Multidimensional inequality and multidimensional generalized entropy measures: An axiomatic derivation

Kai-yuen Tsui

Department of Economics, Chinese University of Hong Kong, Shatin, Hong Kong
(e-mail: b040785@mailserv.cuhk.edu.hk)

Abstract. This paper generalizes the axiomatic approach to the design of income inequality measures to the multiattribute context. While the extension of most axioms considered desirable for inequality indices is straightforward, it is not entirely clear when a situation is more unequal than another when each person is characterised by a vector of attributes of wellbeing. We explore two majorization criteria which are partial orders ranking distributions of attributes by their degree of inequality. The two criteria are motivated by the Pigou-Dalton Transfer Principle in the unidimensional context and its equivalent formulation. These criteria gauge inequality loosely speaking with respect to the dispersion of the multidimensional distribution of the attributes. They, however, fail to address a different dimension of multivariate inequality pertaining to an increase in the correlation of the attributes. In this connection, this paper introduces a correlation-increasing majorization criterion proposed by Boland and Proschan (1988). Finally, in conjunction with other axioms commonly invoked in the literature on inequality, the majorization criteria lead inexorably to the class of multidimensional generalized entropy measures.

1. Introduction

Inequality among a group of people, nations or regions has often been measured in terms of income (see, e.g., Kolm 1976a, b; Atkinson 1970; Sen

1973; Cowell and Kuga 1981; Cowell 1980; Shorrocks 1980, 1984; Ebert 1988a, b). However, many economists have argued that income is a poor indicator of welfare and should be supplemented by other attributes of well-being such as indicators of health and literacy. The basic needs approach in development economics contends that rapid economic growth may not necessarily be accompanied by improvements in health and education, (see, e.g., Streeten 1981). In recent years, Sen has launched a fundamental critique of utilizing income as the sole measure of well-being, (e.g., Sen 1985, 1987, 1992). In the context of inequality measurement, the views summarized inevitably lead to the use of multidimensional inequality indices.

In the multiattribute context, the first problem confronting researchers in the measurement of inequality is the meaning of multidimensional inequality. The pioneering papers by Kolm (1977) and Atkinson and Bourguignon (1982) explored the theoretical foundation of multidimensional inequality.[1] They however stopped short of explicitly deriving inequality indices.[2] A small but expanding literature on multidimensional inequality indices has gradually emerged in recent years.[3]

One objective of this paper is to extend the axiomatic treatment of income inequality indices to the multiattribute context. When is one situation more unequal than another in the multiattribute context? This paper explores two multivariate majorization criteria, namely, Uniform Pigou-Dalton and Uniform Majorization, that are generalizations of univariate majorizations. They are partial orders ranking distributions of attributes by their degrees of inequality. Unlike their univariate counterparts, these criteria are not equivalent.

As shown below, the two multivariate majorizations fail to rank two matrices of attributes as more (or less) unequal if there is an increase in the correlation of the attributes among the persons in question. Let $X = \begin{pmatrix} 2 & 4 & 6 \\ 1 & 5 & 2 \end{pmatrix}$ and $Y = \begin{pmatrix} 2 & 4 & 6 \\ 1 & 2 & 5 \end{pmatrix}$ be two distributions of two attributes among three persons; the jth row is the distribution of the jth attributes among the three persons and the ith column is then the amounts of attributes possessed by the ith person. Y is derived from X by switching the amounts of the second attribute between the second person and the third person. After

---

[1] See also Marshall and Olkin (1979, chapter 10); Mosler (1993, 1994); Koshevoy (1995); Koshevoy and Mosler (1996, 1997); Joe and Verducci (1993).

[2] The multidimensional extension of the ethical indices of inequality is already embedded in Kolm (1977) and Atkinson and Bourguignon (1982). Indeed, Kolm (1977) already suggested in a footnote of his paper how to extend the concept of equally distributed equivalent income.

[3] See, e.g., Maasoumi (1986, 1989); Maasoumi and Jeong (1985); Maasoumi and Nickelsburg (1988); Tsui (1995); Koshevoy and Mosler (1996, 1997); a number of articles on multidimensional inequality have been collected in a special issue of the Journal of Econometrics edited by Slottje (1989). For a recent survey, see Maasoumi (1993).

this transfer, the third person has more of the two attributes than the second person. It seems compelling to argue that X is more equal than Y. This paper offers a generalization of such a concept making use of a majorization criterion put forward by Boland and Proschan (1988).

The next important result of this paper is to demonstrate that a set of axioms leads inexorably to a class of indices that is a multidimensional generalization of the class of generalized entropy (GE) measures:

$$I(x_1, \cdots, x_n) = \begin{cases} \frac{1}{nc(c-1)} \sum\limits_{i=1}^{n} \{(x_i/\mu)^c - 1\}, & c \neq 0, 1 \\ \frac{1}{n} \sum\limits_{i=1}^{n} (x_i/\mu) \log(x_i/\mu), & c = 1 \\ \frac{1}{n} \sum\limits_{i=1}^{n} \log(\mu/x_i), & c = 0 \end{cases},$$

where n is the number of individuals, x_i (a scalar) is the income of the ith individual, μ is the mean income and c is a parameter which determines transfer sensitivity; (see, e.g., Bourguignon 1979; Cowell 1980; Cowell and Kuga 1981; Shorrocks 1980, 1984; Shorrocks and Foster 1987; Ebert 1988a, b).

In section two, those axioms are introduced which can be immediately generalized from the unidimensional context. Majorization criteria, i.e., partial orders ranking matrices of attributes in terms of the degree of inequality, are scrutinized and their relationships explored in section three. Functional characterizations of indices commensurate with the majorization criteria are the subject of section four. It is shown that a set of axioms leads to a class of multidimensional generalized entropy measures. We conclude by pointing out further research directions.

2. The axioms

Before introducing the axioms, an explanation of our notation is in order. $\mathbb{R}^p$, $\mathbb{R}^p_+$ and $\mathbb{R}^p_{++}$ denote the Euclidean p-space, the subset of all nonnegative vectors, and the subset of all strictly positive vectors respectively; $\mathbb{N} = \{1, 2, 3, \ldots\}$. Let $k \geq 1$ be the number of attributes, n the number of persons and $\mathfrak{M}(n)$ the set of $k \times n$ matrices with positive elements. Let the ijth element x_{ij} of $X \in \mathfrak{M}(n)$ be the ith attribute of the jth individual; the ith column and the jth row of X are denoted by $\underline{x}^i$ and $\underline{x}_j$ respectively. $\underline{x}^i$ is then a vector of attributes of the ith person, whereas $\underline{x}_j$ summarizes the distribution of the jth attribute among n persons. For each attribute, μ_j is the mean value of the jth attribute and $\underline{\mu} = (\mu_1, \ldots, \mu_k)$.

A *multidimensional* inequality index is defined as a continuous real-valued function $I^n: \mathfrak{M}(n) \to \mathbb{R}, X \mapsto I^n(X), n \geq 2$, satisfying certain axioms, at least one of which is a majorization criterion, i.e., a partial order of matrices of attributes with respect to their degree of inequality. The axioms introduced in this section are motivated by those included in Shorrocks (1984). Most of them are easily generalizable to the multidimensional context. The exception

is the well-known Pigou-Dalton Transfer Principle. When is a matrix X summarizing the distributions of attributes more or less unequal than another matrix Y? This turns out to be a non-trivial question and the next section will scrutinize this issue.

Let us first start with natural extensions of those axioms underlying unidimensional indices in Shorrocks (1984). As usual we assume that a small change in any element of X does not result an abrupt change in $I^n(.)$:

Continuity (CN): $I^n(X)$ is continuous.
Next, the degree of inequality does not depend on the identities of individuals.
Anonymity (AN): $I^n(X) = I^n(X\Pi)$ for all permutation matrices Π.[4]
If every individual has the same bundle of attributes then there is complete equality, i.e.:
Normalization (NM): $I^n((\underline{x}, \underline{x}, \ldots, \underline{x})) = 0$ for any strictly positive $k \times 1$ vector $\underline{x}$.
In conjunction with the majorization criteria in the next section, $I^n(X) \geq 0$. In the unidimensional setting, it is often informative to decompose overall inequality into between-group inequality and within-group inequality. Furthermore, it seems desirable that overall inequality is the sum of the two. The next axiom, which is a generalization of its unidimensional version in Shorrocks (1984), ensures that $I^n(X)$ has such a decomposability property:
Decomposability (DC): $I^n(X^1, X^2) = A(I^{n_1}(X^1), \underline{\mu}_1, n_1, I^{n_2}(X^2), \underline{\mu}_2, n_2)$ for any $X = (X^1, X^2) \in \mathfrak{M}(n)$ where X^1 includes some n_1 columns of X and $\underline{\mu}_1$ is their mean vector; X^2 is made up of the remaining n_2 columns of X, their mean vector being $\underline{\mu}_2$. The aggregative function A is continuous and strictly increasing in $I^{n_1}(X^1)$ and $I^{n_2}(X^2)$.

Another commonly invoked axiom in the measurement of income inequality is replication invariance: duplicating the population without changing the distribution of attributes does not change the value of the inequality index. The multidimensional version of this axiom is:

Replication Invariance (RI): $I^{nr}((X, X, \ldots, X)) = I^n(X)$ for any $r \in \mathbb{N}$.
The next axiom requires that the index is homogenous of degree zero with respect to each row of X, i.e., I^n is row-wise ratio-scale invariant:
Ratio-scale Invariance (RS): $I^n(\Lambda X) = I^n(X)$, $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_k)$, $\lambda_i > 0$, $i = 1, 2, \ldots, k$.
The inequality index I^n is not affected by any change in the units of measurement. I^n is thus analogous to those *relative* inequality measures in the unidimensional context.

Without majorization criteria which are partial orders ranking matrices of attributes by their degree of inequality, the discussion of multidimensional inequality indices is incomplete. The following section will focus on this issue.

---

[4] An $n \times n$ permutation matrix Π interchanges the coordinates of an n-vector $\underline{x}$ so that $\underline{x}\Pi$ is a permutation of $\underline{x}$.

3. Majorization criteria

A pioneering study of majorization criteria in the context of multidimensional inequality is Kolm (1977). Some of the results discussed in this section may be found in that paper. Another important reference is chapter 15 in Marshall and Olkin (1979). This section will also elaborate on the concept correlation-increasing majorization introduced below and its relationships with other majorization criteria.

To motivate our discussion, a brief review of univariate majorization criteria with one attribute seems appropriate. For one attribute, the Pigou-Dalton transfer is defined as follows: in comparing two income vectors $\underline{x} = (x_1, \ldots, x_n)$ and $\underline{y} = (y_1, \ldots, y_n)$, $\underline{y}$ may be derived from $\underline{x}$ by a *Pigou-Dalton (PD) Transfer* if, for some i and j, the following holds: $x_i < x_j, x_m = y_m$ for $m \notin \{i, j\}, y_i = x_i + \delta \leq x_j, y_j = x_j - \delta \geq x_i$ for some $\delta \geq 0$. In other words, there exists a *Pigou-Dalton Matrix* $T = \lambda E + (1 - \lambda)Q, 0 \leq \lambda \leq 1$, where E is an identity matrix and Q is a permutation matrix interchanging two coordinates of $\underline{x}$ (see Marshall and Olkin, 1979, p. 21), such that $\underline{y} = \underline{x}T$. A PD transfer with respect to $(\underline{x}, \underline{y})$ is **strict** if $\underline{y}$ is not a permutation of $\underline{x}$. A one-attribute index I^n is said to satisfy the *Pigou-Dalton Transfer Principle* if $I^n(\underline{y}) < I^n(\underline{x})$ for any strict PD transfer with respect to $(\underline{x}, \underline{y})$. It is to be noted that the PD transfer principle has a number of equivalent formulations; (see, e.g., Marshall and Olkin 1977). For $k \geq 1$ attributes and $X, Y \in \mathfrak{M}(n)$, the following generalizations of univariate majorizations are conceivable:

Uniform PD Majorization (UPD): $(X, Y) \in$ UPD if and only if $Y = XT$ for some matrix T that is a finite product of PD transfer matrices and is not a permutation matrix.

Uniform Majorization (UM): $(X, Y) \in$ UM whenever $Y = XB$, B is a bistochastic matrix and Y cannot be derived by permuting the columns of X.

For $k = 1$, these criteria reduce to the usual univariate majorization (i.e., reverse Lorenz dominance). When $k > 1$, the next result is well-known (see, e.g., Marshall and Olkin 1977, chapter 15; Kolm 1977):

Proposition 1: UPD $\subset$ UM. *For the special case that $n = 2$, UPD is still equivalent to UM*.

UPD and UM measure in some sense the spread of the attributes.[5] As shown below, they cannot rank (Y, X) in the introductory section where Y is

---

[5] Besides UM and UPD, other majorization criteria have been proposed:

Directional Majorization: $(X, Y) \in$ DM if and only if, for any $\underline{a} \in \mathbb{R}^k$, $\underline{a}Y$ strictly Lorenz dominates $\underline{a}X$.

Positive Directional Majorization: $(X, Y) \in$ PDM if and only if for any $\underline{a} \in \mathbb{R}^k_{++}$, $\underline{a}Y$ strictly Lorenz dominates $\underline{a}X$.

See, e.g., Kolm (1977); Bhandari (1988); Koshevoy (1995); Koshevoy and Mosler (1996); Joe and Verducci (1993). It is not difficult to show that UPD $\subset$ UM $\subset$ DM $\subset$ PDM.

derived from X by a correlation-increasing transfer though it seems compelling that X has a more equitable distribution of attributes. To capture this aspect of inequality, the concept **correlation-increasing transfer** (CIT) proposed by Boland and Proschan (1988) is introduced.[6] For any two k-vectors $\underline{x}$ and $\underline{y}$, define the two operators $\wedge$ and $\vee$ as follows: $\underline{x} \wedge \underline{y} = (\min\{x_1, y_1\}, \ldots, \min\{x_k, y_k\})$ and $\underline{x} \vee \underline{y} = (\max\{x_1, y_1\}, \ldots, \max\{x_k, y_k\})$. For $X, Y \in \mathfrak{M}(n)$, Y may be derived from X by a correlation-increasing transfer if, for some column indices i and j, $i < j$, $\underline{y}^i = \underline{x}^i \wedge \underline{x}^j$ and $\underline{y}^j = \underline{x}^i \vee \underline{x}^j$, $\underline{y}^m = \underline{x}^m$, $m \notin \{i, j\}$. A CIT is strict whenever $\underline{y}^i \neq \underline{x}^i$.

As an illustration, let us examine the following three matrices summarizing the distributions of attributes:

$$X_0 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 2 & 1 \end{pmatrix}, \ X_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \\ 1 & 2 & 3 \end{pmatrix}, \ X_2 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}$$

X_1 is derived from X_0 by a CIT between the first person and the third; likewise, the transformation from X_1 to X_2 pertains to a CIT between the second and the third person. It seems intuitively compelling to argue that the distributions of the attributes summarized by X_2 is most unequal followed by X_1 and then by X_0. Motivated by the concept arrangement increasing majorization introduced by Boland and Proschan (1988), such an ethical judgement may be formalized as follows:

Correlation Increasing Majorization (CIM): For $X, Y \in \mathfrak{M}(n)$, $(X, Y) \in$ CIM whenever X may be derived from Y by a permutation of columns and a finite sequence of correlation increasing transfers at least one of which is strict.

In the above example, $(X_2, X_0) \in$ CIM. This kind of majorization is known as an ordering of dependence in statistics (e.g., Shaked 1992) and in economics of risk as "pairwise more risk" (Richard 1975). CIM corresponds to the Atkinson-Bourguignon ordering (Atkinson and Bourguignon 1982) of X and Y with two attributes by the social welfare function $\int u(z_1, z_2) \, dF(z_1, z_2)$ where $\partial u(z_1, z_2)/\partial z_1 \geq 0$, $\partial u(z_1, z_2)/\partial z_2 \geq 0$ and $\partial^2 u(z_1, z_2)/\partial z_1 \partial z_2 \geq 0$ with the bivariate distribution $F(z_1, z_2)$ having the same marginals. The following proposition shows that UM or UPD does not include CIM:

Proposition 2: (a) CIM $\cap$ UM $= \emptyset$ (b) CIM $\cap$ UPD $= \emptyset$.

Proof. Let $(X, Y) \in$ CIM. For (X, Y) to be in UM, $\underline{x}_j B = \underline{y}_j$ for all $j \in \{1, 2, \ldots, k\}$ where $\underline{x}_j$ and $\underline{y}_j$ are the jth row of X and Y respectively. But B can only be a permutation matrix. Thus, $(X, Y) \notin$ UM and (a) is valid. (b) is true because UPD $\subset$ UM.   Q.E.D.

The above discussion highlights two aspects of multidimensional inequality: (1) UPD and UM pertain to the "spread" of the data of any two

---

[6] See also Epstein and Tanny (1980).

matrices of attributes; (2) CIM takes into account the correlation among attributes.

Before ending this section, we relate the above discussion to a different approach to multidimensional inequality. While the approach adopted in this paper may be non-welfarist in the sense that the well-being of a person depends *directly* on her vector of attributes, some researchers have suggested a welfarist approach, e.g., Maasoumi (1986) and Dardanoni (1992). The well-being derived from the attributes is measured through the intermediary of utility, i.e., for any vector $\underline{x}^i$ of the ith person, her well-being is measured by $u(\underline{x}^i)$ where $u: \mathbb{R}_{++}^k \to \mathbb{R}$ belongs to some class of functions $\mathfrak{U}$. Then, inequality is gauged with respect to $(u(\underline{x}^1), \dots, u(\underline{x}^n))$. Two utility-based majorizations may be considered:

UBM($\mathfrak{U}$) $= \{(X, Y): X, Y \in \mathfrak{M}(n),\ (u(\underline{y}^1), \dots, u(\underline{y}^n))$ strictly Lorenz dominates $(u(\underline{x}^1), \dots, u(\underline{x}^n))$ where $u(.) \in \mathfrak{U}\}$;

WUBM($\mathfrak{U}$) $= \{(X, Y): X, Y \in \mathfrak{M}(n),\ (u(\underline{y}^1), \dots, u(\underline{y}^n))$ strictly Lorenz subdominates $(u(\underline{x}^1), \dots, u(\underline{x}^n))$ where $u(.) \in \mathfrak{U}$,

where an n-vector $\underline{u}$ *Lorenz subdominates* $\underline{v}$ if and only if $\sum_{i=1}^{j} \hat{u}_i \leq \sum_{i=1}^{j} \hat{v}_i$, $j = 1, 2, \dots, n$, $\underline{\hat{u}} = (\hat{u}_1, \dots, \hat{u}_n)$ and $\underline{\hat{v}} = (\hat{v}_1, \dots, \hat{v}_n)$ are permutations of $\underline{u}$ and $\underline{v}$ such that $\hat{u}_1 \geq \cdots \geq \hat{u}_n$ and $\hat{v}_1 \geq \cdots \geq \hat{v}_n$. The subdominance is strict if one of the inequalities is strict. Lorenz subdominance corresponds to weak submajorization in Marshall and Olkin (1979, p. 10)

For different classes of $u(.)$, UBM($\mathfrak{U}$) and WUBM($\mathfrak{U}$) have interesting relationships with other majorization criteria introduced before. Let $\mathfrak{C}$ denote the class of functions $\mathbb{R}_{++}^k \to \mathbb{R}$ that are continuous and strictly concave, $\mathfrak{L}$, those that are increasing and L-superadditive, i.e., $u \in \mathfrak{L}$, if and only if u is increasing and $u(\underline{x} \wedge \underline{y}) + u(\underline{x} \vee \underline{y}) \geq u(\underline{x}) + u(\underline{x})$, and $\mathfrak{A} := \{v \in \mathfrak{L} : v(\underline{x}) = f(x_1) + f(x_2) + \cdots + f(x_k)\}$. The following proposition is from Kolm (1977):

Proposition 3: (Kolm 1977) $UM = WUBM(\mathfrak{C})$.

The next proposition clarifies the relationship between CIM on the one hand and the two utility-based majorizations on the other.

Proposition 4: (a) $CIM \subset WUBM(\mathfrak{L})$, (b) $CIM \subset UBM(\mathfrak{A})$.

Proof. For (a), the proof is motivated by Marshall and Olkin (1979, p. 167). Let $(X, Y) \in$ CIM such that $X = Z^m, Z^{m-1}, \dots, Z^0 = Y$, where $Z^t, t = 0, 1, 2, \dots, m$, are a sequence of matrices such that Z^{t+1} may be derived from $Z^t, t = 0, \dots, m-1$, through a CIT. Then, for $i < j$, $\underline{z}^{t+1,i} = \underline{z}^{ti} \wedge \underline{z}^{tj}$, $\underline{z}^{t+1,j} = \underline{z}^{ti} \vee \underline{z}^{tj}$ and $\underline{z}^{t+1,m} = \underline{z}^{t,m}, m \notin \{i, j\}$. Since $u(.)$ is increasing in its arguments, $u(\underline{z}^{t+1,j}) = u(\underline{z}^{ti} \vee \underline{z}^{tj}) \geq \max\{u(\underline{z}^{ti}), u(\underline{z}^{tj})\}$. Furthermore, since $u(.) \in \mathfrak{L}$, $u(\underline{z}^{ti} \wedge \underline{z}^{tj}) + u(\underline{z}^{ti} \vee \underline{z}^{tj}) \geq u(\underline{z}^{ti}) + u(\underline{z}^{tj})$. There is at least one t such that the inequality above is strict. Thus, $(u(\underline{z}^{ti}), u(\underline{z}^{tj}))$ Lorenz subdominates $(u(\underline{z}^{ti} \wedge \underline{z}^{tj}), u(\underline{z}^{ti} \vee \underline{z}^{tj}))$. By a result in Marshall and Olkin (1979, A.7.a(i), p. 122), if $u_m = v_m$, $m \notin \{i, j\}$, then (u_i, u_j) Lorenz subdominates (v_i, v_j)

implies that $\underline{u}$ Lorenz-subdominates $\underline{v}$. Since $\underline{z}^{t+1,m} = \underline{z}^{tm}$, for all $m \notin \{i, j\}$, it follows that $(u(\underline{z}^{t1}), \dots, u(\underline{z}^{tn}))$ Lorenz subdominates $(u(\underline{z}^{t+1,1}), \dots, u(\underline{z}^{t+1,n}))$. Repeating the argument for each pair of Z^{t+1} and Z^t, $(u(\underline{y}^1), \dots, u(\underline{y}^n))$ Lorenz subdominates $(u(\underline{x}^1), \dots, u(\underline{x}^n))$. The subdominance is strict because at least one CIT with respect to the sequence $\{Z^t, t = 0, 1, 2, \dots, m\}$ is strict.

With regard to (b), since $u(\underline{x}) = f_1(x_1) + \cdots + f_k(x_k)$ and the elements of each row vector of Y is just some permutation of the elements in each row of X, $\sum_{i=1}^{n} \sum_{j=1}^{k} f(y_{ji}) = \sum_{i=1}^{n} \sum_{j=1}^{k} f(x_{ji})$. Thus, $(u(\underline{y}^1), \dots, u(\underline{y}^n))$ strictly Lorenz dominates $(u(\underline{x}^1), \dots, u(\underline{x}^n))$ (see also Marshall and Olkin 1979, Theorem F14, p. 167).   Q.E.D.

4. Functional characterizations of majorization criteria

The main objectives of this section are: (i) to understand the role of each majorization criterion in restricting the functional form of $I^n(X)$, (ii) to axiomatically derive the class of multidimensional generalized entropy measures. With regard to (i), a multidimensional inequality index should satisfy the axiom that I^n is strictly increasing with respect to UM, UPD or CIM. Such axioms will be denoted by the same notation as the corresponding majorizations.

Before adding any majorization criterion, it is possible to show that the other axioms restrict the index to a specific class of functions. The proof is omitted because it is almost identical to the one for the one-attribute case in Shorrocks (1984). Details may however be found in Tsui (1997).[7] The result is stated below as a lemma:

Lemma 1: $I^n : \mathfrak{M}(n) \to \mathbb{R}$ *satisfies CN, AN, NM, DC and RI if and only if there exist continuous functions ϕ and F such that, for every $X \in \mathfrak{M}(n)$ having mean vector $\underline{\mu} := (\mu_1, \dots, \mu_k)$,*

$$F(I^n(X), \underline{\mu}) = \frac{1}{n} \sum_{i=1}^{n} [\phi(\underline{x}^i) - \phi(\underline{\mu})] \tag{1}$$

and F is strictly increasing in $I^n(X)$ and $F(0, \underline{\mu}) = 0$.

Next, we combine UM with the other axioms, resulting in the following theorem:

Theorem 1: *$I^n(X)$ satisfies UM, CN, AN, NM, DC and RI $\Leftrightarrow I^n(X)$ satisfies UPD, CN, AN, NM, DC and RI $\Leftrightarrow$ eq. (1) holds with ϕ strictly convex.*

Proof. Suppose that $I^n(X)$ satisfies UM, CN, AN, NM, DC. By Lemma 1, eq. (1) holds. Let B be a bistochastic matrix such that each element in B is $1/n$ and $Y = XB$ where Y cannot be derived by permuting the columns of XB, then $I^n(Y) < I^n(X)$. By eq. (1),

---

$$F(I^n(Y), \underline{\mu}) - F(I^n(X), \underline{\mu}) = 1/n \sum_{i=1}^{n} \left[\phi(\underline{y}^i) - \phi(\underline{x}^i) \right] < 0.$$

Since $\underline{y}^i = \underline{\mu}$ for all i, $\phi(\mu_1, \ldots, \mu_k) < 1/n \sum_{i=1}^{n} \phi(\underline{x}^i)$. Thus, ϕ is strictly convex. Conversely, Lemma 1 shows that eq. (1) satisfies all the axioms except UM. If $\phi(.)$ is a strictly convex function and $Y = XB$ where B is a bistochastic matrix and XB is not a column-wise permutation of Y, $\phi(\underline{y}^1) + \cdots + \phi(\underline{y}^n) < \phi(\underline{x}^1) + \cdots + \phi(\underline{x}^n)$ (Marshall and Olkin 1979, Theorem B1, p. 433).

If UM is replaced by UPD, eq. (1) still holds with ϕ still strictly convex. Let $Y = XT$, where $T = \lambda E + (1 - \lambda)Q$ is a Pigou-Dalton matrix with E being an identity matrix, Q is a permutation matrix interchanging the i and j coordinates of a vector and $\lambda = 1/2$. We assume that the transfer is strict so that $(X, Y) \in$ UPD. Then, $\underline{y}^i = \underline{y}^j = (\underline{x}^i + \underline{x}^j)/2$ and $\phi((\underline{x}^i + \underline{x}^j)/2) < (\phi(\underline{x}^i) + \phi(\underline{x}^j))/2$. Thus, ϕ is strictly convex. Conversely, if eq. (1) holds, CN, AN, NM, DC and RI are satisfied. If $(X, Y) \in$ UPD, $Y = XT$. Since the product of Pigou-Dalton matrices is a bistochastic matrix, $\phi(\underline{y}^1) + \cdots + \phi(\underline{y}^n) < \phi(\underline{x}^1) + \cdots + \phi(\underline{x}^n)$ (Marshall and Olkin 1979, Theorem B1, p. 433). Thus, $(X, Y) \in$ UPD $\Rightarrow I^n(X) > I^n(Y)$. We have shown that the three statements are equivalent.   Q.E.D.

If ϕ has second partial derivatives ϕ_{ij}, strict convexity of ϕ is equivalent to the following conditions:

$$D_j = \sum_{\sigma \in \Sigma_j} \text{sgn}(\sigma) \prod_{i=1}^{j} \phi_{i\sigma(i)} > 0, \tag{2}$$

for all $j \in \{1, 2, \ldots, k\}$, where Σ_j denotes the set of permutations of $\{1, 2, \ldots, j\}$, $\text{sgn}(\sigma) = +1$ if the permutation is even and $\text{sgn}(\sigma) = -1$ if the permutation is odd.

The next theorem shows that CIM restricts $\phi(.)$ to be *strictly* L-superadditive, i.e., $\phi(\underline{x} \wedge \underline{y}) + \phi(\underline{x} \vee \underline{y}) > \phi(\underline{x}) + \phi(\underline{y})$ if $\underline{x} \wedge \underline{y}$ is not equal to $\underline{x}$ or $\underline{y}$.

Theorem 2: $I^n(X)$ *satisfies CIM, UM, CN, AN, NM, DC and RI* $\Leftrightarrow I^n(X)$ *satisfies CIM, UPD, CN, AN, NM, DC and RI* $\Leftrightarrow$ *eq.* (1) *holds with* $\phi(.)$ *strictly convex and strictly L-superadditive.*

Proof. In view of Theorem 1, we only have to prove the part on the strict L-superadditivity of ϕ. Let $(Y, X) \in$ CIM, where Y is derived from X via a *strict* CIT such that for $i < j$, $\underline{y}^i = \underline{x}^i \wedge \underline{x}^j$, $\underline{y}^j = \underline{x}^i \vee \underline{x}^j$ and $\underline{y}^m = \underline{x}^m$ for $m \notin \{i, j\}$. Then, using eq. (1), $\phi(\underline{x}^i \wedge \underline{x}^j) + \phi(\underline{x}^i \vee \underline{x}^j) > \phi(\underline{x}^i) + \phi(\underline{x}^j)$, i.e., ϕ is strictly L-superadditive.

Conversely, if eq. (1) holds, CN, AN, NM, DC and RI holds by Lemma 1. With regard to CIM, let there be a sequence of matrices $Y = Z^m$, $Z^{m-1}, \ldots, Z^0 = X$, such that Z^t is derived from Z^{t-1}, $t = 1, 2, \ldots, m$, via a CIT, with at least one CIT strict. For any t, let $i < j$ such that $\underline{z}^{t,i} = \underline{z}^{t-1,i} \wedge \underline{z}^{t-1,j}$, $\underline{z}^{t,j} = \underline{z}^{t-1,i} \vee \underline{z}^{t-1,j}$ and $\underline{z}^{t,p} = \underline{z}^{t-1,p}$ for $p \notin \{i, j\}$. Since ϕ is strictly L-superadditive, $\phi(\underline{z}^{t,i}) + \phi(\underline{z}^{t,i}) > \phi(\underline{z}^{t-1,i}) + \phi(\underline{z}^{t-1,j})$ whenever Z^t is derived from

Z^{t-1} by a strict CIT. It is then not difficult to see that $I^n(Y) > I^n(X)$ since at least one of the CITs is strict. Thus CIM is satisfied. Q.E.D.

If $\phi(.)$ has second partial derivatives, *strict* L-superadditivity is equivalent to

$$\phi_{ij} = \frac{\partial^2 \phi}{\partial x_i \partial x_j} > 0, \quad i, j = 1, 2, \ldots, k, \ i \neq j \tag{3}$$

(see Marshall and Olkin (1979), p. 150 and the references cited there).

If $I^n(X)$ satisfies UM, the index is also increasing with respect to WUBM($\mathfrak{C}$). It is increasing with respect to WUBM($\mathfrak{L}$) if $I^n(X)$ also satisfies CIM . Thus, the ethical judgements on inequality of non-welfarist inequality indices satisfying UM (or UPD) and CIM are consistent with utility-based inequality indices for certain classes of utility functions.

$\phi(.)$ in eq. (1) may assume many functional forms. Introducing RS inexorably leads to the class of multidimensional generalized entropy measures.

Theorem 3: *$I^n(X)$ satisfies RS, UM (or UPD), CN, AN, NM, DC and RI if and only if there exists a transformation F such that, for every $X \in \mathfrak{M}(n)$ having mean vector $\underline{\mu}$, either*

$$F(I^n(X)) = \frac{\rho}{n} \sum_{i=1}^{n} \left[\prod_{j=1}^{k} \left(\frac{x_{ji}}{\mu_j} \right)^{c_j} - 1 \right], \tag{4a}$$

where $\sum_{\sigma \in \Sigma_j} \mathrm{sgn}(\sigma) \sum_{i=1}^{i} \rho c_{i\sigma(i)} > 0$, $c_{i\sigma(i)} = c_i c_{\sigma(i)}$, *if* $i \neq \sigma(i)$ *and* $c_{i\sigma(i)} = c_i(c_i - 1)$, *if* $i = \sigma(i)$;
or

$$F(I^n(X)) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{x_{mi}}{\mu_m} \right) \left[\sum_{j=1}^{k} a_{mj} \log \left(\frac{x_{ji}}{\mu_j} \right) \right], \tag{4b}$$

where $m \in \{1, 2, \ldots, k\}$ with the following restrictions on the parameters: (i) *for* $j < m$, $(-1)^j \prod_{i=1}^{j} a_{mi} > 0$, (ii) *for* $j \geq m$, $\sum_{\sigma \in \Sigma_j} \mathrm{sgn}(\sigma) \Delta_j(\sigma) > 0$ *and either* $\Delta_j(\sigma) = (-1)^{j-1} \prod_{i=1}^{j} a_{mi}$ *or* $\Delta_j(\sigma) = (-1)^{j-2} a_{mp}^2 \prod_{i \neq m, p} a_{mi} > 0$; *or*

$$F(I^n(X)) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{k} \delta_j \log \left(\frac{\mu_j}{x_{ji}} \right) \tag{4c}$$

where $\delta_j > 0$ for all j.

Proof. The proof is similar to that for the one-attribute case in Shorrocks (1984). Readers are referred to Tsui (1997) for details. CN, AN, NM, DC and RI lead to eq. (1). Adding RS restricts eq. (1) to (4a)–(4c) above. The transformation F is independent of $\underline{\mu}$ because of RS. The restrictions of the parameters ensure that ϕ in Theorem 1 is strictly convex. Since the functional forms of ϕ corresponding to (4a)–(4c) have second partial derivatives, the ranges of the parameters are derived using eq. (2). The result continues to hold if UM is replaced by UPD because of Theorem 1.   Q.E.D.

To clarify the restrictions on the parameters, it is illuminating to review the case when $k = 2$.

Corollary 1: When $k = 2$, (4a) reduces to

$$F(I^n(X)) = \frac{\rho}{n} \sum_{i=1}^{n} \left\{ \left(\frac{x_{1i}}{\mu_1}\right)^{c_1} \left(\frac{x_{2i}}{\mu_2}\right)^{c_2} - 1 \right\} , \qquad (5a)$$

where $(c_1, c_2) \neq (0,0), (0,1), (1,0)$, $\rho c_1(c_1 - 1) > 0$, $c_1 c_2(1 - c_1 - c_2) > 0$; *(4b) becomes*

$$F(I^n(X)) = \begin{cases} \frac{1}{n} \sum_{i=1}^{n} (x_{1i}/\mu_1)[a_{11} \log(x_{1i}/\mu_1) + a_{12} \log(x_{2i}/\mu_2)] \\ \frac{1}{n} \sum_{i=1}^{n} (x_{2i}/\mu_2)[a_{21} \log(x_{1i}/\mu_1) + a_{22} \log(x_{2i}/\mu_2)] \end{cases}$$

where $a_{11} > 0$, $a_{12}(a_{11} + a_{12}) < 0$, $a_{21} < 0$ *and* $a_{21}(a_{22} + a_{21}) < 0$; *lastly, (4c) becomes*

$$F(I^n(X)) = \frac{1}{n} \sum_{i=1}^{n} [\delta_1 \log(\mu_1/x_{1i}) + \delta_2 \log(\mu_2/x_{2i})] \qquad (5c)$$

where $\delta_1 > 0$ *and* $\delta_2 > 0$.

Proof. By eq. (2), $D_1 = \phi_{11} > 0$ and $D_2 = \phi_{11}\phi_{22} - \phi_{12}^2 > 0$. The restrictions on the parameters are then derived accordingly.   Q.E.D

If CIM is incorporated, (4b) and (4c) are eliminated.

Theorem 4: $I^n(X)$ *satisfies CIM, UM (or UPD), RS, CN, AN, NM, DC, RI and if and only if (4a) holds with the additional restriction that* $\rho c_i c_j > 0$.

Proof. (4a) holds with the additional restriction that $\rho c_i c_j > 0$ because of eq. (3) in Theorem 4. (4b) and (4c) are incompatible with CIM and are thus eliminated. The theorem is still valid if UM is replaced by UPD.   Q.E.D.

For $k = 2$, only (5a) remains with the additional restriction that $\rho c_1 c_2 > 0$.

5. Conclusion

The study of multidimensional inequality indices has gradually been gathering momentum. Much however remains to be done in this area. The discussion in this paper hopefully will stimulate further research on this subject. The class of multidimensional entropy measures should be a welcome addition to the toolkit of those who are interested in the empirical investigation of multidimensional inequality.

The approach of this paper is non-welfarist. It is controversial whether one should measure inequality with respect to attributes per se or in terms of utilities? This question entails a difficult philosophical question: "inequality

of what?" (see, e.g., Sen 1985, 1987, 1992). By exploring utility-based majorizations, we have shown that our class of multidimensional indices is compatible with these majorizations for certain types of functions. More work on the welfarist and nonwelfarist approach hopefully will shed more light on the measurement of multidimensional inequality.

References

Aczel J (1966) Lectures on functional equations and their applications. Academic Press, New York

Atkinson AB (1970) On the measurement of inequality. J. Econ. Theory 2: 244–263

Atkinson AB, Bourguignon F (1982) The comparison of multi-dimensioned distributions of economic status. Rev. Econ. Studies 12: 183–201

Bhandari SK (1988) Multivariate majorization and directional majorization: positive results. Sankhya: The Indian Journal of Statistics 50 (2): 199–204

Boland P, Proschan F (1988) Multivariate arrangement increasing functions with application in probability and statistics. J. Multivariate Analysis 25: 286–298

Bourguignon F (1979) Decomposable income inequality measures. Econometrica 47: 901–902

Cowell FA (1980) On the structure of additive inequality measures. Rev. Econ. Studies 47: 521–531

Cowell FA, Kuga K (1981) Additivity and the entropy concept: an axiomatic approach to inequality measurement. J. Econ. Theory 25: 131–143

Dardanoni V (1992) On multidimensional inequality measurement. Paper presented at the International Meeting on Personal Income Distribution, Inequality and Poverty, University of Siena

Ebert U (1988a) Measurement of inequality: an attempt at unification and generalization. Soc. Choice Welfare 5: 147–169

Ebert U (1988b) On the decomposition of inequality: partition into nonoverlapping sub-groups. In: W. Eichhorn (ed.) Measurement in economics: theory and application of economic indices. Physica-Verlag, Heidelberg

Epstein L, Tanny SM (1980) Increasing generalized correlation: a definition and some economic consequences. Canad. J. Econ. 13: 16–34

Kolm S (1976a) Unequal inequalities I. J. Econ. Theory 12: 416–442

Kolm S (1976b) Unequal inequalities II. J. Econ. Theory 13: 82–111

Kolm S (1977) Multidimensional equalitarianisms. Qu. J. Econ. 91: 1–13

Koshevoy G (1995) Multivariate Lorenz majorization. Soc. Choice Welfare 12: 93–102

Koshevoy G, Mosler K (1996) The Lorenz surface of a multivariate distribution. J. Amer Statist. Assoc. 91: 873–882

Koshevoy G, Mosler K (1997) Multivariate Gini indices. J. Multivariate Analysis 60: 252–276

Joe H, Verduccin J (1993) Multivariate majorization by positive combinations. In: Stochastic inequalities, IMS Lecture Notes – Monograph Series 22: 159–181

Maasoumi E (1986) The measurement and decomposition of multi-dimensional inequality Econometrica 54: 771–779

Maasoumi E (1989) Composite indices of income and other developmental indicators: a general approach. Research on Economic Inequality, vol. 1. JAI Press, Greenwich

Maasoumi E (1993) A compendium to information theory in economics and econometrics. Econometric Rev. 12: 137–181

Maasoumi E, Jeong JH (1985) The trend and the measurement of world inequality over extended periods of accounting. Econ. Lett. 19: 295–301

Maasoumi E, Nickelsburg G (1988) Multivariate measures of well-being and an analysis of inequality in the Michigan data. J. Business Econ. Statist. 6: 327–333

Marshall AW, Olkin I (1979) Inequalities: theory of majorization and its applications. Academic Press, San Diego

Mosler K (1993) Multidimensional welfarisms. In: W. Eichhorn (ed.) Models and measurement of inequality and welfare. Springer, Berlin, Heidelberg, New York

Mosler K (1994) Majorization in economic disparity measures. Linear Algebra Appl. 199: 91–114

Richard SF (1975) Multivariate risk aversion, utility dependence, and separable utility functions. Manag. Sci. 22: 12–21

Sen AK (1973) On economic inequality. Clarendon Press, Oxford

Sen AK (1985) Commodities and capabilities. North Holland, Amsterdam

Sen AK (1987) The standard of living. Cambridge University Press, Cambridge

Sen AK (1992) Inequality re-examined. Clarendon Press, Oxford

Shaked M (1982) A general theory of some positive dependence notions. J. Multivariate Analysis 12: 199–218

Shorrocks AF (1980) The class of additive decomposable inequality measures econometrica 48: 613–625

Shorrocks AF (1984) Inequality decomposition by population subgroups Econometrica 52: 1369–1385

Shorrocks AF (1988) Aggregation issues in inequality measurement. In: W. Eichhorn (ed.) Measurement in economics. Physica-Verlag, Heidelberg

Shorrocks AF, Foster JE (1987) Transfer sensitive inequality measures. Rev. Econ. Studies 54: 485–497

Slottje DJ (ed) (1989) A special issue on "economic issues on analysing inequality". J. Econometrics 42: 1–155

Streeten P (1981) First things first: meeting basic human needs in developing countries. Oxford University Press for the World Bank, New York

Tsui KY (1995) Multidimensional generalizations of the relative and absolute inequality indices: the Atkinson-Kolm-Sen approach. J. Econ. Theory 67: 251–265

Tsui KY (1997) Multidimensional inequality and multidimensional generalized entropy measures: an axiomatic derivation. mimeo, revised, April