Oxford Poverty & Human Development Initiative (OPHI)
Oxford Department of International Development
Queen Elizabeth House (QEH), University of Oxford

OPHI WORKING PAPER NO. 26

Rank Robustness of Composite Indices±

James Foster*, Mark McGillivray** and Suman Seth***

January 2009

Abstract
Many common multidimensional indices take the form of a 'composite index' or a weighted average of several dimension-specific achievements. Rankings arising from such an index are dependent upon an initial weighting vector, and any given judgment could, in principle, be reversed if an alternative weighting vector was employed. This paper examines a variable-weight robustness criterion for composite indicators that views a comparison as robust if the ranking is not reversed at any weight vector within a given set. We characterize the resulting robustness relations for various sets of weighting vectors and illustrate how they moderate the complete ordering generated by the composite indicator. We propose a measure by which the robustness of a given comparison may be gauged and illustrate its usefulness using data from the Human Development Index. In particular, we show how some country rankings are fully robust to changes in weights while others are quite fragile. We investigate the prevalence of the different levels of robustness in theory and practice and offer insight as to why certain datasets tend to have more robust comparisons.

Keywords: Composite indicator, Multidimensional index, Human Development Index, Weighting vector, Robustness, Positive association, Rank correlation, Kendall's tau

JEL Classifications: I31, O12, O15, C02

Citation: Foster, J., McGillivray, M. and Seth, S. (2009). 'Rank robustness of composite indices', OPHI Working Paper 26, University of Oxford.

± This paper has benefited from discussions with and comments on earlier drafts from a number of people. They include Sabina Alkire, Farhad Noorbakhsh, Nora Markova, Tony Shorrocks and participants of the Oxford Poverty and Human Development Initiative (OPHI) "Workshop on Weighting" held on 27-28 May June 2008 at the University of Oxford and the participants of the American Economic Association Meeting 2009 held at San Francisco. The usual disclaimer applies. Correspondence to James Foster, james.e.foster@vanderbilt.edu.

* Department of Economics, Box 1611 Station B, Vanderbilt University, Nashville, TN 37235, USA +1 615 322 2192, James.E.Foster@vanderbilt.edu and Oxford Poverty & Human Development Initiative (OPHI), Queen Elizabeth House (QEH), Department of International Development, 3 Mansfield Road, Oxford OX4 1SD, UK +44 1865 271915.
** Chief Economist, Australian Agency for International Development (AusAID), GPO Box 887, Canberra ACT 2601, Australia, +61 2 6206 43, Mark.McGillivray@ausaid.gov.au.
*** Department of Economics, Vanderbilt University, VU Station B#351819 2301, Vanderbilt Place, Nashville, TN 37235, USA, +1 615 383 6910, Suman.Seth@vanderbilt.edu.

1. Introduction

Remarkable attention is given to rankings arising from various indicators. This is especially true of country rankings. People are naturally curious as to how their country compares to others, national pride is often at stake, and national governments are often quick to claim credit for a high or higher than expected ranking if it can be linked, dubiously or otherwise, to public policy. More generally, the media, business groups, civil society, sections of the research community and international organisations regularly monitor and report on country rankings of indices assessing a variety of phenomena such as sustainability, corruption, rule of law, national income, economic policy efficacy, institutional performance, happiness, human well-being, transparency, globalisation, human freedom, peace or vulnerability.

It is widely recognised that many of the preceding phenomena are multidimensional. This, combined with the availability of more and better data, has in recent decades led to the increased use of composite indices. These indices by their very nature combine in various ways indicators of achievement in the dimensions of the phenomenon in question. Many of the rankings that attract greatest interest arise from indices of this type.

The interest in national governments and others in rankings arising from composite indices is, however, blind to long held concerns regarding their construction. A central concern is the weighting of dimension specific achievements. In a perfect world the weight vectors would be based on information on a meta production function for the phenomenon in question. An absence of accepted information on these functions has resulted in one of three weighting schemes. The most common is to select weights arbitrarily, typically by taking the simple arithmetic mean of the indicators in question.[1] Using this mean is interpreted as assigning equal weights to each dimension. The proponents of this equal weight approach acknowledge that is deficient as in reality the dimensions will almost certainly have differential importance, but argue that there is no accepted basis or guidance for doing otherwise. In this sense the equal weight approach is seen as the least deficient available weighing scheme, one that is likely to attract the least disagreement.[2]

Ambiguity over the numerical values of weights employed by composite indices naturally leads one to question the ranking arising from these measures.

[1] Other approaches are either normative or statistical. The normative approach involves setting weights either in accordance with individual or societal norms, the former often being those of the designers of the index in question. The second is statistical, being purely data-driven. Many different such approaches have been proposed. The most popular being principal components analysis, with the first principal component extracted from the dimension achievement indicators serving as the composite index. Both approaches are fundamentally flawed, the former because of a lack of guidance as to whose norms should be used and the latter because of a difficulties in interpretation.

[2] For example, the proponents of the Environmental Sustainability Index (ESI) argued for equal weights on the grounds that "that no objective mechanism exists to determine the relative importance of the different aspects of environmental sustainability." Other composite indices, used in environmental, well-being and related fields that employ equal weights include the Child Well-being Index, Commitment to Development Index, Economic Resilience Index, Economic Vulnerability Index, Environmental Performance Index, Environmental Sustainability Index, Gender Empowerment Measure, Gender-related Development Index, Genuine Progress Measure, Global Peace Index, Human Development Index, Human Poverty Index, Index of Economic Freedom, Global Peace Index and the Physical Quality of Life Index. In most of the above cases the index is formed by taking the simple arithmetic mean of the component indicators.

Specifically, one can ask to what extent these rankings are dependent upon the initial weighting vector, and whether any given judgment could be reversed if an alternative weighting vector was employed. Such is the focus on this paper. Using a dominance-based analytical framework, it examines a variable-weight robustness criterion for composite indicators that views a comparison as robust if rankings are not reversed at any weight vector within a given set. The paper characterizes the resulting robustness relations for various sets of weighting vectors and illustrates how they moderate the complete ordering generated by the composite indicator in question. It proposes a measure by which the robustness of a given comparison may be gauged and illustrate its usefulness using data from the Human Development Index (HDI). The HDI is a very well known and widely used measure of well-being at the level of nations and the rankings it provides are the subject of intense international interest.[3] The paper shows how some country rankings are fully robust to changes in weights while others are quite fragile, and investigates the prevalence of the different levels of robustness in theory and practice and offer insight as to why certain datasets tend to have more robust comparisons. It from the outset be emphasised that the fundamental purpose of the paper is not to discourage the reporting or use of multidimensional indices and the rankings they provide. Rather, it is to facilitate more incisive interpretation of these rankings.

The remainder of paper is structured as follows. Section II provides with a description of the mathematical concepts, notations and definitions used throughout the paper. A formal treatment of the notion of dominance and its relation to rank robustness is provided in Section III. Section III also defines and characterizes a partial ordering, analogous to that of Foster and Shorrocks (1988) that facilitates the construction of a measure of robustness. Section IV constructs a rank robustness measure. Section V looks at the prevalence of robust comparisons, highlighting how the number of ambiguous comparisons across an entire sample of observation depends on the association between the dimension indicators used in the index in question. The HDI and a number of other indices are used in this section to illustrate key points. The paper concludes in Section VI. Special attention is given with some reflections on the future design of composite indices, in particular trade offs between rank robustness and empirical redundancy.

2. Concepts, Definitions and Notation

This section provides the mathematical definitions and notation used in the paper. The number of *dimensions* to be summarized in a composite index is denoted by an integer $D \geq 2$. An *achievement vector* $x = (x_1,\ldots,x_D) \in R^D$ indicates the score achieved in each of the D dimensions, while $X \subseteq R^D$ gives the set of all of achievement vectors that are possible. A *dataset* is a finite set $\hat{X}$ of $\hat{n}$ elements drawn from X. For any $a,b \in R^D$, let $|a| = \Sigma_{d=1}^{D} a_d$ denote the sum of a's components

[3] The annual publication of the Human Development Report is a much awaited international event owing almost entirely to the HDI country ranking it contains. This is evident from a 2006 article in the *New York Times*, which with a not insignificant dose of fanfare reported that for "the sixth year in a row, Norway was ranked first on the United Nations' human development index as the country providing its citizens with the best chance of living a long and prosperous life." (New York Times, 2006).

and let $a \cdot b = \Sigma_{d=1}^{D} a_d b_d$ represent the inner product of a and b. The expression $a \geq b$ indicates that $a_d \geq b_d$ for $d = 1, \ldots, D$; this is the *vector dominance* relation. If $a \geq b$ with $a \neq b$, then this situation is denoted by $a > b$; while $a >> b$ indicates that $a_d > b_d$ for $d = 1, \ldots, D$. The *least upper bound* of a and b, denoted by $a \vee b$, is the vector having max $\{a_i, b_i\}$ as its i^{th} coordinate; the *greatest lower bound* of a and b, denoted by $a \wedge b$, is the vector having min $\{a_i, b_i\}$ as its i^{th} coordinate.

***** Figure 1 Here *****

The *unit simplex* is defined as $S = \{s \in R^D: s \geq 0 \text{ and } |s| = 1\}$ and is depicted in Figure 1 above for $D = 3$. It contains all possible ways of weighting the various achievements. The *vertices* of the simplex are given by $v_d = e_d$ for $d = 1, \ldots, D$, where e_d is the usual basis element that places full weight on the single achievement d. The *centroid* or central point of the simplex $v^0 = (1/D,\ldots,1/D)$ places equal weight on all achievements. Clearly, v^0 is the simple average of the vertices of S, or $v^0 = \Sigma_{d=1}^{D} v_d / D$. For example, when $D = 3$, the vertices of the simplex S are $v_1 = (1,0,0)$, $v_2 = (0,1,0)$, and $v_3 = (0,0,1)$, while the centroid is $v^0 = (⅓,⅓,⅓)$.

In what follows, we construct smaller versions of S by proportionally contracting the vertices of S toward a given point w^0 of S. For any $r \in (0,1]$ and $d = 1,2,\ldots, D$, define $v_d^r = (1 - r)w^0 + rv_d$, and let S^r be the regular simplex generated by the vertices $v_1^r, v_2^r, \ldots, v_D^r$; equivalently, let S^r be the convex hull of $\{v_1^r, v_2^r, \ldots, v_D^r\}$. For example, if $r = 0.25$ and $w^0 = v^0$ then $v_1^r = (0.5, 0.25, 0.25)$, $v_2^r = (0.25, 0.5, 0.25)$, and $v_3^r = (0.25, 0.25, 0.5)$ as illustrated in Panel I of Figure 2. The resulting simplex S^r is outlined by an equilateral triangle whose vertices are one-fourth of the way to the vertices of S from the centroid v^0. If $w^0 = (0.6, 0.2, 0.2)$, then $v_1^r = (0.7, 0.15, 0.15)$, $v_2^r = (0.45, 0.4, 0.15)$, and $v_3^r = (0.45, 0.15, 0.4)$, and S^r is the regular simplex depicted in Panel II. Note that as r drops to 0, the simplex S^r shrinks to the point w^0, while if $r = 1$, we have $S^r = S$.

***** Figure 2 Here *****

Lemma 1 Let v^0 is the centroid of the unit simplex S. For any $w^0 \in S$ such that $w^0 \neq v^0$, $S^r(w^0)$ and $S^r(v^0)$ are equal in volume for all $r \in (0,1]$.

Proof The d^{th} vertex of the simplex $S^r(w^0)$ and $S^r(v^0)$ are given by the vectors $v_d^r(w^0) = (1 - r)w^0 + rv_d$ and $v_d^r(v^0) = (1 - r)v^0 + rv_d$, respectively. The difference between the d^{th} vertex of these two simplexes are given by the vector $\delta_d^r = v_d^r(w^0) - v_d^r(v^0) = (1 - r)(w^0 - v^0)$. Therefore, $\delta_d^r = (1 - r)(w^0 - v^0)$ for all d and $S^r(w^0)$ is obtained from $S^r(v^0)$

by shifting all vertices of the later by δ_d^r. Hence, $S^r(w^0)$ and $S^r(v^0)$ are equal in volume for all $r \in (0,1]$.[4]

3. Robust Comparisons

While there are many conceivable ways of aggregating achievements, we focus here on a common form of multidimensional index based on a weighted average of individual levels. A *composite indicator* $C: X \times S \to R$ applies the entries of a *weighting vector* $w \in S$ to the respective entries in the achievement vector $x \in X$ and sums to obtain the general form $C(x;w) = w \cdot x$. To implement this approach in practice one must select a specific weighting vector. In what follows, it is assumed that an *initial weighting vector* $w^0 \in S$ has already been chosen; this fixes the specific composite indicator $C_0: X \to R$ defined by $C_0(x) = C(x;w^0)$ for all $x \in X$. The associated strict ordering of achievement vectors will be denoted by $\boldsymbol{C_0}$, so that $x \, \boldsymbol{C_0} \, y$ if and only if $C_0(x) > C_0(y)$.

Arguably the world's best known composite multidimensional index is the HDI. It first appeared in the United Nations Development Program (UNDP) *Human Development Report 1990* (UNDP, 1990). HDI values have since been published annually for more than 170 countries. The HDI provides information on $D = 3$ achievements, namely, health, education and income.[5] The weighting used in the HDI is v^0, or equal weighting, and the resulting composite index has the form $C_0(x) = v^0 \cdot x = (⅓)x_1 + (⅓)x_2 + (⅓)x_3$. Table 1 provides data for the ten countries with the highest HDI levels in 2004 as given in UNDP (2006). The countries are listed in order of HDI from highest (Norway), to second highest (Iceland), and so forth, until the tenth highest country (Netherlands). The associated information on country comparisons is also represented in matrix form in Table 2. Whenever a 'column country' has a higher HDI value than a 'row country', this is indicated with '$\boldsymbol{C_0}$' in the associated cell. Note that every cell below the diagonal is filled, reflecting the fact that the HDI, like any composite index, generates a complete ordering once a specific weighting vector has been chosen.

***** Table 1 Here *****

***** Table 2 Here *****

However, it should be borne in mind that cross-country HDI comparisons are entirely contingent on the chosen weighting vector v^0 and reveal little about the robustness of judgments as the weights are varied. Indeed, look at the twin examples of Australia versus Sweden and Ireland versus Canada. The HDI of Australia exceeds that of Sweden by about 0.006, and Ireland is higher than Canada by the same margin.

[4] Note that if $r = 1$, then $\delta_d^1 = 0$ and $S^1(w^0) = S^1(v^0)$.

[5] The health index is based on life expectancy; the education index is based on enrolment and literacy rates; and the income index based on per capita Gross Domestic Product. For detailed derivation of the dimension specific indices, see the technical note (UNDP, 2006, p. 394).

It is an easy matter to show that Australia has a higher level of the composite indicator than Sweden for all weighting vectors $w \in S$. On the other hand, the pair-wise ranking of Ireland and Canada can be easily reversed: for example, the weighting vector w' obtained from v^0 by a 0.05 shift of weight from each of the education and income dimensions to the health dimension is one such example.

In symbols, let C be the composite indicator over the $D = 3$ dimensions of health, education and income, and let x, y, x', and y' denote the respective achievement vectors for Australia, Sweden, Ireland and Canada. Then for Australia and Sweden we have $C(x;v^0) > C(y;v^0)$ and this ranking is *never* reversed at any $w \in S$; while for Ireland and Canada, $C(x';v^0) > C(y';v^0)$ and this ranking *is* reversed for the alternative weighting vector w' given above. In sum, $\boldsymbol{C_0}$ comparisons that appear to be identical in Tables 1 and 2 can be differentially robust, and this in turn may have some bearing on our interpretation of such comparisons.

To differentiate among $\boldsymbol{C_0}$ comparisons (such as those found in Table 2) we formulate a binary relation over achievement vectors in X that will indicate when an initial comparison is robust. Our construction begins with a set $W \subseteq S$ of 'reasonable' weighting vectors containing the initial vector w^0. We say that *x robustly dominates y given W*, written $x \, \boldsymbol{C_W} \, y$, if and only if $C(x,w^0) > C(y,w^0)$ and $C(x,w) \geq C(y,w)$ for all $w \in W$. In other words, the composite indicator is higher for x than y when the weighting vector is w^0, and this ranking is never reversed for any other weighting vector in W. The relation $\boldsymbol{C_W}$ is clearly transitive, so that if $x \, \boldsymbol{C_W} \, y$ and $y \, \boldsymbol{C_W} \, z$, then $x \, \boldsymbol{C_W} \, z$. However, it will often be the case that $x \, \boldsymbol{C_W} \, y$ does *not* hold despite $x \, \boldsymbol{C_0} \, y$ being true, since $\boldsymbol{C_W}$ requires there to be *no* reversal for any vector in W. As W expands, the likelihood of a reversal rises, and fewer robust comparisons can be made; in other words, if $W \subseteq W'$ then $x \, \boldsymbol{C_{W'}} \, y$ implies $x \, \boldsymbol{C_W} \, y$, but not vice versa. This paper explores the robustness relation $\boldsymbol{C_W}$ for various specifications of the set W, provides characterizations of the relations, and offers insight on their applicability.

A. Full Robustness

B.

If $W = S$, the associated relation $\boldsymbol{C_W}$ applies when an initial judgment $x \, \boldsymbol{C_0} \, y$ is never reversed at *any* configuration of weights. In this case we say that the comparison is *fully robust* and shall denote the relation $\boldsymbol{C_W}$ by $\boldsymbol{C_1}$. Requiring unanimity over all of S is quite demanding and consequently $\boldsymbol{C_1}$ is the least applicable among all robustness relations; however, when it applies, the associated ranking of achievement vectors is maximally robust.

The examples of Australia versus Sweden and Ireland versus Canada from Table 1 suggest a simple characterization of $\boldsymbol{C_1}$. Notice that Australia is higher than Sweden in each of the three dimensions, and hence $\boldsymbol{C_1}$ holds in this case. In contrast, Ireland is higher than Canada in two dimensions and lower in one, which is why $\boldsymbol{C_1}$ does not apply (i.e., the ranking is reversed when the weight is high enough on the 'reversed' dimension). We have the following result.

Theorem 1 Let $x \, \boldsymbol{C_0} \, y$ for $x, y \in X$. Then $x \, \boldsymbol{C_1} \, y$ if and only if $x \geq y$.

Proof Suppose that $x \, \boldsymbol{C_0} \, y$ is true. If $x \geq y$ holds, then clearly $C(x;w) = w \cdot x \geq w \cdot y = C(y;w)$ for all $w \in S$, and thus $x \, \boldsymbol{C_1} \, y$. Conversely, if $x \, \boldsymbol{C_1} \, y$ holds, then setting $w = v_d$ in $C(x;w) \geq C(y;w)$ yields $x_d \geq y_d$ for all d, and hence $x \geq y$. □

In order to check whether a given ranking $x\ \boldsymbol{C_0}\ y$ is fully robust, one need only verify that the achievement levels in x are at least as high as the respective levels in y.[6] The following corollary provides two alternative sets of sufficient conditions for $\boldsymbol{C_1}$.

Corollary Let $x, y \in X$. Then $x\ \boldsymbol{C_1}\ y$ holds if (i) $x >> y$ or if (ii) $w^0 >> 0$ and $x > y$.

Proof Both sets of conditions entail $x \geq y$, hence by Theorem 1 we need only verify that both imply $x\ \boldsymbol{C_0}\ y$. If $x >> y$, then for any $w \in S$ we have $w{\cdot}x > w{\cdot}y$, hence $x\ \boldsymbol{C_0}\ y$. If $w^0 >> 0$ and $x > y$, it follows immediately that $w^0{\cdot}x > w^0{\cdot}y$, and hence $x\ \boldsymbol{C_0}\ y$.

One interesting implication of the Theorem 1 is that the relation $\boldsymbol{C_1}$ is 'meaningful' when variables are ordinal and no basis of comparison between them has been fixed.[7] Suppose that each variable x_d in x is independently altered by its own monotonically increasing transformation $f_d(x_d)$ and let $y = (f_1(x_1), \ldots, f_D(x_D))$ be the resulting transformed achievement vector.[8] We can show that $x\ \boldsymbol{C_1}\ x'$ implies $y\ \boldsymbol{C_1}\ y'$ for the respective transformed vectors y and y'. Indeed, if $x\ \boldsymbol{C_1}\ x'$, we know that $x\ \boldsymbol{C_0}\ x'$ holds by definition, while $x \geq x'$ is true by Theorem 1. Transforming the variables yields $y \geq y'$ and thus

(1) $w{\cdot}y \geq w{\cdot}y'$ for all $w \in S$

since $w \geq 0$. By $x\ \boldsymbol{C_0}\ x'$ it follows that $w^0{\cdot}x > w^0{\cdot}x'$ is true; thus, for some d we have $w^0_d x_d > w^0_d x'_d$, and hence $x_d > x'_d$ with $w^0_d > 0$. Through the transformation we obtain $y_d > y'_d$ and this, when combined with $w^0_d > 0$ and (1), yields $w^0{\cdot}y > w^0{\cdot}y'$, or $y\ \boldsymbol{C_0}\ y'$. By (1), then, we have $y\ \boldsymbol{C_1}\ y'$. In other words, if $\boldsymbol{C_1}$ holds for any given cardinalization of the ordinal variables, it holds for all cardinalizations. Note that while $\boldsymbol{C_0}$ on its own is *not* meaningful in this context (since transforming variables can lead to $y'\ \boldsymbol{C_0}\ y$ even though initially we had $x\ \boldsymbol{C_0}\ x'$), the fully robust relation $\boldsymbol{C_1}$ has the property that $x\ \boldsymbol{C_1}\ x'$ if and only if $y\ \boldsymbol{C_1}\ y'$, and hence is an appropriate technology for ordinal variables.

Now returning to the case of the HDI, we might ask how many of the 45 $\boldsymbol{C_0}$ comparisons given in Table 2 are fully robust. The answer, provided in Table 3, is that just *four* comparisons exhibit vector dominance of achievement vectors and hence $\boldsymbol{C_1}$. This is perhaps not unexpected due to the narrow differences in HDI values among the highest ranked countries. The picture for the entire dataset reveals greater applicability for the relation $\boldsymbol{C_1}$: for the 177 countries, there are a total of 10,875 fully robust comparisons out of a possible 15,576 comparisons, implying that just under 69.8% of all comparisons are fully robust. Why are there so many fully robust HDI comparisons? We return to this point below in Section V.

***** Table 3 Here *****

B. Limited Robustness

---

[6] Note that $x\ C_0\ y$ precludes the possibility that $x = y$, and hence $x\ C_1\ y$ actually entails $x > y$.

[7] Roberts () – meaningful.

[8] The resulting function $f{:}X \rightarrow R^D$ defined by $f(x) = (f_1(x_1), \ldots, f_D(x_D))$ is called a *monotonically increasing transformation* below.

Whereas the previous section took W to be the entire simplex S, we now consider using the smaller simplex S^r of weighting vectors. This will lead to a less demanding robustness relation than $\boldsymbol{C_1}$, but one that is more generally applicable. Recall that the simplex S^r is the convex hull of vertices $\{v_1^r,...,v_D^r\}$ located a fixed proportion r of the way from w^0 to the vertices of S. When $w^0 = v^0$, the resulting simplex S^r is a smaller version of S with v^0 at its center; for general w^0, the set S^r is a scaled down version of S that preserves the relative position of w^0. In either case, the the size of S^r is always the same for fixed $r \in (0,1]$. Substituting $W = S^r$ in the definition of $\boldsymbol{C_W}$ yields the r^{th} *order robustness relation*, denoted here by $\boldsymbol{C_r}$ and defined as follows: $x\ \boldsymbol{C_r}\ y$ if and only if $x\ \boldsymbol{C}_0\ y$ and $C(x;w) \geq C(y;w)$ for all $w \in S^r$. This relation retains all the properties of the general robustness relation, and since the sets S^r are nested for a fixed w^0, we know that that $x\ \boldsymbol{C_r}\ y$ implies $x\ \boldsymbol{C_{r'}}\ y$ whenever $r > r'$.

Now suppose that $x\ \boldsymbol{C_0}\ y$ holds for the pair x and y of achievement vectors. What additional conditions on x and y are needed to ensure that $x\ \boldsymbol{C_r}\ y$? One easy-to-verify set of *necessary* conditions is for $C(x;w) \geq C(y;w)$ at each vertex $w = v_d^r$ of S^r. Indeed, define $x^r = (x_1^r,..., x_D^r)$ where $x_d^r = C(x; v_d^r) = v_d^r \cdot x$ and let y^r be the analogous vector derived from y. Then the necessary condition can be stated as $x^r \geq y^r$. The next theorem shows that, in fact, this is also sufficient.

Theorem 2 Let $x\ \boldsymbol{C_0}\ y$ for $x, y \in X$. Then $x\ \boldsymbol{C_r}\ y$ if and only if $x^r \geq y^r$.

Proof We need only verify that $x\ \boldsymbol{C_0}\ y$ and $x^r \geq y^r$ imply $x\ \boldsymbol{C_r}\ y$. Pick any $w \in S^r$, and note that since S^r is the convex hull of its vertices, w can be expressed as a convex combination of $v_1^r,...,v_D^r$, say $w = \alpha_1 v_1^r+\ldots+\alpha_D v_D^r$ where $\alpha_1+\ldots+\alpha_D = 1$ and $\alpha_d \geq 0$ for $d = 1,\ldots,D$. But then $C(x;w) = w\cdot x = \alpha_1 v_1^r \cdot x+\ldots+\alpha_D v_D^r \cdot x = \alpha_1 x_1^r+\ldots+\alpha_D x_D^r$, and similarly $C(y;w) = \alpha_1 y_1^r+\ldots+\alpha_D y_D^r$; therefore $x^r \geq y^r$ implies $C(x;w) \geq C(y;w)$. Since w was an arbitrary element of S^r, it follows that $x\ \boldsymbol{C_r}\ y$. □

Theorem 2 shows that to evaluate whether a given comparison $x\ \boldsymbol{C_0}\ y$ exhibits r^{th} order robustness, one need only compare the associated vectors x^r and y^r. If each component of x^r is at least as large as the respective component of y^r, then the comparison is robust according to $\boldsymbol{C_r}$; if x^r has at least one component lower than the respective entry of y^r, then, the original ranking is not robust. Note that when $r = 1$, we have $x^r = x$, and Theorem 2 reduces to Theorem 1.[9]

***** Table 4 Here *****

***** Table 5 Here *****

Table 4 illustrates this approach for $r = 0.25$ for the ten highest HDI countries, with the final three columns listing the entries of the associated vector x^r. The table reveals a host of comparisons that can be made using $\boldsymbol{C_r}$ for this value of r. In particular, the ranking between Iceland and USA, which was determined not to be fully robust, is r^{th} order robust as is apparent by the dominance of the last three entries

[9] Note that since $v_d^r = (1\text{-}r)w^0 + rv_d$, it follows that $x_d^r\{\ \} = v_d^r \cdot x = (1\text{-}r)C_0(x) + rx_d$; in other words, the vector x^r is a convex combination of the vector $(C_0(x), \ldots, C_0(x))$ of average achievements and x itself.

for Iceland over the respective entries for USA. On the other hand, the comparison between Ireland and Canada, which Table 1 showed was not fully robust, is also not robust in the present case, since the fourth column entry for Canada is higher than the third column entry for Ireland. Table 5 lists all comparisons among the top ten countries that can be made using the robustness relation $\boldsymbol{C_r}$ for $r = 0.25$; fully 46.7% of the possible comparisons can now be made. For the entire list of 177 countries, 91.8% of the comparisons exhibit r^{th} order robustness for this level of r.

C. Graphical Depiction

Figure 3 uses data from Table 1 to provide graphical representations of the conditions associated with the various robustness relations. In each panel, the two-dimensional simplex S is depicted in the horizontal plane at the base of the graph, as are its three vertices v_d and the initial weighting vector v^0 at its center. The smaller simplex S^r and its vertices v_d^r are also represented within S (for the case $r = 0.25$). Now suppose that a given country with achievement vector x has been selected. For any weighting vector w in the simplex, the level of the composite indicator $C(x;w)$ is graphed as the *height* above the vector w. Thus, the heights at v_1, v_2, v_3, and v^0 are, respectively, x_1, x_2, x_3, and the HDI. The linearity of C in w ensures that these points and the remaining $C(x;w)$ values form a tilted 'achievement simplex' with vertices as high as the dimensional achievements and a center as high as the country's HDI level.

***** Figure 3 Here *****

Comparisons for three pairs of countries - Australia and Sweden, Iceland and USA, and Ireland and Canada - are depicted. In Panel 1 the achievement simplex of Australia is completely above the achievement simplex for Sweden, reflecting the vector dominance of the respective achievement vectors in Table 1. Australia has the higher HDI and it is clear from the graph that there is no weighting vector for which it has a lower composite indicator level than Sweden. This is an example where the relation $\boldsymbol{C_1}$, or complete robustness, holds.

The second panel depicts a rather different scenario for Iceland and USA: the achievement simplexes intersect and $\boldsymbol{C_1}$ cannot hold. Iceland performs better than the USA in terms of both health and education, but USA's achievement is higher for income. More weight on health and education makes Iceland's level of the composite indicator higher than that of the USA, whereas more weight on income makes USA's level higher. Dominance *does* hold if we restrict consideration to the smaller simplex S^r. When the intersection is projected down to S, the resulting (dashed) line does not cross S^r and hence all weights in S^r follow the original HDI ranking in selecting Iceland above USA. Indeed, Table 4 confirms that Iceland has higher levels of the composite indicator than USA at each of the vertices of S^r. Consequently, while $\boldsymbol{C_1}$ does not hold, $\boldsymbol{C_r}$ certainly does for $r = 0.25$.

The final panel depicts the case of Ireland and Canada, which has the same HDI difference as Australia and Sweden and intersecting achievement simplexes like Iceland and USA, but has different robustness characteristics than each. While Ireland's education and income variables are higher than Canada's, the health index has the opposite orientation, and $\boldsymbol{C_1}$ cannot hold. If we project the intersection of the respective achievement simplexes on S, we obtain a dashed line that cuts S^r, implying that $\boldsymbol{C_r}$ does not hold. The Ireland-Canada comparison is not robust even for $r = 0.25$. This is also evident from Table 4 since Ireland has higher levels of the composite indicator at two of the vertices of S^r (namely, v_2^r and v_3^r) and a lower level at the remaining one (v_1^r).

4. Measuring Robustness

Up to now, our method of evaluating the robustness of a comparison $x\ \boldsymbol{C_0}\ y$ has been to fix a set S^r of 'reasonable' weighting vectors and then to confirm that the initial ranking is not reversed at any member of S^r, in which case the associated r^{th} order robustness relation C_r applies. Theorem 2 provides simple conditions for checking whether $x\ \boldsymbol{C_r}\ y$ holds. The present section augments this approach by formulating a robustness measure that associates with any ranking $x\ \boldsymbol{C_0}\ y$ a number r^* between 0 and 1 to indicate its level of robustness.

We construct r^* using two statistics - one that might be expected to move in line with robustness and another that is likely to work against it. The first of these is $\Delta_0 = C(x;w^0) - C(y;w^0) > 0$, or the difference between the composite value of x and the composite value of y at the initial weighting vector w^0. Intuitively, Δ_0 is an indicator of the strength of the dominance of x over y at the initial weighting vector. The second is $\Delta_m = \max_{w\in S}[C(y;w) - C(x;w), 0]$, or the maximal 'contrary' difference between the composite values of y and x. Note that when the original comparison is fully robust, then $C(y;w) - C(x;w) \leq 0$ for all $w \in S$ and there is no contrary difference. Consequently $\Delta_m = 0$. On the other hand, when the comparison is not fully robust, then $C(y;w) - C(x;w) > 0$ for some $w \in S$, and hence $\Delta_m = \max_{w\in S}[C(y;w) - C(x;w)] > 0$. The quantity Δ_m is the worst-case estimate of how far the original difference could be reversed at some other weighting vector.

The measure of robustness we propose is given by $r^* = \Delta_0/(\Delta_0+\Delta_m)$. Notice that when the initial comparison $x\ C_0\ y$ is fully robust, then $\Delta_m = 0$ and hence $r^* = 1$. Alternatively, suppose that the initial comparison is *not* fully robust so that $\Delta_m > 0$. Then it is clear that r^* is strictly increasing in the magnitude of the initial comparison Δ_0, and strictly decreasing in the magnitude of the contrary worst-case evaluation Δ_m. In addition, if Δ_0 tends to 0 while Δ_m remains fixed, the measure of robustness r^* will also tend to 0. These characteristics accord well with an intuitive understanding of how Δ_0 and Δ_m affects robustness.

Practical applications of r^* may be hampered by the fact that Δ_m requires a maximization problem to be solved, namely, $\max_{w\in S}[C(y;w) - C(x;w)]$. However, by the linearity of $C(y;w) - C(x;w) = (y - x)\cdot w$ in w, the problem has a solution at some vertex v_d of S. At the vertex $w = v_d$ the difference $C(y;w) - C(x;w)$ becomes $y_d - x_d$, and so Δ_m can be calculated as $\Delta_m = \max_d(y_d - x_d)$, or the maximum coordinate difference between y and x. The measure r^* can be readily derived using this equivalent definition for Δ_m. For example, recall the case of Ireland and Canada, whose respective achievement vectors x' and y' are found in Table 1. The initial HDI difference $\Delta_0 = C(x';v^0) - C(y';v^0)$ is 0.956 – 0.950 = 0.006. The maximal coordinate difference $y'_d - x'_d$ is in dimension $d = 1$ so that $\Delta_m = 0.919 - 0.882 = 0.037$. Aggregation gives us an r^* of 0.14. In contrast, the comparison of Australia with Sweden yields the same HDI difference $\Delta_0 = 0.006$, but $\Delta_m = 0$ since all dimensional differences support the initial ranking of Australia over Sweden; hence $r^* = 1$. Finally, the example of Iceland and USA produces $\Delta_0 = 0.012$ and $\Delta_m = 0.031$ and hence a robustness level of about $r^* = 0.28$.

Now what is the relationship between the robustness measure r^* and the relations C_r developed in the previous section? The following theorem provides the answer.

Theorem 3 Suppose that $x\ C_0\ y$ for $x, y \in X$ and let r^* be the robustness level associated with this comparison. Then the r^{th} order robustness relation $x\ C_r\ y$ holds for $0 < r \leq r^*$ and does not for $r^* < r \leq 1$.

Proof Let $x\ C_0\ y$ and suppose that $0 < r \leq r^*$. By definition of r^*, we have $r \leq \Delta_0/(\Delta_0+\Delta_m)$ and hence $r\Delta_m \leq (1\text{-}r)\Delta_0$. Pick any $d = 1,\ldots,D$. Then using the definitions of Δ_0 and Δ_m we see that $r(y_d - x_d) \leq (1\text{-}r)(w^0{\cdot}x - w^0{\cdot}y)$ and hence $rv_d{\cdot}y + (1\text{-}r)w^0{\cdot}y \leq rv_d{\cdot}x + (1\text{-}r)w^0{\cdot}x$. Consequently, $v^r_d{\cdot}y \leq v^r_d{\cdot}x$, and since this is true for all d, it follows that $x^r \geq y^r$ and hence $x\ C_r\ y$ by Theorem 2. Alternatively, suppose that $x\ C_0\ y$ and yet $r^* < r \leq 1$. Then $(1\text{-}r)\Delta_0 < r\Delta_m$ so that $(1\text{-}r)(w^0{\cdot}x - w^0{\cdot}y) < r(y_d - x_d)$ for some d, and hence $v^r_d{\cdot}y > v^r_d{\cdot}x$ or $y^r_d > x^r_d$ for this same d. It follows, then, that $x^r \geq y^r$ cannot hold, and neither can $x\ C_r\ y$ by Theorem 2. □

Theorem 3 shows that the measure of robustness r^* is closely related to the robustness relations C_r. Indeed, r^* is the largest r for which $x\ C_r\ y$ holds or, equivalently, for which S^r has no weighting vector that reverses the initial ranking. Panel 1 of Figure 4 depicts the dashed line of intersection where Iceland and USA have the same value of the composite indicator. Also depicted are the simplexes corresponding to three values of r. The smallest simplex ($r = r''$) contains only weighting vectors that yield a strictly higher value for Iceland. The largest simplex ($r = r'$) is cut by the dashed line and hence it contains a region where USA has higher values. The middle simplex ($r = r^*$) contains vectors for which Iceland has the higher value, and a single vector (the vertex that just touches the dashed line) for which the values are the same. A smaller r would leave room for the simplex to expand without reversing the initial comparison; a larger r would lead to a reversal. Consequently, $r = r^*$ is the robustness level of the comparison. Panel 2 illustrates the analogous construction for a case were initial weights are not equal. Note that the r^* value found here is larger, showing that r^* may well depend upon the initial weighting vector w^0.

***** Figure 4 Here *****

***** Table 6 Here *****

To summarize, we have defined a measure of robustness r^*, shown how it can be calculated in practice, and provided an alternative interpretation in terms of the $\boldsymbol{C_r}$ relations and their associated simplexes. Table 6 illustrates this methodology for the ten-country HDI example via a 'robustness profile' that lists the robustness value r^* (in percentage terms) for each of the 45 possible comparisons. This includes the information given in Table 3 (which highlights the four entries with $r^* = 100\%$) and Table 5 (which depicts the 21 comparisons with $r^* \geq 25\%$), and can easily identify the $\boldsymbol{C_r}$ comparisons for any given r. Note that the average value of r^* in Table 3 is only 35%, which reflects the fact that the HDI levels are quite similar for the high HDI countries and so the Δ_0 values are smaller. The next section will apply the methods to datasets associated with the HDI and other well-known composite indicators to evaluate the applicability of the C_r relations in practice.

5. The Prevalence of Robust Comparisons

The focus now shifts from individual comparisons to the entire collection of comparisons associated with a given dataset $\hat{X}$ and an initial weighting vector w^0. The first question is how to judge the overall robustness of the dataset. One option would be to use an aggregate measure (such as the mean) that is strictly increasing in each comparison's robustness level. However, rather than settling on a specific measure we use a 'prevalence function' based on the entire cumulative distribution of robustness levels, and employ a criterion analogous to first order stochastic dominance to indicate greater robustness. We then apply the methodology to several datasets and investigate how various changes affect the prevalence function.

We begin with an initial weighting vector w^0 and a dataset $\hat{X}$ containing $\hat{n}$ observations. Without loss of generality, we enumerate the elements of $\hat{X}$ as x^1, x^2, …, $x^{\hat{n}}$ where $C_0(x^1) \geq C_0(x^2) \geq \ldots \geq C_0(x^{\hat{n}})$. The analysis can be simplified by assuming that no two observations in $\hat{X}$ have the same composite value, so that $C_0(x^1) > C_0(x^2) > \ldots > C_0(x^{\hat{n}})$.[10] There are $\hat{k} = \hat{n}(\hat{n}-1)/2$ ordered pairs of observations x^i and x^j with $i < j$, and each comparison $x^i \, \boldsymbol{C_0} \, x^j$ has an associated robustness level r^*_{ij}. Let $P = [r^*_{ij}]$ represent the *robustness profile* of $\hat{X}$ (given w^0), which lists the level of robustness r^*_{ij} for every ordered pair in a manner similar to Table 6.

The *mean* robustness level in profile P is given by $\bar{r} = \Sigma_i \Sigma_{j>i} \, r^*_{ij}/\hat{k}$; it is the average level of robustness of the $\hat{k}$ many comparisons. Of course, a higher mean level $\bar{r}$ does not necessarily ensure that the prevalence of $\boldsymbol{C_r}$ is higher for any given r. An alternate approach is to summarize robustness levels in a way that reflects the *entire* distribution, and not just the mean. For any given dataset $\hat{X}$ and initial weighting vector w^0, define the *prevalence function* p:[0,1]→[0,1] to be the function which associates with each $r \in [0,1]$ the share $p(r) \in [0,1]$ of the $\hat{k}$ comparisons whose robustness levels are at least r. In other words, $p(r)$ is the proportion of comparisons for which the $\boldsymbol{C_r}$ relation applies.[11] Suppose that p and q are the prevalence functions for $\hat{X}$ (given w^0) and $\hat{Y}$ (given u^0), respectively. We say that $\hat{X}$ *has greater robustness than* $\hat{Y}$ if $p(r) \geq q(r)$ for all $r \in [0,1]$, with $p(r) > q(r)$ for some $r \in [0,1]$. In words, no matter the level of robustness r, the share of all comparisons that exhibit r^{th} order robustness is no lower for $\hat{X}$ than $\hat{Y}$, and for some r it is higher.

[10] This is true for each of the examples presented below.
[11] At $r = 0$ the complete relation $\boldsymbol{C_0}$ is used and hence $p(0) = 1$.

The two are said to *have the same robustness* if their prevalence functions are the same.

A. Some Examples

Figure 5 depicts the prevalence functions obtained from several datasets associated with well-known composite indicators. The first two are from the 1998 and 2004 Human Development Index[12], which uses equal weights across three dimensions (health, education and income) to rank 177 countries. Next is the 2008 Index of Economic Freedom (IEF) created by the Wall Street Journal and the Heritage Foundation to measure the degree of economic freedom across countries[13]. There are 10 dimensions, spanning the spectrum from business freedom to labor freedom, all weighted equally in w^0. The IEF dataset covers 157 countries for the year 2007.

***** Figure 5 Here *****

The final indicator is the 2008 Environmental Performance Index (EPI), developed by Yale University, Columbia University, the World Economic Forum and the Joint Research Centre of the European Commission to complement the environmental targets of the UN Millennium Development Goals. The EPI is a composite indicator whose variables may be viewed at various levels of aggregation[14]. For purposes of illustration, we consider three versions here. EPI_{10} has ten variables and the initial weights are not all equal. EPI_6 combines variables (and initial weights) to obtain a six variable version with unequal weights. EPI_2 aggregates further to obtain two variables with equal weight on each. As the initial weighting vectors are 'consistent', each version of the EPI produces identical values at its respective initial weighting vector; but due to the different numbers of variables, each has distinct robustness characteristics.[15] The EPI dataset covers 149 countries during the year 2007.

Several initial observations can be made from the prevalence functions given in Figure 5. Each graph is downward sloping, reflecting the fact that as r rises, the set S^r expands, and hence the number of comparisons that can be made by $\boldsymbol{C_r}$ is lower (or no higher). As r falls to 0, each function rises to 100% comparability for $\boldsymbol{C_r}$; in the

---

[12] Note that the Human Development Indices for the years 1998 and 2004 are obtained from UNDP (2000 and 2006), respectively.

[13] The ten dimensions of economic freedoms are Business Freedom, Trade Freedom, Fiscal Freedom, Government Size, Monetary Freedom, Investment Freedom, Financial Freedom, Property rights, Freedom from Corruption, Labor Freedom. For each dimension, the score is normalized between zero and hundred. The final score is obtained by simple average of these scores.

[14] EPI is composed of twenty five-dimensions of performance on environment. However, at the objective level all dimensions are summarized in two categories with equal weights: environmental health and ecosystem vitality. At the policy level, all twenty-five dimensions are summarized into six dimensions at. At the policy level, the weight vector used is (0.5, 0.025, 0.075, 0.075, 0.075, 0.25). Further, these six dimensions are sub-categorized into ten dimensions with the weight vector (0.25, 0.125, 0.125, 0.025, 0.075, 0.075, 0.025, 0.025, 0.025, 0.25). To have detailed information on indicators, see http://epi.yale.edu/Methodology.

[15] For consistent initial weighting vectors, the weight on an aggregated variable is sum of the weights on the variables that were aggregated. See the next subsection.

other direction, the value of $p(r)$ at $r = 1$ is the percentage of the comparisons that are fully robust. There is a wide variation in $p(1)$ across datasets. It is clearly highest for the HDI examples, with $p(1)$ being about 69.8% in 2004 and 73.2% in 1998; it is 47.4% for the two variable EPI; and it is much lower for the remaining indicators[16] (4.2% and 1.5% in the case of EPI_6 and EPI_{10}, respectively, and 6.5% for the EFI).

For r between 0 and 1, the HDI comparisons are also more robust than the comparisons of the EPI and the EFI, and of the two HDI datasets, 1998 exhibits greater robustness than 2004. For the EPI examples, a higher level of aggregation and hence a lower number of variables, leads to greater robustness. However, EPI_2 is still less robust than either of the HDIs, which have three variables. The shapes of the $p(r)$ functions are different, with some being essentially linear, and others exhibiting pronounced curves. Drawing on these examples, we now examine the prevalence of robustness from a more theoretical perspective. What transformations allow the resulting datasets to be compared in terms of robustness?

B. Fixed Robustness and Transformations

We begin with transformations of the data that leave $p(r)$ unchanged and hence yield pairs of datasets with the same robustness properties. A *monotonically increasing transformation* of X is a function $f{:}X\rightarrow R^D$ that can be written as $f(x) = (f_1(x_1),\ldots, f_D(x_D))$ 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 R. We say that $\hat{Y}$ *is obtained from* $\hat{X}$ *by a common-slope affine transformation* (respectively, *by a monotonically increasing transformation*) if $\hat{Y} = \{f(x)\text{: } x \in \hat{X}\}$ 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 $C(x';w^0) > C(x;w^0)$ and $C(y';w^0) < C(y;w^0)$ where y' and y are transformations of x' and x, respectively, which implies that the robustness profiles of $\hat{Y}$ and $\hat{X}$ 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 $\beta = (\beta_1, \ldots, \beta_D)$, 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 $\hat{Y}$ and $\hat{X}$ 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. On the other hand, 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

[16] For EPI_6 and EPI_{10}, we also calculate the prevalence function using initial equal weights. We find the dominance relation to hold between EPI_6 and EPI_{10} by EPI_6 is more robust that EPI_{10} for all r. The relationship is explained in Figure 6.

transformation (as noted in the discussion following Theorem 1), and hence the prevalence $p(1)$ of full robustness does not change. These results are summarized in the following theorem.[17]

Theorem 4 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 $p(r)$.[18]

In the example of the HDI, the normalized income, education and health 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 4, 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 income, education and health variables directly.

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:$X \rightarrow R^{D'}$ for some $D' > D$ such that $f(x) = (f_1(x_1),\ldots, f_D(x_D))$, where each $f_d(x_d)$ is the k_d-fold *replication* $(x_d,x_d,\ldots,x_d)$ for some integer $k_d \geq 1$. We say that that $\hat{Y}$ *is obtained from* $\hat{X}$ *by a replicating transformation* if $\hat{Y} = \{f(x): x \in \hat{X}\}$ 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 do. We say u^0 is *consistent* with w^0 if, for each $d = 1,\ldots, \hat{n}$ the weight w_d^0 on x_d is equal to the sum of the k_d entries in u^0 associated with $f_d(x_d) = (x_d,x_d,\ldots,x_d)$. So for example, if $D = 2$ and f replicates each entry two times, then $w^0 = (½, ½)$ is consistent with $u^0 = (½, 0, ¼, ¼)$. We have the following result.

Theorem 5 If $\hat{Y}$ is obtained from $\hat{X}$ by a replicating transformation, and u^0 is consistent with w^0, then $\hat{Y}$ and $\hat{X}$ have the same prevalence function $p(r)$.

Proof 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 $C(x;v_d^r) = v_d^r \cdot x = (1-$

[17] The result on monotonic transformations would be true even if the initial weighting vectors are different – just so both are strictly positive. The role played by common-slope affine transformations is similar to assumptions used in social choice theory: see Blackorby, Donaldson, and Weymark (1984).
[18] The first part of Theorem 4 will generate same prevalence function $p(1)$ even if use the dominance criterion proposed by Cherchye, Ooghe, and Puyenbroeck (2005).

$r)C(x;w^0) + rx_d = (1\text{-}r)C(y;u^0) + ry_{d'} = v^r_{d'} \cdot y = C(y; v^r_{d'})$. 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) $C(x'; v^r_d) \geq C(x; v^r_d)$ if and only if $C(y'; v^r_{d'}) \geq C(y; v^r_{d'})$. Since (ii) holds for each d and every associated d', it follows from Theorem 2 that $x' \boldsymbol{C_r}\, x$ if and only if $y' \boldsymbol{C_r}\, y$, and $p(r)$ is the same for both. □

In other words, 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 ⅓, say where each variable receives a weight of ⅙, 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 5 *does* seem to 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 EPI_6 is constructed from EPI_{10}. The first and fifth variables in EPI_6 are each obtained by combining three distinct variables in EPI_{10} (namely, variables 1-3 and variables 7-9), while the remaining variables are unchanged. Weights from the initial weighting vector u^0 for EPI_{10} are used to construct each new variable in EPI_6 as a weighted average of the source variables from EPI_{10}, and the weight on the new variable is the sum of the corresponding weights in u^0. The new w^0 is thus consistent with u^0. Now consider a ten variable replication of EPI_6 that repeats variable 1 three times and variable 5 three times and let the initial weighting vector be u^0. By Theorem 5, this intermediate dataset has precisely the *same* robustness profile and prevalence function as EPI_6. 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 EPI_{10}, 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 the key driver of robustness, and this is explored further in the next section.

C. Robustness and Positive 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 positive 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[19] 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. In Figure 5 we saw that both HDI datasets have high levels of robustness, and that the prevalence function is higher for 1998 than for 2004. Indeed, the Kendall tau[20] coefficients (τ) 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 2004 are even higher, at 0.59, 0.70, and 0.60. Both intuition and empirical evidence suggest a link between positive association and robustness. We now turn to the theoretical justification for such a link.

For simplicity, assume that the dataset $\hat{X}$ has the property that within each dimension, all observed values of the variable are distinct.[21] Given any two dimensions c and d, let A_{cd} be the number of *concordant* pairs of observations in which one of the two observations has higher values in both dimensions c and d. Let B_{cd} be the number of *discordant* pairs in which one observation is higher in one dimension and the second is higher in the other. Then Kendall's tau correlation coefficient for dimensions c and d is defined as $\tau_{cd} = (A_{cd} - B_{cd})/(A_{cd} + B_{cd})$. Note that the denominator of this expression is $\hat{k} = \hat{n}(\hat{n}-1)/2$ while $B_{cd} = \hat{k} - A_{cd}$, so that $\tau_{cd} = 2A_{cd}/\hat{k} - 1$.

Now consider the special case where there are only two variables, and so there is a single coefficient $\tau = \tau_{12}$ and number $A = A_{12}$ of concordant pairs. In this special case, the number of concordant pairs is precisely the number of fully robust pairs, so the share of fully robust comparisons is $p(1) = A/\hat{k}$. Therefore, $\tau = 2\,p(1) - 1$ and we have the following result.

Theorem 6 If $D = 2$, then $p(1) = (\tau+1)/2$.

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. If $\tau = 1$ so that the variables have perfect positive correlation, then $p(1) = 1$. If $\tau = -1$, and perfect negative correlation obtains, then $p(1) = 0$. The case of $\tau = 0$ of independence implies $p(1) = ½$, so that half the comparisons would be fully robust in this case. The example of EPI_2 has $\tau = -0.053$ and hence $p(1) = 0.474$, by Theorem 6.

Now consider the general case of $D \geq 2$. Full agreement across all dimensions entails concordance in any two dimensions, hence $p(1) \leq A_{cd}/\hat{k} = (\tau_{cd}+1)/2$ for any pair $c > d$. We have the following result.

Theorem 7 Let $\tau_{\min} = \min_{c>d} \tau_{cd}$ be the minimum value of Kendall's tau coefficient across all pairs of variables. Then $p(1) \leq (\tau_{\min} + 1)/2$.

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,

[19] Note that if there are more than two dimensions then not all of them can ever be perfectly negatively associated. In other words, there does not exist any concept called perfect negative association such as perfect positive association.
[20] Kendall's tau is a measure of association or correlation based on ranks of the variables concerned. See Kendall and Gibbons (1990).
[21] This rules out ties and simplifies the definition of Kendall's tau coefficient.

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 entire 2004 HDI dataset, $\tau_{min} = 0.55$, and thus according to Theorem 4, we have $p(1) \leq 0.78$. 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 -0.3395, yielding upper bounds on $p(1)$ of 0.43, 0.38, and 0.33 respectively. The true values for $p(1)$ are actually much lower, at 0.042, 0.015, and 0.065, respectively.

When there are several dimensions, pairwise correlations can provide only partial information on the magnitude of $p(1)$. An interesting alternative is to adjust the definition of Kendall's tau to obtain a multidimensional measure of association that corresponds exactly to $p(1)$. Let A be the number of pairs of observations in which one of the two observations has higher values in all dimensions, and let B be the number of pairs for which dimensions disagree. Given any dataset $\hat{X}$ having an arbitrary number of dimensions $D > 0$, we define *Kendall's coefficient of positive association* by $\tau = (A - B)/(A + B)$, 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 $\hat{k} = \hat{n}(\hat{n}-1)/2$ while $B = \hat{k} - A$, so that $\tau = 2A/\hat{k} - 1 = 2p(1) - 1$ and $p(1) = (\tau+1)/2$.

In the two-dimensional case, the coefficient τ reduces to the standard two-dimensional Kendall's tau; for more dimensions it requires agreement across all dimensions before counting the comparison in favor or positive association. So for example, the positive association measures for the HDI datasets in 1998 and 2004 are, respectively, $\tau = 0.464$ 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 -0.916, 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 positive association and robustness.

An alternative route is to make use of the general notion of 'increasing association' found in Boland and Proschan (1988), among other sources.[22] 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 noncomparable 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 8 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 $\hat{X}$.

[22] The notion of increasing correlation in the literature of multidimensional inequality and poverty was first introduced by Atkinson and Bourguignon. Tsui (1999, 2002) introduced correlation increasing majorization based on 'basic rearrangement' due to Boland and Proschan (1988).

Proof Fix the initial vector w^0 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 1 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 $y >> y'$. By the Corollary 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) – (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 $\hat{X}$. □

One natural implication of the theorem is that association increasing rearrangements lead to a higher value for Kendall's coefficient of positive association τ. It is also easy to see that none of the pairwise coefficients τ_{cd} will fall, and that at least one will rise. Consequently, this form of transformation is especially useful for illustrating the connection between full robustness and positive association.

Theorem 8 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 prevalence of r^{th} order robustness. Suppose that $\hat{X}$ is made up of the four vectors $x = (30,80)$, $x' = (100,30)$, $x'' = (90,100)$, and $x''' = (80,120)$. With equal initial weights we see that $C_0(x) = 55$, $C_0(x') = 65$, $C_0(x'') = 95$ and $C_0(x''') = 100$, and yet only two comparisons $x''\ C_0\ x$ and $x'''\ C_0\ x$ are fully robust. Let $\hat{Y}$ be made up of the four vectors $y = (30,30)$, $y' = (100,80)$, $y'' = x''$, and $y''' = x'''$, 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 $y'' \, C_0 \, y$, $y''' \, C_0 \, y$ and $y' \, C_0 \, y$ 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 vectors used to indicate C_r are (45,65), (79,51), (93,97), and (92,108) for $\hat{X}$ and (30,30), (94,86), (93,97), and (92,108) for $\hat{Y}$. The number of 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 an association increasing rearrangement in this case. Note that the rearrangement results in a vector y' that is not comparable to the other two unchanged vectors, y'' and y''', and this is preserved in the r^{th} order ranking; whereas, the noncomparability of x' with x'' and x''' does not survive the averaging underlying C_r. Since this example has two dimensions, it also follows that Theorem 4 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 positive 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.

6. Conclusion

Rankings arising from composite indicators receive remarkable attention. Yet they are dependent upon an initial weighting vector, and any given judgment could, in principle, be reversed if an alternative weighting vector was employed. This leads one to question rankings provided by composite indices, especially if there is ambiguity over the numerical values of the weights they employ. Many well known and widely used indices are characterised in this way.

Using a dominance-based analytical framework, this paper examined a variable-weight robustness criterion for composite indicators that views a comparison as robust if the ranking is not reversed at any weight vector within a given set. It characterized the resulting robustness relations for various sets of weighting vectors. An illustration of how these relationships moderate the complete ordering generated by the composite indicator. A measure by which the robustness of a given comparison may be gauged was then proposed, and illustrated using the Human Development Index (HDI). The paper also demonstrated how some rankings are fully robust to changes in weights while others are quite fragile. Finally, the paper investigated the prevalence of the different levels of robustness in theory and practice and offer insight as to why certain datasets tend to have more robust comparisons.

It was emphasised at the outset of the paper that its intention was not to discredit or discourage the use of composite indices, but to facilitate better use of them. The paper helps in this regard by reducing the undue emphasis placed on ranking that were not robust to weight vector, hopefully placing greater emphasis on those rankings that have higher robustness. It promotes this outcome by allowing end users of composite indices to discern between the two cases, thereby making the HDI and other composite indicators more useful and less misleading.

Two findings of the paper are worth highlighting further. Both are suggestive of additional research. The first relates to the interesting empirical observation is the near linearity of the prevalence functions associated with the HDI datasets. In other words, increasing r by a given amount decreases the prevalence of robust rankings by fixed amount, independent of the initial level of r. This means that in the case of the HDI, the entire shape of $p(r)$ is determined by the percentage of fully robust comparisons, $p(1)$. Hence if one were to remove from consideration all fully robust comparisons, the conditional prevalence functions would be virtually identical. Put differently, among all comparisons that are not fully robust, the percentage of comparisons having robustness level r or less is r itself; so, for example, only 5% of these comparisons have robustness level of 0.05 or less (or, equivalently, 0.95 or more). It would be interesting to explore this regularity further.

The second finding relates to previous literature on the HDI. This literature has examined the statistical associations between the HDI and pre-existing well-being indicators, including its individual components (McGillivray, 1991, 2005, McGillivray and White, 1993, Booysen, 2002, McGillivray and Noorbakhsh, 2004, Cahill, 2005). The underlying concern of this literature is the empirical redundancy of the HDI, the new information it provides *vis-à-vis* these indicators. Redundancy is an increasing function of the statistical association of the HDI with respect to these indicators. It follows that the HDI would be considered to be fully redundant with respect to a pre-existing well-being indicator if the chosen correlation coefficient between them was unity. McGillivray (1991) introduced the notion of "redundancy of composition" in critiquing the construction of the HDI. High redundancy of composition, in the context of rankings, was considered to be an undesirable property on the grounds of parsimony. That is, on these grounds makes little sense to combine a set indicators if any one of them provides basically the same rankings as their composite owing to high correlations among them. McGillivray found that the HDI was subject to a high degree of rank redundancy of composition given a high degree of rank correlation between the index and its components. While according to the HDI redundancy literature a high degree of correlation between its components and the HDI itself is bad, the current study finds that it is good as rank robustness appears to be an increasing function of the extent of this correlation. Future research could address this apparent contradiction. In the absence of strong conceptual guidelines as to the choice of components of a composite index, or in the choice of indicators to measure achievement components selected with such guidance, one might speculate as to whether an optimal trade off exists between rank robustness and the redundancy of composition. Future research could address this issue.

Four further directions for future research ought to be emphasised. The first is to develop and integrate statistical robustness into the analysis. The second involves a link with a theoretical literature that addresses uncertainty. The structure of the general robustness relation defined above for a given set W of weighting vectors is closely related to discussions of "Knightian uncertainty" (Bewley, 1986) and ambiguity in which an individual decision maker has a set of prior probability distributions instead of a single prior. One way in which the approach of this paper differs from this literature is that it privileges the initial weighting structure instead of treating it as just one of many. The criterion discussed by Bewely requires a strict improvement at all possible probability vectors. This paper's approach allows the comparison to be weak at the other non-distinguished vectors in the set.[23] Nonetheless, there is a fundamental link between the two that would be interesting to explore. Thirdly, the interpretation given in the paper is that each dimension of the

[23] Bewely further privileges a specific option that he calls the status quo in order to explain certain observed behaviors that are inconsistent with traditional decision theory. Our approach treats all options symmetrically.

achievement vector is the measured amount of a given achievement. One could instead view the dimensions as being obtained from the underlying achievements by some transformation based on, for example, the utility or welfare from the specific dimension. Such an approach might well be adapted to deal with this case, and with other departures from the linearity inherent in composite indicators. Finally, there is clearly a link between the fully robust criterion outlined in this paper and first order stochastic dominance in the multidimensional setting. How does *r*th degree robustness relate to multidimensional stochastic dominance? Is there an analog in the framework employed by this paper to second order multidimensional stochastic dominance? It would be interesting to pursue this direction as well.

References

Anand, S. and A. Sen (2000), "The income component of the human development index", *Journal of Human Development* 1, 83–106

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.

Booysen, F. (2002), "An overview and evaluation of composite indices of development", *Social Indicators Research* 59, 115–151

Cahill M. B. (2005), "Is the human development index redundant?" *Eastern Economic Journal* 31(1), 1-5

Chakravarty, S. R. (2003), "A generalized human development index". *Review of Development Economics* 7, 99–114

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 Tom V Puyenbroeck (2008), "Robust human development rankings", *Journal of Economic Inequality* 8(4), 287-321

Foster J. E. and A. F. Shorrocks (1988), "Poverty orderings and welfare dominance", *Social Choice and Welfare* 5, 179-198. Reprinted in W. Gaertner and P.K. Pattanaik, eds., *Distributive Justice and Inequality* (Springer Verlag, New York), 1988.

Foster J. E. and A.F. Shorrocks (1988), "Poverty orderings", *Econometrica* 56, 173-177.

Joe H. (1997), *Multivariate Models and Dependence Concepts*, Chapman and Hall.

Joe H. (1990), "Multivariate Concordance", *Journal of Multivariate Analysis* 35, 12-30,

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

Mayer, Susan E. and Christopher Jencks (1989), "Poverty and the distribution of material hardship", *Journal of Human Resources* 24(1), 88-114.

McGillivray, M. (2005), "Measuring non-economic well-being achievement", *The Review of Income and Wealth* 51(2), 337-364.

McGillivray, M. (1991), "The Human Development Index: yet another redundant composite development Indicator?", *World Development* 19(10), 1461-68

McGillivray M. and Noorbakhsh F. *Composite Indices of Human Well-being: past, present and future*. Research Paper No. 2004/63, World Institute for Development Economics Research, United Nations University, Helsinki.

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

New York Times (2006), "The Week: The other winners and losers", *New York Times*, November 12, 2006.

Roberts F. S., (1979), *Measurement theory: with applications to decision making, utility, and the social sciences*. Encyclopedia of Mathematics and its Applications, vol. 7, Addison-Wesley, Reading, Mass.,

Saisana, M., Saltelli, A., Tarantola, S., (2005), "Uncertainty and sensitivity analysis as tools for the quality assessment of composite indicators" *Journal of the Royal Statistical Society*, Ser. A 168, 1–17

UNDP, (1990, 2000, 2006), *Human Development Report*, Oxford University Press, New York.

The Wall Street Journal and the Heritage Foundation (2008), *The Index of Economic Freedom*.

Figure 1: Unit simplex

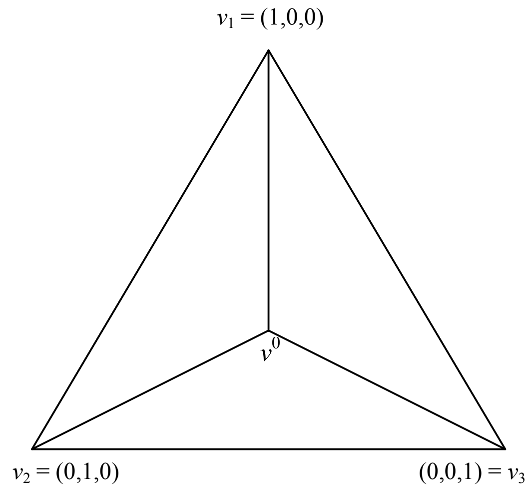

Figure 2: Constructing S^r

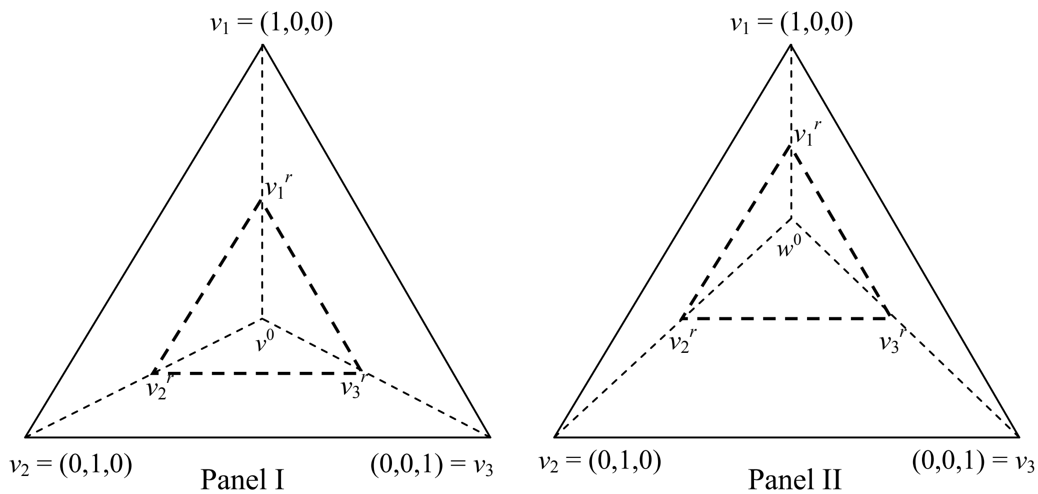

Figure 3: The Robustness Relations

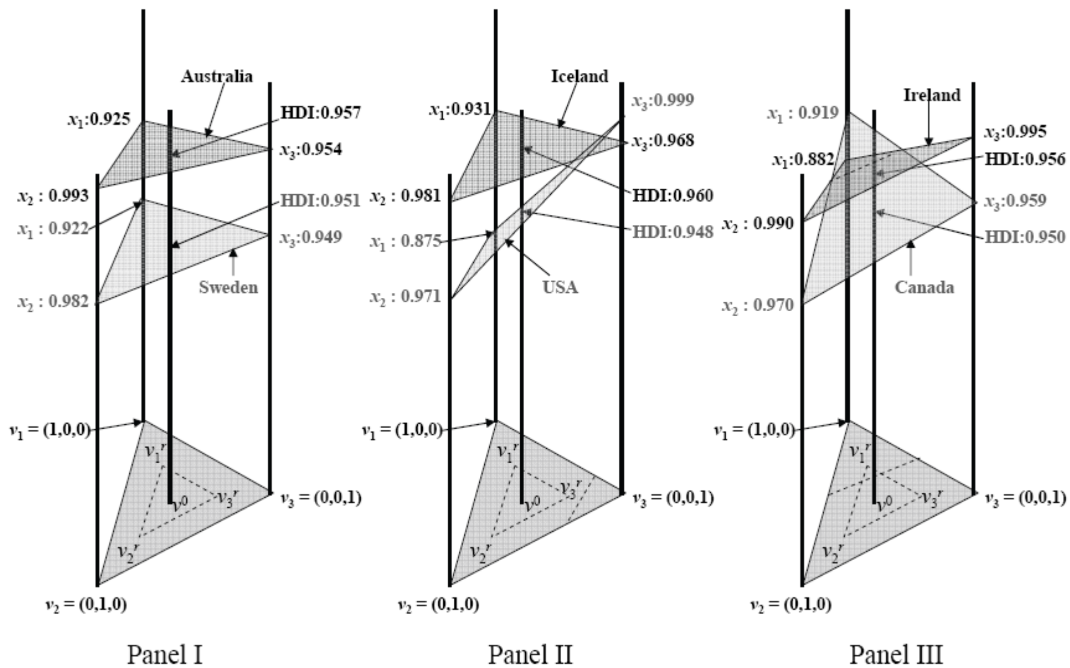

Figure 4: Iceland and USA

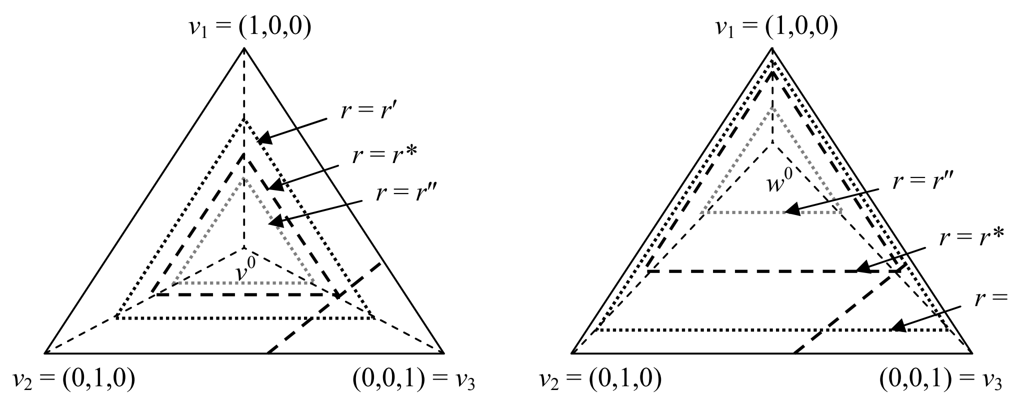

Figure 5: Prevalence Function for all Indices

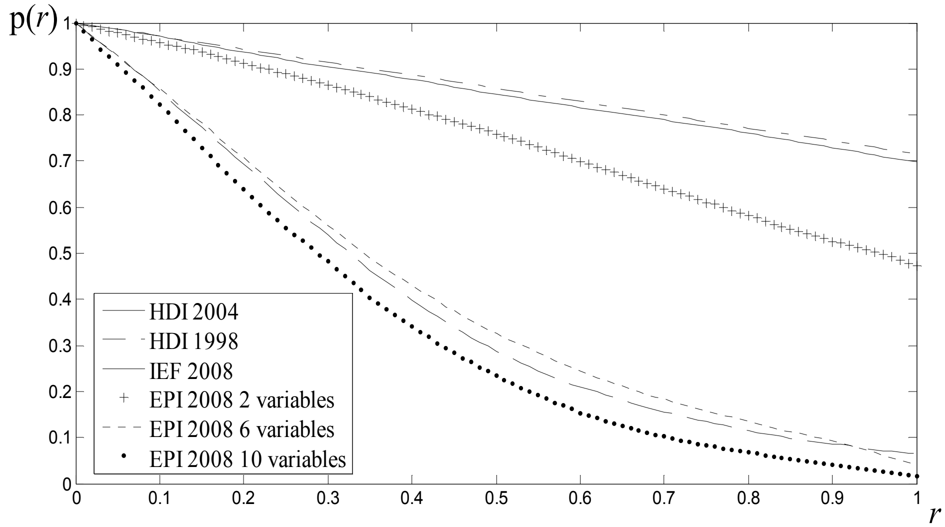

Figure 6: Prevalence function for the EPI: Equal versus Unequal Initial Weights

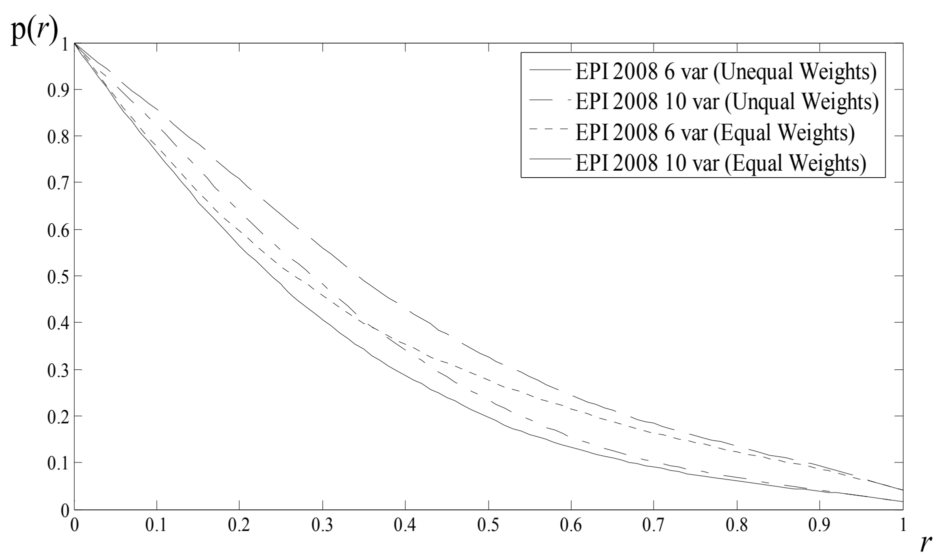

Table 1: Human Development Index and Components: The Top 10 Countries in 2004

Rank	Country	HDI	Health	Education	Income
1	Norway	0.965	0.909	0.993	0.993
2	Iceland	0.960	0.931	0.981	0.968
3	Australia	0.957	0.925	0.993	0.954
4	Ireland	0.956	0.882	0.990	0.995
5	Sweden	0.951	0.922	0.982	0.949
6	Canada	0.950	0.919	0.970	0.959
7	Japan	0.949	0.953	0.945	0.948
8	United States	0.948	0.875	0.971	0.999
9	Switzerland	0.947	0.928	0.946	0.968
10	Netherlands	0.947	0.892	0.987	0.962

Table 2: HDI Comparisons

Country		Norway	Iceland	Australia	Ireland	Sweden	Canada	Japan	USA	Switzerland	Netherlands
	Rank	1	2	3	4	5	6	7	8	9	10
Norway	1										
Iceland	2	C_0									
Australia	3	C_0	C_0								
Ireland	4	C_0	C_0	C_0							
Sweden	5	C_0	C_0	C_0	C_0						
Canada	6	C_0	C_0	C_0	C_0	C_0					
Japan	7	C_0	C_0	C_0	C_0	C_0	C_0				
USA	8	C_0	C_0	C_0	C_0	C_0	C_0	C_0			
Switzerland	9	C_0	C_0	C_0	C_0	C_0	C_0	C_0	C_0		
Netherlands	10	C_0	C_0	C_0	C_0	C_0	C_0	C_0	C_0	C_0	

Table 3: Fully Robust Comparisons

Country		Norway	Iceland	Australia	Ireland	Sweden	Canada	Japan	USA	Switzerland	Netherlands
	Rank	1	2	3	4	5	6	7	8	9	10
Norway	1										
Iceland	2										
Austral.	3										
Ireland	4										
Sweden	5			C_1							
Canada	6		C_1								
Japan	7										
USA	8										
Switzerland	9		C_1								
Netherlands	10	C_1									

Table 4: Human Development Index and Robustness for r = 0.25

Rank	**Country**	**HDI**	x_1^r	x_2^r	x_3^r
1	Norway	0.965	0.951	0.972	0.972
2	Iceland	0.960	0.953	0.965	0.962
3	Australia	0.957	0.949	0.966	0.956
4	Ireland	0.956	0.937	0.964	0.966
5	Sweden	0.951	0.944	0.959	0.951
6	Canada	0.950	0.942	0.955	0.952
7	Japan	0.949	0.950	0.948	0.948
8	United States	0.948	0.930	0.954	0.961
9	Switzerland	0.947	0.942	0.947	0.952
10	Netherlands	0.947	0.933	0.957	0.951

Table 5: Robust Comparisons for r = 0.25

Country		Norway	Iceland	Australia	Ireland	Sweden	Canada	Japan	USA	Switzerland	Netherlands
	Rank	1	2	3	4	5	6	7	8	9	10
Norway	1										
Iceland	2										
Australia	3	C_r									
Ireland	4	C_r									
Sweden	5	C_r	C_r	C_r							
Canada	6	C_r	C_r	C_r							
Japan	7	C_r	C_r								
USA	8	C_r	C_r		C_r						
Switzerland	9	C_r	C_r	C_r							
Netherlands	10	C_r	C_r	C_r	C_r	C_r					

Table 6: Measure of Robustness (in Percentage)

Country		Norway	Iceland	Australia	Ireland	Sweden	Canada	Japan	USA	Switzerland	Netherlands
	Rank	1	2	3	4	5	6	7	8	9	10
Norway	1										
Iceland	2	20									
Australia	3	35	19								
Ireland	4	86	14	4							
Sweden	5	53	94	100	11						
Canada	6	61	100	60	14	14					
Japan	7	28	34	23	9	7	2				
USA	8	77	28	17	67	5	3	1			
Switzerland	9	49	100	41	16	17	20	6	2		
Netherlands	10	100	68	57	47	25	13	4	7	1	