Institute for International Economic Policy Working Paper Series
Elliott School of International Affairs
The George Washington University

Composite Indices: Rank Robustness, Statistical Association and Redundancy

IIEP-WP-2011-19

James Foster
George Washington University

Mark McGillivray
Deakin University and University of Oxford

Suman Seth
University of Oxford

April 2011

Institute for International Economic Policy
1957 E St. NW, Suite 502
Voice: (202) 994-5320
Fax: (202) 994-5477
Email: **iiep@gwu.edu**
Web: **www.gwu.edu/~iiep**

Composite Indices: Rank Robustness, Statistical Association and Redundancy

James E. Foster,
George Washington University and *University of Oxford*[†]

Mark McGillivray
Deakin University and University of Oxford[‡]

Suman Seth
University of Oxford[§]

Abstract

This paper evaluates the robustness of rankings obtained from composite indices that combine information from two or more components via a weighted sum. It examines the empirical prevalence of robust comparisons using the method proposed by Foster, McGillivray and Seth (2010). Indices examined are the Human Development Index, the Index of Economic Freedom and the Environmental Performance Index. Key theoretical results demonstrate links between the prevalence of robust comparisons, Kendall's tau rank correlation coefficient, and statistical association across components. Implications for redundancy among index components are also examined.

JEL Classifications: I31, O12, O15, C02.

Key Words: composite index, multidimensional index, Human Development Index, rank robustness, positive association, Kendall's tau, redundancy, prevalence function.

Acknowledgements: This paper has benefitted from discussions with Sabina Alkire and from the useful comments of two anonymous referees and the Associate Editor of this journal. The authors are grateful for these comments. We thank to Alison Kennedy of the UNDP for making the HDI data available for our use. The usual disclaimer applies.

Correspondence to James E. Foster at: fosterje@gwu.edu.

[†] Professor of Economics and International Affairs, George Washington University, 1957 E Street, NW Suite 502, Washington, D.C. 20052, +1 202 994 8195, fosterje@gwu.edu, and Research Associate, Oxford Poverty and Human Development Initiative, Department of International Development, 3 Mansfield Road, Oxford OX4 1SD, UK +44 1865 271915.

[‡] Research Professor of International Development, Alfred Deakin Research Institute, Deakin University, Geelong 3217, Australia, +613 5527 8011. mark.mcGillivray@deakin.edu.au, and Research Associate, Oxford Poverty and Human Development Initiative (OPHI), Department of International Development, 3 Mansfield Road, Oxford OX4 1SD, UK +44 1865 271915..

[§] Research Officer, Oxford Poverty and Human Development Initiative (OPHI), Department of International Development, 3 Mansfield Road, Oxford OX4 1SD, UK +44 1865 271915, suman.seth@qeh.ox.ac.uk.

I. Introduction

It is common for indices based on multiple dimensions to take the form of a simple weighted sum of component indicators. Indices of this type are found in many different contexts, including the assessment of levels of human development, globalization, environmental sustainability, and freedom, and their associated rankings often receive enormous attention.[1] This is particularly true of human development or well-being indices, which are frequently cited by some national governments as evidence of the success of their policies and criticized or stubbornly ignored by others. A well-known index of this type is the traditional Human Development Index (HDI), values of which have been published annually by the United Nations Development Program since 1990 (UNDP, 1990-2008).[2]

Despite their usefulness, composite indices are subject to a number of limitations, and questions can arise about the reliability of the orderings they provide.[3] One central issue concerns the choice of weights, which embody the relative importance attached to each component in the index. Weights can be set using a variety of approaches, including normative judgments, statistical methods, and rules of thumb.[4] A careful examination reveals, however, that there is usually a multiplicity of weighs consistent with the underlying principles or methods employed, and that the final selection is to some extent arbitrary. Many indices simply employ equal weights, and hence are arithmetic means of the achievements or components on which they are based. The HDI is one example of an index having this weighting scheme. The selection of equal weights is often justified as a least imperfect option, one that is likely to attract the least criticism, or a useful initial position when information on the relative importance of components is lacking.[5] While these arguments have some practical merit, they may not carry enough force to rule out alternative weighting structures that place somewhat higher relative weights on certain components. And if other plausible weighting structures are found to yield different orderings, it would follow that the original orderings are not unambiguous or robust.

---

[1] In general, a composite index is formed when several indicators are aggregated into a single index; see Nardo *et al.* (2005) for example of composite indices. The aggregation across indicators need not be linear, but we focus on the most common type of composite index in the present paper. The term composite indicator is often used instead of composite index. We have opted for the latter term for two reasons. First, in multidimensional analysis, the term indicator is used to denote a component of a dimension (See, UNDP, 2010, p. 215). Secondly, the well-known composite indices that are subject of our discussion use the term 'index' instead of 'indicator', such as the Human Development Index, the Environmental Performance Index, the Global Peace Index.

[2] The latest Human Development Report (HDR 2010) presents a new HDI based upon the geometric mean rather than a simple average of its three indicators. Our approach can be applied to this new index as well as any other indices that are monotonic transformations of composite indices, such as the Human Poverty Index (HPI), the Gender-related Development Index (GDI), and the Inequality-adjusted Human Development Index (IHDI) – all developed by the UNDP. See Foster et al. (2009) and the discussion below on data transformations.

[3] See Saisana *et al.* (2005), Permanyer (2009), Cherchye *et al.* (2008), and Foster *et al.* (2009).

[4] See Decancq and Lugo (2008) and Nardo *et al.* (2005) for a discussion on various approaches for setting weights.

[5] See, for example, Esty *et al.* (2005: 66).

Formatted: French (France)

This applies not just to equal weights, but to any situation in which doubts remain about the relative importance of each component. For example, weights based on statistical methods allow arbitrariness to enter via the choice of method (and data) used to derive the weight.

The possibility of arbitrary weights suggests a need for evaluating the robustness of comparisons generated by composite indices, an issue that has recently been addressed in a number of studies. Nardo *et al.* (2005) and Saisana *et al.* (2005), for example, emphasize that there are many sources of uncertainty, including the choice of weights, that together lead to a distribution of values around the original composite value. They show how this distribution might be estimated using Monte Carlo methods and use a similar technique to evaluate the extent to which alternative orderings deviate from the original ordering. McGillivray and Noorbakhsh (2007) evaluate the effect of changing weights on the HDI by calculating rank correlations between the original HDI country ranks and the ranks found using alternative weights. Cherchye et al. (2008) consider a simultaneous change in the weights, the normalizations of dimensional variables, and the functional form of the composite index, and find conditions under which an original comparison will be preserved. Their use of alternative aggregation procedures goes beyond the question considered here and results in somewhat stronger but less applicable criteria.[6]

Foster, McGillivray and Seth (2010) focus purely on changing weights for composite indices, and differentiate between the case where a ranking is *reversed* by plausible changes to the initial vector of weights from the case where a ranking is *robust* to all such changes.[7] The key to their approach is a set of allowable weighting vectors about the original weighting vector, whose shape is motivated by the epsilon-contamination model from decision theory. They provide necessary and sufficient conditions under which a given comparison is robust to an allowable change in weights, and then show how the size of the largest set allowing robust comparisons yields a natural measure of the robustness of a given comparison. An application to HDI data suggests that the approach can be helpful in evaluating the robustness of comparisons involving composite indices.

The present paper explores in greater depth the empirical usefulness of this approach. It begins by analyzing the prevalence of robustness for several well known composite indices and their respective datasets. We note that comparisons from certain datasets are much more likely

[6] They also rely upon a rather specific form for normalizing the dimensional variables, which lessens the generality of their results.

[7] See also our working paper Foster *et al.* (2009), which includes results from both the present paper and Foster *et al.* (2010).

to be robust than comparisons from others, and explore several characteristics of a dataset that could plausibly be linked to the prevalence of robust comparisons. Of particular interest is the relationship between robustness and the statistical association between component variables. We provide theorems that point to a close relationship between the prevalence of full robustness and the Kendall Tau measure of rank correlation (Kendall and Gibbons, 1990), while a third establishes a link with "association increasing rearrangements" (Boland and Proschan, 1988). These results shed new light on the role of inter-dimensional association in multidimensional measurement. Previous research argued that high correlations or association between component variables are undesirable as they are indicative of statistical redundancy, which occurs when any one component provides largely the same ranking as the index as a whole (McGillivray, 1991 and McGillivray and White, 1993). The current paper finds that from a rank robustness perspective high association between components is a positive rather than negative attribute of composite indices.

The paper proceeds as follows. Section II presents the robustness approach of Foster et al. (2010) along with the needed notation and definitions. Section III examines the prevalence of robustness for three well-known composite indices. Section IV provides several theorems on the prevalence of robustness and, in particular, investigates how the statistical association between components can affect robustness. Section V considers the link between redundancy of dimensions and the rank robustness of the associated composite indexs. Section VI concludes.

II. Rank Robustness

We begin with some notation and definitions. Let $D \geq 2$ denote the number of dimensions under consideration, and suppose that a and b are two D-dimensional vectors. The *vector dominance* relations are defined as follows: $a \geq b$ means that $a_d \geq b_d$ for all $d = 1, \ldots, D$; the expression $a > b$ indicates that $a \geq b$ and $a \neq b$; whereas $a >> b$ means that $a_d > b_d$ for all $d = 1, \ldots, D$. The *least upper bound* of a and b, denoted by $a \vee b$, is the vector having $\max\{a_d, b_d\}$ as its d^{th} coordinate; the *greatest lower bound* of a and b, denoted by $a \wedge b$, is the vector having $\min\{a_d, b_d\}$ as its d^{th} coordinate.

Let $X \subset \clubsuit^D$ be the set of all possible *achievement vectors* under consideration and let $\Delta = \{w \in \clubsuit^D: w \geq 0 \text{ and } w_1+\cdots+w_D = 1\}$ denote the simplex of associated *weighting structures* or *vectors*. The d^{th} coordinate of $x \in X$ gives the achievement level in dimension d; the respective coordinate of $w \in \Delta$ gives the weight attached to the d^{th} dimension. A *composite index* C: $X \times \Delta \rightarrow \clubsuit$ is a linear function $C(x;w) = w \cdot x = w_1x_1 + \cdots + w_Dx_D$, which aggregates the dimensional

achievements in $x \in X$ using the weights in $w \in \Delta$. We assume that an *initial weighting vector* $w^0 \in \Delta$ satisfying $w^0 >> 0$ has already been selected; this fixes the *specific* composite index C_0: $X \to \clubsuit$ defined as $C_0(x) = C(x;w^0)$ for all $x \in X$. The HDI, for example, can be viewed as a composite index C_0 over $D = 3$ human development achievement levels $x = (x_1, x_2, x_3)$ of health, education, and income in a country, where weights are given by $w^0 = (1/3, 1/3, 1/3)$, or the equal weighting structure. In words, the HDI gauges a country's level of development according to a simple arithmetic mean of variables that measure "a long and healthy life", "knowledge" and "a decent standard of living" (UNDP, 2009, p. 208).[8] The associated strict ordering of achievement vectors in X will be denoted by $\boldsymbol{C_0}$, so that $x \, \boldsymbol{C_0} \, y$ holds if and only if $C_0(x) > C_0(y)$.

The conclusion $x \, \boldsymbol{C_0} \, y$ indicates that x has a higher composite value than y at the initial weighting vector, but does not ensure that this ranking will be preserved at other plausible vectors. For example, in the 2004 HDI data, we see that Ireland's value exceeds Canada's for $w^0 = (1/3, 1/3, 1/3)$, but the ranking reverses if the weights are changed to $w = (1/2, 1/4, 1/4)$. In contrast, Australia's HDI value exceeds Sweden's by the same margin, and yet the ranking is never reversed at any other weighting vector in Δ. The robustness analysis of Foster et al. (2010) is designed to address this issue – to discern the relative robustness of a given comparison and to derive an intuitive measure of robustness. We now outline some of its key methods and findings.

Following the lead of Bewley's (2002) approach to Knightian uncertainty, C(x;w) is compared to C(y;w) not just for $w = w^0$, but for all w in a *set* of plausible weighting vectors about w^0. Consider the set $\Delta_\varepsilon = (1-\varepsilon)\{w^0\} + \varepsilon\Delta$ for $\varepsilon \in [0, 1]$, and note that $\Delta_0 = \{w^0\}$ and $\Delta_1 = \Delta$, while for $\varepsilon \in (0, 1)$ the set Δ_ε is a scaled down simplex between these two extremes. The construction of this set is motivated by the *epsilon-contamination model* of ambiguity, where ε measures the lack of confidence one has in the initial weighting vector, and the extent to which other vectors are considered.[9] When $\varepsilon = 0$, there is complete confidence in w^0 and no other weighting vectors are considered; when $\varepsilon = 1$, there is no confidence in w^0 and every other vector in Δ is considered. In between, the weighting vectors in Δ_ε are considered, where the size of Δ_ε is increasing in ε. For any given $\varepsilon \in [0, 1]$, the associated robustness condition is given by $C_0(x) > C_0(y)$ and $C(x;w) \geq C(y;w)$ for all w in Δ_ε, and is denoted by $x \, \boldsymbol{C_\varepsilon} \, y$. When the condition holds, it indicates that the comparison is robust given the set Δ_ε and the level of confidence it represents. There is also a straightforward method for checking when $x \, \boldsymbol{C_\varepsilon} \, y$ holds, namely $x^\varepsilon > y^\varepsilon$ where $x^\varepsilon =$

[8] Additional details of the HDI are provided below and in UNDP (2009).
[9] See for example, Nishimura and Ozaki (2006).

$(1-\varepsilon)(C_0(x), \ldots, C_0(x)) + \varepsilon x$ and $y^{\varepsilon} = (1-\varepsilon)((C_0(y), \ldots, C_0(y)) + \varepsilon y$. Note that $x \; \boldsymbol{C_0} \; y$ reduces to $C_0(x) > C_0(y)$, as before, while the *full robustness* condition $x \; \boldsymbol{C_1} \; y$ is equivalent to vector dominance $x > y$.

As ε rises from 0 to 1, the set Δ_ε expands from $\{w^0\}$ to Δ and the condition $x \; \boldsymbol{C_\varepsilon} \; y$ becomes more stringent. For any ranking $x \; \boldsymbol{C_0} \; y$ let r be the maximum value of ε for which $x \; \boldsymbol{C_\varepsilon}$ y holds. This is a natural measure of robustness for $x \; \boldsymbol{C_0} \; y$ that corresponds to the largest set Δ_ε (and the lowest possible level of confidence ε) for which there will be no reversals of the ranking. Alternatively, let $A = C_0(x) - C_0(y) > 0$ be the difference in composite values of x and y using the initial weighting vector w^0. In the context of the HDI, this is analogous to the difference in HDI values for two countries, and represents the margin by which x dominates y according to C_0. Let $B = \max_{w \in \Delta} [C(y;w) - C(x;w), 0]$ be the *maximum contrary difference* between composite values as w ranges across Δ. This represents the maximal margin by which y could dominate x according to C if weights were allowed to vary freely. The measure of robustness can be equivalently defined as $r = A/(A + B)$. Intuitively, when $B = 0$ so that full robustness $x \; \boldsymbol{C_1} \; y$ holds, then $r = 1$; when B becomes large relative to A, then the measure of robustness r falls towards 0. The maximum possible contrary difference B is also the maximum coordinate-wise difference between y and x, and hence r is straightforward to calculate. For example, Ireland's achievements in health, education, and income in 2004 are given by $x = (0.882, 0.990, 0.995)$ while the respective achievement vector for Canada is $y = (0.919, 0.970, 0.959)$. Ireland's HDI exceeds that of Canada by $A = (0.956 - 0.950) = 0.006$, while the maximum contrary difference is $B = (0.882 - 0.919) = -0.037$. The measure of robustness for this comparison is $r = 0.006/(0.006 + 0.037) = 0.139$. Australia's HDI likewise exceeds that of Sweden by 0.006, but in this case there is vector dominance of achievement vectors which leads to $B = 0$ and hence $r = 1$.

The robustness measure r can be used in a number of ways. First, it can be applied to a specific composite index to evaluate the robustness level of a *given* ranking (say, rejecting its conclusion if robustness is too low) or, alternatively, to determine the relative levels of robustness of *several* comparisons. Second, it can be used across different composite indices to compare their aggregate robustness properties. Foster, McGillivray, and Seth (2010) focuses on the first type of application; this paper explores the second. Of particular interest is the distribution (or prevalence) of robustness values for a given composite index (and dataset). In practice, are certain composite indices more robust than others and, if so, why? We now turn to the definitions and the concepts needed to address this type of question.

III. Prevalence of Robustness

The practical implementation of a composite index requires both an initial weighting vector w^0 and appropriate data. Let $\hat{X} \in \clubsuit^{ND}$ denote the *dataset* of achievement vectors, where N is the total number of objects (such as countries) being evaluated and, as before, D is the number of dimensions. For each $i = 1, \ldots, N$, the i^{th} row of the dataset is the vector of achievements associated with object i; it can be denoted by $\hat{x}^i \in \clubsuit^D$. For example, the dataset for the 2004 HDI dataset has $N = 177$ countries and $D = 3$ dimensions, where each row vector $\hat{x}^i$ contains country i's achievements in the dimensions of health, education, and income (UNDP, 2006). Without loss of generality, we assume that no two objects have identical composite index levels, and reorder them such that $C_0(\hat{x}^1) > C_0(\hat{x}^2) > \ldots > C_0(\hat{x}^N)$ or equivalently $\hat{x}^i \; \boldsymbol{C_0} \; \hat{x}^j$ for all $i < j$. This assumption is satisfied for each of the composite indices considered in this paper.

Consider any pair $\hat{x}^i$ and $\hat{x}^j$ for which the ranking $\hat{x}^i \; \boldsymbol{C_0} \; \hat{x}^j$ holds. Note that there are $k = N(N-1)/2$ many pairs (i,j) such that $1 \le i < j \le N$, and hence k many initial rankings $\hat{x}^i \; \boldsymbol{C_0} \; \hat{x}^j$. Let r_{ij} denote the level of robustness of $\hat{x}^i \; \boldsymbol{C_0} \; \hat{x}^j$, and let $P = [r_{ij}]_{i<j}$ be the associated *robustness profile*, which lists the robustness levels for all k comparisons. The *prevalence function* p:[0,1] → [0,1] associates with each robustness value $r \in [0,1]$ the share $p(r) \in [0,1]$ of the k comparisons whose robustness levels are at least r. Equivalently, $p(r)$ is the share of the comparisons for which the robustness ordering $\boldsymbol{C_r}$ holds, given dataset $\hat{X}$ and initial weighting vector w^0. Now suppose that $q(r)$ is the prevalence function for some other dataset $\hat{Y}$ and initial weighting vector u^0. We say that $\hat{X}$ *has greater robustness than* $\hat{Y}$ if $p(r) \ge q(r)$ for all $r \in [0,1]$, with $p(r) > q(r)$ for some $r \in [0,1]$. In words, for every possible level of robustness r, the share of comparisons having robustness levels of r or more is as high in $\hat{X}$ as in $\hat{Y}$, and for some r it is higher. The two are said to *have the same robustness* if their prevalence functions are the same. It is also possible that neither $\hat{X}$ nor $\hat{Y}$ will have greater (or equal) robustness across all robustness values; instead, one prevalence can be higher for some range of values, while the second is higher for another.

We now examine the prevalence functions for several datasets having countries as the objects of analysis. The first is the HDI for the years 1998 and 2004, as obtained from the UNDP (2000, 2006). As mentioned above, the HDI contains three components, capturing national

achievements in health, education, and per capita income, respectively, and uses the arithmetic (or equal weighted) mean as its composite index C_0.[10] Each component has been normalized to range between zero and one, and hence the HDI takes values in the same range. Each year's HDI generates a ranking of 177 countries. The second composite index is the Index of Economic Freedom (IEF) obtained from the Heritage Foundation (2008). The IEF is based on achievements in ten dimensions relevant to economic freedom.[11] Each component index has been normalized to range between zero and one hundred, and the IEF is formed by taking the arithmetic mean of the ten dimensions as its composite index C_0. We examine IEF for 2007, which ranks 157 countries. The third composite index is the Environmental Performance Index (EPI). The EPI is based on 25 component indices. A number of versions of the EPI exist, each differentiated by the level of aggregation of the components. We examine four versions: EPI2, EPI6, EPI8 and EPI10. EPI2 is based on two equally weighted summary measures of *environmental health* and *ecosystem vitality*, respectively. EPI6, EPI8 and EPI10 are based on a mix of summary and individual indices of environmental health, air pollution, the impact of water, biodiversity and habitat, productive natural resources and climate, and are obtained by aggregating six, eight and ten of these component indices, respectively.[12] Full descriptions of the EPI can be found in Esty *et al.* (2008). The EPIs considered here rank 149 countries for the year 2007.

Prevalence functions for these composite indices are shown in Figure 1 with $p(r)$ presented in percentage terms. Each function is downward-sloping, reflecting the fact that as r rises, the number of comparisons that can be made by $\boldsymbol{C_r}$ is lower (or no higher). As r falls to zero, all functions achieve the 100% comparability arising from $\boldsymbol{C_0}$; in the other direction, the value of $p(r)$ at $r = 1$ is the percentage of the comparisons involving vector dominance, and hence are fully robust. There is, interestingly, a wide variation in $p(1)$ across each composite index under consideration. It is clearly highest for the HDI, with $p(1)$ being 69.8 percent for the 2004 dataset and 71.5 percent for 1998. Put differently, 69.8 percent and 71.5 percent of pair-

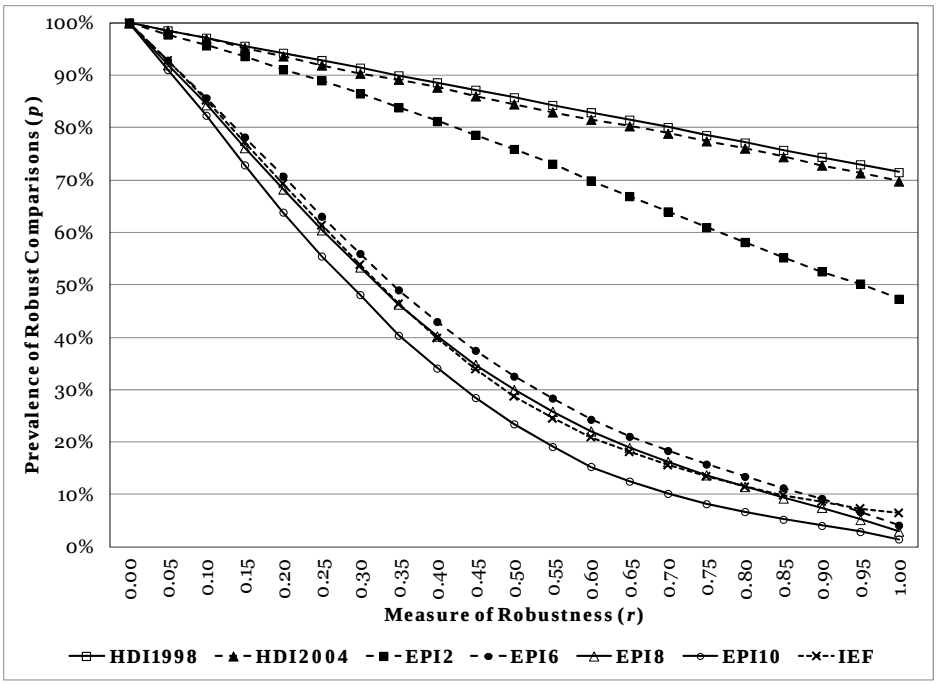
[10] National achievement in per capita income is measured by PPP GDP per capita, that in health is measured by years of life expectancy, and in education by a weighted average of adult literacy and mean years of schooling (UNDP, 2006).

[11] The ten dimensions are government size, property rights, freedom from corruption and freedoms with respect to business, trade, fiscal, monetary, investment, financial and labor.

[12] For the six-component EPI, we consider the five sub-components of the ecosystem vitality component: air pollution, water (ecosystem), production of natural resources, biodiversity and habitat, and climate change. The initial weighting structure in this case is (0.5, 0.025, 0.075, 0.075, 0.075, 0.25). For the eight-component FPI, we further consider three sub-components of the environmental health component: environment burden of diseases, water (health), and pollution. The initial weighting structure in this case is (0.25, 0.125, 0.125, 0.025, 0.075, 0.075, 0.075, 0.25). Finally, for the ten-component EPI is obtained by dividing the biodiversity and habitat sub-component into three further sub-sub-components: forestry, fishery, and agriculture. The initial weighting structure in this case is: (0.25, 0.125, 0.125, 0.025, 0.075, 0.075, 0.025, 0.025, 0.025, 0.25).

wise HDI comparisons are fully robust in 2004 and 1998, respectively.[13] The value of $p(1)$ for EPI2 rankings is 47.3%. It is much lower for the remaining indices, with 4.2%, 3.0%, 1.5% and 6.5% being the $p(1)$ values for EPI6, EPI8, EPI10 and IEF, respectively.

Figure 1: Prevalence Functions for Several Composite Indices



For all r between zero and one, it is clear from Figure 1 that the prevalence of robustness is greater for the HDI than for the EPI and the EFI. The 1998 HDI prevalence function is also higher than the 2004 HDI prevalence function. The EPI10 exhibits the lowest prevalence of robust comparisons. An additional feature of Figure 1 is that the shapes of the $p(r)$ functions are quite different, with those associated with the HDI being essentially linear while others exhibit pronounced curvatures.

IV. Prevalence, Transformation, and Statistical Association

The above examples suggest a number of factors that might affect robustness, including the number of variables, the particular normalization or scale of the variables, and the correlation (or association) among variables. This section provides a theoretical exploration of

[13] The 1998 figures found Foster *et al.* (2009) are slightly different, since they use the rounded values of country achievements as published in the Human Development Reports. The current paper and Foster et al (2010) use more refined figures obtained directly from the Human Development Report Office.

some of these factors. We identify certain basic transformation of data that leave the prevalence functions fixed; the resulting analysis sheds light on the potential roles of the number of variables and their normalizations. We then consider changes that increase the association among variables, and show how they lead to increases in the prevalence of fully robust comparisons, although not necessarily in the overall prevalence function.

Fixed Robustness and Transformations

Our first transformations yield pairs of datasets that have similar robustness properties. A *monotonically increasing transformation* of X is a function $f : X \to \clubsuit^D$ that can be written as $f(x) = (f_1(x_1),..., f_D(x_D))$ where each function $f_d(x_d)$ is monotonically increasing; a *common-slope affine transformation* of X has the additional property that each function $f_d(x_d)$ can be written as $f_d(x_d) = \alpha x_d + \beta_d$ for some $\alpha > 0$ and β_d in $\clubsuit$. Denoting the n^{th} row of datasets $\hat{X}$ and $\hat{Y}$ by $\hat{x}^n$ and $\hat{y}^n$, respectively, we say that $\hat{Y}$ *is obtained from* $\hat{X}$ *by a common-slope affine transformation* (respectively, *by a monotonically increasing transformation*) if $\hat{y}^n = f(\hat{x}^n)$ for all $n = 1,..., N$ for some transformation f having the appropriate property.

Applying a monotonically increasing transformation to a dataset preserves the orderings of achievements within each dimension, but can disrupt the weighted averages across dimensions. In particular, it is possible that $C_0(x') > C_0(x)$ and $C_0(y') < C_0(y)$ where y' and y are transformations of x' and x, respectively, which implies that the robustness profiles of $\hat{X}$ and $\hat{Y}$ can be rather different for the same w^0. On the other hand, if we restrict consideration to common-slope affine transformations, we see that $C(y;w) = w \cdot y = \alpha w \cdot x + w \cdot \beta$ where $\beta = (\beta_1, ..., \beta_D)$, and hence $C(x';w) \geq C(x;w)$ if and only if $C(y';w) \geq C(y;w)$, where y' and y are the respective transformations of x' and x. In this case, $\hat{X}$ and $\hat{Y}$ have the same robustness profile and hence the same prevalence function $p(r)$ given w^0. So, for example, if every dimension is scaled up or down in the same proportion, this will leave $p(r)$ unchanged, as will simply adding a different constant to each dimension. Multiplying each dimension by a *different* positive constant alters the implicit weighting across dimensions, potentially changing the rankings of transformed observations. Using an arbitrary monotonic increasing transformation likewise can alter rankings and lead to different prevalence functions for the transformed dataset. Note, though, that fully robust comparisons are preserved under a monotonic transformation, and hence the

prevalence $p(1)$ of full robustness does not change.[14] These results are summarized in the following theorem.

Theorem 1: Suppose that the initial weighting vector is fixed. If $\hat{Y}$ is obtained from $\hat{X}$ by a monotonically increasing transformation, then $\hat{Y}$ and $\hat{X}$ share the same prevalence value $p(1)$. If $\hat{Y}$ is obtained from $\hat{X}$ by a common-slope affine transformation, then they share the same prevalence function $p(r)$.

Proof: The proof is given in the appendix.

In the example of the HDI, the normalized health, education, and income variables used to construct index values are actually monotonic transformations of underlying variables involving a nonlinear function in the case of income, and affine transformations with different slopes across the three variables. Consequently, the specific shapes of the transformations can influence HDI comparisons as well as their measured robustness levels. However, as indicated in Theorem 1, these transformations do not influence fully robust comparisons and $p(1)$. If one restricted consideration to $\boldsymbol{C_1}$ comparisons, there would be no need to select the "right" transformations or even to transform variables at all: one could use the original health, education, and income variables directly.[15]

A second form of transformation replaces each variable in the achievement vector with one or more copies of that variable. A *replicating transformation* of X is a function $f : X \to \clubsuit^{D'}$ for some $D' > D$ such that $f(x) = (f_1(x_1),..., f_D(x_D))$, where each $f_d(x_d)$ is the k_d-fold *replication* $(x_d, x_d, ..., x_d)$ for some integer $k_d \geq 1$. Denoting the n^{th} row of datasets $\hat{X}$ and $\hat{Y}$ by $\hat{x}^n$ and $\hat{y}^n$, respectively, we say that $\hat{Y}$ *is obtained from* $\hat{X}$ *by a replicating transformation* if $\hat{y}^n = f(\hat{x}^n)$ for all $n = 1,..., N$ for some transformation f of this type. Transformed achievement vectors have higher dimension D' and, consequently, the associated weighting vectors must be adjusted to account for this. Now, which initial weighting vector u^0 for $\hat{Y}$ would correspond to the original w^0 for $\hat{X}$? One option is to divide the weight w_d^0 equally among the associated dimensions in u^0; however, it turns out that any allocation of the weight w_d^0 across its associated dimensions will

[14] The result on monotonic transformations would be true even if the initial weighting vectors were different. The role played by common-slope affine transformations is similar to assumptions used in social choice theory. See, for example, Blackorby, Donaldson, and Weymark (1984). While transformations have been addressed in the literature on composite indices (see, for example, McGillivray and Norrbakhsh, 2007), it is an important issue that is not sufficiently explored and as such is deserving of further research.

[15] The first part of Theorem 1 will generate same prevalence value $p(1)$ even if we use the dominance criterion proposed by Cherchye, Ooghe, and Puyenbroeck (2008). For discussion of transformations in the context of human development indices, see Alkire and Foster (2010).

do. We say u^o is *consistent* with w^o if, for each $d = 1,..., D$, the weight w_d^o on x_d is equal to the sum of the k_d entries in u^o associated with $f_d(x_d) = (x_d, x_d, ..., x_d)$. So for example, if $D = 2$ and f replicates each entry two times, then $w^o = (1/2, 1/2)$ is consistent with $u^o = (1/6, 1/3, 1/4, 1/4)$. We have the following result.

Theorem 2: If $\hat{Y}$ is obtained from $\hat{X}$ by a replicating transformation, and u^o is consistent with w^o, then $\hat{Y}$ and $\hat{X}$ have the same prevalence function $p(r)$.

Proof: The proof is given in the appendix.

In other words, according to the theorem, appending copies of one or more existing variables leaves the comparisons and the robustness properties of a dataset unaffected, as long as the effective weight on each variable is unchanged. As an example, consider what would happen if the education variable in an HDI dataset were replicated to obtain a *four* variable dataset. Using equal weights of ¼ for the four dimensional dataset would likely alter rankings since this would, in effect, increase the aggregate weight on education. However, if the total weight on the two education variables is maintained at 1/3, say where each variable receives a weight of 1/6, then all comparisons and robustness levels would be the same as before.

One implication of this is that the number of variables *per se* does not have an independent impact on a dataset's robustness. In contrast, the empirical evidence provided by Figure 1 might suggest that a greater number of variables is associated with lower robustness. The evidence is particularly striking for the three EPI examples, where the aggregation of variables, and hence the decrease in the number of variables, clearly leads to increased robustness – even though they use the same underlying data. Is this due to the decreased number of variables?

Let us examine how EPI6 is constructed from EPI10. The first and fifth variables in EPI6 are each obtained by combining three distinct variables in EPI10 (namely, variables 1 to 3 and 7 to 9), while the remaining variables are unchanged. Weights from the initial weighting vector u^o for EPI10 are used to construct each new variable in EPI6 as a weighted average of the source variables from EPI10, and the weight on the new variable is the sum of the corresponding weights in u^o. The new w^o is thus consistent with u^o. Now consider a ten variable replication of EPI6 that repeats variable 1 three times and variable 5 three times and let the initial weighting vector be u^o. By Theorem 2, this intermediate dataset has precisely the *same* robustness profile and prevalence function as EPI6. It is not the number of variables that is driving the observed decrease in robustness. Instead, its source is found in the transformation from the intermediate

dataset to EPI10, by which the perfectly correlated triplets are converted to variables that are less positively associated. The fall in robustness is due to disagreements among the new variables, rather than the higher number of variables *per se*. Association among variables is likely a key driver of robustness. We now turn to a discussion of the relationship between robustness and association.

Increased Robustness and Statistical Association

What factors generally lead to greater robustness? At an intuitive level, the possibility of fully robust comparisons is related to the degree of correlation or association among the dimensional variables. For example, if two of the achievements are perfectly negatively correlated, so that when one rises, the second falls, then it is impossible for vector dominance and hence $\boldsymbol{C}_1$ to hold. On the other hand, if there is complete positive association between all variables, so that when any variable rises, all rise, then every achievement vector is comparable by vector dominance, and $\boldsymbol{C}_1$ is universally applicable.[16] We saw in Figure 1 that both HDI datasets have high levels of robustness, and that the prevalence function is higher for 1998 than for 2004. Kendall's tau correlation coefficients for 2004 are 0.55 for health and education, 0.66 for health and income, and 0.58 for income and education, which indicates strong, positive association among variables; the respective values for 1998 are even higher, at 0.58, 0.69, and 0.59. Both intuition and empirical evidence suggest a link between inter-dimensional association and robustness. Is there a theoretical justification for such a link?

Suppose, for simplicity, the dataset $\hat{X}$ has the property that within each dimension all values are distinct.[17] Given any two dimensions c and d, let E_{cd} denote the number of *concordant* pairs of observations in which one of the two observations has higher values in both dimensions c and d; and let F_{cd} be the number of *discordant* pairs in which one observation is higher in one dimension and lower in the other. Then *Kendall's tau correlation coefficient* for dimensions c and d is defined as $\tau_{cd} = (E_{cd} - F_{cd})/\ k$ where $k = N(N - 1)/2$. Since there are no ties within dimensions, $F_{cd} = k - E_{cd}$, so that $\tau_{cd} = 2(E_{cd}/k) - 1$. Kendall's Tau may in this situation also be expressed as $\tau_{cd} = (E_{cd} - F_{cd})/(E_{cd} + F_{cd})$.[18]

Now consider the special case where there are only two variables, and so there is a single coefficient $\tau = \tau_{12}$ and number of concordant pairs $E = E_{12}$. In this special case, the number of

[16] Note that if there are more than two dimensions then it is impossible for all pairs of variables to be perfectly negatively correlated; in other words, there is no analogous notion of perfect negative association in higher dimensions.

[17] This simplifies the definition of Kendall's tau correlation coefficient. Note that this is different to our previous assumption that *composite* values are distinct.

[18] The expression $(E_{cd} - F_{cd})/(E_{cd} + F_{cd})$ is also known as Goodman and Kruskal's Gamma in the statistical literature.

concordant pairs is precisely the number of fully robust pairs, so the share of fully robust comparisons is $p(1) = E/k$. Therefore, $\tau = 2p(1) - 1$ and we have the following result.

Theorem 3: Suppose that $D = 2$ for a given dataset $\hat{X}$. Then the share $p(1)$ of fully robust comparisons is determined by Kendall's tau correlation coefficient τ according to the formula $p(1) = (\tau+1)/2$.

In the case of two variables, there is a direct relationship between $p(1)$ and the level of correlation as measured by Kendall's tau. Whenever $\tau = 1$ so that the variables have perfect positive correlation, we must have $p(1) = 1$. If $\tau = -1$, and perfect negative correlation obtains, then $p(1) = 0$. The case of independence ($\tau = 0$) implies $p(1) = ½$, so that half the comparisons are fully robust. For example, EPI_2 has $\tau = -0.053$, which via Theorem 3 yields $p(1) = 0.473$ as noted above.

Now consider the general case of $D \geq 2$. Full agreement across *all* dimensions entails concordance in any *two* dimensions, hence $p(1)$ is bounded above by E_{cd}/k for any pair c and d. This is the intuition behind next result.

Theorem 4: Let $\tau_{min} = \min_{cd} \tau_{cd}$ be the minimum value of Kendall's tau correlation coefficient across all pairs of variables c and d in dataset $\hat{X}$. Then the share $p(1)$ of fully robust comparisons satisfies $p(1) \leq (\tau_{min} + 1)/2$.

Proof: The proof is given in the appendix.

This result shows that the smallest Kendall's tau coefficient, appropriately transformed, provides us with an upper bound for the proportion of comparisons that are fully robust. If $\tau_{min} = 1$, so that all pairs of variables move together in full accord, then $p(1) = 1$ and the bound is tight. If $\tau_{min} = -1$, say, when a pair of variables exhibits a perfect negative correlation, then no comparison is robust and $p(1) = 0$ is equal to this bounding value. For $0 < \tau_{min} < 1$, the actual value of $p(1)$ can be equal to or below the bound. For example, for the 2004 HDI dataset, $\tau_{min} = 0.55$, and thus according to Theorem 4, we have $p(1) \leq 0.775$. As noted above, the actual prevalence of fully robust comparisons is $p(1) = 0.698$. For EPI_6, EPI_{10} and EFI, the respective values of τ_{min} are -0.147, -0.237, and -0.340, yielding upper bounds on $p(1)$ of 0.43, 0.38, and 0.33 respectively. The true values for $p(1)$ are 0.042, 0.015, and 0.065, respectively.

When there are several dimensions, pair-wise correlations can provide only partial information on the magnitude of $p(1)$. An interesting alternative is to adjust the definition of Kendall's tau itself to obtain a multidimensional measure of association that corresponds exactly

to $p(1)$. Let E be the number of pairs of observations in which one of the two observations has higher values in all dimensions, and let F be the number of pairs for which not all dimensions agree. Given any dataset $\hat{X}$ with an arbitrary number of dimensions $D > 0$, we define the *coefficient of multivariate association* by $\tau = (E - F)/(E + F)$, or the number of fully robust comparisons minus the number that are not fully robust, over the total number of comparisons. With dimensional ties ruled out, the total number of comparisons is once again $k = N(N - 1)/2$, while $F = k - E$, so that $\tau = 2E / k - 1 = 2p(1) - 1$ and $p(1) = (\tau+1)/2$. Our coefficient reduces to the standard Kendall's Tau in the two-dimensional case, but otherwise measures association in terms of agreement across all dimensions. The coefficients of multivariate association for the HDI datasets in 1998 and 2004 are, respectively, $\tau = 0.43$ and $\tau = 0.396$, while for the EFI it drops to $\tau = -0.87$. The coefficient for the EPI dataset rises from -0.97, to -0.94, to -0.916, to -0.053 as we move from largest to smallest number of dimensions. This is a useful way of restating a robustness property of datasets using more familiar terminology, while emphasizing the fundamental link between statistical association and robustness.

An alternative route makes use of the general notion of "increasing association" found in Boland and Proschan (1988), among other sources.[19] We say that dataset $\hat{Y}$ is obtained from dataset $\hat{X}$ by an *association increasing rearrangement* if for some $x \neq x'$ we have: (a) neither $x \geq x'$ nor $x' \geq x$ holds; (b) $y = x \vee x'$ and $y' = x \wedge x'$; and (c) $y'' = x''$ for all $x'' \neq x, x'$. In other words, the datasets are identical apart from a pair of non-comparable observations in $\hat{X}$ that were made comparable in $\hat{Y}$ by placing all the higher values in one observation (the least upper bound) and all the lower values in another (the greatest lower bound). We have the following result.

Theorem 5: Suppose that the initial weighting vector is fixed. If dataset $\hat{Y}$ is obtained from dataset $\hat{X}$ by a series of association increasing rearrangements, then the share $p(1)$ of fully robust comparisons is higher for $\hat{Y}$ than for $\hat{X}$.

Proof: The proof is given in the appendix.

One natural implication of the theorem is that an association increasing rearrangement must lead to a higher value for the coefficient of multivariate association τ. It is also easy to see that none of the pair-wise coefficients τ_{cd} will fall, and that at least one will rise. Consequently,

[19] In the literature on multidimensional inequality and poverty, increasing association was first introduced by Atkinson and Bourguignon(1982). Tsui (1999, 2002) based the notion of correlation increasing majorization on the 'basic rearrangement' used by Boland and Proschan (1988).

this form of transformation is especially useful for illustrating the connection between full robustness and multidimensional association.

Theorem 5 provides information on the share $p(1)$ of fully robust comparisons, but not on $p(r)$ for $r < 1$. The following example shows how greater association across variables need not translate to increased overall prevalence. Suppose that $\hat{X}$ is made up of the four vectors $\hat{x}^1$ = (30,80), $\hat{x}^2$ = (100,30), $\hat{x}^3$ = (90,100), and $\hat{x}^4$ = (80,120). With equal initial weights we see that $C_o(\hat{x}^1) = 55$, $C_o(\hat{x}^2) = 65$, $C_o(\hat{x}^3) = 95$ and $C_o(\hat{x}^4) = 100$, and yet only two comparisons $\hat{x}^3$ C_o $\hat{x}^1$ and $\hat{x}^4$ C_o $\hat{x}^1$ are fully robust. Let $\hat{Y}$ be made up of the four vectors $\hat{y}^1$ = (30,30), $\hat{y}^2$ = (100,80), $\hat{y}^3 = \hat{x}^3$, and $\hat{y}^4 = \hat{x}^4$, so that $\hat{Y}$ is obtained from $\hat{X}$ by an association increasing rearrangement. Then the number of fully robust comparisons rises to three, since now $\hat{y}^2$ C_o $\hat{y}^1$, $\hat{y}^3$ C_o $\hat{y}^1$ and $\hat{y}^4$ C_o $\hat{y}^1$ hold. Clearly, $p(1)$ rises as a result of the association increasing rearrangement.

What about the prevalence $p(r)$ at other values of r? For example, let r = 0.40, and note that the respective x^r vectors used in evaluating $\boldsymbol{C_r}$, as defined in Section II, are (45,65), (79,51), (93,97), and (92,108) for $\hat{X}$ and (30,30), (94,86), (93,97), and (92,108) for $\hat{Y}$. Checking each collection for vector dominance, we find that the number of $\boldsymbol{C_r}$ comparisons in $\hat{X}$ is four, while only *three* C_r comparisons are possible in $\hat{Y}$, and hence $p(r)$ is *negatively* affected by the association increasing rearrangement. Note that the rearrangement results in a vector $\hat{y}^2$ that is not comparable to the other two unchanged vectors, $\hat{y}^3$ and $\hat{y}^4$, and this is preserved in $\boldsymbol{C_r}$; whereas, the non-comparability of $\hat{x}^2$ with $\hat{x}^3$ and $\hat{x}^4$ does not survive the averaging underlying $\boldsymbol{C_r}$. Since this example has two dimensions, it also follows that Theorem 3 applies, and Kendall's tau coefficient is higher in $\hat{Y}$ than $\hat{X}$. Consequently, $p(r)$ can strictly fall when there is greater association, or when the tau coefficient between the two dimensions rises. While it is clear that $p(1)$ is linked to association among variables, the specific mix of factors that determine the placement and shape of $p(r)$ for $r \in (0,1)$ has yet to be determined.

V. Robustness and Redundancy

The results of the previous section show that greater association increases the prevalence of fully robust comparisons and, in this sense, is a desirable attribute of a multidimensional dataset. There is an alternative literature that takes a rather different view of high positive association, and we will now briefly examine these arguments in light of our findings.

A number of previous studies have critiqued the HDI based on the statistical association between the three components used to construct the composite index (McGillivray 1991, 2005; McGillivray and White 1993; Cahill 2005). McGillivray (1991), in particular, provided an argument based on a notion of "redundancy of composition", which arises when there is a strong positive correlation between a composite index and one of its components. Using the data from the UNDP (1990), McGillivray (1991) found that the Spearman rank correlation coefficient between the rankings generated by the per-capita GDP and the rankings generated by the HDI was 0.893. High redundancy of composition is considered to be an undesirable property on the grounds of parsimony: if a single component provides basically the same ranking as the composite index, why not use the former instead of the latter? A second argument invokes the notion of "multidimensionality" of the index: if each pair of component variables is highly correlated, then the index could hardly be characterized as multidimensional, and once again, a single dimension may be all that is needed.

The force of these arguments is mitigated somewhat by our robustness results. To be sure, when the variables are highly correlated in a given dataset, the index may well be tracked by a single component[20] and may act like a unidimensional measure; but the comparisons it makes will tend to be robust. Note that this favorable conclusion (like the critiques) is contingent on the actual dataset employed. At a different point in time, or over a specific subset of observations,[21] the correlations may be dramatically different and the conclusions could be reversed. So the terms "redundant", "multidimensional" and "robust" should not be associated with a given composite index, but rather jointly to the index and a specific dataset. In addition, once a robustness perspective is adopted, the parsimony or multidimensional arguments carry less force: if we replace the original variables with a single one, we lose all information on robustness, since a single variable always generates an unambiguous ranking.

There remains an interesting and unresolved tension between the need for a composite index to improve upon unidimensional alternatives and the desire for the comparisons it makes to be robust. This question has implications for the choice of a specific variable to represent a given dimension in practice. This choice should of course in principle be guided by theory. Yet theory will not necessarily guide the selection of a variable to measure achievement in a dimension. There are, for example, many different variables measuring health and education

[20] It is easy to demonstrate formally that the higher the correlations between components on a composite index the higher will be the correlation between the index itself and any one of its components.

[21] Suppose we are interested in the group of thirty least developed countries according to the HDI. The Kendall's tau rank correlation coefficients between the 2004 HDI and its three components are 0.18, 0.41, and 0.38, respectively, and the Kendall's tau coefficients between each pair of the three components are merely -0.31, -0.01, and 0.08. A similar pattern is found in other groups of interest.

status and a choice needs to be made between them if health and education dimensions are included in the index in question. Is it preferable to select a variable that has low correlation with the other variables to improve the multidimensional integrity of the index? Or might it be better to seek out a variable that has high correlation with the others to ensure more robust comparisons? Further guidance on how to address this tension lies beyond the scope of the present paper, but is an important area for future research. More generally, there is a strong case for design of composite indices to at least take some explicit account the properties of the variables under consideration for inclusion in the index in question as one of the criteria considered. It would appear from reading the literature that these properties are largely, if not totally, ignored.

VI. Conclusion

This paper has analyzed the robustness of rankings obtained from composite indices – the multidimensional indices that combine information on two or more component indices using a weighted average. It examined the empirical prevalence of robust comparisons for three well-known and widely used indices: the Human Development Index, the Index of Economic Freedom and the Environmental Performance Index. The rank robustness of the Human Development Index was found to be the highest, with 73% of pair-wise 1998 country rankings of this index being fully robust. The Environmental Performance Index was the least robust, with no more than 6.5% of its pair-wise rankings being fully robust. The paper then examined the link between various characteristics of the dataset and the prevalence of robust comparisons. One characteristic found to be relevant was the statistical association among index components, and many results were proved linking robustness and association. In particular, maximal robustness is obtained when components are perfectly positively associated. The paper briefly touched upon a dilemma concerning the design of composite indices. According to the above results, highly positive correlations among component variables are desirable as they can enhance rank robustness. But according to previous research such correlations are to be avoided on the grounds of redundancy. Should the design of a composite index be focused on rank robustness or on the avoidance of redundancy, or should we try to attain an optimal balance between the two? This question has been left to future research.

One further question raised by our paper concerns the shape of prevalence functions and the implied empirical distribution of robust comparisons. It is evident that for both years the HDI prevalence functions are approximately linear (more precisely, affine), as is the function associated with EPI2. The other prevalence paths have a strictly convex shape. A question is:

what is it about the former composite indices and their datasets that produce a linear form? Linearity ensures that, if consideration is restricted to comparisons that are not fully robust, the empirical distribution of robustness levels is approximately uniform. In other words, the robustness level r is also the share of these comparisons having a robustness of r or below, and the share of comparisons having, say, $r = 0.95$ or above is $1-r = 0.05$. This is certainly a notable regularity, and it would be useful to identify its source. Additional structure on the nature of this association, such as is available with a copula, may be helpful in this regard.

Finally, this paper has focused on the rank robustness of a number of well-known indices. Yet there are many more indices that receive very widespread attention in research and policy circles. In their recent Human Development Report (UNDP, 2010), the UNDP has replaced the old indices with four new indices: the new Human Development Index, the Inequality-Adjusted Human Development Index (IHDI), the Multidimensional Poverty Index (MPI) and the Gender Inequality Index (GII). Although the Foster *et al.* criterion proposed in this paper can be applied to some of these indices, but it may not be directly applicable to other measures such as the MPI and the GII. Future research could also examine the rank robustness of these indices, further developing robustness assessment techniques as appropriate for these tasks.

Appendix

Proof of Theorem 1.

(i) Consider x and x' in $\hat{X}$ such that $x > x'$. Let us denote the corresponding pair of transformed vectors in $\hat{Y}$ by y and y'. Then it is obvious that $y > y'$. Thus, all fully robust comparisons in $\hat{X}$ remain fully robust in $\hat{Y}$. Conversely, if vectors x and x' in $\hat{X}$ are not fully robust, then the corresponding transformed vectors cannot be fully robust. Consequently, $\hat{Y}$ and $\hat{X}$ share the same prevalence value $p(1)$. (ii) Recall that the measure of robustness between x and x' in $\hat{X}$ with $x\ C_0\ x'$ is $r = A/(A + B)$, where $A = C_0(x) - C_0(x')$ and $B = \max_{w\in\Delta}[C(x';w) - C(x;w), 0]$. If $y = \alpha\cdot x + \beta$ and $y' = \alpha\cdot x' + \beta$ in $\hat{Y}$, then $C(y;w) = \alpha C(x;w) + \beta$ and $C(y';w) = \alpha C(x';w) + \beta$ for all w. Consequently, $A' = \alpha A$ and $B' = \alpha B$, where $A' = C_0(y) - C_0(y')$ and $B = \max_{w\in\Delta}[C(y';w) - C(y;w)$; and $r' = A'/(A' + B') = r$. Hence, $\hat{X}$ and $\hat{Y}$ will have identical prevalence. □

Proof of Theorem 2.

Suppose that y is a replicated achievement vector associated with x, so that $y = f(x)$ for a replicating transformation f. Given the initial weighting vector w^0 and a consistent weighting vector u^0, it is clear that $C(y;u^0) = u^0\cdot f(x) = w^0\cdot x = C(x;w^0)$. Now let $r \in (0,1]$ and select any $d = 1,\ldots,D$ along with an index value d' of one of its copies. Let v_d^r denote the dimension d vertex of the simplex S^r in R^D and let $v_{d'}^r$ denote the dimension d' vertex of the simplex S^r in $R^{D'}$. It is clear that $C(x;v_d^r) = v_d^r\cdot x = (1\text{-}r)C(x;w^0) + rx_d = (1\text{-}r)C(y;u^0) + ry_{d'} = v_{d'}^r\cdot y = C(y;v_{d'}^r)$. Hence, where y' and y are the respective transformations of x' and x, we have (i) $C(x';w^0) \geq C(x;w^0)$ if and only if $C(y';u^0) \geq C(y;u^0)$, and (ii) $C(x';v_d^r) \geq C(x;v_d^r)$ if and only if $C(y';v_{d'}^r) \geq C(y;v_{d'}^r)$. Since (ii) holds for each d and every associated d', it follows from Theorem 3 of Foster *et al.* (2010) that $x'\mathbf{C}_r\ x$ if and only if $y'\mathbf{C}_r\ y$, and $p(r)$ is the same for both. □

Proof of Theorem 4.

Let τ_{cd} be the Kendall's Tau correlation coefficient between variables c and d. Let E_{cd} be the number of concordant pairs and F_{cd} be the number of discordant pairs. If a third variable d' is added, then it is not possible to have more than E_{cd} concordant pairs. If the variable d' is perfectly positively associated with either dimension c or dimension d, then the number of concordant pairs remain unchanged. The number of concordant pair is lower, otherwise. Thus, if there are more than two variables and $\tau_{cd} = \tau_{\min}$, the number of concordant pair cannot be more than E_{cd}. Hence, the proof follows using the relation between $p(1)$ and τ.

Proof of Theorem 5.

Fix the initial vector w^o and let $\hat{Y}$ be obtained from $\hat{X}$ by a single association increasing rearrangement involving x, x', y, and y' as defined in (a)-(c) above. If we can show that $p(1)$ rises, then we are done. To do this, we need only focus on comparisons involving at least one of the vectors x and x' in $\hat{X}$, since the remaining vectors are unchanged. Consider first the comparison involving both x and x'. By (b) we know that neither $x \geq x'$ nor $x' \geq x$ holds, and hence by Theorem 2 of Foster *et al*. (2010) neither $x \boldsymbol{C_1} x'$ nor $x' \boldsymbol{C_1} x$ can be true. However, by construction $y > y'$ and since by assumption no achievements in any given dimension of x and x' can be equal, we must have $y >> y'$. Again, by the same theorem it follows that $y \boldsymbol{C_1} y'$ holds, which represents a gain of one comparison for $\hat{Y}$ as compared to $\hat{X}$.

Now consider a case-by-case analysis of comparisons involving vectors x and x' and any given unchanged vector x''. (i) Suppose that x'' can be compared to both of x and x' using $\boldsymbol{C_1}$. The case where $x \boldsymbol{C_1} x''$ and $x'' \boldsymbol{C_1} x'$ simultaneously hold is impossible, since it implies $x \geq x'$ in contradiction to (a). Similarly the case where $x' \boldsymbol{C_1} x''$ and $x'' \boldsymbol{C_1} x$ both apply contradicts $x' \geq x$, and is likewise impossible. On the other hand, if $x'' \boldsymbol{C_1} x$ and $x'' \boldsymbol{C_1} x'$ hold, then $x'' \geq x$ and $x'' \geq x'$ must both be true, and hence $x'' >> x$ and $x'' >> x'$ since no two vectors in $\hat{X}$ can have equal entries in a given dimension. By construction, then, $y'' >> y$ and $y'' >> y'$, which yields $y'' \boldsymbol{C_1} y$ and $y'' \boldsymbol{C_1} y'$, by the Corollary. Similarly, $x' \boldsymbol{C_1} x''$ and $x \boldsymbol{C_1} x''$ yields $y' \boldsymbol{C_1} y''$ and $y \boldsymbol{C_1} y''$, and so in all possible cases y'' can be compared to both of y and y' using $\boldsymbol{C_1}$. Clearly, $\hat{Y}$ and $\hat{X}$ have the same number of fully robust comparisons of this type. (ii) Suppose that x'' can be compared to exactly one of x and x' using $\boldsymbol{C_1}$. If the comparison is $x \boldsymbol{C_1} x''$, then $x >> x''$ and hence by construction $y >> y''$, which implies $y \boldsymbol{C_1} y''$. In a similar fashion, if the comparison is $x' \boldsymbol{C_1} x''$, then we also conclude $y \boldsymbol{C_1} y''$. Alternatively, if the comparison is $x'' \boldsymbol{C_1} x$, then $x'' >> x$ and hence by construction $y'' >> y'$, which implies $y'' \boldsymbol{C_1} y'$. By the same argument, if the comparison is $x'' \boldsymbol{C_1} x'$, then we conclude $y'' \boldsymbol{C_1} y'$ once again. So in each circumstance, y'' can be compared to at least one of y and y' using $\boldsymbol{C_1}$ and hence $\hat{Y}$ has at least as many fully robust comparisons of this type as $\hat{X}$. (iii) Suppose that x'' can be compared to neither of x and x' using $\boldsymbol{C_1}$. Then, trivially, $\hat{Y}$ has at least as many fully robust comparisons of this type as $\hat{X}$. Consequently, the number of fully robust comparisons across cases (i) to (iii) is at least as high for $\hat{Y}$ as for $\hat{X}$; and given the original single comparison gain by $\hat{Y}$ over $\hat{X}$, it follows that $p(1)$ must be strictly higher for $\hat{Y}$ than for $\hat{X}$. □

References

Alkire, S. and Foster, J. (2010), "Designing the Inequality-Adjusted Human Development Index (IHDI)", Working Paper 37, Oxford Poverty and Human Development Initiative, University of Oxford, Oxford.

Atkinson, A.B., and F. Bourguignon (1982), "The Comparison of Multi-Dimensioned Distributions of Economic Status", *Review of Economic Studies*, 49(2), 183-201.

Blackorby, C., D. Donaldson, and J. A. Weymark, (1984), "Social Choice with Interpersonal Utility Comparisons: A Diagrammatic Introduction", *International Economic Review*, 25(2), 327–56.

Boland, P. J., and F. Proschan (1988), "Multivariate Arrangement Increasing Functions with Applications in Probability and Statistics", *Journal of Multivariate Analysis*, 25(2), 286–298.

Cahill M. B. (2005), "Is the Human Development Index Redundant?" *Eastern Economic Journal* 31(1), 1-5

Cherchye L., W. Moesen, N. Rogge, T. Van Puyenbroeck, M. Saisana, A. Saltelli, R. Liska, S. Tarantola (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

Cherchye L., E. Ooghe and T. V. Puyenbroeck (2008), "Robust Human Development Rankings", *Journal of Economic Inequality*, Vol. 6(4), 287-321

Decancq, K. and Lugo, M. A. (2010), *Weights in Multidimensional Indices of Well-Being: An Overview*, forthcoming in Econometric Reviews.

Esty, D., M. Levy, T. Srebotnjak, and A. de Sherbinin (2005), *Environmental Sustainability Index: Benchmarking National Environmental Stewardship*. Yale Center for Environmental Law and Policy, Yale University, New Haven.

Esty, D. C., C. Kim, T. Srebotnjak, M. A. Levy, A. de Sherbinin and V. Mara (2008), *2008 Environmental Performance Index,* Yale Center for Environmental Law and Policy, Yale University, New Haven.

Foster J. E., M. McGillivray and S. Seth (2009), *Rank Robustness of Composite Indices*, Working Paper 26, Oxford Poverty and Human Development Initiative, University of Oxford, Oxford.

Foster J. E., M. McGillivray and S. Seth (2010), Rank Robustness of Composite Indices: Dominance and Ambiguity, Paper Presented at the 31st General Conference of The International Association for Research in Income and Wealth, St. Gallen, Switzerland, August 22-28.

Heritage Foundation (2008), *Index of Economic Freedom,* The Heritage Foundation, Washington, DC.

Kendall M. and J.D. Gibbons, (1990), *Rank Correlation Methods*, Oxford University Press, Fifth Edition, Oxford.

McGillivray, M. (1991), "The Human Development Index: Yet Another Redundant Composite Development Indicator?", *World Development* 19(10), 1461-68.

McGillivray, M. and H. White (1993), "Measuring Development? The UNDP's Human Development Index", *Journal of International Development* 5(2), 183-92.

McGillivray, M. (2005), "Measuring Non-economic Well-being Achievement", *The Review of Income and Wealth*, 51(2), 337-364.

McGillivray, M. and F. Noorbakhsh (2007), "Composite Indexes of Well-being: Past, Present and Future", in M. McGillivray (editor), *Human Well-being: Concept and Measurement*, Palgrave-Macmillan, 2007.

Nardo M., M. Saisana, A. Saltelli, S. Tarantola, A. Hoffman and E. Giovannini (2005), *Handbook on Constructing Composite Indicators: Methodology and User's Guide,* Joint Research Centre (JRC) of the European Commission and OECD, Paris.

Nishimura, K. G. and H. Ozaki (2006). "An axiomatic approach to ε-contamination", Economic Theory, Vol. 27, 333–340.

Permanyer, I. (2009), "Uncertainty and Robustness in Composite Indices Rankings", paper presented at the Spanish Economic Association Annual Congress, Valencia, Spain.

Saisana, M., A. Saltelli and S. Tarantola (2005), "Uncertainty and Densitivity Analysis as Tools for the Quality Assessment of Composite Indicators" *Journal of the Royal Statistical Society*, Series A 168, 1–17.

Tsui, K.-Y. (1999), "Multidimensional Inequality and Multidimensional Generalized Entropy Measures: An Axiomatic Derivation", *Social Choice and Welfare*, 16(1), 145-157.

Tsui, K.-Y. (2002), "Multidimensional Poverty Indices", *Social Choice and Welfare*, 19(1), 69-93

United Nations Development Program (UNDP) (1990, 1998), *Human Development Report*, Oxford University Press, New York.

United Nations Development Program (UNDP) (2006, 2009, 2010), *Human Development Report*, Palgrave-Macmillan, New York.