Available online at www.sciencedirect.com
SCIENCE DIRECT®
Mathematical and Computer Modelling 41 (2005) 301–311
MATHEMATICAL AND COMPUTER MODELLING
www.elsevier.com/locate/mcm

Comparing Partial and Truncated Conglomerates from a Concentration Theoretic Point of View

L. EGGHE
LUC, Universitaire Campus, B-3590 Diepenbeek, Belgium
and
UA, IBW, Universiteitsplein 1, B-2610 Wilrijk, Belgium
leo.egghe@luc.ac.be

R. ROUSSEAU
KHBO, Industrial Sciences and Technology
Zeedijk 101, B-8400 Oostende, Belgium
and
UA, IBW, Universiteitsplein 1, B-2610 Wilrijk, Belgium
and
LUC, Universitaire Campus, B-3590 Diepenbeek, Belgium
ronald.rousseau@khbo.be

(Received and accepted August 2003)

Abstract—When studying numerical properties of a population (technically: a conglomerate) it often happens that not all data are known. It might be that the total number of objects (persons) in the population is known, but that data on a number of them is missing. It even happens frequently that the total number of objects (N) is unknown. Referring to the population as 'sources' and to the property under investigation as 'items' or as 'the production', the whole dataset of this conglomerate can be represented as an N-vector. In this article N-vectors representing sources and their respective productions are studied from the point of view of concentration theory. Partial vectors (N is known, but data concerning the least productive sources are missing) and truncated vectors (N is unknown) are compared in two ways. First-order comparisons study vectors, while second-order comparisons study differences between vectors. In the case of first-order comparisons, it is shown that truncated vectors may be incomparable, while partial ones are always completely comparable. Similarly for second-order comparisons, partial vectors can be compared and yield a totally ordered double sequence, while truncated ones may be incomparable. Finally, we describe how to make second-order comparisons for vectors with a different number of sources.

Keywords—Truncation, Partial conglomerates, Generalized bibliographies, Concentration.

1. INTRODUCTION

Consider a set Ω of sources. These sources may or may not have produced a number of items. As a generic name for this framework we use the term 'generalized bibliography' or conglomerate [1]. Examples are the members of a university department as sources and their publications (during

0895-7177/05/$ - see front matter
doi:10.1016/j.mcm.2003.08.009

one year) as the corresponding items; or companies and the number of personnel; or journals, and their impact factors; and so on.

If we have a conglomerate with N sources we will rank these depending on the number of items they have produced. This yields a row of N numbers. We will refer to such a row as an *N-vector*, or a *vector of length (dimension) N*, or simply a *vector*. Next, we will consider the following question: what happens if you 'reduce' or 'cut' a vector? This happens when not all data are known. It might be that the total number of sources is known, but that data on a number of them is missing. It even happens frequently that the total number of sources (N) is unknown. This leads to a smaller vector. There are, of course, different ways of making a vector 'smaller', but we will consider only two cases. The first way is to keep the first i-components of the vector and replacing the other components by zeros (this is only possible if N is known). This yields a partial vector. A partial vector of an N-vector is again an N-vector. We will refer to the original N-vector as the *parent* vector. The second way is to keep only the first i-components ($0 < i < N$) and make no replacements. This operation is called truncation. The length of a truncated vector is always smaller than that of the parent vector. If we truncate after the i^{th} component, the resulting vector is called the i-truncation of the parent vector. In practice we often encounter truncated vectors with unknown parent.

We introduce the following mathematical notation for these notions. Let X be a vector, then $X = (x_1, x_2, \ldots, x_N)$, where x_j denotes the number of items produced by the j^{th} source, S_j, and $x_1 \geq x_2 \geq \cdots \geq x_N$. We denote by $X_i = (x_1, x_2, \ldots, x_i, 0, \ldots, 0)$ the i^{th} partial N-vector. Similarly, $X_{i,t} = (x_1, x_2, \ldots, x_i)$ denotes the i-truncation of X.

Truncated and partial vectors will be studied using concentration theory. Recall that, basically, concentration can be described as *the relative apportionment of items among the sources present.* The study of concentration and its opposite, diversity, has many implications in fields such as economics (e.g., geographical concentration of firms), sociology (concentration of wealth), ecology (biodiversity), and informetrics (as a parameter to describe the unequal scientific production among countries, institutes, or authors). Concentration measures can be considered as scientometric indicators [2]. Concentration can best be studied by using Lorenz curves, or variations thereof [3–6]. We assume that the reader is familiar with the construction of a classical Lorenz curve and will not repeat it here.

2. RESEARCH PROBLEM

First, we will make so-called first-order comparisons. By this we mean the following: considering two truncated conglomerates of the same parent: which of the two is the most concentrated? Similarly, considering two partial conglomerates of the same parent, can we say which of the two is the most concentrated? We will show that truncated ones may be incomparable, while partial ones are always completely comparable.

Next, we will make second-order comparisons, applying relative concentration as introduced in [6]. By the term second-order comparison we mean that we will compare the difference between the $(j-1)^{\text{th}}$ and the j^{th} partial vector, with the difference between the j^{th} and the $(j+1)^{\text{th}}$ partial vector. In a similar way, we will consider differences of truncated vectors.

In practice, our results yield information about situations (or studies) where one 'forgets' or otherwise removes the least productive sources. What is the influence (on concentration) of this removal or reduction? Examples are: not considering 'unimportant' journals in citation studies; or not paying attention to one-man companies in a study of company sizes.

3. A STRICT ORDER IN THE SET OF PARTIAL VECTORS OF A FIXED-PARENT CONGLOMERATE

We have already shown in [7] that the partial vectors of a fixed-parent conglomerate form a completely ordered subset in the partially ordered set of all N-vectors with the same Lorenz

curves. We will not repeat the argument here, but will present a small adaptation for the case of weighted Lorenz curves. Weighted Lorenz curves occur when one is interested in how different the concentration is with respect to a standard. This standard can be an internal or an external standard [2]. This happens, e.g., if one wants to compare the publication output of countries taking population (the standard) into account. Weighted Lorenz curves are constructed as follows [2,8,9]. Let $\mathrm{SV} = (s_1, s_2, \ldots, s_N)$ denote the standard vector and let $X = (x_1, x_2, \ldots, x_N)$ denote the distribution vector that we want to compare with this standard. Note that now indices must correspond. If, e.g., X denotes numbers of publications and SV denotes population then x_i and s_i must refer to the same country C_i. We assume, moreover, that none of the components of SV is zero. In order to construct the Lorenz curve for comparisons with a standard, the components of both vectors are ordered in such a way that

$$\frac{x_1}{s_1} \geq \frac{x_2}{s_2} \geq \cdots \geq \frac{x_N}{s_N}. \tag{1}$$

Next we normalize the vectors X and SV, leading to vectors A_X and W, where

$$a_i = \frac{x_i}{\sum_{j=1}^{N} x_j} \qquad \text{and} \qquad w_i = \frac{s_i}{\sum_{j=1}^{N} s_j}. \tag{2}$$

Note that normalizing does not change the order. Finally, the weighted Lorenz curve is defined as the broken line connecting the origin $(0,0)$ and the points with components

$$\left(\sum_{j=1}^{i} w_j, \sum_{j=1}^{i} a_j \right)_{i=1,\ldots,N}. \tag{3}$$

For a fixed standard these Lorenz curves again introduce a partial order in the set of equivalent N-vectors, i.e., those with the same Lorenz curve.

Note that, when comparing with a standard and considering partial vectors, the vector SV, or W, does not change. It is only X, and hence A_X, that is changed. It is now not difficult to see that, for $i = 1, \ldots N-1$, L_{i+1} (the weighted $(i+1)$ partial Lorenz curve) is at no point situated strictly above L_i. Indeed, let $X_i = (x_1, x_2, \ldots, x_i, 0, \ldots, 0)$ be the i^{th} partial N-vector (with $(N-i)$ zeros), and let $X_{i+1} = (x_1, x_2, \ldots, x_i, x_{i+1}, 0, \ldots, 0)$ be the $(i+1)^{\text{th}}$ partial N-vector (with $(N-i-1)$ zeros), then the Lorenz curves L_i and L_{i+1} are constructed as follows. L_i connects the points

$$(0,0), \left(w_1, \frac{x_1}{\sum_{n=1}^{i} x_n} \right), \ldots, \left(\sum_{n=1}^{i-1} w_n, \frac{\sum_{n=1}^{i-1} x_n}{\sum_{n=1}^{i} x_n} \right), \left(\sum_{n=1}^{i} w_n, 1 \right), \ldots, (1,1),$$

while L_{i+1} connects the points

$$(0,0), \left(w_1, \frac{x_1}{\sum_{n=1}^{i+1} x_n} \right), \ldots, \left(\sum_{n=1}^{i-1} w_n, \frac{\sum_{n=1}^{i-1} x_n}{\sum_{n=1}^{i+1} x_n} \right), \left(\sum_{n=1}^{i} w_n, \frac{\sum_{n=1}^{i} x_n}{\sum_{n=1}^{i+1} x_n} \right), \left(\sum_{n=1}^{i+1} w_n, 1 \right), \ldots, (1,1).$$

The point is that considering partial vectors *does not change the ranking* of the first i components for the construction of a weighted Lorenz curve, while the ranking of the last ones plays no role (as they yield the same weighted Lorenz curve). Indeed, if for $0 < j < i$

$$\frac{a_j}{w_j} = \frac{1}{w_j} \frac{\sum_{n=1}^{j} x_n}{\sum_{n=1}^{N} x_n} \geq \frac{a_{j+1}}{w_{j+1}} = \frac{1}{w_{j+1}} \frac{\sum_{n=1}^{j+1} x_n}{\sum_{n=1}^{N} x_n}, \tag{4}$$

then also (for $0 < j < i$)

$$\frac{1}{w_j}\frac{\sum_{n=1}^{j} x_n}{\sum_{n=1}^{i} x_n} \geq \frac{1}{w_{j+1}}\frac{\sum_{n=1}^{j+1} x_n}{\sum_{n=1}^{i} x_n}. \tag{5}$$

It is now clear that for every $i \in \{1, \ldots, N-1\}$, L_{i+1} is at no point situated strictly above L_i. Hence, these partial vectors form a completely ordered subset.

Truncated vectors, on the other hand, are not 'well behaved'. Truncation can make a vector intrinsically incomparable with its parent. A simple illustration is given by: $X = (3,1,1)$ and $X_{2,t} = (3,1)$. These two vectors are clearly incomparable.

Recall that it is shown in [10] that adding a source only gives a smaller vector if the production of this source is equal to the average production. The relation between a vector and its truncation can be seen as that of a vector and a vector with one source added, where this added source has a production at most equal to the smallest (and not the average) production. Hence, a truncated vector is never strictly larger than its parent, and only equal if the parent is the equality vector. Of course, besides being incomparable a truncated vector can be strictly smaller than its parent.

One might wonder what would happen if we truncate the most productive sources. We will call such an operation a forward-truncation. An example would be the case that one studies only rare birds and does not count birds that occur in large flocks, or one does not consider rich and very rich people, but only average and poor ones. This kind of truncation does not give nice results either: the parent $Y = (4,4,1)$ and its forward-truncation $Y^t = (4,1)$ are incomparable.

This ends the first-order comparison. Note that the 'indeterminate' case (truncation) occurs more often in practice that the 'well-behaved' case (partial vectors). In the next sections we will study second-order comparisons.

4. SECOND-ORDER COMPARISONS: RELATIVE CONCENTRATION

Partial N-vectors of the same parent, such as X_i and X_{i+1}, have the same length (N), and hence we may apply the theory of symmetric relative concentration [6]. Recall [6] that the Lorenz curve of symmetric relative concentration (referred to as the Egghe-Lorenz curve in [2]) is constructed as follows: let $X = (x_i)_{i=1,\ldots,N}$ and $Y = (y_i)_{i=1,\ldots,N}$ be two N-vectors and let $A_X = (a_i)_{i=1,\ldots,N}$ and $B_Y = (b_i)_{i=1,\ldots,N}$ denote their relative vectors (sum of all components equal to one). Then the components of the difference vector $D = (d_i)_{i=1,\ldots,N}$ with $d_i = a_i - b_i$ are ranked from largest to smallest. Next, we put

$$s_k = \sum_{j=1}^{k} d_j = \sum_{j=1}^{k} (a_j - b_j). \tag{6}$$

The Lorenz curve for symmetric relative concentration is then obtained by joining the origin $(0,0)$ and the points with coordinates

$$\left(\frac{k}{N}, s_k\right)_{k=1,\ldots,N}. \tag{7}$$

Such Lorenz curves will be denoted here by script $\mathcal{L}$s. Let now

$$A_{X_i} = \left(\frac{x_1}{\sum_{k=1}^{i} x_k}, \frac{x_2}{\sum_{k=1}^{i} x_k}, \ldots, \frac{x_i}{\sum_{k=1}^{i} x_k}, 0, \ldots, 0\right)$$

and

$$A_{X_{i+1}} = \left(\frac{x_1}{\sum_{m=1}^{i+1} x_m}, \frac{x_2}{\sum_{m=1}^{i+1} x_m}, \ldots, \frac{x_i}{\sum_{m=1}^{i+1} x_m}, \frac{x_{i+1}}{\sum_{m=1}^{i+1} x_m}, 0, \ldots, 0 \right).$$

Then

$$A_{X_{i+1}} - A_{X_i} = \left(\frac{-x_1 x_{i+1}}{\left(\sum_{m=1}^{i+1} x_m\right)\left(\sum_{k=1}^{i} x_k\right)}, \ldots, \frac{-x_i x_{i+1}}{\left(\sum_{m=1}^{i+1} x_m\right)\left(\sum_{k=1}^{i} x_k\right)}, \frac{x_{i+1}}{\left(\sum_{m=1}^{i+1} x_m\right)}, 0, \ldots, 0 \right).$$

Putting

$$B_i = \frac{-x_{i+1}}{\left(\sum_{m=1}^{i+1} x_m\right)\left(\sum_{k=1}^{i} x_k\right)}$$

yields, after rearranging components in decreasing order,

$$A_{X_{i+1}} - A_{X_i} = \left(\frac{x_{i+1}}{\sum_{m=1}^{i+1} x_m}, 0, \ldots, 0, -x_i B_i, -x_{i-1} B_i, \ldots, -x_1 B_i \right)$$

whose j^{th} component we further rewrite as

$$\left(A_{X_{i+1}} - A_{X_i}\right)_j = \beta_j, \qquad j = 1, \ldots, N.$$

Egghe [4] has shown that any continuous, convex function ϕ yields a measure of symmetric relative concentration C_ϕ, defined as

$$C_\phi = \frac{1}{N} \sum_{j=1}^{N} \phi(N\beta_j). \tag{8}$$

An example, cf. [6], is the relative weighted squared coefficient of variation, denoted as V_r^2, where $\phi(x)$ is the function $x^2 - 1$

$$V_r^2 = \frac{1}{N} \sum_{j=1}^{N} (N\beta_j)^2 - 1 = N \sum_{j=1}^{N} \beta_j^2 - 1. \tag{9}$$

A general continuous, convex function ϕ yields the following symmetric, relative concentration value:

$$C_\phi = \frac{1}{N} \left[\phi\left(\frac{N x_{i+1}}{\sum_{m=1}^{i+1} x_m} \right) + (N - i - 1)\phi(0) + \phi\left(\frac{-x_i x_{i+1} N}{\left(\sum_{m=1}^{i+1} x_m\right)\left(\sum_{k=1}^{i} x_k\right)} \right) + \cdots + \phi\left(\frac{-x_1 x_{i+1} N}{\left(\sum_{m=1}^{i+1} x_m\right)\left(\sum_{k=1}^{i} x_k\right)} \right) \right].$$

If we denote

$$\frac{x_{i+1}}{\sum_{m=1}^{i+1} x_m} \text{ as } a'_{i+1} \text{ and } \frac{x_j}{\sum_{k=1}^{i} x_k} \text{ as } a_j, \qquad j = 1, \ldots, i,$$

then

$$C_\phi = \frac{1}{N}\left(\phi\left(Na'_{i+1}\right) + (N-i-1)\phi(0) + \phi\left(-a'_{i+1}a_iN\right) + \cdots + \phi\left(-a'_{i+1}a_1N\right)\right). \tag{10}$$

In particular, the relative weighted squared coefficient of variation yields

$$V_r^2 = N\left(a'_{i+1}\right)^2\left(1 + \sum_{j=1}^{i} a_j^2\right) - 1. \tag{11}$$

We will next show that the symmetric relative concentration of X_{i+2} with respect to X_{i+1} is always smaller than that of X_{i+1} with respect to X_i.

Comparing X_i and X_{i+1} and applying the construction of the Egghe-Lorenz curve yields a curve denoted as $\mathcal{L}_i$, with abscissa k/N corresponding to ordinates s_k where

$$s_k = \begin{cases} a'_{i+1}, & \text{for } k = 1, \ldots, N-i, \\ a'_{i+1}\left(1 - \sum_{j=N-k+1}^{i} a_j\right), & \text{for } k = N-i+1, \ldots, N. \end{cases}$$

We obtain a similar curve, denoted as $\mathcal{L}_{i+1}$, based on a vector with abscissas k/N, and ordinates s'_k,

$$s'_k = \begin{cases} a''_{i+2}, & \text{for } k = 1, \ldots, N-i+1, \\ a''_{i+2}\left(1 - \sum_{j=N-k+1}^{i+1} a'_j\right), & \text{for } k = N-i+2, \ldots, N, \end{cases}$$

where

$$a''_{i+2} = \frac{x_{i+2}}{\sum_{l=1}^{i+2} x_l}.$$

THEOREM. $\mathcal{L}_{i+1} < \mathcal{L}_i$.

PROOF. We first consider the first $N-i-1$ ordinates

$$a''_{i+2} = \frac{x_{i+2}}{\sum_{l=1}^{i+2} x_l} < \frac{x_{i+1}}{\sum_{m=1}^{i+1} x_m} = a'_{i+1},$$

where the inequality holds because the $(x_i)_i$ are decreasing. Then we have for the $(N-i)^{\text{th}}$ coordinate

$$a''_{i+2}\left(1 - a'_{i+1}\right) < a''_{i+2} < a'_{i+1}$$

leaving only the last i coordinates to check. The j $(\leq i)$ last ones, i.e., the coordinates with index $N-j+1$, i.e., with abscissas $(N-j+1)/N$, have the following form:

$$\begin{aligned} a''_{i+2}\left(1 - a'_{i+1} - \sum_{k=0}^{i-j} a'_{i-k}\right) &= a''_{i+2}\left(a'_1 + \cdots + a'_{j-1}\right) \\ &= \frac{x_{i+2}}{\sum_{l=1}^{i+2} x_l} \sum_{k=1}^{j-1}\left(\frac{x_k}{\sum_{m=1}^{i+1} x_m}\right) < \frac{x_{i+1}}{\sum_{m=1}^{i+1} x_m} \sum_{k=1}^{j-1}\left(\frac{x_k}{\sum_{n=1}^{i} x_n}\right) \\ &= a'_{i+1}\left(a_1 + \cdots + a_{j-1}\right) = a'_{i+1}\left(1 - \sum_{l=0}^{i-j} a_{i-l}\right). \end{aligned}$$

This last expression is exactly the ordinate of $\mathcal{L}_i$ with the same index $N-j+1$. This proves the theorem.

We conclude from this result that X_{i+1} and X_{i+2} are more similar than X_i and X_{i+1}. This corresponds with our intuition. We may say that the double sequence $(X_i, X_{i+1})_{i=1,\ldots,N-1}$ is completely ordered.

5. TRUNCATION AND RELATIVE CONCENTRATION

If $X = (x_1, x_2, \ldots, x_N)$, then we will call the vector $(x_1, \ldots, x_i)$ the i-truncation of X, denoted as $X_{i,t}$. The vector $(x_1, \ldots, x_i, 0)$ will be called the expanded i-truncation of x, denoted as: $X_{i,e}$.

In this section, we will compare $X_{i+1,t} = (x_1, \ldots, x_{i+1})$ and $X_{i,e} = (x_1, \ldots, x_i, 0)$ on the one hand, and $X_{i+2,t} = (x_1, \ldots, x_{i+2})$ and $X_{i+1,e} = (x_1, \ldots, x_{i+1}, 0)$ on the other. Yet, similar to the first-order comparison, truncation destroys complete-ordering. We just present one example.

Let $X = (3, 1, 1)$, and take $i = 1$. Then $X_{i+1,t} = (3, 1)$, $X_{i,e} = (3, 0)$, $X_{i+2,t} = (3, 1, 1) = X$, and $X_{i+1,e} = (3, 1, 0)$. Relative vectors are, respectively: $(3/4, 1/4)$, $(1, 0)$, $(3/5, 1/5, 1/5)$, and $(3/4, 1/4, 0)$. Taking differences yields the vectors $(-1/4, +1/4)$ and $(-3/20, -1/20, 4/20)$ or after reordering from largest to smallest: $(1/4, -1/4)$ and $(4/20, -1/20, -3/20)$. Corresponding Egghe-Lorenz curves cross, and hence, these two vectors are incomparable.

6. VARIABLE NUMBER OF SOURCES

Let $X = (x_1, x_2, \ldots, x_N)$ and let $Y = (y_1, y_2, \ldots, y_M)$, with M in general different from N. In this section, we will explain how to study the relative symmetric concentration between an N-vector such as X and an M-vector such as Y. We present, in particular, the necessary formulae in order to make comparisons in the case of a variable number of sources. Such comparisons are necessary in dynamic studies of conglomerates. Indeed, in real applications the number of sources is usually not constant, but is time dependent. Our approach is essentially an application of the general theory of Lorenz curves (and similar curves) described in [4].

Consider L_X and L_Y, the classical Lorenz curves of X and Y. Recall that any such curve can be identified with a function having this curve as its graph. This is usually done without mentioning it, but we like to point out the fact that we make this identification for these Lorenz curves. Form the difference function and its graph $L_Y - L_X$. This graph is a polygonal curve that goes up and down (possibly several times) and begins and ends at the value zero. Consider now the set of all slopes of this curve and rank these from highest to smallest (which is always negative, unless $L_X = L_Y$). Use these slopes to draw a new curve, denoted as $\mathcal{L}_{Y-X}$. Note that $\mathcal{L}_{Y-X}$ is different from $L_Y - L_X$. The curve $\mathcal{L}_{Y-X}$ is a concave weighted (!) polygonal curve beginning in $(0, 0)$ and ending in $(1, 0)$. Consequently, Egghe's general theory for concentration measures [4] is applicable here.

Considering the pairs of vectors (X, Y) and (X', Y') we can compare their difference curves $\mathcal{L}_{Y-X}$ and $\mathcal{L}_{Y'-X'}$. Assume now that

$$\mathcal{L}_{Y-X} < \mathcal{L}_{Y'-X'},$$

then we know that there exist measures C, such that

$$C(\mathcal{L}_{Y-X}) < C(\mathcal{L}_{Y'-X'}).$$

Indeed, if $(a_i)_{i=1,\ldots,K}$ denotes the normalized vector and $W = (w_i)_{i=1,\ldots,K}$ denotes the weight vector then, for any convex function ϕ,

$$C = \sum_{i=1}^{K} \phi\left(\frac{a_i}{w_i}\right) w_i \tag{12}$$

is such a measure [4].

LEMMA. *Let $D = (d_1, \ldots, d_N)$ be a vector weighted by $W = (w_1, \ldots, w_N)$, such that $\sum_{j=1}^{N} d_j = 0$ and assume that $(d_j/w_j)_{j=1,\ldots,N}$ is decreasing. Let $\mathcal{L}_D$ be the corresponding Egghe-Lorenz curve. Then also $(-d_{N-j+1}/w_{N-j+1})_{j=1,\ldots,N}$ is decreasing. The corresponding Egghe-Lorenz curve, denoted as $\mathcal{L}_{D'}$ is that of $D' = (d'_1, \ldots, d'_N) = (-d_N, \ldots, -d_1)$, weighted by $W' = (w'_1, \ldots, w'_N) = (w_N, \ldots, w_1)$. The curve $\mathcal{L}_{D'}$ is nothing but the mirror image of $\mathcal{L}_D$ with respect to the line $x = 1/2$.*

PROOF. We first note that $(-d_{N-j+1}/w_{N-j+1})_{j=1,\ldots,N}$ is decreasing, and hence, $\mathcal{L}_{D'}$ is an acceptable Lorenz curve. It suffices now to prove symmetry for the vertices of the polygonal curve. Let $(\sum_{j=1}^{i} w_j, \sum_{j=1}^{i} d_j)$ be the i^{th} vertex of $\mathcal{L}_D$. The symmetric point (with respect to $x = 1/2$) has abscissa $\sum_{k=i+1}^{N} w_k = \sum_{j=0}^{N-i-1} w_{N-j} = \sum_{j=0}^{N-i-1} w'_j$. So all that is left to show is that $\sum_{j=0}^{N-i-1} d'_j = \sum_{k=1}^{i} d_k$. Using the fact that $d'_j = -d_{N-j}$ yields

$$\sum_{j=0}^{N-i-1} d'_j = \sum_{j=0}^{N-i-1} -d_{N-j} = -\sum_{k=i+1}^{N} d_k = \sum_{k=1}^{i} d_k \qquad \left(\text{as } \sum_{k=1}^{N} d_k = 0\right).$$

The following corollaries generalize Proposition II.1.2 and Corollary II.1.3 of [6].

COROLLARY 1. *The Lorenz curve $\mathcal{L}_{X-Y}$ is the mirror image of $\mathcal{L}_{Y-X}$ with respect to the line $x = 1/2$. Consequently, the area under $\mathcal{L}_{X-Y}$ is equal to the area under $\mathcal{L}_{Y-X}$.*

COROLLARY 2. *$\mathcal{L}_{X-Y} < \mathcal{L}_{X'-Y'}$ if and only if $\mathcal{L}_{Y-X} < \mathcal{L}_{Y'-X'}$.*

THEOREM. *Using the notation of the preceding lemma we have the following equivalent assertions:*

(i) *$\mathcal{L}_D = \mathcal{L}_{D'}$.*
(ii) *$\mathcal{L}_D$ and $\mathcal{L}_{D'}$ are both symmetric curves with respect to the line $x = 1/2$.*
(iii) *W is a symmetric weight vector, i.e., for every $i = 1, \ldots, N : w_i = w_{N-i+1}$ and D is an antisymmetric vector, i.e., for every $i = 1, \ldots, N : d_i = -d_{N-i+1}$.*

PROOF. The equivalence between (i) and (ii) is trivial. We next show that (i) implies (iii).

If $\mathcal{L}_D = \mathcal{L}_{D'}$ then it is clear that the weight vector W is symmetric. Moreover,

$$\begin{aligned} d_1 &= -d_N, \\ d_1 + d_2 &= -d_N - d_{N-1}, \\ &\vdots \\ \sum_{j=1}^{N} d_j &= -\sum_{j=0}^{N-1} d_{N-j}, \end{aligned}$$

which implies the antisymmetry of the D-vector. Similarly, (iii) implies (i).

We will now apply this lemma and the general theory of Lorenz curves [4] in the case of i-truncations, i.e., to vectors truncated after the i^{th} component.

Let $X_{i,t} = (x_1, \ldots, x_i)$ and $X_{i+1,t} = (x_1, \ldots, x_{i+1})$ with, as always, components ranked in decreasing order. Note that $L_{X_{i,t}}$ is a polygonal curve, unweighted on intervals of length $1/i$ while $L_{X_{i+1,t}}$ is also an unweighted polygonal curve, but this time on intervals of length $1/(i+1)$. Consequently, $L_{X_{i+1,t}-X_{i,t}}$ is a weighted polygonal Lorenz curve. We will now describe this Lorenz curve. We know that on the interval $[j/i, (j+1)/i]$ the first one is a line segment joining

$$\left(\frac{j}{i}, \sum_{k=1}^{j} \frac{x_k}{\sum_{n=1}^{i} x_n}\right) \text{ or in short } \left(\frac{j}{i}, A_j\right), \text{ with } \left(\frac{j+1}{i}, \sum_{k=1}^{j+1} \frac{x_k}{\sum_{n=1}^{i} x_n}\right) \text{ or in short } \left(\frac{j+1}{i}, A_{j+1}\right),$$

for $j = 0, \ldots, i-1$. The second one connects, for $j = 0, \ldots, i$, $(j/(i+1), \sum_{k=1}^{j}(x_k/(\sum_{m=1}^{i+1} x_m)))$, with $((j+1)/(i+1), \sum_{k=1}^{j+1} x_k/(\sum_{m=1}^{i+1} x_m))$ or in short, $(j/(i+1), A'_j)$ with $((j+1)/(i+1), A'_{j+1})$, and this on the interval $[j/(i+1), (j+1)/(i+1)]$.

We note that: $\forall j = 0, \ldots, i-1 : (j+1)/i \in](j+1)/(i+1), (j+2)/(i+1)]$. Considering $L_{X_{i,t}}$ as a function we may write: $L_{X_{i,t}}((j+1)/i) = A_{j+1}$ while $L_{X_{i+1,t}}((j+1)/i)$ belongs to the line segment connecting $((j+1)/(i+1), A'_{j+1})$ with $((j+2)/(i+1), A'_{j+2})$. The equation of this line is

$$y - A'_{j+1} = \frac{A'_{j+2} - A'_{j+1}}{1/(i+1)} \left(x - \frac{j+1}{i+1} \right)$$

or

$$y = A'_{j+1} + (i+1)\left(A'_{j+2} - A'_{j+1}\right)\left(x - \frac{j+1}{i+1}\right).$$

Hence,

$$\begin{aligned} L_{X_{i+1,t}}\left(\frac{j+1}{i}\right) &= A'_{j+1} + (i+1)\left(A'_{j+2} - A'_{j+1}\right)\left(\frac{j+1}{i} - \frac{j+1}{i+1}\right) \\ &= A'_{j+1} + \left(A'_{j+2} - A'_{j+1}\right)\frac{j+1}{i}. \end{aligned}$$

Putting $\Delta L = L_{X_{i+1,t}} - L_{X_{i,t}}$ we obtain

$$\Delta L\left(\frac{j+1}{i}\right) = A'_{j+1} - A_{j+1} + \left(A'_{j+2} - A'_{j+1}\right)\frac{j+1}{i}. \tag{13}$$

Similarly, $\forall j \leq i : (j+1)/(i+1) \in [j/i, (j+1)/i[$. Hence, $L_{X_{i+1,t}}((j+1)/(i+1)) = A'_{j+1}$, while $L_{X_{i,t}}((j+1)/(i+1))$ belongs to the line segment connecting $(j/i, A_j)$ with $((j+1)/i, A_{j+1})$. The equation of this line is

$$y - A_{j+1} = \frac{A_{j+1} - A_j}{1/i}\left(x - \frac{j+1}{i}\right)$$

or

$$y = A_{j+1} + i\left(A_{j+1} - A_j\right)\left(x - \frac{j+1}{i}\right).$$

Consequently,

$$\begin{aligned} L_{X_{i,t}}\left(\frac{j+1}{i+1}\right) &= A_{j+1} + i\left(A_{j+1} - A_j\right)\left(\frac{j+1}{i+1} - \frac{j+1}{i}\right) \\ &= A_{j+1} - \left(A_{j+1} - A_j\right)\frac{j+1}{i+1} \end{aligned}$$

and

$$\Delta L\left(\frac{j+1}{i+1}\right) = A'_{j+1} - A_{j+1} + \left(A_{j+1} - A_j\right)\frac{j+1}{i+1}. \tag{14}$$

Equations (13) and (14) determine all vertices of $L_{X_{i+1}} - L_{X_i}$. Next, we want to determine the slopes of all segments in $L_{X_{i+1}} - L_{X_i}$

(I) $$\frac{\Delta L\left((j+1)/(i+1)\right) - \Delta L\left(j/i\right)}{(j+1)/(i+1) - j/i}, \qquad \text{with } j = 0, \ldots, i-1,$$

(II) $$\frac{\Delta L\left((j+1)/i\right) - \Delta L\left((j+1)/(i+1)\right)}{(j+1)/i - (j+1)/(i+1)}, \qquad \text{with } j = 0, \ldots, i-1.$$

Again using (13) and (14) we find

$$\Delta L\left(\frac{j}{i}\right) = A'_j - A_j + (A'_{j+1} - A'_j)\frac{j}{i},$$

$$\Delta L\left(\frac{j+1}{i}\right) = A'_{j+1} - A_{j+1} + (A'_{j+2} - A'_{j+1})\frac{j+1}{i},$$

$$\Delta L\left(\frac{j+1}{i+1}\right) = A'_{j+1} - A_{j+1} + (A_{j+1} - A_j)\frac{j+1}{i+1}.$$

Then

$$\Delta L\left(\frac{j+1}{i+1}\right) - \Delta L\left(\frac{j}{i}\right) = (A'_{j+1} - A'_j) - (A_{j+1} - A_j) + (A_{j+1} - A_j)\frac{j+1}{i+1} - (A'_{j+1} - A'_j)\frac{j}{i} \quad (15)$$

and

$$\Delta L\left(\frac{j+1}{i}\right) - \Delta L\left(\frac{j+1}{i+1}\right) = (A'_{j+2} - A'_{j+1})\frac{j+1}{i} - (A_{j+1} - A_j)\frac{j+1}{i+1}. \quad (16)$$

In this way we can determine all slopes and rank them in decreasing order. This leads to $\mathcal{L}_{(X_{i+1,t}-X_{i,t})}$.

An Example

Let $X_{i,t} = (5,3,2)$ and $X_{i+1,t} = (5,3,2,1)$. Then $L_{X_{i,t}}$ consists of line segments connecting the points $(0,0) - (1/3, 5/10) - (2/3, 8/10) - (1,1)$, while $L_{X_{i+1,t}}$ consists of the line segments connecting $(0,0) - (1/4, 5/11) - (2/4, 8/11) - (3/4, 10/11) - (1,1)$. At the points

$$0, \frac{3}{12}, \frac{4}{12}, \frac{6}{12}, \frac{8}{12}, \frac{9}{12}, 1$$

the first Lorenz curve takes the values

$$0, \frac{3}{8}, \frac{5}{10}, \frac{13}{20}, \frac{8}{10}, \frac{17}{20}, 1$$

while the second one takes the values

$$0, \frac{5}{11}, \frac{6}{11}, \frac{8}{11}, \frac{28}{33}, \frac{10}{11}, 1.$$

In these points differences between the two Lorenz curves are (we are calculating now values of $L_{X_{i+1,t}} - L_{X_{i,t}}$)

$$0, \frac{7}{88}, \frac{1}{22}, \frac{17}{220}, \frac{16}{330}, \frac{13}{220}, 0.$$

Note that we knew already that all these differences must be positive. Consequently the slopes of the consecutive line segments of this difference curve are

$$\frac{7/88}{3/12} = \frac{7}{22}; \qquad \frac{-3/88}{1/12} = -\frac{9}{22}; \qquad \frac{7/220}{2/12} = \frac{21}{110};$$

$$\frac{-19/660}{2/12} = -\frac{57}{330}; \qquad \frac{7/660}{1/12} = \frac{7}{55}; \qquad \frac{-13/220}{3/12} = -\frac{13}{55}.$$

Ranking these slopes from largest to smallest yields the following points of $\mathcal{L}_{(X_{i+1,t}-X_{i,t})}$:

$$(0,0) - \left(\frac{3}{12}, \frac{7}{88}\right) - \left(\frac{5}{12}, \frac{49}{440}\right) - \left(\frac{6}{12}, \frac{161}{1320}\right) - \left(\frac{8}{12}, \frac{41}{440}\right) - \left(\frac{11}{12}, \frac{3}{88}\right) - (1,0).$$

The corresponding curve shows no symmetry whatsoever.

7. CONCLUSION

We have shown, in the case of first-order comparisons, that truncated vectors may be incomparable, while partial ones are always completely comparable. Similarly for second-order comparisons, partial vectors can be compared and yield a totally ordered double sequence, while truncated ones may be incomparable. Finally, we described how to make second-order comparisons for vectors with a different number of sources.

REFERENCES

1. L. Egghe and R. Rousseau, Aging, obsolescence, impact, growth and utilization: Definitions and relations, *Journal of the American Society of Information Science* **51** (11), 1004–1017, (2000).
2. R. Rousseau, Concentration and evenness measures as macro-level scientometric indicators, In *Research and University Evaluation*, (Edited by G.-H. Jiang), pp. 72–89, (in Chinese), Red Flag Publishing House, Beijing, (2001); English translation available from the author.
3. M.O. Lorenz, Methods of measuring concentration of wealth, *Journal of the American Statistical Association* **9**, 209–219, (1905).
4. L. Egghe, Construction of concentration measures for general Lorenz curves using Riemann-Stieltjes integrals, *Mathl. Comput. Modelling* **35** (9/10), 1149–1163, (2002).
5. L. Egghe, Sampling and concentration values of incomplete bibliographies, *Journal of the American Society of Information Science and Technology* **53**, 271–281, (2002).
6. L. Egghe and R. Rousseau, Symmetric and asymmetric theory of relative concentration and applications, *Scientometrics* **52**, 261–290, (2001).
7. L. Egghe and R. Rousseau, The core of a scientific subject: An exact definition using concentration and fuzzy sets, In *Proceedings of the 8^{th} International Conference on Scientometrics and Informetrics*, (Edited by M. Davis and C.S. Wilson), pp. 147–156, BIRG (UNSW), Sydney, (2001).
8. H. Theil, *Economics and Information Theory*, North-Holland, Amsterdam, (1967).
9. G.P. Patil and C. Taillie, Diversity as a concept and its measurement, *Journal of the American Statistical Society* **77**, 548–561, (1982).
10. L. Egghe and R. Rousseau, Evolution of information production processes and its relation to the Lorenz dominance order, *Information Processing and Management* **29**, 499–513, (1993).