ix	Ind x	Ind xi	Ind xii	Ind xiii	Ind xiv	Ind xv	Ind xvi	Ind xvii	Ind xviii	Ind xix	Ind xx	Ind xxi	Ind xxii	Ind xxiii
Austria (10)	100.00%	0	0	0	0.2271	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.165	0	0.6079	0	0
Belgium (11)	100.00%	0	0	0	0.3549	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.2899	0	0.3552	0	0
Germany (15)	100.00%	0	0.0774	0	0	0	0	0.2027	0.0388	0.0396	0	0.3602	0	0	0	0	0.0462	0	0	0.235	0	0	0	0
Denmark (2)	100.00%	0	0.0382	0	0.0074	0	0	0.1939	0.0514	0.0326	0	0.1781	0	0	0	0	0.102	0	0	0.3964	0	0	0	0
Spain (20)	100.00%	0	0	0.009	0	0.0424	0	0	0.0148	0	0	0.1834	0	0	0	0	0.1858	0.007	0	0.4563	0	0.0164	0.085	0
Finland (4)	100.00%	0	0	0.1357	0.1114	0	0	0	0.1003	0	0	0.1337	0	0	0	0	0.1888	0.0015	0	0.2691	0.0193	0.0092	0.031	0
France (12)	100.00%	0	0	0.0641	0.0058	0	0	0	0.0725	0	0	0.3222	0	0	0	0	0.1264	0	0	0.2488	0.0839	0.0022	0.0741	0
Greece (24)	100.00%	0	0.0171	0	0	0.0352	0	0	0	0	0.0001	0.2449	0	0	0.0169	0	0.2841	0.0281	0	0.3527	0	0	0.0161	0.0048
Ireland (9)	100.00%	0	0	0	0	0.0009	0	0	0	0	0.041	0.0362	0	0.0222	0.2568	0	0.3297	0	0	0.2981	0	0	0.0152	0
Italy (23)	100.00%	0.0515	0	0.0069	0	0	0	0	0	0	0.0045	0.1581	0	0	0	0	0	0	0.0431	0.7029	0	0	0.0331	0
Luxembourg (3)	100.00%	0	0	0	0	0	0	0	0.1225	0.011	0	0.2246	0	0	0	0	0.1359	0.0489	0.0942	0.2396	0.0441	0	0.0792	0
Netherlands (8)	100.00%	0	0	0	0.0943	0	0.0385	0	0.0324	0.0126	0.081	0.1797	0	0	0	0	0.0342	0	0.1242	0.2394	0.1322	0	0.0314	0
Portugal (27)	83.02%	0	0	0	0	0	0	0	0	0	0	0	0	0.0098	0	0	0	0	0.7242	0.0696	0	0.0266	0	0
Sweden (1)	100.00%	0	0	0.025	0	0	0	0	0	0	0	0	0	0.0925	0	0	0	0	0.8754	0	0	0.0072	0	0
United Kingdom (7)	100.00%	0.096	0.3531	0	0	0	0	0	0	0	0	0.3621	0	0	0	0	0.16	0	0	0	0	0.0288	0	0
Cyprus (19)	100.00%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.2045	0	0.7955	0	0
Czech Republic (21)	88.57%	0	0.0434	0	0	0	0	0.2076	0.0203	0	0	0.1899	0	0	0	0	0.189	0	0	0.2355	0	0	0	0
Hungary (26)	100.00%	0	0.0163	0	0	0.0243	0	0	0.0099	0	0.0002	0.4262	0	0	0.0041	0	0.2897	0	0	0.2095	0	0	0.0191	0.0008
Estonia (17)	100.00%	0	0.0668	0	0	0	0.0035	0.095	0	0.0086	0	0.1871	0	0	0	0	0.5009	0	0	0.1332	0	0	0	0.0049
Lithuania (25)	100.00%	0.0023	0	0.0011	0	0	0	0.0229	0.2143	0	0	0.2448	0	0	0	0	0.2835	0	0.0597	0.1715	0	0	0	0
Latvia (22)	100.00%	0	0	0.0039	0	0	0	0	0.0635	0	0	0.2212	0	0	0	0	0.4215	0.0147	0.0484	0.2003	0.0219	0	0.0048	0
Malta (18)	100.00%	0	0.0037	0	0	0	0	0	0.1794	0.0599	0	0.2318	0	0	0	0	0	0	0.1007	0.3042	0.0695	0.018	0.0328	0
Poland (29)	100.00%	0	0	0	0.0823	0	0.0081	0	0	0	0	0.2924	0	0	0	0	0.2517	0.038	0	0.2076	0	0.1006	0.0194	0
Slovenia (16)	100.00%	0	0	0	0.0821	0	0.1368	0	0	0	0	0.1078	0	0	0.0599	0	0.2015	0.0117	0	0.2967	0	0.1036	0	0
Slovakia (28)	98.43%	0	0	0	0	0	0	0	0	0	0	0.2307	0	0	0	0	0.3131	0	0	0.3293	0	0.1112	0	0
EU15 (13)	100.00%	0	0	0.0032	0	0.0492	0	0	0.0292	0	0	0.2962	0.0338	0.0203	0	0	0.1237	0	0	0.2643	0.0994	0	0.0808	0
EU25 (14)	100.00%	0	0	0.0362	0.0013	0	0	0	0.0601	0	0	0.272	0	0	0	0	0.1212	0.0022	0	0.3042	0.1015	0.0252	0.0759	0
USA (5)	100.00%	0.0032	0	0.0344	0	0	0	0	0	0	0.0187	0.3904	0	0	0	0	0	0	0.0302	0.4928	0	0	0.029	0.0012
Japan (6)	100.00%	0.0574	0.3239	0	0	0	0	0	0	0	0	0.2532	0	0	0	0	0.2217	0	0	0.1439	0	0	0	0