Session Number: Parallel Session 2C
Time: Monday, August 23, PM

Paper Prepared for the 31st General Conference of
The International Association for Research in Income and Wealth

St. Gallen, Switzerland, August 22-28, 2010

Rank Robustness of Composite Indices: Dominance and Ambiguity

James E. Foster, Mark McGillivray, and Suman Seth

For additional information please contact:

Name: Suman Seth

Affiliation: Oxford Poverty & Human Development Initiative

Email Address: Suman.Seth@qeh.ox.ac.uk,

Rank Robustness of Composite Indices: Dominance and Ambiguity*

James E. Foster, *George Washington University* and *University of Oxford*†
Mark McGillivray, *Australian Agency for International Development*‡
Suman Seth, *Oxford Poverty & Human Development Initiative*§

Abstract

Many common multidimensional indices take the form of a "composite index" which aggregates linearly across several dimensions using a vector of weights. Judgments rendered by composite indices are contingent on the selected vector of weights. A comparison could be reversed at another plausible vector; or, alternatively, the comparison might be robust to variations in weights. This paper presents general robustness criteria to discern between these two situations. We define a robustness quasiordering requiring unanimity for a set of weighting vectors, and utilize methods from Bewley's (2002) model of Knightian uncertainty to characterize this quasiordering. We then focus on a particular set of weighting vectors suggested by the epsilon-contamination model of ambiguity; this allows the degree of confidence in the initial weighting vector to vary analogous to Ellsberg (1961). We provide a practical vector-valued representation of the resulting "epsilon robustness" quasiordering, and propose a related numerical measure by which the robustness of any given comparison can be gauged. Our methods are illustrated using data on the Human Development Index from the 2006 Human Development Report.

JEL Classifications: I31, O12, O15, C02.

Key Words: Composite index, robust comparisons, dominance, ambiguity, epsilon-contamination, Knightian uncertainty, Human Development Index.

July, 2010

* This paper has benefited from discussions with Sabina Alkire, Bart Lipman, Nora Markova, Farhad Noorbakhsh, Tony Shorrocks, John Weymark the participants of the Oxford Poverty and Human Development Initiative (OPHI) "Workshop on Weighting" May 2008 at the University of Oxford, the 2009 American Economic Association Annual Meeting, Johns Hopkins University Theory Seminar, and the 10th International Meeting of the Society for Social Choice and Welfare, 2010. 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.

† 1957 E Street, NW Suite 502, George Washington University, Washington, D.C., 20052, USA +1 202 994 8195, fosterje@gwu.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.

§ Oxford Poverty & Human Development Initiative (OPHI), Queen Elizabeth House (QEH), Department of International Development, 3 Mansfield Road, Oxford OX4 1SD, UK, Suman.Seth@qeh.ox.ac.uk, +44 (0)7725989603.

1. Introduction

Composite indices, or weighted linear aggregation methods, are commonly used in social and economic assessments involving multiple dimensions. A prominent example is given by the Human Development Index, which can be viewed as an average of three dimensional achievement levels – one representing income, a second education and a third health.[1] Another example is provided by annual college rankings that combine multiple attributes, including tests scores and educational facilities, into a composite index to measure quality of a university or an academic department. The associated rankings generate substantial interest, and may have significant impact on resource allocation and other economically relevant outcomes.

Each ranking generated by a composite index is, however, contingent on the choice of initial weights; a slight variation in weights may well lead to a different judgment between a given pair of alternatives. And while there are a number of plausible methods for selecting initial weights, none is so compelling or precise as to exclude all alternative weights. Given the possibility that any judgment rendered by a composite index could be reversed, it would be useful to have additional information on the robustness of comparisons with respect to changes in the weights. This is the motivation for the present paper, which presents new and tractable methods for evaluating the robustness of rankings generated by composite indices.

We consider two related approaches - one that is based on quasiorderings and a second based on a numerical measure of robustness. The first approach is similar to techniques used in the evaluation of poverty, in which poverty comparisons are subjected to robustness checks over a range of lines; and is also closely linked the way the Lorenz quasiordering tests the robustness of comparisons generated by a single inequality measure.[2] A "robustness quasiordering" is defined based on an a priori specification of a set of weighting vectors. A comparison made by the composite index is said to be robust if it is not reversed for any weighting vector in the set. The analysis draws from structures

[1] Country-based composite indices have proliferated of late, and include indices of sustainability, corruption, rule of law, economic policy efficacy, institutional performance, happiness, human well-being, transparency, globalisation, human freedom, peace and vulnerability.

[2] See Atkinson (1970, 1987), Foster and Shorrocks (1988a,b), and Foster and Sen (1997) for related discussions. Other examples include stochastic dominance (Bawa 1975), the usual Pareto dominance ranking, and social choice models with partial comparability (Sen 1970).

found in the literatures on multiple prior models of ambiguity (Gilboa and Schmeidler, 1989) and on Knightian uncertainty (Bewley, 2002). Motivated by a result from Bewley, we characterize this specific form of quasiordering from among all possible relations that might be used to check robustness. We further show a straightforward link between the quasiordering and the Gilboa-Schmeidler maximin criterion.

In order to implement the robustness approach in practice, a specific set of weighting vectors must be selected. One convenient possibility is suggested by the epsilon-contamination model from decision theory: the set of weighting vectors that can be expressed as a convex combination of the initial weighting vector and any other weighting vector, where the coefficients on each are respectively 1-ε and ε. The coefficient 1-ε is interpreted as the level of confidence in the initial vector; the "contamination" parameter ε is a direct measure of the size of the set around this vector. Greater confidence in the initial weighting vector is reflected in a lower level of ε-contamination and a smaller set. Our main result characterizes the resulting ε-robustness relation and demonstrates that it has a tractable vector-valued representation; in other words, one can check robustness by mapping alternatives to a vector space ordered by vector dominance.[3]

A second approach seeks to obtain a continuous measure of the robustness of a given comparison, rather than employing a zero-one test. We construct a measure comprised of two elements: the difference between the levels of the composite index of the two alternatives; and the maximal "contrary" difference across all weighting vectors. We show that the measure has an intuitive interpretation as the maximal level of contamination ε for which comparison is ε-robust, and note its relationship to the maximin function.

The purpose of this paper is to present methods that can be used in practice to help refine our understanding of rankings generated by composite indicators. We therefore present an extended illustration of our methods based on country data obtained from the 2006 Human Development Report. The Human Development Index or HDI is a

[3] See Foster (1993, 2010) and Foster and Sen (1997, p. 205-207) for a detail discussion on vector valued representation.

composite index that aggregates over three dimensions representing income, education and health achievements in a given country using equal weights. We find that a significant proportion of pairwise comparisons across countries are fully robust, while other comparisons are quite sensitive to variations in the weighting vector. We fix the value of ε and provide examples of comparisons that satisfy and fail our ε-robustness test. We calculate a table of robustness levels for comparisons among ten countries with the highest HDI levels, and then examine the overall prevalence of the various levels of robustness in the 2006 data and two previous years. Our example suggests that these techniques can be readily employed to help interpret rankings generated by composite indices.

The rest of the paper is structured as follows. Section 2 provides the notation and definitions used in the paper. A formal treatment of the general robustness quasiorderings as well as ε-robustness is provided in Section 3. Section 4 constructs the robustness measure and demonstrates its relationship to ε-robustness. Section 5 provides an application to inter-country comparisons of the Human Development Index. The paper concludes in Section 6.

2. Notation and Definitions

Let $X \subseteq R^D$ denote the nonempty set of alternatives to be ranked, where each alternative is represented as a vector $x \in X$ of achievements in $D \geq 2$ dimensions. For $a, b \in R^D$, 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 $a \neq b$, this situation is denoted by $a > b$; while $a >> b$ indicates that $a_d > b_d$ for $d = 1, \ldots, D$. Let $\Delta = \{w \in R^D\colon w \geq 0 \text{ and } w_1 + \ldots + w_D = 1\}$ be the *simplex of weighting vectors*. A *composite index* $C\colon X \times \Delta \rightarrow R$ combines the dimensional achievements in $x \in X$ using a weighting vector $w \in \Delta$ to obtain an aggregate level $C(x;w) = w{\cdot}x = w_1x_1 + \ldots + w_Dx_D$. In what follows, it is assumed that an *initial weighting vector* $w^0 \in \Delta$ satisfying $w^0 >> 0$ has already been chosen; this fixes the specific composite index $C_0\colon X \rightarrow R$ defined as $C_0(x) = C(x;w^0)$ for all $x \in X$. The associated strict partial ordering of achievement vectors will be denoted by $\boldsymbol{C_0}$, so that $x\ \boldsymbol{C_0}\ y$ holds if and only if $C_0(x) > C_0(y)$. For every $d \in \{1,\ldots,D\}$, we denote the D-dimensional basis vector by v_d, whose dth element is equal to one and the rest of the elements are zero. For example, $v_1 = (1,0,\ldots,0)$, $v_2 = (0,1,0,\ldots,0)$, and so forth.

3. Robust Comparisons

We construct a general criterion for determining when a given comparison $x\ \boldsymbol{C_0}\ y$ is robust. Let $W \subseteq \Delta$ be a nonempty set of weighting vectors. Define the *weak robustness relation* $\boldsymbol{R_W}$ on X by $x\ \boldsymbol{R_W}\ y$ if and only if $C(x,w) \geq C(y,w)$ for all $w \in W$. If both $x\ \boldsymbol{C_0}\ y$ and $x\ \boldsymbol{R_W}\ y$ hold for $w^0 \in W$, then we say that x *robustly dominates* y (*given* w^0 and W), and denote this by $x\ \boldsymbol{C_W}\ y$. In words, the level of the composite index is higher for x than y at w^0, and this ranking is not reversed at any other weighting vector in W. If instead $x\ \boldsymbol{C_0}\ y$ holds, but $x\ \boldsymbol{R_W}\ y$ does not, then this indicates that the ranking $C(x,w^0) > C(y,w^0)$ is *not robust* (relative to the given W) since the initial inequality is reversed at another weighting vector, say, $C(x,w^1) < C(y,w^1)$ for $w^1 \in W$.

The relations $\boldsymbol{R_W}$ and $\boldsymbol{C_W}$ are closely linked with other dominance criteria, including Sen's (1970) approach to partial comparability in social choice and Bewley's (1986) multiple prior model of Knightian uncertainty. Bewley's presentation, in particular, suggests a natural characterization of $\boldsymbol{R_W}$ among all binary relations $\boldsymbol{R}$ on X. Consider the following properties, each of which is satisfied by $\boldsymbol{R_W}$.

Quasiordering (Q): $\boldsymbol{R}$ is transitive and reflexive.
Monotonicity (M): (i) If $x > y$ then $x\ \boldsymbol{R}\ y$; (ii) if $x >> y$ then $y\ \boldsymbol{R}\ x$ cannot hold.
Independence (I): Let $x, y, z, y', z' \in X$ where $y' = \alpha x + (1-\alpha)y$ and $z' = \alpha x + (1-\alpha)z$ for $0 < \alpha < 1$. Then $y\ \boldsymbol{R}\ z$ if and only if $y'\ \boldsymbol{R}\ z'$.
Continuity (C): The sets $\{x \in X \mid x\ \boldsymbol{R}\ z\}$ and $\{x \in X \mid z\ \boldsymbol{R}\ x\}$ are closed for all $z \in X$.

Axiom Q allows $\boldsymbol{R}$ to be incomplete. Axiom M ensures that $\boldsymbol{R}$ follows vector dominance when it applies, and rules out the converse ranking when vector dominance is strict. Axiom I is a standard independence axiom, which requires the ranking between y and z to be consistent with the ranking of y' and z' obtained from y and z, respectively, by a convex combination with another vector x. Finally, Axiom C ensures that the upper and lower contour sets of $\boldsymbol{R}$ contain all their limit points. We have the following characterization, the proof of which is given in the appendix.[4]

Theorem 1: Suppose that X is closed, convex and has a nonempty interior. Then a binary relation $\boldsymbol{R}$ on X satisfies axioms Q, M, I, and C if and only if there exist a non-empty, closed and convex set $W \subseteq \Delta$ such that $\boldsymbol{R} = \boldsymbol{R}_W$.

Thus any robustness relation satisfying the four axioms is generated by pair-wise comparisons of the composite index over some fixed set W of weighting vectors.

The relation $\boldsymbol{R}_W$ has an interesting interpretation in terms of the well-known maxmin criterion of Gilboa and Schmeidler (1989) for multiple priors. Suppose we know that $x\ \boldsymbol{R}_W\ y$ for some nonempty, closed set $W \subseteq \Delta$. By linearity of the composite index, this can be expressed as $C(x - y,w) \geq 0$ for all $w \in W$, or as $\min_{w\in W} C(x - y,w) \geq 0$. The Gilboa-Schmeidler evaluation function $G_W(z) = \min_{w\in W} C(z,w)$ represents the maxmin criterion, which ranks a pair of options x and y by comparing $G_W(x)$ and $G_W(y)$, or the respective minimum values of the composite indicator on the set W. Our robustness ranking $x\ \boldsymbol{R}_W\ y$ is obtained by applying G_W to the *net vector* $(x - y)$ and checking whether the resulting value is nonnegative. Indeed, $x\ \boldsymbol{R}_W\ y$ if and only if $G_W(x - y) \geq 0$.[5]

Theorem 1 shows that under the given axioms, the selection of a robustness criterion reduces to the choice of an appropriate set W of multiple weighting vectors used in $\boldsymbol{R}_W$. But which W should be used? As we argue below, the answer depends in part on the confidence one places in the initial weighting vector w^0. If one has confidence that w^0 is the most appropriate weighting vector, then this would be reflected in the selection of a smaller set W containing w^0. The limiting case of $W = \{w^0\}$ indicates utmost confidence in w^0 and hence entails no robustness test at all: $x\ \boldsymbol{C}_0\ y$ is equivalent to $x\ \boldsymbol{C}_W\ y$. On the other hand, a larger W would suggest less confidence in w^0, a more demanding robustness test $\boldsymbol{R}_W$, and correspondingly fewer robust comparisons according to $\boldsymbol{C}_W$. Clearly $\boldsymbol{C}_{W'}$ is a

[4] All proofs are found in the Appendix.

[5] The maxmin criterion applies when $G_W(x) - G_W(y) \geq 0$, while our robustness criterion holds when $G_W(x\text{-}y) \geq 0$. The maxmin criterion generates a complete relation, but requires comparisons of $C(x,w)$ with $C(y,w')$ for some $w \neq w'$, which is not easily interpreted in the present context. See Ryan (2009) for related discussions of Bewley (1986) and Gilboa and Schmeidler (1989).

subrelation of $\boldsymbol{C_W}$ whenever $W \subseteq W'$. We now investigate the robustness relations for some natural specifications of the set W of allowable weighting vectors.[6]

Full Robustness

We begin with the limiting case where W is the set Δ of all possible weighting vectors, and denote the associated robustness relations by $\boldsymbol{R_1}$ and $\boldsymbol{C_1}$. When $x\ \boldsymbol{C_1}\ y$ holds we say the comparison $x\ \boldsymbol{C_0}\ y$ is *fully robust* since it is never reversed at *any* configuration of weights. Of course, requiring unanimity over all of Δ is quite demanding and consequently $\boldsymbol{C_1}$ is the least complete among all such relations; however, when it applies the associated ranking of achievement vectors is maximally robust.

Consider the vertices of Δ, 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. Clearly $C(x,v_d) = x_d$, which suggests a link between the robustness relations and vector dominance. Indeed, we have the following characterizations of $\boldsymbol{R_1}$ and $\boldsymbol{C_1}$.

Theorem 2: Let $x, y \in X$. Then (*i*) $x\ \boldsymbol{R_1}\ y$ if and only if $x \geq y$ and (*ii*) $x\ \boldsymbol{C_1}\ y$ if and only if $x > y$.

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.

One interesting implication of Theorem 1 is that judgments made by $\boldsymbol{C_1}$ are "meaningful" even 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 $x' = (f_1(x_1),\ldots, f_D(x_D))$ be the resulting transformed achievement vector. It is clear that $x > y$ if and only if $x' > y'$, and consequently, by Theorem 2 we have $x\ \boldsymbol{C_1}\ y$ if and only if $x'\ \boldsymbol{C_1}\ y'$. 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 (as $y'\ \boldsymbol{C_0}$

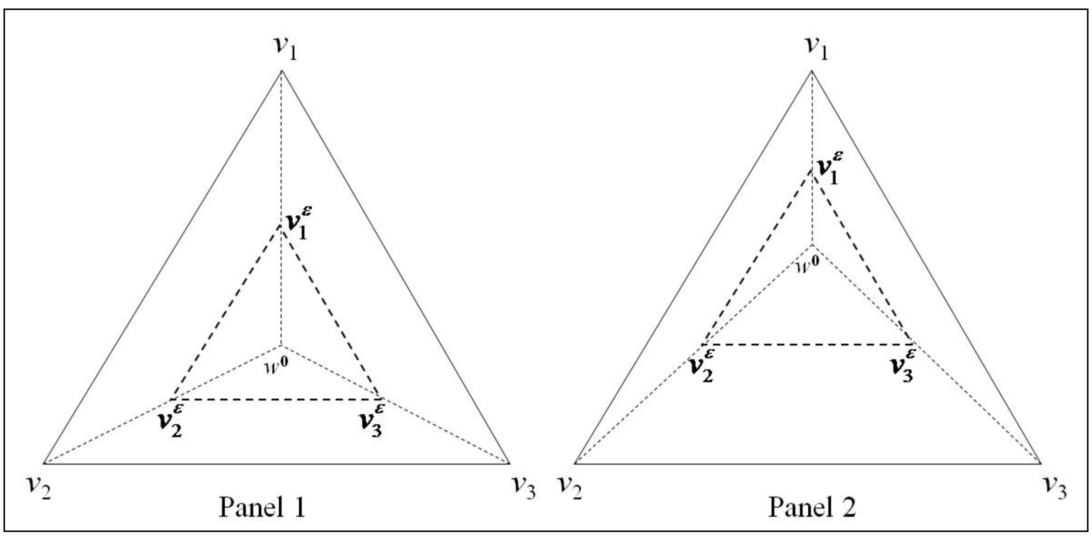
[6] Since $\boldsymbol{C_W}$ is the intersection of $\boldsymbol{C_0}$ and $\boldsymbol{R_W}$, it is a strict partial order (transitive and irreflexive) satisfying conditions I and M.
[7] For a formal discussion of "meaningful statements" see Roberts (1979).

x' is entirely consistent with $x\ \boldsymbol{C_0}\ y$), the fully robust relation $\boldsymbol{C_1}$ is preserved and hence is appropriate for use with ordinal variables.

Epsilon Robustness

Now consider $\Delta_\varepsilon \subseteq \Delta$ defined by $\Delta_\varepsilon = (1\text{-}\varepsilon)\{w^0\} + \varepsilon\Delta$ for $0 \le \varepsilon \le 1$, which is made up of vectors of the form $(1\text{-}\varepsilon)w^0 + \varepsilon w$, where $w \in \Delta$. Parameter value $\varepsilon = 0$ yields $\Delta_0 = \{w^0\}$ and hence the "no robustness" case, while $\varepsilon = 1$ yields $\Delta_1 = \Delta$ or full robustness. Each Δ_ε with $0 < \varepsilon < 1$ is a scaled down version of Δ located so that w^0 is in the same relative position in Δ_ε as it is in Δ. Figure 1 provides examples of Δ_ε for the case of $D = 3$ and $\varepsilon = ¼$, where Panel 1 has $w^0 = (1/3,1/3,1/3)$ and Panel 2 has $w^0 = (3/5,1/5,1/5)$. As noted in the Figure, ε is a measure of the relative size of Δ_ε. Moreover, for a given w^0 the sets are nested in such a way that $\Delta_\varepsilon \subset \Delta_{\varepsilon'}$ whenever $\varepsilon' > \varepsilon$.

Figure 1: Multiple Weighting Vectors: The ε-Robustness Set Δ_ε



The set Δ_ε of weighting vectors can be motivated using the well-known epsilon contamination model of multiple priors commonly applied in statistics and decision theory.[8] In that context, w^0 corresponds to an initial subjective distribution and Δ_ε contains probability distributions that are convex combinations of w^0 and the set of all

[8] See for example, Carlier, Dana, and Shahidi (2003); Chateauneuf, Eichberger, and Grant (2006); Nishimura and Ozaki (2006); Carlier and Dana (2008); Asano (2008); and Kopylov (2009).

objectively possible distributions, where (1-ε) represents the decision maker's level of confidence in w^0 and ε is the extent of the "perturbation" from w^0. The Gilboa-Schmeidler evaluation function G_W then reduces to a form invoked by Ellsberg (1961), namely $G_\varepsilon(z) = (1-\varepsilon)C(z,w^0) + \varepsilon \min_{w\in\Delta} C(z,w)$ using our notation.

Substituting Δ_ε in the definitions of $\boldsymbol{R_W}$ and $\boldsymbol{C_W}$ yields the *ε-robustness relations* $\boldsymbol{R_\varepsilon}$ and $\boldsymbol{C_\varepsilon}$. Since the sets Δ_ε are nested for a given w^0, it follows that $x\ \boldsymbol{C_\varepsilon}\ y$ implies $x\ \boldsymbol{C_{\varepsilon'}}\ y$ whenever $\varepsilon > \varepsilon'$. The rankings clearly require $C(x,w) \geq C(y,w)$ for all w in Δ_ε and hence at each of its vertices $v_d^\varepsilon = (1-\varepsilon)w^0 + \varepsilon v_d$. Define $x^\varepsilon = (x_1^\varepsilon,\ldots,x_D^\varepsilon)$ where $x_d^\varepsilon = C(x,v_d^\varepsilon) = v_d^\varepsilon \cdot x$, and let y^ε be the analogous vector derived from y. The following result characterizes $\boldsymbol{R_\varepsilon}$ and $\boldsymbol{C_\varepsilon}$.

Theorem 3: Let $x, y \in X$. Then (i) $x\ \boldsymbol{R_\varepsilon}\ y$ if and only if $x^\varepsilon \geq y^\varepsilon$ and (ii) $x\ \boldsymbol{C_\varepsilon}\ y$ if and only if $x^\varepsilon > y^\varepsilon$.

Theorem 3 shows that to evaluate whether a given comparison $x\ \boldsymbol{C_0}\ y$ is ε-robust, one need only compare the associated vectors x^ε and y^ε. If each component of x^ε is at least as large as the respective component of y^ε, then the comparison is ε-robust; if any component is larger for y^ε than x^ε, then the comparison is not. Checking whether the x^ε vector dominates y^ε is equivalent to requiring the inequality $C(x,w) \geq C(y,w)$ to hold for each vertex $w = v_d^\varepsilon$ of the set Δ_ε. Note further that x^ε is a convex combination of the vectors $(C_0(x),\ldots,C_0(x))$ and x, namely, $x^\varepsilon = (1-\varepsilon)(C_0(x),\ldots,C_0(x)) + \varepsilon x$, so that when $\varepsilon = 1$ we obtain the condition $x \geq y$ in Theorem 2, while when $\varepsilon = 0$, the condition reduces to a simple comparison of $C_0(x)$ and $C_0(y)$.

4. Measuring Robustness

Our method of evaluating the robustness of comparison the $x\ \boldsymbol{C_0}\ y$ fixes a set Δ_ε of weighting vectors and confirms that the ranking at w^0 is not reversed at any other $w \in \Delta_\varepsilon$, in which case the associated ε-robustness relation applies. Theorem 3 provides simple conditions for checking whether $x\ \boldsymbol{C_\varepsilon}\ y$ holds. The present section augments this approach by formulating a robustness measure that associates with any comparison $x\ \boldsymbol{C_0}\ y$ a number $r \in [0,1]$ that indicates 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 $A = C(x;w^0) - C(y;w^0) > 0$, or the difference between the composite value of x and the composite value of y at the initial weighting vector w^0. Intuitively, A is an indicator of the strength of the dominance of x over y at the initial weighting vector. The second is $B = \max_{w\in\Delta}[C(y;w) - C(x;w), 0]$, 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 $w \in \Delta$ and there is no contrary difference. Consequently, $B = 0$. On the other hand, when the comparison is not fully robust, then $C(y;w) - C(x;w) > 0$ for some $w \in \Delta$, and hence $B = \max_{w\in\Delta}[C(y;w) - C(x;w)] > 0$. B is the worst-case estimate of how far the original difference at w^0 could be reversed at some other weighting vector.

We propose $r = A/(A+B)$ as a measure of robustness. Notice that when the initial comparison $x\ C_0\ y$ is fully robust, then $B = 0$ and hence $r = 1$. Alternatively, when the initial comparison is *not* fully robust and $B > 0$, the measure r is strictly increasing in the magnitude of the initial comparison A, and strictly decreasing in the magnitude of the contrary worst-case evaluation B. In addition, if A tends to 0 while B remains fixed, the measure of robustness r will also tend to 0. These characteristics accord well with an intuitive understanding of how A and B might affect robustness.

Practical applications of r may be hampered by the fact that it requires a maximization problem to be solved, namely, $\max_{w\in\Delta}[C(y;w) - C(x;w)]$. 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 where the difference $C(y;w) - C(x;w)$ becomes $y_d - x_d$. Consequently, $B = \max_d(y_d - x_d)$, or the maximum coordinate-wise difference between y and x. The measure r can be readily derived using this equivalent definition.

Now what is the relationship between the robustness measure r and the relation C_ε developed in the previous section? The following theorem provides the answer.

Theorem 4: Suppose that $x\ C_0\ y$ for $x, y \in X$ and let r be the robustness level associated with this comparison. Then the ε-robustness relation $x\ C_\varepsilon\ y$ holds if and only if $\varepsilon \leq r$.

Raising ε leads to a more demanding robustness criterion and a more incomplete relation C_ε. Theorem 4 identifies r as the *maximal* ε for which $x\ C_\varepsilon\ y$ holds, and hence the largest set Δ_ε for which the original comparison is not reversed. Alternatively, it corresponds to the lowest level of confidence (1-ε) for which the Gilboa-Schmeidler (or Ellsburg) evaluation function of the net achievement vector (x–y) is always nonnegative; i.e., the largest ε for which $G_\varepsilon(x–y) = (1-\varepsilon)C(x–y,w^0) + \varepsilon \min_{w\in\Delta}C(x–y,w) \geq 0$.

5. Application

We illustrate our methods using data from the 2004 Human Development Index (HDI) dataset as published in the 2006 Human Development Report.[9] The HDI is a composite index $C(x;w^0)$ constructed by taking the simple average of three dimension-specific indicators (of education, health and income) and hence $w^0 = (1/3,1/3,1/3)$ is the initial weighting vector. Table 1 provides information on the top ten countries according to the HDI, including their rankings and HDI values.[10] This yields the C_0 relation over these 10 countries, but says nothing about the robustness of any given judgment.

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

Rank	Country	HDI
1	Norway	0.965
2	Iceland	0.960
3	Australia	0.957
4	Ireland	0.956
5	Sweden	0.951
6	Canada	0.950
7	Japan	0.949
8	United States	0.948
9	Switzerland	0.947
10	Netherlands	0.947

[9] Our underlying dataset was obtained directly from the UNDP and is less severely rounded off than the published data.
[10] Due to rounding off, the HDI levels of Switzerland and Netherlands appear to be equal; in fact, Switzerland has a slightly higher HDI than Netherlands.

Table 2 focuses on three specific comparisons; the middle columns provide the dimensional achievements x_1, x_2, and x_3 needed to ascertain whether full robustness C_1 obtains. The achievement vector for Australia dominates the achievement vector for Sweden, and hence by Theorem 2 this comparison is fully robust. However, the comparison for Iceland and USA reverses in the income dimension, while the Ireland and Canada comparison has a reversal in health, and so neither of these rankings is fully robust. Observe that the HDI margin between Australia and Sweden (0.006) is identical to the margin for Ireland and Canada, and yet the robustness characteristics of the two comparisons are quite different. The HDI margin between Iceland and USA is twice as large (0.012) and yet it too is not fully robust.

Table 2: Robustness of Three HDI Comparisons

Rank	Country	HDI	Hel x_1	Edu x_2	Inc x_3	Hel $x_1^{0.25}$	Edu $x_2^{0.25}$	Inc $x_3^{0.25}$
3	Australia	0.957	0.925	0.993	0.954	0.949	0.966	0.956
5	Sweden	0.951	0.922	0.982	0.949	0.944	0.959	0.951
2	Iceland	0.960	0.931	0.981	0.968	0.953	0.965	0.962
8	USA	0.948	0.875	0.971	0.999	0.930	0.954	0.961
4	Ireland	0.956	0.882	0.990	0.995	0.937	0.964	0.966
6	Canada	0.950	0.919	0.970	0.959	0.942	0.955	0.952

The final columns of Table 2 give the entries of the associated x^{ε} vectors for ε = 0.25 in order to ascertain ε-robustness of the comparisons. A quick evaluation in terms of vector dominance reveals that both the Australia/Sweden and the Iceland/USA comparisons are ε-robust, but the reversal in the Ireland/Canada comparison implies that ε-robustness does not hold for this ranking when $\varepsilon = 0.25$. By Theorem 3 we know that there are weighting vectors in Δ_ε at which Canada has a larger composite index level than Ireland.

The levels of robustness can also be calculated for each of these comparisons. The Australia/Sweden comparison is fully robust, with $A = 0.006$ and $B = 0$, and hence r = 100%. The Iceland/USA comparison has $A = 0.012$ and $B = 0.031$, and hence r = 28%. In contrast, the Ireland/Canada ranking has $A = 0.006$ and $B = 0.037$, and therefore r = 14%. Table 3 presents the level of robustness of pair-wise comparisons for the top ten countries in the HDI ranking. For every cell below the diagonal the "column country" of the cell has a higher ranking according to $\boldsymbol{C_0}$ than the "row country". The number in the cell indicates the level of robustness of the associated comparison, expressed in percentage terms. Out of the 45 pair-wise comparisons, four are fully robust as denoted by r = 100%, while 20 of them, or 44.4 percent, are robust at r = 25%. For the entire dataset of 177 countries for the same year, 69.7 percent of the comparisons are fully robust and about 92 percent are robust for r = 25%.

Table 3: Measure of Robustness (%)

Country		NOR	ISL	AUS	IRL	SWE	CAN	JPN	USA	SWI	NLD
	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	–

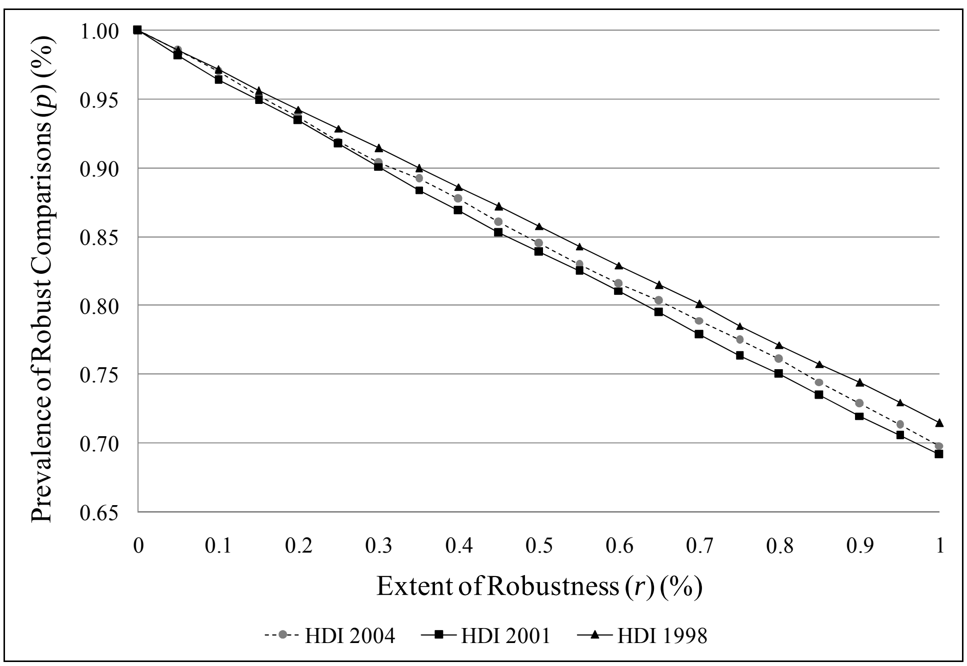
Prevalence of Robust Comparisons

The let us shift our focus 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.

(The theoretical part has been shortened and some notations changed) Suppose the initial weighting vector is w^0 and there is a dataset $\hat{X}$ containing n observations.

Without loss of generality, we enumerate the elements of $\hat{X}$ as $x^1, x^2, \ldots, x^n$ where $C_0(x^1) \geq C_0(x^2) \geq \ldots \geq C_0(x^n)$. The analysis can be simplified by assuming that no two observations in $\hat{X}$ have the same initial composite value, so that $C_0(x^1) > C_0(x^2) > \ldots > C_0(x^n)$.[11] There are $m = n(n-1)/2$ 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 3. We summarize robustness levels in P in a way that reflects the *entire* distribution. 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 m 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.[12]

Figure 2: Prevalence Functions of HDI for Various Years



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

Figure 2 depicts the prevalence functions obtained from HDI datasets for three different years, which uses equal weights across three dimensions to rank 177, 175, and 174 countries, respectively.[13] Several initial observations can be made from the prevalence functions given in Figure 2. Each graph 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 0, 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 that can be compared using $\boldsymbol{C_1}$ and hence is fully robust. There is a wide variation in $p(1)$ across datasets. It is reasonably large for all the HDI examples, with $p(1)$ being about 69.8% in 2004, 69.2% in 2001, and 71.5% in 1998. The shapes of the $p(r)$ functions are essentially linear for all three HDI dataset. These regularities of prevalence functions are worth examining from a more theoretical perspective. If we set a target of 25 percent robustness, then on an average 92 percent to 93 percent of the HDI comparisons are robust.

6. Conclusion

Rankings arising from composite indices 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 when there is a disagreement over the set of weights they employ.

This paper examines a variable-weight robustness criterion for composite indices, drawing from structures found in the literatures on multiple prior models of ambiguity (Gilboa and Schmeidler, 1989) and on Knightian uncertainty (Bewley, 2002). The idea is to check how robust a ranking is to the variation or contamination in the initially chosen weights or the prior. A ranking is considered robust if the ranking is not reversed for a set of feasible weights around the initially chosen weights and not robust if the ranking is reversed for any weights in the set. This idea is analogous to the concept of partial ordering. It is argued that the size and the shape of the set should depend on the confidence one places on the initial choice. If one is supremely confident about the initial

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

choice, then the set reduces to a singleton set containing the initially chosen wights only and there is no reason for checking robustness. On the other hand, if one is not confident about the initial choice at all, then the set contains all possible weights and fewest robust comparisons can be made.

This paper proposes an intermediate approach where one is partially confident about the initial choice and a smaller set of weights around the initial weights is used for checking robustness of ranking. It characterizes the resulting robustness relations for various sets of weighting vectors. An illustration of how these relationships moderate the complete ordering generated by the composite indices is provided. A measure by which the robustness of a given comparison may be gauged is then proposed, and illustrated using the Human Development Index (HDI).

Few other studies have also delved into the issues of robustness of ranking. Chercheye et al. (2008) analysed the issue in terms of HDI ranking using an approach based on Generalized Lorenz ordering. Our approach significantly differs from. The two main areas where their approach diverges from that of ours are that their approach is applicable to a particular type of normalization only and assumes that dimensions are anonymous to each other. We, on the other hand, are more interested in dimension specific pair-wise comparison. There is another branch of studies that uses sensitivity analysis to verify the strength of comparisons. The sensitivity analysis is different from the robustness analysis in the sense that it estimates confidence intervals around each composite index depending on different scenarios. If the confidence intervals of two composite indicators do not overlap, an unambiguous comparison is possible. See for example Saisana et al. (2005).

In the current paper, we primarily focus on the ranking of composite indices that are linear on dimensional achievements. However, it can be easily shown that our approach may be extended to the composite indices that are weighted average of the monotonic transformation of the dimensional achievements. Examples include the indices such as the Human Poverty Index and the classes of indices proposed by Bourguignon (2003), and Foster, López-Calva, and Székely (2005).

References

Arrow K., A. Sen, K. Suzumura (2010), "Freedom, Opportunity and Wellbeing", forthcoming in Handbook of Social Choice and Welfare, North Holland.

Asano, T. (2008), "Portfolio inertia and epsilon-contaminations", *Theory and Decision*, Vol. 68, 341-365.

Atkinson, A. B. (1970), "On the measurement of inequality", *Journal of Economic Theory*, Vol. 2, 244-263.

Atkinson, A. B. (1987), "On the Measurement of Poverty", *Econometrica*, Vol. 2, 244-263.

Bawa, V. (1975), "Optimal rules for ordering uncertain prospects", *Journal of Financial Economics*, Vol. 2, 95-121.

Bewley, T. F. (2002), "Knightian decision theory. Part I", *Decisions in Economics and Finance*, Vol. 25, 79-110.

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

Carlier, G. and R. A. Dana (2008), "Two-persons efficient risk-sharing and equilibria for concave law-invariant utilities" *Economic Theory*, Vol. 36, 189-223.

Carlier, G., R. A. Dana, and N. Shahidi (2003), "Efficient insurance contracts under epsilon-contaminated utilities", *The GENEVA Papers on Risk and Insurance – Theory*, Vol. 28, 59-71.

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

Chateauneuf, A., J. Eichberger, and S. Grant (2006), "Choice under uncertainty with the best and worst in mind", *Journal of Economic Theory*, Vol. 137, 538-567.

Cherchye L., W. Moesen, N. Rogge, T. V. 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*, Vol. 59**,** 239–251.

Cherchye L., E. Ooghe and T. V. Puyenbroeck (2008), "Robust human development rankings", *Journal of Economic Inequality*, Vol. 6, 287-321.

Ellsberg, D. (1961), "Risk, ambiguity, and the savage axioms", *Quarterly Journal of Economics*, Vol. 75, 643–669.

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.

Foster J. E. (1983), "Notes on Effective Freedom", mimeographed; presented at the Stanford Workshop on Economic Theories of Inequality, sponsored by the McArthur Foundation.

Foster J. E. and A. Sen (1997), "On Economic Inequality: After a Quarter Century", Annex to the enlarged edition of *On Economic Inequality*, by Amartya Sen, Clarendon Press, 1997.

Foster J. E. and A. F. Shorrocks (1988a), "Poverty orderings and welfare dominance", *Social Choice and Welfare*, Vol. 5, 179-198. 1988.

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

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

Kopylov, I. (2009), "Choice deferral and ambiguity aversion", *Theoretical Economics*, Vol. 4, 199-225.

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

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

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

McGillivray, M. (1991), "The Human Development Index: yet another redundant composite development Indicator?", *World Development*, Vol. 19, 1461-1468.

McGillivray M. and F. Noorbakhsh (2004), "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 H. White (1993), "Measuring development? The UNDPs Human Development Index", *Journal of International Development*, Vol. 5, 183-192.

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

Permanyer I. (2007), "On the Robust Measurement of Well-being in a Gender Perspective" Ph.D. Dissertation, Mimeo, Universitat Autònoma de Barcelona.

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.

Ryan, M. J., (2009), "Generalizations of SEU: a geometric tour of some non-standard models", *Oxford Economic Papers*, Vol. 61, 327-354.

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

Sen, A. (1970), "Collective Choice and Social Welfare", San Francisco: Holden Day.

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

Heritage Foundation (2008), "Index of Economic Freedom", Washington, D.C: The Heritage Foundation.

Appendix

***Proof of Theorem* 1**. Let $\boldsymbol{R}$ be a binary relation on a set X that is closed, convex, and has some z in its interior.

If $\boldsymbol{R} = \boldsymbol{R}_W$ for some non-empty, closed, and convex $W \subseteq \Delta$, then it is immediate that $\boldsymbol{R}_W$ satisfies Q, M, I, and C.

Conversely, suppose that $\boldsymbol{R}$ satisfies Q, M, I, and C. Define $U = \{x \in X: x \boldsymbol{R} z\}$ be the upper contour set of $\boldsymbol{R}$ at z. We know that $z \in U$ by Q and U is closed by C. Moreover, we can show that U is convex. Pick any $x, y \in U$. Let $x' = \alpha x + (1 - \alpha)y$ for some α with $0 < \alpha < 1$. Then, where $z' = \alpha z + (1 - \alpha)y$, we have $x', z' \in X$ and by axiom I it follows that $x' \boldsymbol{R} z'$. Moreover, by a second application of I, it follows from $y \boldsymbol{R} z$ that $z' \boldsymbol{R} z$. Therefore, by Q we have $x' \boldsymbol{R} z$ and so U is convex.

Since, z is in the interior of X, there exists $\varepsilon > 0$ such that $N_\varepsilon = \{x \in R^D: \|x - z\| \leq \varepsilon\} \subseteq X$. Define $U_\varepsilon = U \cap N_\varepsilon$ and note that it is compact, convex, and contains z, so that the set $K_\varepsilon = \{z\} - U_\varepsilon$ is compact, convex, and contains 0. Let $K = \text{Cone } K_\varepsilon$ be the cone generated by K_ε. It is immediate that K is closed, compact, and contains 0. We can state that K has the property that for $x, y \in X$ we have $x \boldsymbol{R} y$ if and only if $y - x \in K$. To see this, let $x, y \in X$ and select $\alpha > 0$ small enough that z' satisfying $z = \alpha y + (1 - \alpha) z'$ lies in N_ε and $x' = \alpha x + (1 - \alpha) z'$ is also in N_ε. Clearly, $z - x' = \alpha(y - x)$ for $\alpha > 0$. So if $x \boldsymbol{R} y$, we know that $x' \boldsymbol{R} z$ by I, and hence $z - x' \in K$ which implies $y - x \in K$. On the other hand, if $y - x \in K$, then since $z - x' \in K$, we have $x' \boldsymbol{R} z$ so that $x \boldsymbol{R} y$ by I, establishing the result.

Now let $P = \{p \in R^D: p{\cdot}k \leq 0 \text{ for all } k \in K\}$ be the polar cone of K, so that by standard results on polar cones, P is closed and convex. It is clear that $P \subseteq R^D_+$, since by monotonicity, we have $-v_d \in K$ and so $p{\cdot}(-v_d) \leq 0$ and $p_d \geq 0$, where v_d is the D-dimensional usual basis vector for co-ordinate d. In addition, we can show that P contains at least one element $p \neq 0$. Indeed, it is clear from M that K contains no $k >> 0$ (otherwise, we would have $x << z$ with x ***R*** z). Then, $K \cap R^D_{++} = \emptyset$ and since both sets are convex, we can apply the Minkowski separation theorem to find $p^0 \neq 0$ in P. Let $W = \Delta \cap P$, so that cone $W = P$. Clearly, K is the polar cone of both P and W, hence, $K = \{t \in R^D: w{\cdot}t \leq 0$ for all $w \in W\}$.

We now show that $\boldsymbol{R} = \boldsymbol{R}_W$. If x ***R*** y, then $y - x \in K$ and so $w(y - x) \leq 0$ for all $w \in W$, hence $x \boldsymbol{R}_W y$. Conversely, if $x \boldsymbol{R}_W y$, then by definition we have $w(y - x) \leq 0$ for all $w \in W$, hence $x - y \in K$ or x ***R*** y. □

***Proof of Theorem* 2**. 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 $w \in \Delta$, 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 $x \geq y$. □

***Proof of Theorem* 3**. We need only verify that $x \boldsymbol{C_0} y$ and $x^\varepsilon \geq y^\varepsilon$ imply $x \boldsymbol{C}_\varepsilon y$. Pick any $w \in \Delta_\varepsilon$, and note that since Δ_ε is the convex hull of its vertices, w can be expressed as a convex combination of $v_1^\varepsilon,...,v_D^\varepsilon$, say $w = \alpha_1 v_1^\varepsilon + \ldots + \alpha_D v_D^\varepsilon$ where $\alpha_1 + \ldots + \alpha_D = 1$ and $\alpha_d \geq 0$ for $d = 1,\ldots,D$. But then $C(x;w) = w{\cdot}x = \alpha_1 v_1^\varepsilon{\cdot}x + \ldots + \alpha_D v_D^\varepsilon{\cdot}x = \alpha_1 x_1^\varepsilon + \ldots + \alpha_D x_D^\varepsilon$, and similarly $C(y;w) = \alpha_1 y_1^\varepsilon + \ldots + \alpha_D y_D^\varepsilon$; therefore $x^\varepsilon \geq y^\varepsilon$ implies $C(x;w) \geq C(y;w)$. Since w was an arbitrary element of Δ_ε, it follows that $x \boldsymbol{C}_\varepsilon y$. □

***Proof of Theorem* 4**. Let $x\ C_0\ y$ and suppose that $0 < \varepsilon \leq r$. By the definition of r, we have $\varepsilon \leq A/(A + B)$ and hence $\varepsilon B \leq (1 - \varepsilon)A$. Pick any $d = 1,\ldots,D$. Then using the definitions of A and B, we see that $\varepsilon(y_d - x_d) \leq (1 - \varepsilon)(w^0{\cdot}x - w^0{\cdot}y)$ and hence $\varepsilon v_d{\cdot}y + (1 - \varepsilon)w^0{\cdot}y \leq \varepsilon v_d{\cdot}x + (1 - \varepsilon)w^0{\cdot}x$. Consequently, $v_d^\varepsilon{\cdot}y \leq v_d^\varepsilon{\cdot}x$, and since this is true for all d, it follows that $x^\varepsilon \geq y^\varepsilon$ and hence $x\ C_\varepsilon\ y$ by Theorem 3.

Conversely, suppose that $x\ C_0\ y$ and yet $r < \varepsilon \leq 1$. Then $(1 - \varepsilon)A < \varepsilon B$ so that $(1 - \varepsilon)(w^0 \cdot x - w^0 \cdot y) < \varepsilon(y_d - x_d)$ for some d, and hence $v_d^\varepsilon \cdot y > v_d^\varepsilon \cdot x$ or $y_d^\varepsilon > x_d^\varepsilon$ for this same d. It follows, then, that $x^\varepsilon \geq y^\varepsilon$ cannot hold, and neither can $x\ C_\varepsilon\ y$ by Theorem 3. □