BENCHMARKING SUSTAINABLE DEVELOPMENT: A SYNTHETIC META-INDEX APPROACH *

LAURENS CHERCHYE
CATHOLIC UNIVERSITY OF LEUVEN
Campus Kortrijk and Center for Economic Studies
E. Sabbelaan 53
B-8500 Kortrijk
Belgium
Tel. +32 56 24 61 09
Fax. +32 56 24 69 99
E-mail. Laurens.Cherchye@kulak.ac.be

TIMO KUOSMANEN
WAGENINGEN UNIVERSITY
Department of Social Sciences
Hollandseweg 1
6706 KN Wageningen
The Netherlands
Tel. +31 317 484 738
Fax. +31 317 484 933
E-mail. Timo.Kuosmanen@wur.nl

Working Paper

December, 2002

* This study forms a part of the research program on Nonparametric Methods in Economics of Production, Natural Resources, and the Environment. For further information, see the program homepage:
http://www.sls.wageningen-ur.nl/enr/staff/kuosmanen/program1/
The financial support for this program from the *Emil Aaltonen Foundation*, Finland, is gratefully acknowledged.
The first author thanks the Fund for Scientific Research – Flanders (Belgium) (FWO-Vlaanderen) for his postdoctoral fellowship.
We owe special thanks to Rob Dellink, Marcus Stronzik and Frank Vöhringer for constructive comments. The usual disclaimer applies.

ABSTRACT
The need for monitoring countries' overall performance in *Sustainable Development* (SD) is widely recognized, but the methods for aggregating vast amounts of empirical data remain rather crude. This paper examines the so-called 'benefit-of-the-doubt' weighting method as a tool for identifying benchmarks without imposing strong normative judgement about SD priorities. The weighting method involves linear optimization techniques, and allows countries to emphasize and prioritize those SD aspects for which they perform relatively well. Using this method, we construct a *meta-index of SD* (MISD), which combines 14 existing aggregate SD indices (developed by well-established organizations and/or expert teams) into a single synthesizing overall SD index. Within a sample of 154 countries, our index identifies 6 benchmark countries (3 high-income countries and 3 upper-middle-income countries), but also a number of seriously under-performing countries. We view this approach as a first step towards more systematic international comparisons, aimed at facilitating diffusion of the best practices and policies from the benchmark countries to the less developed world.

Key Words: *Sustainable Development, Integrated Assessment, Benchmarking, Benefit of the Doubt Weighting, Data Envelopment Analysis*

JEL classification: Q01, O57, C43, C61

1. INTRODUCTION

Benchmarking is a well-established tool for measuring the performance of business and public sector organizations (see e.g. Cox and Thompson, 1998; and Auluck, 2002, for discussion). The benchmarking practice typically starts with the identification of peers (e.g., competing firms in the same sector, firms in other industries, or other comparable organizational units) which exemplify the best practice in some activity, function or process. These best-practice peers represent reference points against which actual performance is evaluated. Reference points are often selected from external comparison partners; external benchmarking usually works effectively in drawing attention to areas of under-performance that may be ignored in internal audits.

Benchmarking is now widely applied in various types of *sustainable development* (SD) projects, mainly in the field of public administration and at the level of local communities. The benchmarking practice is typically based on *performance indices,* which aggregate various performance dimensions into a single numerical figure. Consequently, a whole literature has emerged on the construction of an operational *index of sustainable development* (ISD), which should be easy to understand and use in the context of political decision-making.[1] In this respect, major research efforts are currently targeted at developing ISDs at the local, national and international level; for example, the International Institute for Sustainable Development lists more than 200 voluntarily submitted ISD initiatives (see iisd.ca/measure/compinfo.htm). At the global scale, well known ISD initiatives include Prescott-Allen's (2001) Wellbeing Index, the Ecological Footprint of Wackernagel et al.

[1] An alternative approach for measuring sustainable development tries to *correct* national accounts (and their main aggregates like the GDP, GNI, and NI) for the cost of depleting environmental and natural resources (which is ignored in the standard national accounting system), following the classical work of Nordhaus and Tobin (1972). See e.g. Gerlagh et al. (2002) for a detailed discussion and references.

(2002), the Environmental Sustainability Index of the World Economic Forum (WEF, 2002), and the Human Development Index of the United Nations Development Program (UNDP, 2001), among many others.

Despite the generally recognized importance of a well-defined ISD for effective policy making and the considerable research effort devoted to the construction of an *ideal* ISD, we are still far from reaching consensus on the standard indicators and benchmarking methodologies. One immediate explanation for the observed heterogeneity of ISDs proposed in the literature pertains to the vague definition of the SD concept; see, e.g., Lélé (1991) for a critical discussion of various interpretations. For example, the most frequently cited definition, which comes from the Brundlandt Commission report (BCR) (World Commission on Environment and Development, 1987), describes SD as *"development that meets the needs of the present without compromising the ability of future generations to meet their own needs."* This definition most clearly illustrates the diverse, multi-dimensional character of SD. As a consequence, any operational ISD is necessarily limited in scope, capturing only a selection of SD performance dimensions. No operational ISD is perfectly objective: the indicators selection, together with the weight assigned to the different indicators, implicitly reflect the normative values of those who developed the ISD.

A second problem concerning the practical construction of an encompassing ISD relates to the choice of operational indicators associated with the selected SD dimensions. These indicators generally provide imperfect proxies for what we would really like to measure. We inevitably have to trade-off alternative 'proxy indicators' in terms of multiple criteria such as reliability, relevance, validity, cost, and coverage of data. This makes that there currently exist numerous ISDs, which differ according to the selection of SD dimensions and/or the indicators to represent those dimensions. Consequently, the results reported in the aforementioned studies are all but unisonous.

Despite these difficulties, we think that the benchmarking approach offers new, hitherto unexplored possibilities for promoting sustainable policies and practices at the international and national levels. Even though many important aspects related to the 'organizational learning' side of benchmarking do not directly apply in the international context, explicit international benchmarks could provide some –necessary?- political pressure for governments to pay more attention to SD. Perhaps most importantly, benchmarking could promote SD through facilitating diffusion of experiences and expertise from the leading countries in SD to the less-developed and under-performing countries. An obvious first step towards this end consists in identifying the SD benchmark countries.

This paper proposes a *meta-index* of sustainable development (MISD), which combines existing knowledge into a single synthesizing index of SD. We believe that all existing ISD efforts provide useful information about at least some SD aspects, in terms of the dimensions/indicators that are selected. On the other hand, each ISD can be criticized in that it only partially captures overall SD, as the number of dimensions/indicators that are included is necessarily limited. Hence, the basic motivation of our MISD is that we want to combine and structure the information captured in the existing ISDs, rather than add yet another index to an already quite long list. In line with the principles of benchmarking, our index is a comparative index: we cannot infer whether any particular country is genuinely on the sustainable development path or not; we can solely identify the best performance in *relative* terms.

Our proposal of a meta-index is similar in spirit to the research method known as "meta-analysis", a statistical approach towards reviewing and summarizing the literature (see e.g. Stanley, 2001). Our approach resembles meta-analysis in that we also aim at fully exploiting the information obtained in earlier studies on the subject under investigation. Still, our analysis differs from a typical meta-analysis in that we do not aim at a quantitative synthesis of the literature *stricto sensu*. Rather, we present a methodology for constructing a new (synthesizing) ISD that combines the ISD results reported in earlier studies.

Our main challenge in constructing the MISD pertains to the aggregation of the constituent ISDs. Clearly, the aggregation method has a decisive impact on the index values, and hence it should be based on explicitly stated, scientifically sound premises. In this paper, we propose a so-called "*benefit of the doubt*" weighting method as a potentially useful aggregation method. More specifically, in the absence of an *a priori* weighting scheme, we *endogenously* select those weights that yield the highest MISD value for each country under investigation. Putting it differently, as it is *a priori* not clear which ISD is the most appropriate to evaluate SD, for each country we attach higher weights to those ISDs for which the country under evaluation performs relatively well.

This weighting method has been successfully applied for similar aggregation problems in the context of macro-level policy performance assessment, where synthetic indicators have been used to merge performance indicators for heterogeneous economic dimensions like GDP growth, inflation, unemployment, and balance of payment surplus/deficit (after Melyn and Moesen, 1991; see, e.g., Cherchye, 2001, for a recent discussion). We argue that the multi-dimensional nature of the method easily allows extension to the more complex setting of SD. As it is closely related to Samuelson's (1948) revealed preference theory, shadow pricing in the context of activity analysis and linear programming (Dantzig, 1949; Koopmans, 1951), and the Data Envelopment Analysis (DEA) technique for productivity and efficiency analysis (Charnes et al., 1978),[2] the foundation of this method is extensively studied and theoretically sound.

Two closely related recent applications of the benefit of the doubt weighting are worth explicit mention: Zaim, Färe, and Grosskopf (2001) evaluate the well-being of individuals in different countries using the DEA technique. Mahlberg and Obersteiner (2001) apply the same approach to the indicators underlying the UNDP's Human Development Index. Whereas these authors concentrate on some specific aspects of SD, we extend the scope towards overall SD performance. Furthermore, we construct our MISD by explicitly solving a weight-selection problem, while Zaim et al. and Mahlberg and Obersteiner focus on the dual problem of measuring minimal distance to the empirical best-practice frontier. Finally, we propose a number of methodological innovations regarding the weighting procedure itself, which pertain to the bounding of the acceptable weight domain and the dealing with missing data.

The remainder of this study unfolds as follows. In the next section, we present our methodology for constructing the MISD. To describe our data sources, Section 3 reviews existing ISDs, with special attention for the selection of SD dimensions that underlies each proposal. In addition, we present a classification of existing ISDs, based on Munasinghe's (1993) triangle. Section 4 presents and discusses the empirical MISD values. In Section 5 we compare our MISD results to those of more standard approaches, and we analyze the impact

[2] Carrington, Coelli, and Groom (2002) recently investigated the usefulness of DEA for identifying international benchmarks for regulating natural monopolies like regional utility providers.

of the different ISD components on the meta-index values. Finally, in Section 6 we summarize our main conclusions and we set out a number of avenues for further research.

2. META-INDEX OF SUSTAINABLE DEVELOPMENT

Consider the general case of a cross-section of *m* ISDs for *n* countries, and let y_{ij} be the value of ISD *i* in country *j*. We assume all ISDs satisfy the following two properties (possibly after some appropriate normalization): 1) $y_{ij} \in [0,1]\ \forall i, j$; 2) $y_{ij} > y_{ik} \Rightarrow$ country *j* performs better than country *k* for ISD *i*.

Our objective is to merge these individual ISDs into a single-valued MISD, defined as the weighted average of the *m* ISDs. Given that each ISD has been developed by a team of experts, it is reasonable to assert that we cannot rate any ISD to be superior to the other ISDs by any objective grounds. This means that we are generally unable to specify *a priori* any generally acceptable weights to be accorded to each ISD. (We return to the lack of agreement among experts on the issue of SD priorities below.)

BENEFIT-OF-THE-DOUBT WEIGHTING

To resolve this weighting problem, we propose to resort to a so-called "benefit-of-the-doubt" weighting method. In this method we apply weights which maximize the index value for each country, subject to the constraint that no other country yields the index value greater than one when applying those same weights. Formally, the general Meta-Index of Sustainable Development (MISD) for country *j* is defined as the weighted average

$$\mu_j \equiv \sum_{i=1}^{m} y_{ij} \cdot w_{ij}^{*} ,$$

where the weights $w_{ij}^{*} = \arg\max_{w_{ij}} \mathrm{M}$ are obtained from the optimal solution to the linear programming problem

(objective: *weighted ISD sum*) $$\mathrm{M} = \max_{w_{ij}} \sum_{j=1}^{n} \left(\sum_{i=1}^{m} y_{ij} w_{ij} \right) v_j ;$$

s.t.

(*scaling constraint*) $$\sum_{i=1}^{m} y_{ij} w_{ik} \leq 1 \qquad \forall j, k = 1,...,n ;$$

(*non-negativity constraint*) $$w_{ij} \geq 0 \qquad \forall i = 1,...,m \ ; \ j = 1,...,n .$$

The linear programming problem has the following interpretation. The objective function reveals that the ISD weights for each country are endogenously selected to maximize the weighted sum of country-specific MISD values. As in the classic index theory, each country *j* ($j = 1,...,n$) is weighted by the *a priori* specified weight v_j for that country. The interpretation of the weights v_j is analogous to that of expenditure share or volume based weights in the construction of price indices.[3] Most importantly, each w_{ij} ($i = 1,...,m$; $j = 1,...,n$) represents the weight accorded to ISD *i* for computing the MISD value for country *j*. Unlike the v_j, the w_{ij} are not fixed *a priori*, but are endogenously selected in way that maximizes the index value of the country. To guarantee an index with an intuitive degree interpretation, we impose that no country in the sample can achieve an SD index value greater than one under these weights; see the *scaling constraint*. Finally, the individual ISD weights cannot be negative, and hence the MISD is a non-decreasing function of the ISDs; see the *non-negativity constraint*. All this implies that $0 \leq \mu_j(\cdot) \leq 1$ for each country *j*, where higher values can be interpreted as better overall SD performance.

The interpretation of the benefit-of-the-doubt weighting (or the selection of *most favorable* weights for each country) is immediate: highest relative weights will be accorded to those ISDs for which the country *j* performs best (in relative terms) when compared to other countries in the sample. This prevents policy makers from claiming that an unfair weighting scheme is employed for evaluating their country; any other weight profile can only worsen the position of the country *vis-à-vis* the other countries in the sample. In a way, the proposed methodology allows the policy makers of each country to define "their own weights"; the method reveals the optimal priority orderings for each evaluated country.[4] The result $\mu_j(\cdot) = 1$ means that there exists at least one weighting scheme under which country *j* yields the highest attainable MISD value over all countries in the sample. Alternatively, $\mu_j(\cdot) < 1$ gives the proportion of the actual MISD value (under optimal weights) over the highest attainable value in the sample of countries under investigation.

Of course, a possible criticism of this benefit-of-the-doubt approach is that it makes SD performance "look better" than what it really is, since the selected weights can deviate from the "true" (but unknown) priorities. Still, given the complexity of biological and physical systems that underlie objective priorities in terms of SD, it is very unlikely that experts will ever agree on appropriate weights/priority orderings (compare with Ludwig et al., 1993). Therefore, we opt for a second best route in this paper, where we let "the data speak for themselves" and determine the weights endogenously rather than to resort to specific *a priori* weights for each ISD.

Finally, while the use of specific *a priori values* for the ISD weights is problematic, it may well be that there is consensus on "generally acceptable" *a priori restrictions* regarding the acceptable domain of ISD weight values, which are stronger than the mere non-negativity restriction in the above model. Interestingly, the proposed methodology naturally allows for imposing such "general" weight bounds. We next discuss this issue in greater detail, hereby proposing some new approaches for setting upper- and lower-bounds for the weigh domain.

---

[3] In the present context, equality of nations supports assigning equal weights v_j for each country. Alternatively, proportioning weights to the population could better reflect equality of human beings, while weighting by the GDP shares would reflect the economic power.

[4] This idea of our "benefit of the doubt" weights comes very close in spirit to the "natural" weighting idea formulated by Hardin (1968). In fact, we believe the proposed procedure of implicit weighting suggests an attractive and easily implemented approach for addressing the problem put forward by Hardin: "It is when the hidden decisions are made explicit that the arguments begin. The problem for the years ahead is to work out an acceptable theory of weighting."

METHODOLOGICAL EXTENSIONS

In the basic MISD model, the only restriction on the ISD 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 ISD, which would then completely determine the overall SD performance value; the other ISDs would 'not matter' as their relative weight equals zero. Of course, such extreme weighting schemes can hardly be regarded as realistic or relevant. There is, hence, a need for further restricting the endogenously selected ISD weights.

This issue of imposing additional *a priori* weight bounds has attracted considerable attention in the closely related DEA literature; see for example Pedraja-Chaparro et al. (1997) for a review. The conventional approach in that literature is to bound the variability of the weights at the level of individual performance indicators (*in casu* the ISDs). In our analysis, we also impose bounds at the levels of countries and ISD categories (to be introduced below). Suppose for the moment that ISDs can be classified in p mutually exclusive categories $S_1,\ldots,S_p$; each category represents a certain orientation or focus (such as economic development, social/political equity, or environmental sustainability). Imposing weight bounds on these categories involves a relatively straightforward extension of the more standard ISD weight bounds, but is particularly interesting in this specific context, as we explain below. To the best of our knowledge, the type of country weight bounds constitutes a new innovation.

Formally, we distinguish three types of supplementary weight bounds:

(*weight bound: ISDs*) $$\frac{1}{\alpha} \le \frac{w_{hj}}{w_{ij}} \le \alpha \qquad \forall h,i = 1,\ldots,m\,;\; j = 1,\ldots,n\,;$$

(*weight bound: categories*) $$\frac{1}{\beta} \le \frac{\sum_{i \in S_k} w_{ij}}{\sum_{i \in S_l} w_{ij}} \le \beta \qquad \forall k,l = 1,\ldots,p\,;\; i = 1,\ldots,m\,;\; j = 1,\ldots,n\,;$$

(*weight bound: countries*) $$\frac{1}{\gamma} \le \frac{w_{ij}}{w_{ik}} \le \gamma \qquad \forall i = 1,\ldots,m\,;\; j,k = 1,\ldots,n\,.$$

These bounds are incorporated in the original model by simply adding the corresponding constraints to the programming problem. To enhance intuition, we here write the weight restrictions in the ratio form. The Appendix shows how these constraints are normalized to preserve the linear structure of the optimization problem. For simplicity, we here write the lower bound as the reciprocal of the upper bound; in the general case, the upper- and the lower- bound may be set independently.

In the above restrictions, the parameters $\alpha, \beta, \gamma \ge 1$ define upper- and lower weight bounds at respectively the ISD level (see *weight bound: ISDs*), the category level (see *weight bound: categories*) and the country level (see *weight bound: countries*). These weight bounds are motivated as follows:

- weight bounds for ISDs limit the variability of ISD weights by means of the parameter α, and directly exclude "unrealistic" cases where an extremely high relative weight are accorded to only one ISD (or a very limited number of ISDs); lower values of α imply more stringent weight bounds;

- weight bounds for *categories* of ISDs guarantee that different aspects of SD (e.g., economic, environmental and social-political) are adequately represented in our index; again, lower values of β imply more stringent weight bounds;
- weight bounds for *countries* determine the extent to which the ISD weights can vary over countries; like before, lower γ levels imply more stringent weight bounds.

The idea of incorporating category and country bounds originates from the observation that it is often difficult to define weight bounds on the level of individual ISDs (i.e. weight bounds of the standard type) *a priori*. It seems a much simpler task to put intuitive limits on the weight variability at the level of ISD categories or countries. Indeed, categorical weights directly reflect the importance of the key components of SD in the eventual MISD value. On the other hand, country weight bounds simply reflect to what extent categorical and country-specific weight values can differ from the mean weight-levels in the sample.

3. EXISTING ISDs: A SELECTIVE SURVEY

As discussed in the introduction, numerous ISDs have been presented, and it is practically impossible to provide an exhaustive survey of all these proposals. We will restrict our attention to a selection of ISDs, hereby following three criteria induced by the specific scope of this study (i.e., providing a SD-based cross-country comparison that synthesizes existing ISD results). An evident first criterion is availability of calculated values. The second criterion is that the ISD should have large country coverage. Finally, to ensure meaningful and fair comparisons across countries, we require that the data underlying the selected ISDs are obtained by using a uniform methodology across countries, and are thus preferably based on objectively measured quantitative statistics. Table 1 lists the ISD initiatives that meet these three conditions.

Table 1: Summary of Selected ISDs

Index	Source	Year (publication/ reference)	Country coverage	Primary focus
Human Development Index (HDI)	UNDP	2001/1999	162	Economic
Human Poverty Index-1 (HPI-1)	UNDP	2001/1999	162	Economic
Human Poverty Index-1 (HPI-2)	UNDP	2001/1999	162	Economic
Gender-related Development Index (GDI)	UNDP	2001/1999	162	Social-political
Gender Empowerment Measure (GEM)	UNDP	2001/1999	162	Social-political
Human Wellbeing Index (HWI)	Prescott-Allen	2001/n.a	180	Economic
Ecosystem Wellbeing Index (EWI)	Prescott-Allen	2001/n.a	180	Environmental
Environmental Sustainability Index-1 (ESI-1)	WEF	2002 / +-2000	142	Environmental
Environmental Sustainability Index-2 (ESI-2)	WEF	2002 / +-2000	142	Environmental
Environmental Sustainability Index-3 (ESI-3)	WEF	2002 / +-2000	142	Social-political
Environmental Sustainability Index-4 (ESI-4)	WEF	2002 / +-2000	142	Social-political
Environmental Sustainability Index-5 (ESI-5)	WEF	2002 / +-2000	142	Social-political
Health-Adjusted Life Expectancy (HALE)	WHO	2001 / 2000	191	Social-political
Ecological Footprint (EF)	Redefining Progress	2000 / 1996,1998	142	Environmental

Evidently, these criteria necessarily imply that we exclude a number of impressive initiatives that are still at the stage of theoretical exercise, that have only been adopted calculated for a handful of countries (such as the Ecological Rucksacks, the Material Input Per Service unit indices, the Genuine Progress Indicator, and the Indicator for Sustainable Economic Welfare), or that involve a considerable subjective element (such as the Corruption Perception Index, which is based on questionnaire data). Still, it is worth to stress at this point that our methodology is of course easily applied to alternative ISDs, if such would seem recommendable from the specific orientation of the study.

For the sake of brevity, we abstract from a detailed discussion of each ISD. We primarily focus on the SD dimensions that are covered, which is instrumental for our further discussion. In this respect, we will use Munasinghe's (1993) triangle as a framework for classifying the presented ISDs. Munasinghe classifies sustainability issues into three categories: 1) economic issues (efficiency, growth and stability), 2) social-political issues (poverty, consultation/empowerment, culture/heritage), and 3) environmental issues (biodiversity/resilience, natural resources, pollution); we label each ISD that we consider as 'social-political', 'environmental' or 'economic'; see also Table 1.

HUMAN DEVELOPMENT REPORT (2001)

The first five indices that we consider were adopted from the UNDP's (2001) *Human Development Report*: the Human Development Index (HDI); the Human Poverty index for Developing Countries (HPI-1); the Human Poverty Index for Selected OECD Countries (HPI-2); the Gender-related Development Index (GDI); and the Gender Empowerment Measure (GEM). Index values are calculated for 162 countries. These indices are interpreted as follows:

- The HDI is a summary measure of human development. It measures the average achievement of a country in three basic dimensions, viz. a long and healthy life, knowledge and standard of living.
- While the HDI measures average achievement, the HPI-1 and HPI-2 measure deprivations in terms of human development, respectively for developing countries and for a set of selected OECD countries. More specifically, HPI-1 captures vulnerability for death at a relatively early age, exclusion from the world of reading and communications and lack of access to overall economic provisioning. The HPI-2 measures deprivation in the same way as the HPI-1, somewhat differently defined, and includes an additional dimension of social exclusion.
- Next, the GDI adjusts the HDI to reflect the inequalities between men and women in the dimensions captured by the HDI.
- Finally, in contrast to the GDI, the GEM focuses on women's opportunities rather than capabilities, in three dimensions: political participation and decision-making power, economic participation and decision-making and power over economic resources.

The HDI, GDI and GEM are constructed in such a way that higher values indicate better performance. The opposite interpretation holds for the HPI-1 and HPI-2 values. In our below discussion we will use 1 minus the original values to convert these 'bads' into 'goods'; observe that the HPI-1 and HPI-2 are percentage indices, so that this conversion procedure preserves the informational contents of the original indices. The HDI mainly captures economic aspects of SD; to some extent it could be argued that it also (indirectly) includes social-political SD aspects. A similar interpretation holds for HPI-1 and HPI-2. Finally, the GDI and GEM have almost exclusively a social-political orientation.

THE WELLBEING OF NATIONS (2001)

We have further selected two ISDs proposed by Prescott-Allen (2001): the Human Wellbeing Index (HWI) and the Ecosystem Wellbeing Index (EWI).[5] These indices are computed for 180 countries. Their interpretation is as follows:

- The HWI gives an overall measure of socio-economic conditions; its interpretation is similar to that of the HDI index presented above.
- The EWI is a broad measure of the state of the environment.

Again, higher values always indicate better performance. Like the HDI, the HWI can be considered as a measure for economic SD performance. Obviously, the EWI can be regarded as a measure for the environmental aspects of SD.

ENVIRONMENTAL SUSTAINABILITY INDEX (2002)

Next, we consider the indices suggested by the World Economic Forum (WEF; 2002). Whereas Prescott-Allen (2001) presented the EWI, the WEF proposes an overall Environmental Sustainability Index (ESI), which is intended "to measure overall progress towards environmental sustainability." This overall index has been computed for 142 countries.

In our following discussion, we will not directly concentrate on this overall ESI index. Rather, we will consider its five core components, which pertain to different aspects of environmental sustainability:

- The state of the environmental systems (ESI1), which captures air quality, water quantity, water quality, biodiversity and land.
- The stresses on those systems (ESI2), as measured in terms of air pollution, water stress, ecosystem stresses, waste and consumption pressures and population growth.
- Human vulnerability to environmental change (ESI3), in the form of basic human sustenance and environmental health.
- Social and institutional capacity to cope with environmental challenges (ESI4), pertaining to science and technology, the capacity for debate, environmental government and eco-efficiency.
- Global stewardship (ESI5), as reflected in participation in international collaborative efforts, greenhouse-gas emissions, and reducing transboundary environmental pressures.

The five indices are constructed so that higher values reflect better performance. While the first two ESI components are almost exclusively concerned with environmental aspects, the last three components have a more social-political orientation.

WORLD HEALTH REPORT (2001)

The World Health Organization (WHO, 2001) provides an index for the health-adjusted life expectancy (HALE), which was reported for 191 countries in the year 2000. This HALE indicator combines losses from premature death (defined as the difference between the actual age of death and life expectancy at that age in a low-mortality population), and loss of healthy

---

[5] Prescott-Allen (2001) also propose an overall "Wellbeing Index" (WI) which is obtained as an equally weighted average of the HWI and EWI. As our MISD similarly combines the HWI and EWI (using an unequal weighting procedure), we will not directly consider this WI in our following discussion.

life resulting from disability. Clearly, this index primarily captures social-political aspects of SD, while it also indirectly reflects the economic and environmental aspects. Higher values can be interpreted as better SD performance.

Ecological Footprint (1996-1998)

National estimates of the Ecological Footprint (EF) per capita, proposed and discussed by Wackernagel et al. (2002), are calculated by the public policy organization Redefining Progress, and are reported in WWF's (2000) Living Planet Report for 142 countries in 1996; Redefining Progress provides updated figures for a sub-sample of 48 countries (for the year 1998; see http://www.rprogress.org/programs/sustainability/ef/projects/1998_results.html).[6] This footprint statistic measures the land and water area that is required to support a defined human population and material standard indefinitely, using prevailing technology. Clearly, this ISD can also be interpreted as measuring "the burden of human lifestyle to the ecology, i.e. the area of 'average quality' land needed to support one human being by the ecological services he needs." Hence, lower Ecological Footprint values indicate better environmental SD performance. For convenience, we will consider a transformation of the original data in our below discussion so that better performance is associated with higher values. Specifically, we use 1- f/max(f), where f denotes the original Ecological Footprint index.

Correlation Analysis

Before proceeding to the empirical application, we briefly discuss the correlations between the different ISDs considered, so as to give an impression about the underlying tradeoffs between the different dimensions of SD. Our results in Table 2 suggest a number of interesting patterns. First, our results reveal high correlation between economic ISDs (HDI, HPI-1 and HWI) and ISDs with a primary focus on social-political SD dimensions (GDI, GEM, ESI-3 and HALE); the picture is somewhat less outspoken for ESI-4 and mixed for the economic HPI-2 and social-political ESI-5. The nearly perfect correlation between the HDI and the GDI is especially striking; the GDI correlates even stronger with the economic HDI then with the GEM, while the GDI and GEM equally refer to differences between men and women (in terms of capabilities and opportunities, respectively).

While economic and social-political ISDs clearly tend in the same direction, the picture is much more ambiguous when comparing economic and social-political ISDs with environmental ISDs. We almost persistently find a negative correlation between the environmental ISDs (ESI1, ESI2, EWI and Ecological Footprint) on the one hand and the economic or social-political ISDs (again, except for HPI-2 and ESI-5), on the other. The few positive correlations are of low magnitude. In a way, these observations provide an empirical confirmation for our earlier ISD classification, which was originally based on the single-dimensional indicators that underlie each ISD.

Further, in contrast to our findings for the economic and social-political ISDs, the correlation between the different environmental ISDs is generally low; it is even negative in some cases. This supports the position that different environmental dimensions should be considered simultaneously when assessing environmental SD; concentrating on a single environmental ISD may yield normative conclusions that are heavily influenced by the narrow focus of the analysis.

[6] To strike balance between the problem of missing data and the intention of using the most recent data, the arithmetic average of the 1996 and 1998 figures was used in all computations below.

Table 2: ISD (Spearman Rank) Correlation Matrix

	HDI	HPI-1	HPI-2	HWI	EWI	EF	ESI1	ESI2	GDI	GEM	HALE	ESI3	ESI4	ESI5
Economic:														
HDI	100.00%													
HPI-1	87.79%	100.00%												
HPI-2	21.77%	(n.a.)	100.00%											
HWI	95.38%	85.54%	54.70%	100.00%										
Environmental:														
EWI	-24.21%	-31.03%	82.60%	-23.62%	100.00%									
EF	-90.58%	-82.94%	4.72%	-87.89%	27.46%	100.00%								
ESI1	7.00%	9.56%	20.08%	9.69%	14.28%	-12.44%	100.00%							
ESI2	-26.54%	-10.91%	49.66%	-18.73%	9.28%	30.22%	24.31%	100.00%						
Social-political:														
GDI	99.76%	90.56%	14.19%	95.77%	-25.13%	-90.20%	9.22%	-28.47%	100.00%					
GEM	79.49%	74.53%	47.75%	78.37%	-19.01%	-70.34%	12.68%	-39.52%	80.84%	100.00%				
HALE	94.67%	79.70%	15.19%	90.10%	-27.75%	-83.99%	-2.01%	-25.21%	94.91%	74.02%	100.00%			
ESI3	94.50%	84.01%	36.84%	92.99%	-26.94%	-89.51%	7.65%	-24.16%	94.16%	72.61%	92.44%	100.00%		
ESI4	57.82%	44.80%	63.08%	62.50%	-4.12%	-46.28%	22.32%	-19.51%	61.45%	70.48%	54.22%	55.22%	100.00%	
ESI5	-45.06%	-43.84%	70.42%	-37.68%	27.71%	51.47%	22.21%	7.63%	-45.76%	8.11%	-40.16%	-38.38%	18.50%	100.00%

More generally, Table 2 aptly reveals that different SD indicators, referring to particular (economic, social-political or environmental) SD dimensions, can yield very different results in terms of SD performance (e.g., regarding the country ranking). This highlights the need for a synthesizing *meta*-ISD (or MISD), which brings together these different/complementary pieces of information.

4. Empirical Results

Model Specification

Recall that the original MISD model described in Section 2 allows for full weight flexibility. In principle, it is even possible that only a single ISD is weighted in the eventual index value, which makes that the overall SD performance would be completely determined by that ISD. Such extreme weight scenarios seem all the more problematic in view of the relatively large number of performance dimensions in the present study; see the 14 ISDs reviewed in the previous section. Indeed, it seems hardly reasonable to evaluate SD in terms of only a single ISD if we select not less than 14 ISDs in total. For these reasons, we impose some additional weight bounds in our present MISD application, which reflect our *a priori* judgment about the relative importance of different ISDs and ISD categories, and which put normative limits on the variation of weights across countries.

As for the ISD categories, it is widely held that all three dimensions of sustainability (social-political, economic, and environmental) should be equally represented in the index. This would suggest the parameter value $\beta = 1$ in the weight bound of the ISD categories. However, in view of the numerous shortcomings in the international data from which the ISDs are calculated, we allow for some flexibility in the group weighting, and set the bound parameter at $\beta = 1.2$.

Next, we find it reasonable to impose that, for any ISD, the weight of one country should not depart too much from that of another country. Indeed, while the operational conditions and the policy preferences of countries can differ considerably, we think it recommendable from a normative point of view that countries should conform to the most rudimentary ideals and values of SD. Clearly, the tradeoff between conformity to universal weighting and freedom for country-specific deviations is not easy to resolve. Therefore, we opt for a so-called 'conservative' bound and put $\gamma = 3$, implying that the maximum weight (over countries) in any particular ISD can only be 3 times higher than the minimum weight.

Finally, we find it most difficult to set acceptable bounds for the relative ISD weights that are selected for each country. We believe all ISDs should get a positive weight in the index, but we also want to let to data speak for themselves, i.e. to fully exploit the attractive benefit of the doubt interpretation that underlies our MISD model. Again adopting a conservative perspective, we specify the bound parameter as $\alpha = 10$. This means that, for each country, the maximum weight of an ISD is at most 10 times greater than the minimum ISD weight.

In sum, we end up with the following bound specifications:

- $\alpha = 10$, i.e., $0.1 \le \frac{w_{hj}}{w_{ij}} \le 10$ for all ISDs *h*, *i* and countries *j*;
- $\beta = 1.2$, i.e., $0.833 \le \frac{\sum_{i \in S_k} w_{ij}}{\sum_{i \in S_l} w_{ij}} \le 1.2$ for all ISD categories S_k, S_l;
- $\gamma = 3$, i.e., $0.333 \le \frac{w_{ij}}{w_{ik}} \le 3$ for all countries *j*, *k* and all ISDs *i*.

We next calculated the MISD values for a sample of 154 countries. All countries for which at least 6 out of 14 ISDs are reported were included in the sample. The large numbers of missing data, especially for developing countries, causes some difficulties for our analysis. Of course, we could limit attention to those countries for which the complete data is available, but this would yield a sample consisting of only 15 countries. Given that our methodology (which –to recall- directly builds on the observed data) generally requires a large sample, there is no other option than to proceed with the unbalanced data. The missing entries are then simply discarded from the analysis by inserting the value of zero in the data matrix. Clearly, this creates a possible positive bias in our results, since data unavailability may signal problems in that particular area. Still, we think that the minimum number of six ISDs, which all contain valuable information, should suffice to provide a reasonably balanced overall SD picture.

From the technical perspective, the standard DEA weighting model will automatically match the missing entries with a weight of zero, i.e., the missing ISDs for a given country are *ipso facto* excluded from the analysis of that particular country. However, we also need to account for the missing data when defining the weight bounds, to avoid that zero entries arbitrarily influence the results. Interestingly, we can circumvent the problem of missing/zero entries in the data matrix by means of a simple modification of the weight bounds: we multiply the inequality constraints by the product of the corresponding ISDs, which is a constant; see Kuosmanen (2002) for a more detailed discussion. For example, the resulting weight bound for ISDs reads

$$\frac{1}{\alpha}\cdot y_{hj}\cdot y_{ij}\leq\frac{w_{hj}}{w_{ij}}\cdot y_{hj}\cdot y_{ij}\leq\alpha\cdot y_{hj}\cdot y_{ij}.$$

Clearly, if either one of the data entries equals zero, then the inequalities become redundant, and hence the missing entries cannot flaw the relative weights. On the other hand, if both ISD values are strictly positive, then this simple modification has no impact whatsoever on the original inequalities.

MISD RANKINGS

The MISD values were calculated from the optimization problem M discussed in Section 2, using GAMS with the CONOPT2 solver. (See the Appendix for the Linear Programming formulation of the problem.)

Although we treated all countries equally in the pooled sample, we find it most illustrative to view the results from the perspective of the proper peer groups. We therefore classify the countries in high-income (Gross National Income (GNI) per capita greater than 9.266 US dollars in 2000), upper-middle-income (GNI per capita between 2996 - 9.266 US dollars), lower-middle-income (GNI per capita between 755 - 2996 US dollars), and low-income (GNI per capita less than 755 US dollars) countries, according to the classification of the World Bank (www.worldbank.org/data/databytopic/class.htm). We think this classification makes comparison and ranking of countries more meaningful, and allows us to identify more appropriate benchmarks for each country (although it should be stressed that the MISD scores are in principle also comparable across the income groups).

Table 3 lists the MISD rankings of the 28 high-income countries in our sample. Norway, Sweden, and Austria show example as the leading countries on the way towards more sustainable development.[7] Overall, the country rankings do not offer any major surprises. Northern and Western European countries strongly dominate the index. The oil-producing countries (i.e. Kuwait and the United Arab Emirates) distinguish as a low-performing sub-group among the richest countries. Finally, the relatively low score of the United States, which may be somewhat surprising at first, is solely due to the weak performance in terms of the environmental dimensions. A somewhat similar qualification applies for Luxembourg, although missing data may also partly explain the low rank of that country.

We firmly stress that these results should be interpreted with sufficient caution. For example, we cannot directly conclude from these results that the top-ranked countries are on the SD path. Indeed, our MISD is by construction a comparative index, which assesses SD performance of any country relative to that of the other countries in the sample. Our index does not directly account for the burden of technological and economic processes to the World's ecosystem. Still, we strongly believe that this comparative approach – when correctly interpreted - has its own merits. Probably most importantly in that respect, the MISD evaluates SD performance in terms of what is actually achieved (by the countries in our sample), which indeed seems an attractive second-best route in the absence of full information about the true physical, technological and economic possibilities.

[7] In case of ties, we ranked countries using the arithmetic average of ISDs as a secondary criterion. For Norway and Sweden, also the arithmetic averages were equal.

Table 3: MISD Rankings; High-Income Countries

Rank	Country	MISD	Rank	Country	MISD
1	Norway	1.000	**15**	Slovenia	0.944
1	Sweden	1.000	**16**	Greece	0.939
3	Austria	1.000	**17**	Australia	0.922
4	Switzerland	0.991	**18**	Belgium	0.918
5	Finland	0.983	**19**	Israel	0.914
6	Netherlands	0.983	**20**	United Kingdom	0.909
7	Iceland	0.976	**21**	Denmark	0.907
8	Spain	0.964	**22**	Ireland	0.900
9	France	0.960	**23**	New Zealand	0.897
10	Portugal	0.959	**24**	United States	0.804
11	Canada	0.957	**25**	Luxembourg	0.748
12	Japan	0.953	**26**	Singapore	0.699
13	Germany	0.948	**27**	Kuwait	0.692
14	Italy	0.948	**28**	United Arab Emirates	0.605

Table 4 presents the results and rankings of the upper-middle-income group. Also in this group our index identifies 3 leading nations: Costa Rica, Uruguay, and Panama. More generally, we find that the Middle- and South American nations perform especially well. The good performance of the EU-candidates Croatia, Hungary and Slovakia is equally encouraging. In certain areas of SD, the top-ranked countries of this group can act as benchmarks for the countries in the middle-income group. However, also in this case we cannot directly infer from our comparative indices that these countries are on a truly sustainable path; for example, non-governmental organizations have expressed their concerns about the violence against women in all three benchmark countries of this group (see, e.g., the UN Economic and Social Council, 1999a), and it is well-known that pesticides are intensively used in Costa Rican banana plantations (see, e.g., UN Economic and Social Council, 1999b).

Table 4: MISD Rankings; Upper-Middle-Income Countries

Rank	Country	MISD	Rank	Country	MISD
1	Costa Rica	1.000	**14**	Trinidad & Tobago	0.888
2	Uruguay	1.000	**15**	Lebanon	0.887
3	Panama	1.000	**16**	Poland	0.886
4	Croatia	0.984	**17**	South Korea	0.886
5	Chile	0.977	**18**	Botswana	0.854
6	Hungary	0.971	**19**	South Africa	0.852
7	Slovakia	0.966	**20**	Libya	0.847
8	Brazil	0.963	**21**	Oman	0.839
9	Venezuela	0.962	**22**	Gabon	0.825
10	Mexico	0.960	**23**	Saudi Arabia	0.785
11	Turkey	0.930	**24**	Argentina	0.725
12	Estonia	0.920	**25**	Mauritius	0.662
13	Czech Republic	0.892			

The MISD rankings of the lower-middle-income countries are reported in Table 5. Apparently, none of the countries in this group can be distinguished as a global benchmark. Still, Colombia, Peru, and Latvia do come very close the to top-ranked richer nations. Further, and in line with our results in Table 4, the Latin American countries perform relatively well. More generally, we observe that the overall distribution of the MISD values in this group is fairly well comparable to that of the upper-middle-income group. The good overall performance of the countries in the middle-income group as a whole becomes especially apparent by comparing these countries to those of the high-income group: as many as 19 middle-income countries perform better than Japan, and no less than 60 countries (including Russia) outperform the US!

Table 5: MISD Rankings; Lower-Middle-Income Countries

Rank	Country	MISD	Rank	Country	MISD
1	Colombia	0.991	24	El Salvador	0.911
2	Peru	0.976	25	Romania	0.911
3	Latvia	0.974	26	Lithuania	0.907
4	Cuba	0.973	27	Moldova	0.907
5	Armenia	0.970	28	Jamaica	0.906
6	Dominican Rep.	0.969	29	China	0.904
7	Sri Lanka	0.963	30	Morocco	0.900
8	Bolivia	0.961	31	FYR Macedonia	0.892
9	Thailand	0.955	32	Guatemala	0.885
10	Paraguay	0.954	33	Syria	0.879
11	Philippines	0.949	34	Egypt	0.874
12	Ecuador	0.946	35	Namibia	0.871
13	Albania	0.938	36	Papua New Guinea	0.859
14	Jordan	0.932	37	Kazakhstan	0.853
15	Algeria	0.924	38	Russia	0.848
16	Belarus	0.922	39	Turkmenistan	0.800
17	Bulgaria	0.922	40	Iraq	0.774
18	Iran	0.920	41	Bosnia & Herzegovina	0.712
19	Indonesia	0.914	42	Belize	0.690
20	Honduras	0.912	43	Fiji	0.671
21	Viet Nam	0.912	44	Guyana	0.654
22	Malaysia	0.911	45	Cape Verde	0.633
23	Tunisia	0.911	46	Maldives	0.604

The results in Table 6, which pertain to the 55 countries of the low income group, are more disappointing. Eritrea, Lesotho, and Yemen are distinguished as the least developed countries in the entire sample. Still, despite the obvious economic and social problems, a number of countries of this group (e.g., Myanmar) perform relatively well in terms of environmental indicators that capture emissions and material flows. In fact, these findings rather make us doubt whether the original ISD indicators, which are essentially constructed from the perspective of the high-income economies, are well-adapted to give a reasonably balanced picture of the SD performance of low-income nations. Nonetheless, our results do suggest that an important challenge for the developed world consists in improving the economic welfare of the low-income countries without causing excessive harm to the natural environment.

Overall, the country rankings conform well to our prior expectations. The group of leading countries appear among the top in almost all individual ISDs, but a large number of countries come very close to the best performers. Indeed, one of the striking features of this index is that the differences between the highly developed countries and the developing countries does not seem as insurmountable as one might expect.

The above results further suggest that it is difficult to distinguish any obvious groups of especially well or poorly performing countries without reference to secondary criteria like the geographical location or the income level. Therefore, we next take a closer look at a number of summary statistics for different income and geographical groups.

Table 6: MISD Rankings; Low-Income Countries

Rank	Country	MISD	Rank	Country	MISD
1	Myanmar	0.900	29	Mozambique	0.815
2	Nicaragua	0.898	30	Congo, Dem. Rep. (Zaire)	0.813
3	Kyrgyzstan	0.887	31	Gambia	0.801
4	Ghana	0.885	32	Zambia	0.801
5	Bhutan	0.882	33	Senegal	0.800
6	Congo, Rep	0.881	34	Ukraine	0.798
7	Cameroon	0.869	35	Chad	0.793
8	Azerbaijan	0.865	36	Ivory Coast	0.792
9	Zimbabwe	0.859	37	Mali	0.786
10	Bangladesh	0.858	38	Guinea	0.773
11	Laos	0.857	39	Angola	0.769
12	Kenya	0.856	40	Guinea-Bissau	0.767
13	Cambodia	0.855	41	Ethiopia	0.765
14	Tanzania	0.848	42	Burkina Faso	0.760
15	Tajikistan	0.848	43	Rwanda	0.760
16	Uzbekistan	0.844	44	Nigeria	0.751
17	Benin	0.844	45	Burundi	0.734
18	India	0.841	46	Mauritania	0.732
19	Sudan	0.840	47	Niger	0.729
20	Central African Rep.	0.837	48	Sierra Leone	0.703
21	Nepal	0.832	49	Somalia	0.622
22	Madagascar	0.829	50	North Korea	0.619
23	Mongolia	0.822	51	Liberia	0.615
24	Togo	0.822	52	Comoros	0.510
25	Uganda	0.822	53	Yemen	0.483
26	Haiti	0.817	54	Lesotho	0.477
27	Malawi	0.816	55	Eritrea	0.458
28	Pakistan	0.815			

Table 7 summarizes the results at the level of income groups. The shapes of the MISD distributions appear rather similar in all four income groups; the most interesting differences concern the position of the distributions. Not surprisingly, the average MISD values tend to be higher for higher income groups. As noted above, however, the two middle-income classes do not fall very far apart from the high-income group; the average MISD score of the lower-middle-income group is only 0.027 points below that of the high-income group. The difference is much more pronounced for the low-income countries. Overall, we conclude that, abstracting from the poorest countries, high GDP is not a prerequisite for SD.

Table 7: Summary Statistics of the MISD Values; 4 Income Classes

	High	Upper-mid.	Lower-mid.	Low
GNI / cap. (US.$)	> 9266	2996 – 9266	755 – 2995	< 755
number	28	25	46	55
min.	0.605	0.662	0.604	0.458
max.	1.000	1.000	0.991	0.900
average	0.908	0.898	0.881	0.783
st.dev.	0.103	0.087	0.098	0.106
skewness	-1.731	-1.027	-1.579	-1.779

To further scrutinize the particular relationship between SD performance and income, Figure 1 visualizes the correlation between our MISD and GDP per capita (measured in 2000 US dollars; PPP). To facilitate the interpretation, the figure also indicates the identity of the top-ranked countries and a selection of under-performers. Although none of the poorest countries achieves high MISD values, the distribution of MISD values is relatively stable across the observed income range after a minimal thresh-hold of roughly 3,000 US$ per capita (i.e., in the upper-middle income and high income groups). These findings fall in line with our above results and further support the argument that, apart from the low income and lower-middle income countries, the GDP per capita indicator does not adequately reflect the various dimensions of SD; it is important to additionally account for the environmental and social-political aspects. Conversely, the MISD demonstrably contains additional information not captured by the mere GDP figures.

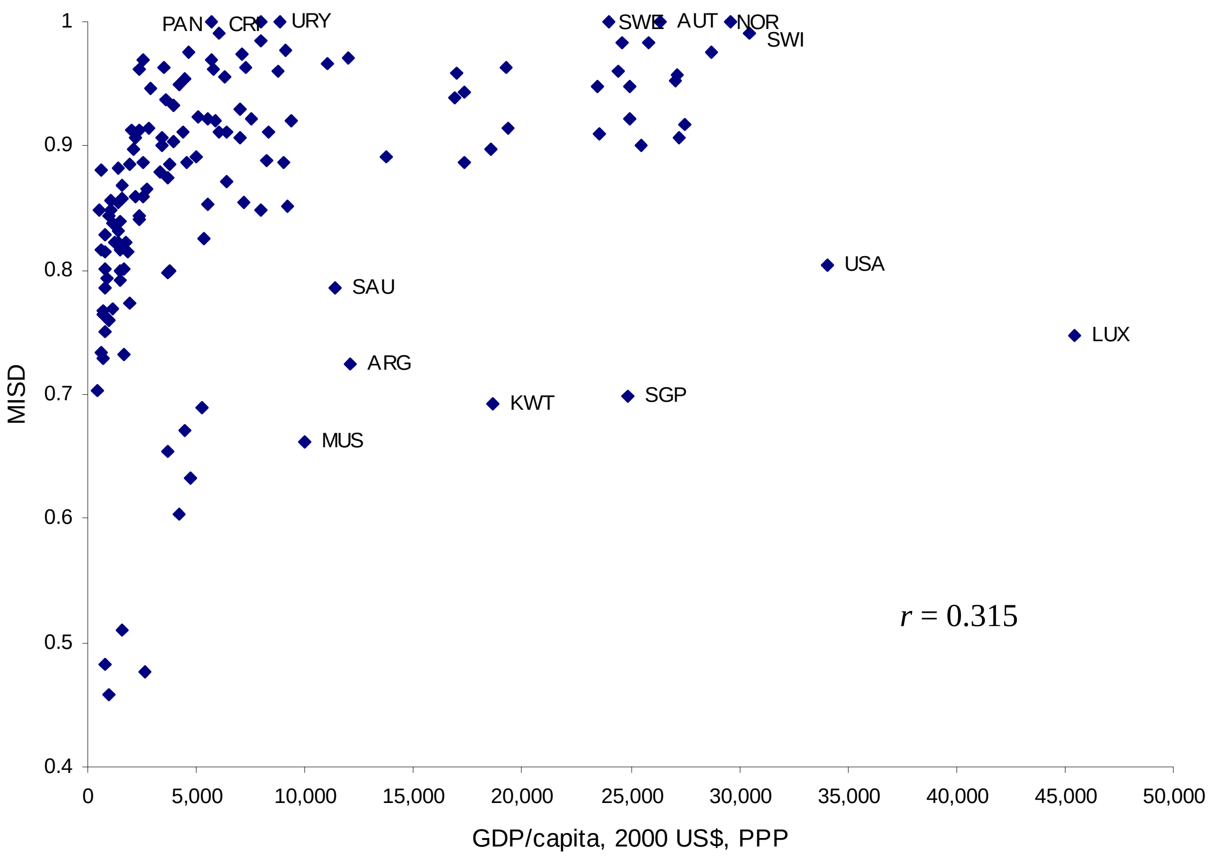

Figure 1: Scatter Plot of the MISD and the GDP per Capita.

Next, we compare the differences between the developed and the developing countries, hereby classifying the developing countries according to their geographical location (following the World Bank classification). The group of developed countries is the same as the high-income group above (these countries are not included in the geographic classification of the World Bank). Table 8 reports the summary statistics for each group of countries.

Table 8: Developed versus Developing Countries; by Region

	Developed countries	Developing countries					
		Europe & Central Asia	Latin America & Caribbean	East Asia & Pacific	Middle East & North Africa	South Asia	Sub-Saharan Africa
number	28	25	24	14	13	7	43
min.	0.605	0.712	0.654	0.619	0.483	0.604	0.458
max.	1.000	0.984	1.000	0.955	0.932	0.963	0.885
average	0.908	0.894	0.913	0.858	0.843	0.828	0.768
st.dev.	0.103	0.064	0.098	0.098	0.119	0.110	0.104
skewness	-1.731	-0.938	-1.650	-1.716	-2.604	-1.502	-1.598

Comparing the average MISD values, we find that there are no great differences between developing countries of Europe, Asia, and Pacific. Interestingly, the Latin American and Caribbean countries seem to outperform the European and Central Asian countries; in fact, the former group of countries does even slightly better than the group of high-income developed countries. Finally, the Sub-Saharan African countries stand out as a group with a significantly lower average MISD value than the other country groups. Indeed, the development problems of this region are well known, but appropriate solutions remain undiscovered.

Next turning to the other summary statistics in Table 8, we observe that the Middle East and North Africa group has the highest standard deviation and the lowest skewness in the MISD distribution. This is entirely due to the weak performance of Yemen; the other countries of this group have a profile that is very similar to that of the neighbor-group Europe and Central Asia.

In fact, when also considering the underlying results for the individual countries, it turns out that the group of developed countries can be considered as the most diverse in terms of the associated performance values. Indeed, this group does not only include the benchmark countries Norway, Sweden, and Austria, but also the low-ranked Arab nations Kuwait and the United Arab Emirates; Kuwait and the United Arab Emirates fall behind all other countries in their region (apart from Yemen).

WEIGHTS

Besides the index values, the optimal solution of the MISD problem also provides us with valuable information about the implicit policy weights *w*. In this respect, recall that our benefit-of-the-doubt weighting scheme uses the weights that maximize the relative index value for each country.

Let us first consider the classification of ISDs in terms of economic, social-political, and environmental indices. We imposed the restriction that the sum of weights in each category should not exceed the sum of weights in any other category by more than 20 per cent. This constraint proved to be binding in the optimal solution. To our surprise, however, the countries of our sample would actually have accorded a higher weight to the environmental category than to the economic or the social-political oriented ISDs, had we not imposed this additional weight bound: the environmental indicators were assigned the maximum weight (37.5% of the total sum of weights), while the minimum weight (31.2%) was given to both the economic indicators and the social-political indicators.

The most plausible explanation for this result lies in the relative nature of our MISD. It appears that the inequalities are more pressing in the economic and social dimensions than in the environmental indicators. In general, the leading high-income countries have made considerable investments in cleaner technologies, which show up as good performance in "emissions per capita" –type of indicators. On the other hand, many of the middle- and low-income nations do well in terms of the "lifestyle" indicators because of their low consumption of resources. Admittedly, great differences in the environmental performance of nations remain. Still, the good and the bad environment are more evenly distributed than the economic wealth. Therefore, the majority of countries benefits if the environmental SD dimensions are emphasized in the assessment. Still, the environment is not the only aspect of

SD; our category weight restrictions guarantee that the economic and social/political SD dimensions are also important in the calculated index values.

To give an idea about the general importance of each ISD in our MISD, Table 9 reports the average weight of each ISD, together with the associated standard deviation. The standard deviations reveal substantial country-specific variations within the specified bounds. Still, the average weights at least provide a rough impression about the overall impact of each ISD. In that respect, especially the Ecological Footprint and the HDI turn out to be rather influential for the calculated MISD values, while the ESI-4 and the GEM have a more moderate impact; the latter only attract about 10 to 15 percent of the weights accorded to the former. It is important to interpret these results correctly: the fact one ISD gets a higher average weight than another does not necessarily mean that is more reliable or important for overall SD, but rather that it is generally advantageous for countries to attach a higher weight to the first ISD in our index.

Table 9: Summary Statistics of the Weights

ISD	Mean	St.dev	Orientation
EF	0.227	0.065	Environmental
HDI	0.206	0.082	Economic
ESI-2	0.153	0.062	Environmental
HALE	0.152	0.064	Social-political
ESI-5	0.127	0.064	Social-political
HPI-1 (developing countries)	0.114	0.093	Economic
GDI	0.044	0.025	Social-political
ESI-1	0.041	0.045	Environmental
HWI	0.039	0.020	Economic
EWI	0.036	0.018	Environmental
HPI-2 (developed countries)	0.033	0.058	Economic
ESI-3	0.032	0.015	Social-political
GEM	0.029	0.010	Social-political
ESI-4	0.027	0.006	Social-political

5. DISCRIMINATORY POWER AND CORRELATION ANALYSIS

In this section, we assess the performance of the proposed MISD by comparing it to the results of the two more standard alternatives: 1) the equally weighted average of the ISDs, and 2) the index resulting from basic benefit-of-the-doubt weighting without additional weight bounds. The first method is frequently used in context of the sustainability indices; see e.g. the UNDP's development indices or Prescott-Allen's (2001) Wellbeing Index. The latter method, widely known as Data Envelopment Analysis (DEA), is often applied in similar weighting problems in decision sciences; from the methodological perspective, the DEA analysis can be viewed as an intermediate step towards our MISD proposal. In our further exposition, the SD index obtained as the arithmetic average of the ISDs will be labeled 'Average' index, and the index obtained from benefit-of-the-doubt weighting without additional weight restrictions will be labeled 'DEA' index; the weight-restricted index advocated in the current paper will again be labeled 'MISD'.

In our empirical assessment, we first consider the discriminatory power of the empirical criteria. Subsequently, we will consider the correlation between the alternative overall SD indices (Average, DEA and MISD) and the correlation between these indices and the constituent ISDs.

DISCRIMINATORY POWER

Obviously, a well-defined MISD should have sufficient discriminatory power. We compare the performance of our MISD to the two standard alternatives by looking at the descriptive statistics associated with each variant (average, standard deviation, minimum and maximum) and the percentage of 'efficient' countries (i.e. countries with a relative performance score of 100%); also, we consider the distribution histogram associated with each index.

In a first step, we compare the Average results with the DEA results. The associated descriptive statistics are given in Table 10; Figure 2 highlights the corresponding distribution histograms. From these results we learn that the DEA model has very low discriminatory power; e.g., Figure 2 shows that almost all countries are efficient or almost efficient. As compared to the Average index, the average performance value raises from 62% to almost 98%. In addition, standard deviation is only 5% and the minimum performance value is 70%. These somewhat disappointing results for the DEA model suggest imposing additional weight bounds as a promising avenue.

Table 10: Summary Statistics; Average, DEA and MISD

	Average	DEA	MISD
min.	0.47385621	0.701016474	0.458
max.	0.901128433	1	1
average	0.657073893	0.977714843	0.853766234
st.dev.	0.102428594	0.048897638	0.11297225
% eff.	0	83	6

This is indeed confirmed by our results, which clearly illustrate the intermediate nature of our MISD. By imposing intuitive, widely acceptable normative weight bounds (on the ISD level, the ISD category level and the country level), this measure excludes cases where the performance score is fully determined by only a single ISD value, which entails considerably higher discriminatory power than the DEA index. Also, it does not resort to the overly restrictive weighting scheme where each ISD gets the same weight, which is reflected in the fact that the MISD values are generally higher than the Average values.

Overall, discriminatory power of our MISD is satisfactory. Only 6 of the 154 countries in our sample are declared as (relatively) efficient, as opposed to as much as 83 countries for the unrestricted DEA index. In addition, while average efficiency naturally increases as compared to the Average index (from 66% to 85%), standard deviation of the MISD remains somewhat above that of the Average index (from 10% to 11%), and higher standard deviation can be interpreted as evidence of more discriminatory power. Finally, the minimal efficiency value is even slightly below that of the Average index (and, even more, is below 50%), which indicates that overall poor performance (i.e. in terms of the economic, environmental as well as social-political ISDs) remains severely penalized. These features are also graphically illustrated in Figure 2.

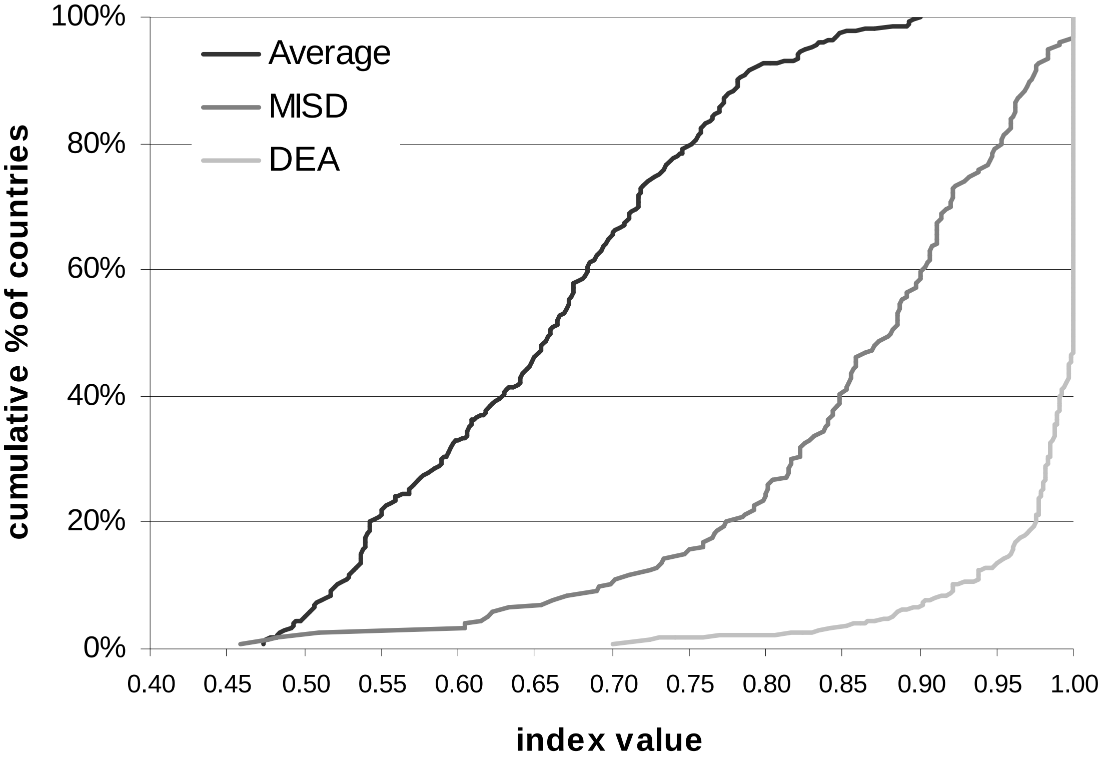

Figure 2: Cumulative Distribution Functions; Average, DEA, and MISD

CORRELATION ANALYSIS

It can be argued that a well-defined MISD should adequately reflect the information captured by each of its ISD components. We therefore next look at the correlation between the three overall SD indices (Average, DEA and MISD) and the corresponding values of the constituent ISDs, reported in Table 4. Note that this also 'penalizes' MISDs of which values are based on a single ISD for which the evaluated country performs relatively well, while the country performs generally poorly for the other ISDs. Of course, this criterion also comprises that a well-defined MISD should give a 'balanced' indication of the 'aggregate' performance in the three dimensions of SD performance (social-political, economic and environmental).

We should first note that this correlation analysis is complementary in nature to the above analysis of discriminatory power. Discriminatory power pertains to the question whether the SD index allows for clearly distinguishing between countries in terms of SD performance, and hence relates to *practical usefulness* of the index. Still, it does not give any information about the 'interpretation' of the associated SD index values. In fact, it can be argued that an index with high discriminatory power adds little value when it does not adequately reflect the information captured in its constituent ISDs. Correlation analysis should give us better insight into the *interpretation* of our MISD.

Ideally, a MISD should correlate at least moderately with all or at least a majority of the underlying ISDs. Still, high correlation is not desirable *per se*: typically, a high correlation with a few ISDs would cost a low or negative correlation with other ISDs, implying a disproportionate representation of ISDs in the overall index; generally, there is a trade-off between the degree of correlation with one ISD and the degree of correlation with another ISD when the two SDIs are negatively interrelated. A well-balanced MISD would require that the correlation coefficients across ISDs should converge to some common level.

As a preliminary step, we first regard the correlation between the Average index, the DEA index and the MISD, reported in Table 11. Our results clearly illustrate the intermediate position of the MISD; it correlates stronger than the DEA index with the Average index and stronger than the Average index with the DEA index. Further, we find that the MISD correlates stronger with the Average index than with the DEA index. At least this suggests that imposing some restrictions on the endogenously determined ISD weights implies a ranking that is considerably different from that obtained from the extreme model that uses no weight bounds at all; the corresponding correlation coefficient amounts to merely 50%. Still, the obtained ranking also differs quite substantially from that resulting from the other extreme model where fixed (equal) weights are used to value each ISD in the index; the corresponding correlation coefficient is no more than 72%.

Let us then have a closer look at the correlations between the presented overall SD indices and the underlying ISDs, also reported in Table 11. We find that the Average model systematically correlates relatively strongly with the economic and social-political ISDs (the only exception is ESI5). The intuition of the result lies in the fact that economic indicators and social-political indicators are strongly positively inter-related, while there appears to be some trade-off between economic and social-political performance on the one hand and environmental performance on the other (see Section 3). This makes that the Average index, which gives an equal weight to each indicator, will tend in the direction of these predominant economic and social-political indicators.

Table 11: MISD-ISD (Spearman Rank) Correlation Matrix

	Average	DEA	MISD
Average	100.00%		
DEA	41.47%	100.00%	
MISD	71.80%	49.82%	100.00%
Economic:			
HDI	84.91%	34.26%	61.36%
HPI-1	74.94%	53.57%	56.60%
HPI-2	76.78%	(n.c.)	52.73%
HWI	87.56%	36.76%	62.22%
Environmental:			
EWI	1.82%	-5.77%	-10.01%
EF	-72.83%	-20.60%	-51.16%
ESI1	29.99%	17.08%	23.48%
ESI2	0.50%	13.79%	18.67%
Social-political:			
GDI	86.03%	36.69%	59.40%
GEM	70.15%	54.39%	35.57%
HALE	82.11%	30.93%	62.40%
ESI3	84.85%	33.16%	67.13%
ESI4	70.95%	45.35%	56.66%
ESI5	-12.21%	13.18%	-0.23%

High correlation with environmental and social-political indicators makes that this index correlates rather poorly with environmental indicators. This poor correlation is most pronounced in the Ecological Footprint (EF) indicator; the associated correlation coefficient is –73%, which is the worst correlation in the table. This falls in line with our earlier point that high correlation with constituent ISDs should not be desirable *per se*, because it necessarily implies poor correlation with other ISDs.

We find more moderate correlation patterns for the DEA index and the MISD index. When inter-comparing these two 'more balanced' measures, we find that the correlation coefficient associated with the MISD is above that corresponding to the DEA index in 9 out of the 14 cases; the difference is even quite pronounced in many cases.

In our opinion, these results give an argument in favor of the MISD as a 'well-balanced' measure of overall SD, when compared to the extreme Average and DEA indices. This argument is further strengthened by identifying, for each ISD, the overall SD measure with the lowest correlation: this is the MISD only in the case of the EWI and the GEM, while it is the Average index in three cases (footprint, ESI2 and ESI5) and the DEA index in the remaining 9 cases (HDI, HPI-1, HWI, ESI1, GDI, HALE, ESI3 and ESI4).

The general conclusion of our empirical assessment is favorable for the MISD: it has satisfactory discriminatory power, and it has a relatively well-balanced pattern of correlation with the constituent ISDs. These findings are all the more attractive in view of the intuitively appealing methodological starting points that underlie the MISD, i.e. benefit-of-the-doubt-weighting (by letting the data speak for themselves) complemented by generally acceptable weight restrictions.

6. CONCLUDING DISCUSSION

Sustainable Development is a complex, multi-dimensional concept. The need for quantifying SD is widely accepted, but the vague definition of SD leaves room for different interpretations and hence for different SD indices. In this paper we approached the quantification problem from the perspective of comparative evaluation and benchmarking. We proposed to synthesize the information of various existing indices of SD and its sub-components in a *meta*-index, which is constructed as an unequally weighted average of the underlying ISDs. The methodology builds on an intuitive benefit of the doubt weighting principle, which allows the countries to accord a higher weight to those SD dimensions in which they perform relatively well.

We should firmly stress that the proposed MISD is essentially *comparative* in its nature. In particular, we cannot directly infer whether a particular country is on the sustainable development path or not; we can only assess overall SD in comparison to other countries. Hence, the comparative indices alone are clearly insufficient for thorough SD monitoring. Still, we strongly believe that they can be particularly useful for identifying and promoting sustainable policies and practices.

In particular, the MISD can be employed for *benchmarking* purposes. Indeed, even if all countries have much to improve, some countries are clearly ahead of others. By taking a highly aggregate perspective, we can identify a handful of countries exemplifying the best practices and policies towards SD, not just in one but in many (if not all) aspects of SD. In this respect, the MISD rankings can stimulate governments and policy-makers both in the top-ranked and the low-ranked countries to engage into active dialogue in order to share knowledge and experiences, and so to speed up the diffusion of sustainable policies and practices from the leading countries of SD towards the less developed nations. We do not claim that all countries should mimic and replicate *every* policy of, e.g., Sweden or Norway (i.e., 2 benchmark countries), especially since we have limited insight into the driving factors behind our MISD results. Still, given our large coverage of existing SD indices, we are confident that the current policy practice in Sweden and Norway can provide useful inspiration for identifying sustainable (economic, social/political or environmental) policies, which could be applied in other countries as well. Essentially, we plead for country-level implementation of the benchmarking approach, which has already become an influential business paradigm and has also seen extensive application in the public sphere at the local level administration.

Still, many challenges remain ahead. First, although we have adopted a highly aggregate meta-level approach in the current paper, the presented technique can equally well be employed for assessing policy performance at a more detailed level; e.g., by focusing solely on environmental sustainability. Indeed, in the area of productivity and efficiency analysis, the DEA method that underlies our MISD is most often applied for performance evaluations at the micro level.

Second, the dynamic nature of SD calls for indices that measure changes in sustainability over time. In this respect, our approach readily extends towards dynamic SD indices of the Malmquist-type; compare with Färe et al. (1994) and Färe and Grosskopf (1996). Such dynamic analysis obviously requires panel data for the SD indicators, which are presently not yet available. This pleads for continuing the initiated efforts for developing SD indicators so that it will become possible in the years to come to monitor SD development over time.

Third, the quality and coverage of data remains an ongoing concern. For example, in our analysis we have been confronted with missing data for many countries and SD indicators. Although it can be expected that better data will become available as research on the specific SD topics continues, improving robustness of the SD indices to missing data deserves further research.

A final avenue for further research relates to our methodology. In particular, while we have opted for minimally restricting the flexibility of the revealed policy weights in the current paper, it seems worthwhile to further devote research efforts to identifying more stringent weight bounds, e.g., reflecting expert opinion consensus. We think that such information could proof most valuable at the level of the individual SD indices, and -to a somewhat lesser extent- at the level of SD categories (economic, environmental and social-political) and countries.

APPENDIX

The programming problem presented in Section 2 is non-linear. In order to turn it into a more standard linear programming problem, we need to adjust the different weight bounds, which are originally expressed in ratio form. A linear programming version of our optimization problem is the following:

$$
\begin{array}{ll}
\max\limits_{w_{ij}} \sum\limits_{j=1}^{n}\left(\sum\limits_{i=1}^{m} y_{ij} w_{ij} \right); & \\
\text{s.t.}\, 1-\sum\limits_{i=1}^{m} y_{ij} w_{ik} \geq 0 & \forall j,k=1,\ldots,n; \\
\left(\alpha w_{ij}-w_{hj}\right)\cdot y_{ij}\cdot y_{hj} \geq 0 & \forall h,i=1,\ldots,m\,;\, j=1,\ldots,n; \\
\gamma \sum\limits_{i\in S_l} w_{ij} - \sum\limits_{i\in S_k} w_{ij} \geq 0 & \forall k,l=1,2,3;\ i=1,\ldots,m\,;\, j=1,\ldots,n; \\
\left(\beta w_{ik}-w_{ij}\right)\cdot y_{ij}\cdot y_{ik} \geq 0 & \forall i=1,\ldots,m\,;\, j,k=1,\ldots,n;
\end{array}
$$

This formulation also incorporates the proposed technical solution for dealing with zero data entries in the weight restrictions (see Section 4).

REFERENCES

Auluck, R. (2002): Benchmarking: A tool for facilitating organizational learning? *Public Administration and Development* 22(2), 109-122.

Carrington, R., T. Coelli, and E. Groom (2002): International Benchmarking for Monopoly Price Regulation: The Case of Australian Gas Distribution, *Journal of Regulatory Economics* 21(2), 191-216.

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

Cherchye, L. (2001): Using Data Envelopment Analysis to Assess Macro-Economic Policy Performance, *Applied Economics* 33, 407-416.

Cox, A., and I. Thompson (1998): On the appropriateness of benchmarking, *Journal of General Management* 23(4), 1-20.

Dantzig, G.B. (1949): Programming of Interdependent Activities, Part II: Mathematical Model, *Econometrica* 17, 200-211.

Färe, R., and S. Grosskopf (1996): *Intertemporal Production Frontiers: With Dynamic DEA*, Kluwer Academic Publishers, Boston.

Färe, R., Grosskopf, S., Norris, M. and Z. Zhang (1994): Productivity Growth, Technical Progress and Efficiency Change in Industrialised Countries, *American Economic Review* 84, 66-83.

Gerlagh, R., R. Dellink, M. Hofkes, and H. Verbruggen (2002): A Measure of Sustainable national Income for The Netherlands, *Ecological Economics* 41, 157-174.

Hardin, G. (1968), The Tragedy of the Commons, *Science* 162, 1243-1248.

Koopmans, T.C. (1951): Analysis of Production as an Efficient Combination of Activities, in: T.C. Koopmans, ed., *Activity Analysis of Production and Allocation: Proceedings of a Conference* (Yale University Press, New Haven).

Kuosmanen, T. (2002): Modeling Blank Entries in Data Envelopment Analysis, EconWPA working paper. Available online at: http://econwpa.wustl.edu:8089/eps/em/papers/0210/0210001.pdf

Lélé, S.M. (1991): Sustainable Development: A Critical Review, *World Development* 19, 607-621.

Ludwig, D., R. Hilborn and C. Waters (1993): Uncertainty, Resource Exploitation, and Conservation: Lessons from History, *Science* 260, 36.

Mahlberg, B, and M. Obersteiner (2001): Remeasuring the HDI by Data Envelopment Analysis, International Institute for Applied Systems Analysis (IIASA), Interim Report IR-01-069, Laxemburg, Austria.

Melyn, W. and W. Moesen (1991): Towards a Synthetic Indicator of Macroeconomic Performance: Unequal Weighting when Limited Information is Available, Public Economics Research Paper 17, Center for Economic Studies, Leuven.

Munasinghe, M. (1993), Environmental Economics and Sustainable Development, World Bank Environment Paper 3, The World Bank, Washington, D.C..

Nordhaus, W.D., and J. Tobin (1972): *Economic Growth: Is Growth Obsolete?* Columbia University Press, New York.

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

Prescott-Allen, R. (2001) *The Wellbeing of Nations*: *A Country-by-Country Index of Quality of Life and the Environment*, Island Press.

Samuelson, P.A. (1948): Consumption Theory in Terms of Revealed Preference, *Economica* 15, 243-253.

Stanley, T.D. (2001): Wheat from Chaff: Meta-Analysis as Quantitative Literature Review, *Journal of Economic Perspectives* 15, 131-150.

UNDP (2001): *Human Development Report 2001: Making new technologies work for human development*, Oxford University Press, Oxford, New York.

UN Economic and Social Council (1999a): Integration of the Human Rights of Women and the Gender Perspective: Violence against Women, Report of the Special Rapporteur E/CN.4/1999/68. Available online at http://www.hri.ca/fortherecord1999/documentation/commission/e-cn4-1999-68.htm

UN Economic and Social Council (1999b): Adverse Effects of the Illicit Movement and Dumping of Toxic and Dangerous Products and Wastes on the Enjoyment of Human Rights, Report of the Special Rapporteur E/CN.4/1999/46/Add.1. Available online at http://www.hri.ca/fortherecord1999/documentation/commission/e-cn4-1999-46-add1.htm

Wackernagel, M., *et al.* (2002): Tracking the ecological overshoot of the human economy, *Proceedings of the National Academy of Science* 99(14), 9266-9271.

WEF (2002): *2002 Environmental Sustainability Index; Main report*, Global Leaders of Tomorrow Environmental Task Force, World Economic Forum; available on-line at http://www.ciesin.columbia.edu/indicators/ESI.

WHO (2001): *World Health Report: Mental Health: New Understanding, New Hope*, France; available on-line at http://www.who.int/whr/2001/main/en/pdf/whr2001.en.pdf

World Commission on Environment and Development (1987): *Our Common Future*, Oxford University Press, Oxford.

WWF (2000): Living Planet Report 2000, Gland, Switzerland.

Zaim, O., Färe, R. and S. Grosskopf (2001): An Economic Approach to Achievement and Improvement Indexes, *Social Indicators Research* 56, 91-118.