UNIVERSITÀ DI SIENA
1240

QUADERNI DEL DIPARTIMENTO
DI ECONOMIA POLITICA E STATISTICA

Mariateresa Ciommi
Ernesto Savaglio

On multidimensional diversity orderings with categorical variables

n. 711 – Luglio 2015

On multidimensional diversity orderings with categorical variables

Mariateresa Ciommi*and Ernesto Savaglio†

June 16, 2015

Abstract

We axiomatically characterize a family of multidimensional diversity orderings which evaluates the dissimilarity of a population of individuals who differ in many personal attributes represented by categorical variables. An empirical application highlights the relevance and the easy applicability of our analysis.

JEL Classification: I30, I32, D63

Key Words: Multidimensional Diversity, Categorical variables, Orderings.

1 Introduction

Motivation. It is controversial if diversity is a desirable requirement for a society. A policy maker could be interested in removing diversity when it is detrimental for the collective welfare as in the case of a society in which there are individuals of different ethnicity with difficulty of integration. On the contrary, when people receive benefits from having different skills, different talents or different points of view, diversity is worth preserving. In order to consider these individual dissimilarities, economic theorists have recently focused their attention on the problem of how to properly measure (social) diversity.[1] The classes of diversity measures proposed typically consider *one individual characteristic* at a time (see, e.g. Alesina, Devleeschauwer, William, Kurlat and Wacziarg [1], Fearon [10] and Patsiurko, Campbell and Hall [15]). However, since people differ in *many relevant characteristics*, more than one (individual) variable needs to be considered in order to compare societies in terms of diversity. Thus, the aim of the present work is to define and characterize a class of *multidimensional diversity orderings* that can be implemented when data on individuals are race, ethnicity, gender, socioeconomic status, class of age, physical abilities, religious or political beliefs, etc.

*Department of Economic and Social Sciences, Università Politecnica delle Marche - m.ciommi@univpm.it

†Department of Economics, University "G.d'Annunzio" Chieti-Pescara - ernesto@unich.it

[1] See Alesina and La Ferrara [2], Gravel [11], Nehring and Puppe [14] and Weitzman [19], among others.

Contents. We model each individual dimension as a categorical variable, a natural restriction that allows us to treat in a unified (and admittedly simplified) framework several types of variables. Each individual is represented by her endowment (i.e. her vector of characteristics) and the diversity among individuals is supposed to be the *dissimilarity* of their endowments. We define two individuals as *similar* if they have the same endowment, otherwise their diversity is just computed by summing the number of dimensions in which they differ. The result of this two-by-two individual comparison is then summarized by a *counting multidimensional diversity index* $(CMDI)$.[2] We observe that the $CMDI$ reduces to a transformation of the celebrated Simpson's diversity index [18] for the case in which individuals differ in only one variable. The $CMDI$ computes the total dissimilarity of the individuals in a society and induces a *class of multidimensional diversity orderings* over the set of all multidimensional distributions of categorical variables. We provide a *minimalist* axiomatic characterization of such a class that involves only two suitable properties: a *Monotonicity* axiom, claiming that adding one individual to a society is a diversity-enhancing operation, and a *Separability* axiom, allowing to preserve the relative ranking between two societies if a new set of individuals, satisfying some reasonable conditions, is added to both communities.

Related Literature. Our work is related to the theoretical economic literature on the measurement of diversity and in particular to the work of Weitzman [19] (that defines the diversity of a set as the pair-wise dissimilarity between its elements and introduces a cardinal numeric measure of distance to compute the latter), of Nehring and Puppe [13] (who generalize Weitzman's purpose) and of Bossert, Pattanaik and Xu [5] (who propose an axiomatic characterization of Weitzman's work). The instrumental role of option-diversity assessments in the evaluation of opportunity sets in terms of freedom of choice, as highlighted by the two early contributions of Pattanaik and Xu [16] and Bossert, Pattanaik and Xu [5], is another major source of the growing concern for diversity rankings. Those works rely on a *distance function* that establishes if objects of different sets are similar or dissimilar. Finally, Pattanaik and Xu [17] characterize an ordinal distance function that induces a *diversity dominance* criterion between sets. All these contributions make use of some *proximity* or dissimilarity notion. We follow this trend by considering a *similarity pseudometric* that for any pair of individual endowments computes the number of characteristics in which they do *not* differ. Then, all the dissimilarities are *counted* by a class of multidimensional diversity indices, depending on a parameter. This class -to repeat- induces a class of diversity orderings that is axiomatically characterized.

It is worth noticing that Bossert, D'Ambrosio and La Ferrara [6] characterize a generalized ethno-linguistic *fractionalization index*, adopting an intuitive multidimensional diversity approach, which uses real and ordinal individual variables to measure fractionalization. Nonetheless, the present work is, up to our knowledge, the first attempt to study and characterize a class of multidimensional *diversity orderings*.

The rest of the paper is organized as follows. In section 2, we introduce the notation and

[2] A *counting approach* is also used in a context of deprivation (Atkinson, [4]), social exclusion (Chakravarty and D'Ambrosio [7]) and multidimensional poverty (Alkire and Foster [3] and Lasso de la Vega [12]).

definitions, discuss the two axioms used to characterize our family of multidimensional diversity orderings and present the main result. In section 3, we show that the class of diversity orderings we introduce is also a useful tool, easy to be implemented, when a policy maker or a practitioner have to evaluate the effectiveness of a public policy concerned with a reduction or a preservation of (social) diversity. Section 4 concludes, while proofs are collected in appendix.

2 On a class of Counting Multidimensional Diversity Orderings

Notation and definitions. We represent a population of $N = \{1, ..., n\}$ individuals differing in $k \geq 1$ categorical variables as e.g. gender, political choice, religion, level of education, mother tongue, race, etc, by a $n \times k$ matrix $\mathbf{X} = [x_{ij}]$. Therefore, $\mathbf{X}$ is a multidimensional discrete distribution whose generic entry $x_{i,j}$ denotes the realization of the j-th variable for the i-th individual. We denote the set of all $n \times k$ matrices by $M(n, k)$ and with x_i and x^j the row-vector and column-vector, respectively. Then, for any $\mathbf{X} \in M(n, k)$ and for any pair of individuals $i, j \in N$,

$$c_{ij} = \left|\{h \mid x_{ih} = x_{jh} \text{ for } h = 1, 2, ..., k\}\right|,$$

is a measure of the individual *similarity* which computes the number of variables in which two individuals i and j do *not* differ. A *coincidence matrix* is defined as the $n \times n$ matrix $\mathbf{C_X} = [c_{i,j}]$ with $c_{ij} \in \{0, 1, ..., k\}$, $c_{ii} = k$ and $c_{ij} = c_{ji}$ for any i and j. We denote by $C = \bigcup_M \mathbf{C_X}$ the set of all the *feasible* coincidence matrices.

A function $S^\alpha(\mathbf{X}, n, k) : C \to \mathbb{N}$ aggregates all the information provided by a coincidence matrix according to the following rule:

$$S^\alpha(\mathbf{X}, n, k) = n^2 k^\alpha - \sum_{i=1}^{n} \sum_{j=1}^{n} c_{i,j}^\alpha \qquad \text{with } \alpha \in \mathbb{N} \setminus \{0, 1\}. \tag{1}$$

It is worth noticing that the multidimensional parametric functional (1) reduces to the famous Herfindahl-Hirschmann index (HHI) of concentration in the unidimensional case.

Remark 1. Indeed, for $k = 1$ (unidimensional diversity), equation (1) can be written as:

$$S^\alpha(\mathbf{X}, n, 1) = n^2[1 - \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} c_{i,j}^\alpha]$$

Suppose now that the population is partitioned in l distinct groups $(n_1,\ n_2, ..., n_l)$ with n_i the number of individuals in the i-th group, for $i \in \{1, ..., l\}$. Since $c_{i,j}^\alpha = c_{i,j}$ for any α, we have:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} = 2 \sum_{i=1}^{l} \frac{n_i(n_i - 1)}{2} + n = \sum_{i=1}^{l} [n_i(n_i - 1)] + \sum_{i=1}^{l} n_i = \sum_{i=1}^{l} n_i^2$$

Consequently,

$$S^\alpha(\mathbf{X}, n, 1) = n^2 \left[1 - \frac{1}{n^2} \sum_{i=1}^{l} \left(\frac{n_i}{n}\right)^2\right],$$

that is n^2 minus the *Simpson Diversity Index* or HHI. $\square$

Define now with $d_{i,j} = k^\alpha - c^\alpha_{i,j}$ a measure of the level of diversity between individual i and j,[3] with $\mathbf{D_X} = [d_{i,j}] \in \mathbb{R}^{n\times n}$ the corresponding *diversity matrix* and with

$$s^\alpha_i(\mathbf{X}, n, k) = nk^\alpha - \sum_{j=1}^{n} c^\alpha_{i,j} = \sum_{j=1}^{n} \left(k^\alpha - c^\alpha_{i,j}\right) = \sum_{j=1}^{n} d_{i,j}$$

the *diversity score* representing the diversity of an individual i with respect to the society in which she lives, for $i = \{1, 2, \dots n\}$.[4] We refer to $\mathbf{s}^\alpha(\mathbf{X}, n, k) = (s^\alpha_1(\mathbf{X}, n, k), \dots, s^\alpha_n(\mathbf{X}, n, k))^T$ as a *diversity profile*. Thus, expression (1) can be restated as follows:

$$S^\alpha(\mathbf{X}, n, k) = \sum_{i=1}^{n}\sum_{j=1}^{n} d_{i,j} \tag{2}$$

We refer to (2) as a class of *Counting Multidimensional Diversity Indices* ($CMDI$s).

In what follows, we characterize the class of *multidimensional diversity orderings* induced by the $CMDI$s. In particular, denote with $(M(n, k), \ \succeq)$ a binary relational system such that, for any $\mathbf{X}, \mathbf{Y} \in M(n, k)$, $\mathbf{X} \succeq \mathbf{Y}$ means that $\mathbf{X}$ shows at least as much diversity as $\mathbf{Y}$, then the class of diversity orderings $\succeq_\alpha$ we characterize is defined as follows:[5]

Definition 1. For any $\mathbf{X} \in M(n, k)$ and $\mathbf{Y} \in M(n', k')$

$$\mathbf{X} \succeq_\alpha \mathbf{Y} \quad \text{if and only if} \quad S^\alpha(\mathbf{X}, n, k) \geq S^\alpha(\mathbf{Y}, n', k') \tag{3}$$

with $\alpha \in \mathbb{N}\backslash\{0, 1\}$.

In words, we say that the society $\mathbf{X}$ is at least as much diverse as $\mathbf{Y}$ if and only if the sum of the individual diversity $d_{i,j}$ among all the individuals in the first society is not smaller than the sum of the individual diversity among all the individuals in the second one. We refer to $\succeq_\alpha$ as a *class of Counting Multidimensional Diversity Orderings,* hereafter $CMDO$s. We characterize the CMDO using only two very compelling properties.

A characterization result. The first plausible axiom for a diversity orderings claims that if two identical societies are equivalent in terms of diversity, then, if we add one individual to a society, the diversity should not decrease. Analytically:

Monotonicity - M. For all $\mathbf{X}, \mathbf{Y} \in M(n, k)$,

$$\begin{cases} \text{if } x_{i,j} = y_{i,j} \text{ for any } i = 1, \dots, n \text{ and } j = 1, \dots, k, \text{ then } \mathbf{X} \sim \mathbf{Y} \\ \text{if } \mathbf{Y} = \mathbf{X}\backslash\left[x_i\right], \text{ then } \mathbf{X} \succeq \mathbf{Y}. \end{cases}$$

We consider now a society $\mathbf{Z} \in M(n, k)$ to which we add a group G of $i = \{1, 2, ..., g\}$ identical individuals to obtain a new society $\mathbf{Z}' = \mathbf{Z} \cup G$ with a population size of $n + g$. If we compute the

[3] Since the value of $d_{i,j}$ depends on α, we refer to $d_{i,j}$ as a α-dependent measure of diversity.
[4] If there is no possibility of confusion we use s^α_i for $s^\alpha_i(\mathbf{X}, n, k)$.
[5] The symmetric and asymmetric component of $\succeq_\alpha$ are denoted with $\sim_\alpha$ and $\succ_\alpha$ respectively.

diversity matrix for $\mathbf{Z}^{'}$, we obtain the following block matrix:

$$\mathbf{D}_{\mathbf{Z}'} = \begin{pmatrix} \mathbf{D}_{\mathbf{Z}} & \mathbf{G}_{\mathbf{Z}} \\ \mathbf{G}_{\mathbf{Z}}^{T} & \mathbf{0} \end{pmatrix} \tag{4}$$

where the block $\mathbf{0}$ is the null matrix representing the diversity among individuals in the group G, $\mathbf{G}_{\mathbf{Z}}$ is the matrix whose elements express the diversity between individuals in $\mathbf{Z}$ and in G and $\mathbf{G}_{\mathbf{Z}}^{T}$ is the transpose of $\mathbf{G}_{\mathbf{Z}}$.[6] Computing the diversity scores for $\mathbf{Z}'$, we distinguish the case in which individual i belongs to society $\mathbf{Z}$:

$$s_i^{\alpha}(\mathbf{Z}', n+g, k) = \sum_{j=1}^{n+g} \left(k^{\alpha} - c_{i,j}^{\alpha}\right) \text{ for } i \in \{1, ..., n\}$$

from the case in which she belongs to the new group G:

$$s_i^{\alpha}(\mathbf{Z}', n+g, k) = \sum_{j=1}^{n+g} \left(k^{\alpha} - c_{i,j}^{\alpha}\right) \text{ for } i \in \{n+1, n+2, ..., n+g\} \tag{5}$$

Since all the individuals in G are identical $\sum_{j=n+1}^{n+g} \left(k^{\alpha} - c_{i,j}^{\alpha}\right) = 0$ for $i \in \{n+1, n+2, ..., n+g\}$, then equation (5) becomes:

$$s_i^{\alpha}(\mathbf{Z}', n+g, k) = \sum_{j=1}^{n} \left(k^{\alpha} - c_{i,j}^{\alpha}\right) + \sum_{j=n+1}^{n+g} \left(k^{\alpha} - c_{i,j}^{\alpha}\right) = \sum_{j=1}^{n} \left(k^{\alpha} - c_{i,j}^{\alpha}\right) \text{ for } i \in \{n+1, n+2, ..., n+g\}.$$

Thus, the aggregate measure of multidimensional diversity for the individuals in the group G with respect to the individuals in the society $\mathbf{Z}$ is:

$$gs_i^{\alpha}(\mathbf{Z}', n+g, k) = g \sum_{j=1}^{n} \left(k^{\alpha} - c_{i,j}^{\alpha}\right) \quad \text{for } i \in \{n+1, n+2, ..., n+g\}.$$

Finally, if $\mathbf{X}'$ and $\mathbf{Y}'$ are two distribution matrices obtained by adding two groups G and F of identical individuals to $\mathbf{X} \in M(n, k)$ and $\mathbf{Y} \in M(n^{'}, k)$ respectively and if

$$gs_i^{\alpha}(\mathbf{X}', n+g, k) = fs_j^{\alpha}(\mathbf{Y}', n\prime + f, k) \text{ for } i \in \{1, 2, ..., g\} \text{ and } j \in \{1, 2, ..., f\},$$

i.e. the value of the (multidimensional) diversity for the two groups of individuals added to $\mathbf{X}$ and $\mathbf{Y}$ is the same, then the multidimensional diversity ranking of $\mathbf{X}'$ and $\mathbf{Y}'$ should be consistent with that of $\mathbf{X}$ and $\mathbf{Y}$, namely:

α-Separability - α-S. For any $\mathbf{X} \in M(n, k)$ and $\mathbf{Y} \in M(n', k)$ and for any two groups G and F of $i \in \{1, 2, ..., g\}$ and $j \in \{1, 2, ..., f\}$ identical individuals, if $gs_i^{\alpha}(\mathbf{X}', n+g, k) = fs_j^{\alpha}(\mathbf{Y}', n'+f, k)$, then

$$\mathbf{X} \succeq \mathbf{Y} \text{ if and only if } \mathbf{X} \cup G = \mathbf{X}' \succeq \mathbf{Y}' = \mathbf{Y} \cup F.$$

[6] Notice that all the columns of $\mathbf{G}_{\mathbf{Z}}$ (and of course all the rows of $\mathbf{G}_{\mathbf{Z}}^{T}$) are equal.

We show that the axioms **M** and α**-S** are necessary and sufficient conditions to characterize the class of Counting Multidimensional Diversity Orderings in (3).

Theorem 1. *A diversity total ordering $\succeq$ satisfies M and α-S if and only if $\succeq = \succeq_\alpha$.*

Theorem 1 allows to rank multidimensional distributions of individuals who differ in several characteristics in terms of diversity. CMDOs adopt an intuitive approach to the identification of individuals as (dis-)similar. Our characterization is minimalistic since it uses only two suitable key properties. Finally, it is worth observing that Theorem 1 can be applied using real data to obtain meaningful results as the illustrative example below shows.

3 Empirical findings

In order to show that the class of multidimensional diversity orderings we characterize is easy to be implemented and could be a useful tool of policy analysis, we compare people aged over 50, living in some European Countries in two different periods, in terms of their diversity.

The fertility decline below the replacement level and the increasing in life expectancy have transformed Europe into one of the oldest region in the world. For instance, the percentage of people aged over 65 was 12.8% in 1985 and it is increased to 17.4% in 2010 (source, Eurostat [9]). European Governments (EG) are particularly concerned with this issue (2012 was indeed declared the European Year of Active Ageing[7]) and expect a more active role in the society from older people (namely more intra-olders solidarity and help) in order to strengthen their independency and responsibility to arrange their own life and to decrease the provision of expensive health public goods and social services.

To proceed, we use microdata from the first (2004/06) and the fourth (2010/12) waves of the Survey of Health, Ageing and Retirement in Europe (SHARE).[8] SHARE collects microdata (cross-national panel) of individuals over 50 years living in 20 countries (19 European countries plus Israel) with respect to several variables related to health, socioeconomic status and social and family network. We use the individual-level information on all household-heads and the diversity ordering is computed according to the following personal characteristics:

Country of birth. The variable collects the country of birth of all the individuals in the sample.

Employment status. According to SHARE definition, each individual is attributed to one of five categories: (i) employed or self-employed (including working for family business), (ii) retired, (iii) permanently sick or disabled, (iv) homemaker, (v) unemployed.

These are typical variables used to analyze individual diversity. We go further adding four

[7] Active ageing is a situation where people continue to participate to the formal labor market, as well as are engaged in other unpaid productive activities (as e.g. care provision to family members and volunteering), and live healthy, independent and secure lives as they age.

[8] In particular, we use data from SHARE Wave 1 release 2.6.0 (DOI: 10.6103/SHARE.w1.260) and SHARE Wave 4 release 1.1.1 (DOI: 10.6103/SHARE.w4.111).

other variables that collect information on the participation of older people to social life (in the month before the interview).

Religious Activity. It collects information on the participation of individuals to a religious organization, such as church, synagogue, mosque and so on.

Voluntary Activity. Interviewers have to declare if the do voluntary or charity work.

Political Participation. Taken part in a political or community-related organization.

Sport. It summarizes information on the frequency of participation to vigorous and moderate physical activity, assuming values from 1 (more than once a week) to 4 (hardly ever, or never).

We rely on the aforementioned variables to compute the multidimensional diversity for 8 European Countries.[9]

The robustness of the country-ranking is checked by a *replacement sub-sampling bootstrap* (see e.g. Efron and Tibshirani [8]) on the sample. In particular, we reduce the effect of random sampling errors in our bootstrap procedure, for each countries and for both the years, by performing $s = 1000$ times the following experiment: (a) we randomly select (re-samples with replacement) a population of $m = 1000$ individuals from the dataset; (b) for each of these samples, we compute $S^{\alpha}(\mathbf{X}, n, k)$,[10] that is our statistics, and we analyze the empirical distribution of the results; (c) finally, we calculate the mean value of each of the s replicates in the experiment and the obtained value is chosen to represent the value of the multidimensional diversity for a given countries in a given year. Table 1 summarizes our results. Looking at the rankings, the most diverse countries are

Table 1: Rankings according to the CMDI

n.	*Wave 1*	*Wave 4*
1	Italy	Denmark
2	Spain	Spain
3	Denmark	Italy
4	France	France
5	Belgium	Austria
6	Austria	Belgium
7	Netherlands	Netherlands
8	Switzerland	Switzerland

Netherlands and Switzerland in both the waves. Whereas, the countries displaying the lower level of (multidimensional) diversity are Italy and Denmark, in wave 1 and wave 4 respectively.

Four countries preserve their positions, namely, Spain, France, Netherlands and Switzerland. Whereas, there are two pairs of countries that switch their relative position, that is Italy-Denmark

[9] Wave 1 and wave 4 account for 11 and 14 countries, respectively. However, we reduce the sample eliminating those countries that participating in only one wave. To the final sample we also remove Germany since the number of people surviving after the elimination of missing values was less then the 10% of the original sample. Hence our sample account for 8 countries.

[10] The CMDI depends on α. Since any increasing transformation is order-preserving, we only compute $S^2(\mathbf{X}, n, k)$.

and Belgium-Austria.

The computation of confidence intervals reveals that the orderings are consistent in both years, since the confidence intervals at 99% for the mean do not overlap. Thus, we check the robustness of the rankings also via a parametric and a non-parametric test. We apply normality tests to verify that all the distributions of the s-data are *normal*, then we apply a pair-wise t-test to the differences between the means of the s-data of any pair of consecutively-ranked countries.[11]

The non-parametric test consists in constructing a new vector whose entries are the differences between the ordered components of each consecutive pair of the s-data vectors. For each of these new vector distributions, we compute the percentiles 2.5 and 97.5. Also in this case, we obtain that the rankings are statistically significant.

4 Conclusive remarks

It is worth noticing that our class of diversity orderings departs from the rest of literature on diversity adopting a multidimensional (and therefore possibly less controversial) approach. Indeed, individuals differ in several dimensions and a diversity measure has to take into account the many aspects of personal diversity in order to be natural and non-controversial. That quite elementary consideration prompted us to extend the standard analysis to a multivariate context. Such an extension of the scope of diversity measurement may substantially improve our understanding of diversity in any given population and may well have far-reaching policy implications. In order to keep the analysis as general as possible, we have compared multidimensional discrete distribution of categorical variables and we have axiomatically characterized a class of diversity orderings in such a multidimensional framework. To the best of our knowledge, there have been no previous attempts to characterize *multidimensional diversity orderings* within that setting. As a by product of our study, we show the class of diversity indices, that induces our orderings, reduces to a transformation of the Simpson index for the unidimensional case. Finally, the recent availability of individual data on different dimensions of diversity makes it possible to apply the orderings we characterize to address several interesting empirical and eventually policy issues.

Acknowledgement. We thank Eugenio Peluso, Stefano Vannucci and the participants to the GRASS Workshop (2011), Ecineq (2013) and ASSET (2013) for useful comments. A special thank is due to Casilda Lasso de la Vega who helped us to improve the quality of our work with many detailed remarks and useful insights. The usual disclaimers apply.

5 Appendix

Proof. To prove that $\succeq_\alpha$ is a total ordering satisfying Monotonicity and $\alpha-$Separability is straightforward.

[11] The results of all these controls can be provided by the authors upon request.

In order to prove sufficiency, consider a society $\mathbf{X}^*$ with n individuals, all equal but one, who differ in only one attribute and a society $\mathbf{X}_*$ composed of two individuals who differ in $\overline{k} > 1$ attributes.

A straightforward computation of $S^\alpha(\mathbf{X}, n, k)$ for $\mathbf{X}^*$ entails that $S^\alpha(\mathbf{X}^*, n, 1) = 2(n-1)$. We prove that any society $\mathbf{X}$ can be associated to a $\mathbf{X}^*$ with the same level of diversity, namely:

Lemma 1 *For all $\mathbf{X} \in M(n,k)$ there exists a $\mathbf{X}^* \in M(n^{'}, 1)$ such that $S^\alpha(\mathbf{X}, n, k) = S^\alpha(\mathbf{X}^*, n^{'}, 1)$*

Proof of Lemma 1. For any $\mathbf{X} \in M(n,k)$,

$$S^\alpha(\mathbf{X}, n, k) = \sum_{i=1}^{n}\sum_{j=1}^{n}\left(k^\alpha - c_{ij}^\alpha\right) \in \{0, 2, 4, ..., n(n-1)k^\alpha\}.$$

Suppose $S^\alpha(\mathbf{X}, n, k) = d$, then consider a $\mathbf{X}^* \in M((d/2)+1, 1)$, i.e. a society with $(d/2)+1$ individuals who differ in only one attribute and such that $d/2$ of the individuals are all equal but one. Hence, the fact that $S^\alpha(\mathbf{X}^*, \frac{d}{2}+1, 1) = 2(\frac{d}{2}) = d$ completes the proof. $\square$

We observe that, for any $\mathbf{X}_*$, $S^\alpha(\mathbf{X}_*, 2, \overline{k}) = 2\overline{k}^\alpha$. Moreover, if the number of individuals in a society $\mathbf{X}^*$ is equal to $\overline{k}^\alpha + 1$, where $\overline{k}$ is the number of attributes of the two-people society $\mathbf{X}_*$, then $S^\alpha(\mathbf{X}^*, n, 1) = S^\alpha(\mathbf{X}_*, 2, \overline{k})$, i.e.:

Lemma 2 *For any $\mathbf{X} \in M(2,k)$, let $\mathbf{X}^* \in M(\frac{S^\alpha(\mathbf{X},2,k)}{2} + 1, 1)$ be the society associated to $\mathbf{X}$ according to Lemma 1. Then, for any multidimensional diversity total ordering $\succeq$ satisfying axiom $\boldsymbol{\alpha} - \mathbf{S}$, $\mathbf{X} \sim \mathbf{X}^*$.*

Proof of Lemma 2. For any $\mathbf{X} \in M(2,k)$, consider the diversity matrices of $\mathbf{X}$ and $\mathbf{X}^*$, namely:

$$\mathbf{D_X} = \begin{pmatrix} 0 & k^\alpha - c^\alpha \\ k^\alpha - c^\alpha & 0 \end{pmatrix} \qquad \mathbf{D_{X^*}} = \begin{pmatrix} 0 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix}$$

where $c = c_{12} = c_{21}$. Take $\mathbf{Y} \in M(2,k)$ and $\mathbf{Z} \in M(2,1)$ such that individuals in both societies are the same, i.e. $c = k$ in $\mathbf{Y}$ and $c = 1$ in $\mathbf{Z}$, then $\mathbf{D_Y} = \mathbf{D_Z}$. Add one individual y_3 to $\mathbf{Y}$ such that $c_{31} = c_{32} = c$, then, by $\boldsymbol{\alpha} - \mathbf{S}$, the diversity matrix associated to $\mathbf{Y'} = \mathbf{Y} \cup \{y_3\}$ is:

$$\mathbf{D_{Y'}} = \begin{pmatrix} 0 & 0 & k^\alpha - c^\alpha \\ 0 & 0 & k^\alpha - c^\alpha \\ k^\alpha - c^\alpha & k^\alpha - c^\alpha & 0 \end{pmatrix}$$

Add a group G of identical individuals to $\mathbf{Z}$ whose individuals are different from those already in $\mathbf{Z}$ and whose number g is equal to $k^\alpha - c^\alpha$, then the diversity matrix for the new society $\mathbf{Z'} = \mathbf{Z} \cup G$

is:

$$\mathbf{D}_{\mathbf{Z}'} = \begin{pmatrix} 0 & 0 & 1 & 1 & \cdots & 1 \\ 0 & 0 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & 1 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

Thus, by $\boldsymbol{\alpha} - \mathbf{S}$, $\mathbf{Y}' \sim \mathbf{Z}'$. If we remove the second individual from both societies in order to obtain $\mathbf{Y}'' = \mathbf{Y}' \backslash \{y_2\}$ and $\mathbf{Z}'' = \mathbf{Z}' \backslash \{z_2\}$ such that $\mathbf{Y}'' \in M(2,k)$ and $\mathbf{Z}'' \in M(1+g,1)$ and we compute the diversity matrices for the two new societies, we have:

$$\mathbf{D}_{\mathbf{Y}''} = \begin{pmatrix} 0 & k^\alpha - c^\alpha \\ k^\alpha - c^\alpha & 0 \end{pmatrix} \qquad \mathbf{D}_{\mathbf{Z}''} = \begin{pmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Since, for $\{y_2\}$ and $\{z_2\}$, $s_2^\alpha(\mathbf{Y}'', 2, k) = k^\alpha - c^\alpha = g = s_2^\alpha(\mathbf{Y}'', 1+g, 1)$, then, by $\boldsymbol{\alpha} - \mathbf{S}$, we obtain the desired result. $\square$

Thus, suppose that $\succeq$ is a total order that satisfies $\mathbf{M}$ and $\mathbf{S}$, then we need to show that for all $\mathbf{X} \in M(n+1,k)$ and its associated matrix $\mathbf{X}^*$:

$$S^\alpha(\mathbf{X}, n+1, k) = S^\alpha(\mathbf{X}^*, n'+1, 1) \Rightarrow \mathbf{X} \sim \mathbf{X}^* \tag{6}$$

By induction, if $n = 1$, $\mathbf{X} \in M(2,k)$ then (6) follows from the Lemma 2. Suppose (6) holds for $n-1$, then (6) holds for all society $\mathbf{X} \in M(n,k)$. Indeed,

$$S^\alpha(\mathbf{X}, n, k) = \sum_{i=1}^{n} \sum_{j=1}^{n} \left(k^\alpha - c_{ij}^\alpha\right) = \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \left(k^\alpha - c_{ij}^\alpha\right) + 2\sum_{j=1}^{n} \left(k^\alpha - c_{nj}^\alpha\right) \tag{7}$$

We denote by $\mathbf{X} \backslash \{x_n\}$ the society $\mathbf{X}$ in which we remove the n-th individual. Then, the individual diversity score for that individual is $s_n^\alpha(\mathbf{X}, n, k) = \sum_{j=1}^{n} \left(k^\alpha - c_{nj}^\alpha\right)$. By Lemma 1, $S^\alpha(\mathbf{X} \backslash \{x_n\}, n-1, k) = S^\alpha((\mathbf{X} \backslash \{x_n\})^*, n, 1)$ and by induction we have $\mathbf{X} \backslash \{x_n\} \sim (\mathbf{X} \backslash \{x_n\})^*$. Suppose now to add to $\mathbf{X} \backslash \{x_n\}$ an individual and to $(\mathbf{X} \backslash \{x_n\})^*$ a group G of identical individuals, i.e. $(\mathbf{X} \backslash \{x_n\}) \cup \{x_n\} = \mathbf{X}$ and $(\mathbf{X} \backslash \{x_n\})^* \cup G = \mathbf{X}^*$. Then, the diversity of the new individual in $\mathbf{X}$ is $1 \cdot s_n^\alpha(\mathbf{X}, n, k)$ while, in the other society, we have exactly $s_n^{\mathbf{X}}$ individuals with the same diversity value (that is 1). Thus, the level of diversity is $s_n^{\mathbf{X}} \cdot 1$. Therefore, by α-Separability, we have that $\mathbf{X} \sim \mathbf{X}^*$.

Now, let $\mathbf{X} \in M(n,k)$ and $\mathbf{Y} \in M(m,k')$. By Lemma 1, it is possible to find two societies $\mathbf{X}^* \in M(n',1)$ with $n' = \frac{S^\alpha(\mathbf{X},n,k)}{2} + 1$ and $\mathbf{Y}^* \in M(m',1)$ with $m' = \frac{S^\alpha(\mathbf{Y},m,k')}{2} + 1$ such that

$$S^\alpha(\mathbf{X}, n, k) = S^\alpha(\mathbf{X}^*, n', 1) \quad \text{and} \quad S^\alpha(\mathbf{Y}, m, k') = S^\alpha(\mathbf{Y}^*, m', 1).$$

Thus, by the previous result, we obtain:

$$\mathbf{X} \sim \mathbf{X}^* \quad \text{and} \quad \mathbf{Y} \sim \mathbf{Y}^*. \tag{8}$$

Using Lemma 1, we have:

$$\begin{aligned} S^\alpha(\mathbf{X}^*, n', 1) &= 2\,(n'-1) \\ S^\alpha(\mathbf{Y}^*, m', 1) &= 2\,(m'-1) \end{aligned} \tag{9}$$

then, by Monotonicity,

$$S^\alpha(\mathbf{X}^*, n', 1) \geq S^\alpha(\mathbf{Y}^*, m', 1) \Longleftrightarrow n' \geq m'$$

and

$$\mathbf{X}^* \succeq \mathbf{Y}^* \quad \Rightarrow \quad S^\alpha(\mathbf{X}^*, n', 1) \geq S^\alpha(\mathbf{Y}^*, m', 1).$$

Hence, equation (8) and Lemma 2 entail:

$$\mathbf{X} \succeq \mathbf{Y} \quad \Rightarrow \quad S^\alpha(\mathbf{X}, n, k) \geq S^\alpha(\mathbf{Y}, m, k')$$

that completes the proof. □

References

[1] Alesina A., Devleeschauwer A. , William E., Kurlat S., Wacziarg R., Fractionalization, *Journal of Economic Growth*, 8, (2003), 155-194.

[2] Alesina A., La Ferrara E., Ethnic Diversity and Economic Performance, *Journal of Economic Literature*, 43,3, (2005), 762-800.

[3] Alkire S., Foster, J. Counting and Multidimensional Poverty Measurement, *Journal of Public Economics*, 95,7-8, (2011), 476-487.

[4] Atkinson A.B., Multidimensional deprivation: Contrasting social welfare and counting approach, *Journal of Economic Inequality* 1, (2003), 51-65.

[5] Bossert W., Pattanaik P.K., Xu Y., Similarity of options and the measurement of diversity, *Journal of Theoretical Politics*,(2003), 405-421.

[6] Bossert, W., D'Ambrosio C., La Ferrara E., A generalized index of fractionalization, *Economica*, 78(312) (2011), 723-750.

[7] Chakravarty S.R., D'Ambrosio C, The measurement of social exclusion, *Review of Income and Wealth*, 52,3, (2006), 377-398.

[8] Efron B, Tibshirani R. (1993), *An Introduction to the Bootstrap*, Chapman and Hall, New York.

[9] EUROSTAT (2012), *Active ageing and solidarity between generations. A statistical portrait of the European Union* 2012EU.

[10] Fearon, J. D., Ethnic and Cultural Diversity by Country, *Journal of Economic Growth*, 8, (2003), 195-222.

[11] Gravel N., *What is diversity?*, Working Paper, May 15th 2007.

[12] Lasso de la Vega M.C., Counting poverty orderings and deprivation curves, *Studies in Applied Welfare Analysis, Research on Economic Inequality*, 18, (2010), 153-172.

[13] Nehring K., Puppe C., A Theory of Diversity, *Econometrica*, 70,3, (2002), 1155-1190.

[14] Nehring, K., Puppe, C., *Diversity,* The Handbook of Rational and Social Choice, ed. by P. Anand, P. Pattanaik and C. Puppe: Oxford University Press, (2009), Chapter 12.

[15] Patsiurko N., Campbell J. L., Hall J. A., Measuring cultural diversity: ethnic, linguistic and religious fractionalization in the OECD, *Ethnic and Racial Studies*, 35, (2012), 195-217.

[16] Pattanaik P.K., Xu Y., On diversity as freedom of choice, *Mathematical Social Sciences*, 40,2 (2000), 123-130.

[17] Pattanaik P.K., Xu Y., Ordinal Distance, Dominance, and the Measurement of Diversity, Andrew Young School of Policy Studies Research Paper No. 06-57, May 2006.

[18] Simpson, E. H. (1949) Measurement of diversity. *Nature*, 163, 688.

[19] Weitzman M. L., On diversity, *The Quarterly Journal of Economics*, 107,2, (1992), 363-405.