Size invariant measures of association: characterization and difficulties[*]

Margherita Negri[†] and Yves Sprumont[‡]

August 7, 2014

Abstract

A measure of association is row-size invariant if it is unaffected by the mutliplication of all entries in a row of a cross-classification table by a same positive number. It is class-size invariant if it is unaffected by the mutliplication of all entries in a class (i.e., a row or a column). We prove that every class-size invariant measure of association assigns to each $m \times n$ cross-classification table a number which depends only on the cross-product ratios of its 2×2 subtables. We propose a monotonicity axiom requiring that the degree of association should increase after shifting mass from cells of a table where this mass is below its expected value to cells where it is above –provided that total mass in each class remains constant. We prove that no continuous row-size invariant measure of association is monotonic if $m \geq 4$. *Keywords:* association, contingency tables, margin-free measures, size invariance, monotonicity, transfer principle.

[†]School of Economics and Finance, University of St Andrews, St Andrews, Fife, KY169AR, Scotland, UK, email: margherita.negri@gmail.com

[‡]Département de Sciences Économiques and CIREQ, Université de Montréal, CP 6128 succursale centre-ville, Montréal QC H3C3J7, Canada, email: yves.sprumont@umontreal.ca

1

1 Introduction

This paper revisits the old problem of measuring the degree of association between two characteristics whose values cannot be quantified or even ordered in a meaningful way. Gender, professional occupation, ethnic origin, political affiliation are all common examples. Each characteristic may fall in a number of predetermined classes and the distribution of the two characteristics in the population under consideration is summarized by a cross-classification table: this is an $m \times n$ matrix whose (i,j)th entry records the mass of individuals whose first characteristic belongs to class i and whose second characteristic belongs to class j. A measure of association is a function which assigns a number to each such matrix. A survey of the literature on the problem of constructing such a measure is offered in Goodman and Kruskal (1954). The measurement of racial segregation is an example of application that has received distinguished attention: see the survey by James and Taeuber (1985)[1].

   We are interested here in association measures that are invariant under multiplication of all entries in any given class –row or column– by a same positive number. Such measures are sometimes called "margin-free". The motivation for this class-size invariance axiom is that the total mass in a class is often determined by factors that have nothing to do with the inherent association between the characteristics. Yule (1912), who originally suggested the axiom, gives the following example. Suppose we are concerned with the effect of a medical treatment on persons suffering from a potentially fatal disease: the first characteristic is whether or not the person received the treatment, the second is whether or not the person died from the disease. Data from two hospitals are recorded in the following tables.

<div align="center">

Hospital 1

	Lived	*Died*	*Total*
Treated	84	4	88
Not treated	3	9	12
Total	87	13	100

Hospital 2

</div>

---

[1]The problem of measuring correlation between *ordinal* characteristics (whose values cannot be measured but can be ordered) is different and has generated a separate literature.

	Lived	Died	Total
Treated	42	2	44
Not treated	14	42	56
Total	56	44	100

These two tables differ only in the proportion of persons who received treatment and the proportion who did not: the second table may be obtained from the first by multiplying the first row by $\frac{1}{2}$ and multiplying the second by $\frac{14}{3}$. It follows that the conditional probabilities of life given treatment (non-treatment) are the same in both hospitals, namely .955 (.250). The degree of association between treatment and life should therefore be considered equal in both hospitals; the fact that the proportion of persons receiving treatment is higher in hospital 1 is irrelevant.

Class-size invariance is well understood in the case of dichotomous characteristics, i.e., when $m = n = 2$. In that case, Edwards (1963) and Goodman (1965) showed that every class-size invariant measure must be a function of the *cross-product ratio* of the cross-classification table, that is, the ratio between the product of its diagonal entries and the product of its off-diagonal entries.

For tables of arbitrary dimensions, Yule and Kendall (1950) observed that any function of the cross-product ratios of the 2×2 subtables defines a class-size invariant measure. Goodman (1969) showed how to use the collection of 2×2 cross-product ratios to study patterns of association. The class-size invariant approach was extended to continuous densities by Plackett (1965), Mosteller (1968), Holland and Wang (1987), and others.

Our first contribution in this paper is to prove the converse of Yule and Kendall's observation: every class-size invariant measure of association assigns to each $m \times n$ cross-classification table a number which depends only on the cross-product ratios of its 2×2 subtables.

Class-size invariance is a demanding property. In many contexts, a one-sided version of the axiom is more compelling. In fact, in Yule's example, the argument that the hospital's choice of the proportion of treated patients has nothing to do with the association between treatment and life justifies *row-size invariance*: multiplying all entries in a row by a same number should leave the degree of association unchanged. The dual axiom of column-size invariance, on the other hand, makes little sense because "multiplying all entries in a column by a same number" is not an operation within the hospital's control. More generally, row-size invariance is appealing when a one-

directional causal relationship is suspected between the row characteristic and the column characteristic. The axiom has been discussed in a number of applications, notably in the literature on segregation measurement: see for instance James and Taeuber (1985) and Reardon and Firebaugh (2002). Class-size invariance may be a better axiom when there is no reason to suspect a one-directional causality between the characteristics.

Our second contribution consists in proving a disturbing incompatibility between row-size invariance (hence, a fortiori, class-size invariance) and a monotonicity condition which, we believe, captures the essence of what an association measure should do. It is generally accepted that (the absolute value of) a meaningful measure should reach its minimum when the characteristics are independent, that is to say, when the proportion of observations in any given cell (i, j) of the cross-classification table is equal to the expected proportion, namely, the product of the proportion of observations in row i by the proportion of observations in column j. We submit that an increase in the mass of observations in cells where this mass already exceeds the expected mass, coupled with a decrease in the mass of observations in cells where it falls short of the expected mass should, if total mass in each class is kept unchanged, increase the degree of association. We prove that no continuous row-size invariant measure satisfies this monotonicity axiom if $m \geq 4$.

2   Framework

Let $M = \{1, ..., m\}$, $m \geq 2$, be the set of relevant classes for the first characteristic and let $N = \{1, ..., n\}$, $n \geq 2$, be the set of relevant classes for the second characteristic. A cross-classification table is an $m \times n$ positive real matrix A. The number a_i^j in the intersection of row i and column j records the mass of agents whose first characteristic belongs to class i and whose second characteristic belongs to class j. We denote the ith row of A by A_i and its jth column by A^j. We write $a_i = \sum_{j \in N} a_i^j$, $a^j = \sum_{i \in M} a_i^j$, and $a = \sum_{i \in M} \sum_{j \in N} a_i^j$. We let $\mathcal{A}(m, n)$ denote the set of positive real $m \times n$ matrices and define an *(association) measure* to be a function $F : \mathcal{A}(m, n) \rightarrow \mathbb{R}_+$.

A number of remarks are in order.

1) The labels $1, ..., m$ attached to rows and the labels $1, ..., n$ attached to columns are used for convenience only. They have no intrinsic meaning: a lower label should not be interpreted as reflecting a "lower value" of the

4

characteristic. No relevant order structure is assumed on the sets M, N : characteristics are only "categorical", not "ordinal".

2) A fundamental role of an association measure is to allow ordinal comparisons across cross-classification tables. Any function $F : \mathcal{A}(m,n) \to \mathbb{R}_+$ generates an *association ordering* $\succsim$ on $\mathcal{A}(m,n)$ via the relation $A \succsim B \Leftrightarrow F(A) \geq F(B)$: the two characteristics are "more closely associated" in table A than in table B if and only if the association measure reaches a higher value at A than at B. We are primarily interested here in this ordering, which is invariant under monotonic transformations of F. Cardinal measures of association are beyond the scope of this paper.

3) We assume that the number of relevant classes for each characteristic is fixed. We do so because the size invariance and monotonicity axioms do not require a variable-dimensions framework. It is straightforward to reformulate our results in such an extended context (taking $\cup_{(m,n) \geq (2,2)} \mathcal{A}(m,n)$ as the domain of an association measure) but this generalization brings no new insight into the problem we are interested in.

4) We assume that the entries in all cross-classification tables under consideration are real numbers rather than integers. This is important for the incompatibility result of Section 4, which involves a continuity argument. On the other hand, we could (at the cost of a somewhat cumbersome reformulation of the size invariance axioms) express our results in a framework where the total mass of observations, a, is fixed and equal to one.

5) We focus on positive matrices. We briefly discuss the construction of class-size invariant measures when zero entries are allowed at the end of Section 3.

3   The class-size invariant measures of association

If $A \in \mathcal{A}(m,n)$, $i \in M$, $j \in N$ and $\lambda \in \mathbb{R}_{++}$, we denote by $(\lambda A_i, A_{-i})$ the matrix obtained by multiplying each entry of the ith row of A by λ and leaving all other entries unchanged. Likewise, $(\lambda A^j, A^{-j})$ is the matrix obtained by multiplying each entry of the jth column of A by λ and leaving all other entries unchanged.

Class-Size Invariance. For all $A \in \mathcal{A}(m, n)$, $i \in M$, $j \in N$ and $\lambda \in \mathbb{R}_{++}$,
$F(\lambda A_i, A_{-i}) = F(A) = F(\lambda A^j, A^{-j})$.

An *elementary submatrix* of $A \in \mathcal{A}(m, n)$ (also called a "tetrad" by Yule and Kendall (1950)) is any 2×2 submatrix whose entries belong to adjacent rows and columns of A. The *cross-product ratio* of an elementary submatrix (or an *elementary cross-product ratio* for short) is the ratio between the product of the diagonal entries of this submatrix and the product of its off-diagonal entries. Theorem 1 below asserts that a measure is class-size invariant if and only if it assigns to each matrix a number that can be written as a function of its elementary cross-product ratios.

More formally: for all $A \in \mathcal{A}(m, n)$, all $i \in M \setminus \{m\}$ and all $j \in N \setminus \{n\}$, define

$$r_i^j(A) = \frac{a_i^j a_{i+1}^{j+1}}{a_i^{j+1} a_{i+1}^j}$$

and let $r(A) \in \mathcal{A}(m-1, n-1)$ denote the matrix $(r_i^j(A))_{i \in M \setminus \{m\}}^{j \in N \setminus \{n\}}$.

Theorem 1. *A measure $F : \mathcal{A}(m, n) \to \mathbb{R}_+$ is class-size invariant if and only if there exists a function $f : \mathcal{A}(m-1, n-1) \to \mathbb{R}_+$ such that $F(A) = f(r(A))$ for all $A \in \mathcal{A}(m, n)$.*

Proof. "If". This follows from the fact that for any $\lambda > 0$, $i \in M$ and $j \in N$, we have $r(\lambda A_i, A_{-i}) = r(A) = r(\lambda A^j, A^{-j})$.

"Only if". Let F be a class-size invariant measure and let $A, B \in \mathcal{A}(m, n)$ be such that $r(A) = r(B)$. We show that $F(A) = F(B)$. We use the following notation: if $C, D \in \mathcal{A}(p, q)$, then $C * D$ is the matrix in $\mathcal{A}(p, q)$ whose (i, j)th entry is the product $c_i^j d_i^j$ of the (i, j)th entries of C and D.

Step 1. There exist $\lambda, \lambda_2, \lambda^2 > 0$ such that

$$\begin{pmatrix} b_1^1 & b_1^2 \\ b_2^1 & b_2^2 \end{pmatrix} = \begin{pmatrix} \lambda & \lambda^2 \\ \lambda_2 & \frac{\lambda_2 \lambda^2}{\lambda} \end{pmatrix} * \begin{pmatrix} a_1^1 & a_1^2 \\ a_2^1 & a_2^2 \end{pmatrix}. \tag{1}$$

To see this, simply define $\lambda = \frac{b_1^1}{a_1^1}$, $\lambda_2 = \frac{b_2^1}{a_2^1}$, and $\lambda^2 = \frac{b_1^2}{a_1^2}$. Trivially, $b_1^1 = \lambda a_1^1$, $b_2^1 = \lambda_2 a_2^1$, and $b_1^2 = \lambda^2 a_1^2$. And since $r(A) = r(B)$ implies

$$\frac{a_1^1 a_2^2}{a_1^2 a_2^1} = \frac{b_1^1 b_2^2}{b_1^2 b_2^1},$$

6

we have
$$b_2^2 = \left(\frac{b_1^2 b_2^1 a_1^1}{a_1^2 a_2^1 b_1^1}\right) a_2^2 = \frac{\lambda_2 \lambda^2}{\lambda} a_2^2,$$
proving (1).

Step 2. There exist $\lambda;\ \lambda_2, ..., \lambda_m;\ \lambda^2, ..., \lambda^n > 0$ such that

$$B = \begin{pmatrix} \lambda & \lambda^2 & \lambda^3 & \cdots & \lambda^n \\ \lambda_2 & \frac{\lambda_2 \lambda^2}{\lambda} & \frac{\lambda_2 \lambda^3}{\lambda} & \cdots & \frac{\lambda_2 \lambda^n}{\lambda} \\ \vdots & & & & \vdots \\ \lambda_m & \frac{\lambda_m \lambda^2}{\lambda} & \frac{\lambda_m \lambda^3}{\lambda} & \cdots & \frac{\lambda_m \lambda^n}{\lambda} \end{pmatrix} * A. \tag{2}$$

For ease of exposition we assume $m \le n$; the case $n \le m$ is treated similarly. From Step 1 we know that there exist $\lambda, \lambda_2, \lambda^2 > 0$ such that $b_i^j = \left(\frac{\lambda_i \lambda^j}{\lambda}\right) a_i^j$ for all $(i, j) \in M \times N$ such that $\max(i, j) \le 2$. Here and below, $\lambda_1 = \lambda^1 = \lambda$. The rest of the proof proceeds by induction on $\max(i, j)$.

Fix an integer k such that $2 < k < n$ and make the following *induction hypothesis*: there exist $\lambda;\ \lambda_2, ..., \lambda_{\min(k-1,m)};\ \lambda^2, ..., \lambda^{k-1} > 0$ such that $b_i^j = \left(\frac{\lambda_i \lambda^j}{\lambda}\right) a_i^j$ for all $(i, j) \in M \times N$ such that $\max(i, j) \le k - 1$. We make two claims. The first is that there exists a number $\lambda^k > 0$ such that

$$b_i^k = \left(\frac{\lambda_i \lambda^k}{\lambda}\right) a_i^k \text{ for } i = 1, ..., \min(k - 1, m). \tag{3}$$

The second claim is that, if $k - 1 < m$, there exists a number $\lambda_k > 0$ such that

$$b_k^j = \left(\frac{\lambda_k \lambda^j}{\lambda}\right) a_k^j \text{ for } j = 1, ..., k. \tag{4}$$

In order to prove the first claim, define $\lambda^k = \frac{b_1^k}{a_1^k}$. By construction, $b_1^k = \lambda^k a_1^k = \frac{\lambda_1 \lambda^k}{\lambda} a_1^k$. Next, fix an integer q such that $1 < q < k$ and make the *auxiliary induction hypothesis* that $b_i^k = \left(\frac{\lambda_i \lambda^k}{\lambda}\right) a_i^k$ for $i = 1, ..., q - 1$. By the assumption $r(A) = r(B)$ we have

$$\frac{a_{q-1}^{k-1} a_q^k}{a_{q-1}^k a_q^{k-1}} = \frac{b_{q-1}^{k-1} b_q^k}{b_{q-1}^k b_q^{k-1}},$$

7

hence,

$$b_q^k = \left(\frac{a_{q-1}^{k-1} b_{q-1}^k b_q^{k-1}}{a_{q-1}^k a_q^{k-1} b_{q-1}^{k-1}} \right) a_q^k. \tag{5}$$

By the (main) induction hypothesis, $b_q^{k-1} = \left(\frac{\lambda_q \lambda^{k-1}}{\lambda} \right) a_q^{k-1}$ and $b_{q-1}^{k-1} = \left(\frac{\lambda_{q-1} \lambda^{k-1}}{\lambda} \right)$ a_{q-1}^{k-1} whereas by the auxiliary induction hypothesis $b_{q-1}^{k-1} = \left(\frac{\lambda_{q-1} \lambda^k}{\lambda} \right) a_{q-1}^k.$ Therefore (5) implies

$$b_q^k = \left(\frac{\lambda_q \lambda^k}{\lambda} \right) a_q^k,$$

proving our first claim.

The proof of the second claim is similar. Suppose $k - 1 < m$. Define $\lambda_k = \frac{b_k^1}{a_k^1}$. Then $b_k^1 = \lambda_k a_k^1 = \frac{\lambda_k \lambda^1}{\lambda} a_k^1$. Fix an integer q such that $1 < q \leq k$ (notice the weak inequality this time) and make the *second auxiliary hypothesis* that $b_k^j = \left(\frac{\lambda_k \lambda^j}{\lambda} \right) a_k^j$ for $j = 1, ..., q - 1$. By assumption,

$$\frac{a_{k-1}^{q-1} a_k^q}{a_k^{q-1} a_{k-1}^q} = \frac{b_{k-1}^{q-1} b_k^q}{b_k^{q-1} b_{k-1}^q},$$

hence

$$b_k^q = \left(\frac{a_{k-1}^{q-1} b_k^{q-1} b_{k-1}^q}{a_k^{q-1} a_{k-1}^q b_{k-1}^{q-1}} \right) a_k^q. \tag{6}$$

By the main induction hypothesis, $b_{k-1}^{q-1} = \left(\frac{\lambda_{k-1} \lambda^{q-1}}{\lambda} \right) a_{k-1}^{q-1}.$ By the second auxiliary induction hypothesis, $b_k^{q-1} = \left(\frac{\lambda_k \lambda^{q-1}}{\lambda} \right) a_k^{q-1}.$ Finally, $b_{k-1}^q = \left(\frac{\lambda_{k-1} \lambda^q}{\lambda} \right) a_k^{q-1}$: this holds by virtue of the main induction hypothesis if $q < k$ and by virtue of the first claim we proved if $q = k$. Therefore (6) implies

$$b_k^q = \left(\frac{\lambda_k \lambda^q}{\lambda} \right) a_k^q,$$

proving our second claim.

Step 3. Because of (2), the matrix B can be obtained by (i) multiplying A_1 and A^1 by $\sqrt{\lambda}$, (ii) multiplying A_i by $\frac{\lambda_i}{\sqrt{\lambda}}$ for each $i = 2, ..., m$, and (iii) multiplying A^j by $\frac{\lambda^j}{\sqrt{\lambda}}$ for each $j = 2, ..., n$. Class-Size Invariance now implies that $F(A) = F(B)$. ∎

We make two brief remarks on Theorem 1.

1) The function f whose existence is asserted in Theorem 1 is unique. The reason is that for each $z \in \mathcal{A}(m-1, n-1)$ there exists a matrix $A \in \mathcal{A}(m, n)$ such that $r(A) = z$.

Of course, the cross-product ratio can be defined for every 2×2 (not necessarily elementary) submatrix of A: for all $i, k \in M$ such that $i < k$ and all $j, l \in N$ such that $j < l$, let

$$r_{i,k}^{j,l}(A) = \frac{a_i^j a_k^l}{a_i^l a_k^j}.$$

One checks that

$$r_{i,k}^{j,l}(A) = \prod_{\substack{i'=i,\ldots,k-1 \\ j'=j,\ldots,l-1}} r_{i'}^{j'}(A). \tag{7}$$

It follows that any function of the cross-product ratios of all 2×2 submatrices of a matrix can be rewritten as a function of its elementary cross-product ratios only.

We may therefore reformulate Theorem 1 as follows. Denote by μ the number of 2×2 submatrices of any matrix $A \in \mathcal{A}(m, n)$. Let $\overline{r}(A) = (r_{i,k}^{j,l}(A))_{1 \le i < k \le m}^{1 \le j < l \le n} \in \mathbb{R}_{++}^{\mu}$ be the vector of cross-product ratios of these submatrices. A measure $F : \mathcal{A}(m, n) \to \mathbb{R}_+$ is class-size invariant if and only if there exists a function $\overline{f} : \mathbb{R}_{++}^{\mu} \to \mathbb{R}_+$ such that $F(A) = \overline{f}(\overline{r}(A))$ for all $A \in \mathcal{A}(m, n)$. Such a function $\overline{f}$, however, is not unique because $\overline{r} : \mathcal{A}(m, n) \to \mathbb{R}_{++}^{\mu}$ is not onto.

2) In many applications cross-classification tables may have some zero entries. Class-size invariant measures may be defined on the domain of non-negative real $m \times n$ matrices having at least one positive 2×2 submatrix. Let $\widetilde{\mathcal{A}}(m, n)$ denote that domain. For each $A \in \widetilde{\mathcal{A}}(m, n)$, let $\widetilde{r}(A)$ denote the vector of cross-product ratios of all the positive 2×2 submatrices of A. For any function $\widetilde{f}$, the measure F defined on $\widetilde{\mathcal{A}}(m, n)$ by $F(A) = \widetilde{f}(\widetilde{r}(A))$ is class-size invariant.

4 The incompatibility between Row-Size Invariance and Monotonicity

In order to construct a good measure of association, it is useful to identify circumstances under which the degree of association between two character-

istics should undoubtedly be deemed to increase. The fundamental postulate of the literature is that association is nil in the case of independence, i.e., in a matrix A such that $a_i^j a = a_i a^j$ for all $i \in M$ and $j \in N$. Taking independence as a benchmark, we say that i, j are *positively associated (in A)* if

$$a_i^j a > a_i a^j, \tag{8}$$

or equivalently $a_i^j > \frac{a_i a^j}{a}$, i.e., if the mass of observations in cell (i, j) exceeds the "expected" mass. We say that i, j are *negatively associated* if the opposite strict inequality holds.

We submit that the degree of association between two characteristics should increase when we shift mass from cells where the mass is already below expectation to cells where it is already above –provided that the total mass in every class (row or column) remains unchanged. Our monotonicity condition focuses on the simplest such shifts.

Monotonicity. Let $i, i' \in M$ and $j, j' \in N$ be such that $i \neq i'$ and $j \neq j'$. Let $A, B \in \mathcal{A}(m, n)$. Suppose that both i, j and i', j' are positively associated in A while both i, j' and i', j are negatively associated. If there exists $\delta > 0$ such that $b_i^j = a_i^j + \delta$, $b_{i'}^{j'} = a_{i'}^{j'} + \delta$, $b_i^{j'} = a_i^{j'} - \delta$, $b_{i'}^j = a_{i'}^j - \delta$, and $b_k^l = a_k^l$ whenever $k \notin \{i, i'\}$ or $l \notin \{j, j'\}$, then $F(B) > F(A)$.

As an illustration, consider the matrix

$$A = \begin{pmatrix} 7 & 27 & 16 \\ 13 & 13 & 24 \\ 30 & 10 & 60 \end{pmatrix}.$$

Note that $i = 2$, $j = 2$ are positively associated in A since $a_2^2 a = 2600 > a_2 a^2 = 2500$: the mass in cell $(2, 2)$ is above its expected value. Similarly, $i = 3$, $j = 3$ are positively associated. On the contrary, $i = 2$, $j = 3$, as well as $i = 3$, $j = 2$, are negatively associated. Shifting one unit of mass from cell $(2, 3)$ to cell $(2, 2)$ and one unit of mass from cell $(3, 2)$ to cell $(3, 3)$ yields

$$B = \begin{pmatrix} 7 & 27 & 16 \\ 13 & \mathbf{14} & \mathbf{23} \\ 30 & \mathbf{9} & \mathbf{61} \end{pmatrix}.$$

Our axiom therefore requires that $F(B) > F(A)$.

Monotonicity involves changes in no less than four distinct cells. This may seem unduly complicated. One may think that more elementary mass

shifts should already be deemed to unambiguously increase the degree of association. Consider for instance the matrix

$$C = \begin{pmatrix} 7 & 27 & 16 \\ 13 & \mathbf{14} & \mathbf{23} \\ 30 & 10 & 60 \end{pmatrix},$$

which is obtained from A by merely shifting one unit of mass from cell $(2, 3)$ to cell $(2, 2)$. Since the mass in cell $(2, 3)$ is below expectation in A and since the mass in cell $(2, 2)$ is above expectation, it is tempting to ask that $F(C) > F(A)$. We think that this conclusion is unwarranted. The reason is that the change from A to C increases the total mass in column 2 (from $a^2 = 50$ to $c^2 = 51$) and decreases the total mass in column 3 (from $a^3 = 100$ to $c^3 = 99$). This in turn changes the expected masses in all cells belonging to columns 2 and 3. In particular, the expected mass in cell $(1, 2)$ increases from $\frac{a_1 a^2}{a} = 12.5$ in A to $\frac{c_1 c^2}{c} = 12.75$ in C. As a consequence, the actual mass in cell $(1, 2)$, which remains unchanged from A to C (at $a_1^2 = c_1^2 = 27$) becomes *closer* to its expected value (of 12.75 in C vs 12.5 in A). Similarly, the expected mass in cell $(1, 3)$ decreases from $\frac{a_1 a^3}{a} = 25$ in A to $\frac{c_1 c^3}{c} = 24.75$ in C, making the actual mass in that cell ($a_1^3 = c_1^3 = 16$) closer to expectation (of 24.75 in C vs 25 in A). These induced changes should tend to *decrease* the degree of association between the two characteristics, counteracting the direct effect of the original changes and leaving the net effect unclear.

In fact, one checks that the most popular measure of association for cross-classification tables of constant total mass (here $a = c$), the chi-square measure, does decrease from A to C :

$$\chi^2(A) = \sum_{(i,j) \in M \times N} \frac{\left(a_i^j - \frac{a_i a^j}{a} \right)^2}{\frac{a_i a^j}{a}} = 34.56$$

$$> \chi^2(C) = \sum_{(i,j) \in M \times N} \frac{\left(c_i^j - \frac{c_i c^j}{c} \right)^2}{\frac{c_i c^j}{c}} \approx 34.36.$$

In contrast, because the changes considered in the monotonicity axiom leave the total mass in every class (row or column) unchanged, the expected mass in every cell also remains constant and, as a consequence, the gap between actual and expected mass is modified *only* in the four cells under consideration. Since *all* these gaps become wider, it is compelling to conclude that

11

the degree of association increases. It is a simple exercise to check that the chi-square measure of association does indeed satisfy Monotonicity. Another famous measure that passes the monotonicity test is the entropy-based index proposed by Theil and Finizza (1971) in the context of measuring segregation,

$$I(A) = \sum_{(i,j) \in M \times N} \frac{a_i^j}{a} \log \left(\frac{a_i^j a}{a_i a^j} \right).$$

To the best of our knowledge, Monotonicity is a new condition. It is a rather weak axiom. A similar but stronger axiom, inspired by the Pigou-Dalton principle, can be found in the segregation measurement literature (see for instance James and Taeuber (1985)):

Transfer Principle. Let $i, i' \in M$ and $j, j' \in N$ be such that $i \neq i'$ and $j \neq j'$. Let $A, B \in \mathcal{A}(m,n)$. Suppose that $\frac{a_i^j}{a^j} > \frac{a_i^{j'}}{a^{j'}}$ and $\frac{a_{i'}^j}{a^j} < \frac{a_{i'}^{j'}}{a^{j'}}$. If there exists $\delta > 0$ such that $b_i^j = a_i^j + \delta$, $b_{i'}^{j'} = a_{i'}^{j'} + \delta$, $b_i^{j'} = a_i^{j'} - \delta$, $b_{i'}^j = a_{i'}^j - \delta$, and $b_k^l = a_k^l$ whenever $k \notin \{i, i'\}$ or $l \notin \{j, j'\}$, then $F(B) > F(A)$.

This means, for instance, that if the proportion of black students is higher in school j than in school j' while the proportion of white students is higher in school j' than in school j, shifting a black student from j' to j and a white student from j to j' increases segregation. This conclusion does not follow from Monotonicity, unless the proportion of black students in j and the proportion of white students in j' exceed their expected value and the proportion of white students in j and the proportion of black students in j' are below their expected value.

Our second theorem is an incompatibility result.

Row-Size Invariance. For all $A \in \mathcal{A}(m,n)$, $i \in M$ and $\lambda \in \mathbb{R}_{++}$, $F(\lambda A_i, A_{-i}) = F(A)$.

Theorem 2. *If $m \geq 4$, no continuous measure $F : \mathcal{A}(m,n) \to \mathbb{R}_+$ satisfies Monotonicity and Row-Size Invariance.*

Proof. Step 1. If $F : \mathcal{A}(m,2) \to \mathbb{R}_+$ satisfies Monotonicity and Row-Size Invariance, then F satisfies the following property:

Strong Monotonicity. Let $i, i' \in M$ be such that $i \neq i'$ and let $A, B \in \mathcal{A}(m,2)$. Suppose that $a_i^1 a_{i'}^2 > a_i^2 a_{i'}^1$. If there exists $\delta > 0$ such that $b_i^1 = a_i^1 + \delta$,

12

$b_{i'}^2 = a_{i'}^2 + \delta$, $b_i^2 = a_i^2 - \delta$, $b_{i'}^1 = a_{i'}^1 - \delta$, and $b_k^j = a_k^j$ for all $k \notin \{i, i'\}$ and $j \in \{1, 2\}$, then $F(B) > F(A)$.

To prove this claim, let $i, i' \in M$ be such that $i \neq i'$. Without loss of generality, say $i = 1$ and $i' = 2$. Let $A, B \in \mathcal{A}(m, 2)$. Suppose $a_1^1 a_2^2 > a_1^2 a_2^1$ and suppose there exists $\delta > 0$ such that

$$
B = \begin{pmatrix} a_1^1 + \delta & a_1^2 - \delta \\ a_2^1 - \delta & a_2^2 + \delta \\ a_3^1 & a_3^2 \\ \vdots & \vdots \\ a_m^1 & a_m^2 \end{pmatrix}.
$$

We must show that $F(B) > F(A)$.

For every $\varepsilon > 0$, let

$$
A(\varepsilon) = \begin{pmatrix} a_1^1 & a_1^2 \\ a_2^1 & a_2^2 \\ \varepsilon a_3^1 & \varepsilon a_3^2 \\ \vdots & \vdots \\ \varepsilon a_m^1 & \varepsilon a_m^2 \end{pmatrix}.
$$

Let $a_i^j(\varepsilon)$ denote the (i, j)th entry of $A(\varepsilon)$ and write $a_i(\varepsilon) = \sum_{j \in N} a_i^j(\varepsilon)$, $a^j(\varepsilon) = \sum_{i \in M} a_i^j(\varepsilon)$, and $a(\varepsilon) = \sum_{i \in M} \sum_{j \in N} a_i^j(\varepsilon)$.

We claim that for ε small enough,

$$
a_1^1(\varepsilon) a(\varepsilon) > a_1(\varepsilon) a^1(\varepsilon). \tag{9}
$$

To see why, define

$$
\begin{aligned}
\alpha(\varepsilon) &= a(\varepsilon) - (a_1^1 + a_1^2 + a_2^1 + a_2^2), \\
\alpha^1(\varepsilon) &= a^1(\varepsilon) - (a_1^1 + a_2^1),
\end{aligned}
$$

and note that $\alpha(\varepsilon) \to 0$ and $\alpha^1(\varepsilon) \to 0$ when $\varepsilon \to 0$. Moreover, we find that

$$
\begin{aligned}
& a_1^1(\varepsilon) a(\varepsilon) - a_1(\varepsilon) a^1(\varepsilon) \\
= & a_1^1 \left[\alpha(\varepsilon) + (a_1^1 + a_1^2 + a_2^1 + a_2^2) \right] - a_1 \left[\alpha^1(\varepsilon) + (a_1^1 + a_2^1) \right] \\
= & (a_1^1 a_2^2 - a_1^2 a_2^1) + \left(a_1^1 \alpha(\varepsilon) - a_1 \alpha^1(\varepsilon) \right).
\end{aligned}
$$

13

Since $a_1^1 a_2^2 - a_1^2 a_2^1 > 0$ and the term in the last parenthesis vanishes as $\varepsilon \to 0$, there exists $\varepsilon > 0$ such that (9) holds.

By a similar argument, the inequality $a_1^1 a_2^2 > a_1^2 a_2^1$ also implies that

$$a_2^2(\varepsilon)a(\varepsilon) > a_2(\varepsilon)a^2(\varepsilon) \tag{10}$$

and

$$a_1^2(\varepsilon)a(\varepsilon) < a_1(\varepsilon)a^2(\varepsilon), \tag{11}$$

$$a_2^1(\varepsilon)a(\varepsilon) < a_2(\varepsilon)a^1(\varepsilon) \tag{12}$$

for $\varepsilon > 0$ small enough.

Now pick $\varepsilon > 0$ satisfying (9), (10), (11) and (12). This means that $i = 1$, $j = 1$, as well as $i = 2$, $j = 2$, are positively associated in $A(\varepsilon)$ whereas $i = 1$, $j = 2$, as well as $i = 2$, $j = 1$, are negatively associated. By Row-Size Invariance and Monotonicity,

$$F(A) = F(A(\varepsilon)) < F \begin{pmatrix} a_1^1 + \delta & a_1^2 - \delta \\ a_2^1 - \delta & a_2^2 + \delta \\ \varepsilon a_3^1 & \varepsilon a_3^2 \\ \vdots & \vdots \\ \varepsilon a_m^1 & \varepsilon a_m^2 \end{pmatrix} = F(B).$$

Step 2. If $F : \mathcal{A}(m, 2) \to \mathbb{R}_+$ satisfies Monotonicity and Row-Size Invariance, then F satisfies the following property:

Local Insensitivity. Let $i, i', i'' \in M$ be distinct. Let $A \in \mathcal{A}(m, 2)$ be such that $a_i^1 a_{i'}^2 > a_i^2 a_{i'}^1$ and $a_i^1 a_{i''}^2 < a_i^2 a_{i''}^1$. For every δ such that $0 < \delta < a_i^2$, denote by A^δ the matrix $B \in \mathcal{A}(m, 2)$ such that $b_i^1 = a_i^1 + \delta$, $b_i^2 = a_i^2 - \delta$, and $b_k^j = a_k^j$ for all $k \neq i$ and $j = 1, 2$. There exists δ^* such that $0 < \delta^* < a_i^{j'}$ and $F(A^\delta) = F(A)$ for all δ such that $0 < \delta < \delta^*$.

To prove this claim, fix distinct $i, i', i'' \in M$. Without loss, say $i = 1$, $i' = 2$, $i'' = 3$. Let $A \in \mathcal{A}(m, 2)$ be such that

$$a_1^1 a_2^2 > a_1^2 a_2^1 \tag{13}$$

and

$$a_1^1 a_3^2 < a_1^2 a_3^1. \tag{14}$$

14

Let $\delta^* = \min\left(\frac{a_1^2 a_3^1 - a_1^1 a_3^2}{a_1^1 + a_1^2 + a_3^1 + a_3^2}, a_1^2\right)$. We show that $F(A^\delta) = F(A)$ for all δ such that $0 < \delta < \delta^*$.

For every δ such that $0 < \delta < a_1^2$, consider the matrix

$$A^\delta = \begin{pmatrix} a_1^1 + \delta & a_1^2 - \delta \\ a_2^1 & a_2^2 \\ \vdots & \vdots \\ a_m^1 & a_m^2 \end{pmatrix}.$$

For all $\lambda > \frac{\delta}{a_2^1}$, Row-Size Invariance and Strong Monotonicity imply

$$
\begin{aligned}
F(A) \;=\; & F\begin{pmatrix} a_1^1 & a_1^2 \\ \lambda a_2^1 & \lambda a_2^2 \\ a_3^1 & a_3^2 \\ \vdots & \vdots \\ a_m^1 & a_m^2 \end{pmatrix} \\[1em]
<\; & F\begin{pmatrix} a_1^1 + \delta & a_1^2 - \delta \\ \lambda a_2^1 - \delta & \lambda a_2^2 + \delta \\ a_3^1 & a_3^2 \\ \vdots & \vdots \\ a_m^1 & a_m^2 \end{pmatrix} \\[1em]
=\; & F\begin{pmatrix} a_1^1 + \delta & a_1^2 - \delta \\ a_2^1 - \frac{\delta}{\lambda} & a_2^2 + \frac{\delta}{\lambda} \\ a_3^1 & a_3^2 \\ \vdots & \vdots \\ a_m^1 & a_m^2 \end{pmatrix}.
\end{aligned}
$$

Since F is continuous, letting $\lambda \to +\infty$ implies

$$F(A) \leq F(A^\delta). \tag{15}$$

Next, fix δ such that $0 < \delta < \delta^*$. Observe that for all $\lambda \geq 1$,

$$\delta < \delta^* \leq \frac{a_1^2 a_3^1 - a_1^1 a_3^2}{a_1^1 + a_1^2 + a_3^1 + a_3^2} \leq \frac{\lambda(a_1^2 a_3^1 - a_1^1 a_3^2)}{(a_1^1 + a_1^2) + \lambda(a_3^1 + a_3^2)},$$

hence, after rearranging,

$$(a_1^1 + \delta)(\lambda a_3^2 + \delta) < (a_1^2 - \delta)(\lambda a_3^1 - \delta).$$

15

For all $\lambda > \max(1, \frac{\delta}{a_3^1})$, Row-Size Invariance and Strong Monotonicity therefore imply

$$
\begin{aligned}
F(A) \;=\; & F\begin{pmatrix} a_1^1 & a_1^2 \\ a_2^1 & a_2^2 \\ \lambda a_3^1 & \lambda a_3^2 \\ a_4^1 & a_4^2 \\ \vdots & \vdots \\ a_m^1 & a_m^2 \end{pmatrix} \\[2mm]
>\; & F\begin{pmatrix} a_1^1 + \delta & a_1^2 - \delta \\ a_2^1 & a_2^2 \\ \lambda a_3^1 - \delta & \lambda a_3^2 + \delta \\ a_4^1 & a_4^2 \\ \vdots & \vdots \\ a_m^1 & a_m^2 \end{pmatrix} \\[2mm]
=\; & F\begin{pmatrix} a_1^1 + \delta & a_1^2 - \delta \\ a_2^1 & a_2^2 \\ a_3^1 - \frac{\delta}{\lambda} & a_3^2 + \frac{\delta}{\lambda} \\ a_4^1 & a_4^2 \\ \vdots & \vdots \\ a_m^1 & a_m^2 \end{pmatrix}.
\end{aligned}
$$

Since F is continuous, letting $\lambda \to +\infty$ yields $F(A) \geq F(A^\delta)$. Combining this inequality with (15) gives $F(A) = F(A^\delta)$, as claimed.

Step 3. No continuous measure $F : \mathcal{A}(m,2) \to \mathbb{R}_+$ satisfies Monotonicity and Row-Size Invariance.

Let $m \geq 4$ and suppose that $F : \mathcal{A}(m,2) \to \mathbb{R}_+$ is a continuous measure satisfying Monotonicity and Row-Size Invariance. We derive a contradiction.
Let $A \in \mathcal{A}(m,2)$ be a matrix such that

$$a_1^1 a_2^2 > a_1^2 a_2^1, \tag{16}$$

$$a_1^1 a_3^2 < a_1^2 a_3^1, \tag{17}$$

$$a_4^2 a_3^1 > a_4^1 a_3^2, \tag{18}$$

$$a_4^2 a_2^1 < a_4^1 a_2^2, \tag{19}$$

16

$$a_1^1 a_4^2 > a_4^1 a_2^1. \tag{20}$$

The above inequalities are compatible: take for instance

$$\begin{pmatrix} a_1^1 & a_1^2 \\ a_2^1 & a_2^2 \\ a_3^1 & a_3^2 \\ a_4^1 & a_4^2 \end{pmatrix} = \begin{pmatrix} 3 & 2 \\ 1 & 2 \\ 2 & 1 \\ 2 & 3 \end{pmatrix}.$$

For each $\delta > 0$ sufficiently small, define

$$A^\delta = \begin{pmatrix} a_1^1 + \delta & a_1^2 - \delta \\ a_2^1 & a_2^2 \\ \vdots & \vdots \\ a_m^1 & a_m^2 \end{pmatrix}.$$

By Step 2, F satisfies Local Insensitivity. Because of inequalities (16) and (17) we may use that property with $i = 1$, $i' = 2$, $i'' = 3$ to conclude

$$F(A) = F(A^\delta) \tag{21}$$

for all $\delta > 0$ sufficiently small.

For each $\delta > 0$ sufficiently small, define

$$A_\delta^\delta = \begin{pmatrix} a_1^1 + \delta & a_1^2 - \delta \\ a_2^1 & a_2^2 \\ a_3^1 & a_3^2 \\ a_4^1 - \delta & a_4^2 + \delta \\ a_5^1 & a_5^2 \\ \vdots & \vdots \\ a_m^1 & a_m^2 \end{pmatrix}$$

and note that (18) and (19) imply

$$\begin{aligned} (a_4^1 - \delta)a_2^2 &> (a_4^2 + \delta)a_2^1, \\ (a_4^1 - \delta)a_3^2 &< (a_4^2 + \delta)a_3^1. \end{aligned}$$

Because of these inequalities, we may use Local Insensitivity with $i = 4$, $i' = 2$, $i'' = 3$ to conclude that

$$F(A_\delta^\delta) = F(A^\delta)$$

17

for each $\delta > 0$ sufficiently small, which, combined with (21), implies that

$$F(A) = F(A_\delta^\delta)$$

for each $\delta > 0$ sufficiently small. Because of inequality (20), this contradicts Strong Monotonicity

Step 4. Let $m \geq 4$ and let $n \geq 3$. Suppose, by way of contradiction, that there exists a continuous measure $F : \mathcal{A}(m,n) \to \mathbb{R}_+$ that satisfies Monotonicity and Row-Size Invariance. We show that there must then exist a continuous measure $\widetilde{F} : \mathcal{A}(m,2) \to \mathbb{R}_+$ that satisfies Monotonicity and Row-Size Invariance, contradicting Step 3.

For each matrix

$$A = \begin{pmatrix} a_1^1 & a_1^2 \\ a_2^1 & a_2^2 \\ \vdots & \vdots \\ a_m^1 & a_m^2 \end{pmatrix}$$

in $\mathcal{A}(m,2)$, define $\widetilde{A} \in \mathcal{A}(m,n)$ by

$$\widetilde{A} = \begin{pmatrix} a_1^1 & a_1^2 & a_1 & \cdots & a_1 \\ a_2^1 & a_2^2 & a_2 & \cdots & a_2 \\ \vdots & \vdots & \vdots & & \vdots \\ a_m^1 & a_m^2 & a_m & \cdots & a_m \end{pmatrix}.$$

Define $\widetilde{F} : \mathcal{A}(m,2) \to \mathbb{R}_+$ by

$$\widetilde{F}(A) = F\left(\widetilde{A}\right).$$

Because F is continuous, so is $\widetilde{F}$. Because F satisfies Row-Size Invariance, so does $\widetilde{F}$: considering for instance a rescaling of the first row, we have, for any $\lambda > 0$,

$$\widetilde{F} \begin{pmatrix} \lambda a_1^1 & \lambda a_1^2 \\ a_2^1 & a_2^2 \\ \vdots & \vdots \\ a_m^1 & a_m^2 \end{pmatrix}.$$

18

$$= F \begin{pmatrix} \lambda a_1^1 & \lambda a_1^2 & \lambda a_1 & \cdots & \lambda a_1 \\ a_2^1 & a_2^2 & a_2 & \cdots & a_2 \\ \vdots & \vdots & \vdots & & \vdots \\ a_m^1 & a_m^2 & a_m & \cdots & a_m \end{pmatrix}$$

$$= F \begin{pmatrix} a_1^1 & a_1^2 & a_1 & \cdots & a_1 \\ a_2^1 & a_2^2 & a_2 & \cdots & a_2 \\ \vdots & \vdots & \vdots & & \vdots \\ a_m^1 & a_m^2 & a_m & \cdots & a_m \end{pmatrix}$$

$$= \widetilde{F} \begin{pmatrix} a_1^1 & a_1^2 \\ a_2^1 & a_2^2 \\ \vdots & \vdots \\ a_m^1 & a_m^2 \end{pmatrix}.$$

It remains to be checked that $\widetilde{F}$ is monotonic. Let $A \in \mathcal{A}(m, 2)$ and assume that, say,

$$a_1^1 a > a_1 a^1, \tag{22}$$
$$a_2^2 a > a_2 a^2, \tag{23}$$
$$a_1^2 a < a_1 a^2, \tag{24}$$
$$a_2^1 a < a_2 a^1. \tag{25}$$

Let δ be such that $0 < \delta < \min(a_1^2, a_2^1)$ and define

$$A(\delta) = \begin{pmatrix} a_1^1 + \delta & a_1^2 - \delta \\ a_2^1 - \delta & a_2^2 + \delta \\ a_3^1 & a_3^2 \\ \vdots & \vdots \\ a_m^1 & a_m^2 \end{pmatrix}.$$

We must prove that $\widetilde{F}(A) < \widetilde{F}(A(\delta))$. From the definition of $\widetilde{A}$ and inequalities (22) to (25), we get

$$\begin{aligned}
\widetilde{a}_1^1 \widetilde{a} &= (n-1)a_1^1 a > (n-1)a_1 a^1 = \widetilde{a}_1 \widetilde{a}^1, \\
\widetilde{a}_2^2 \widetilde{a} &= (n-1)a_2^2 a > (n-1)a_2 a^2 = \widetilde{a}_2 \widetilde{a}^2, \\
\widetilde{a}_1^2 \widetilde{a} &= (n-1)a_1^2 a < (n-1)a_1 a^2 = \widetilde{a}_1 \widetilde{a}^2, \\
\widetilde{a}_2^1 \widetilde{a} &= (n-1)a_2^1 a < (n-1)a_2 a^1 = \widetilde{a}_2 \widetilde{a}^1.
\end{aligned}$$

19

By definition of $\widetilde{F}$ and monotonicity of F,

$$
\begin{aligned}
\widetilde{F}(A) \;=\;& F(\widetilde{A}) \\[4pt]
<\;& F \begin{pmatrix}
\widetilde{a}_1^1 + \delta & \widetilde{a}_1^2 - \delta & \widetilde{a}_1^3 & \cdots & \widetilde{a}_1^n \\
\widetilde{a}_2^1 - \delta & \widetilde{a}_2^2 + \delta & \widetilde{a}_2^3 & \cdots & \widetilde{a}_2^n \\
\widetilde{a}_3^1 & \widetilde{a}_3^2 & \widetilde{a}_3^3 & \cdots & \widetilde{a}_3^n \\
\vdots & \vdots & \vdots & & \vdots \\
\widetilde{a}_m^1 & \widetilde{a}_m^2 & \widetilde{a}_m^3 & \cdots & \widetilde{a}_m^n
\end{pmatrix} \\[4pt]
=\;& F \begin{pmatrix}
a_1^1 + \delta & a_1^2 - \delta & a_1 & \cdots & a_1 \\
a_2^1 - \delta & a_2^2 + \delta & a_2 & \cdots & a_2 \\
a_3^1 & a_3^2 & a_3 & \cdots & a_3 \\
\vdots & \vdots & \vdots & & \vdots \\
a_m^1 & a_m^2 & a_m & \cdots & a_m
\end{pmatrix} \\[4pt]
=\;& \widetilde{F} \begin{pmatrix}
a_1^1 + \delta & a_1^2 - \delta \\
a_2^1 - \delta & a_2^2 + \delta \\
a_3^1 & a_3^2 \\
\vdots & \vdots \\
a_m^1 & a_m^2
\end{pmatrix} \\[4pt]
=\;& \widetilde{F}(A(\delta)),
\end{aligned}
$$

as desired. ∎

The whole analysis above is ordinal. We could have worked directly with association orderings (as defined in Section 2) rather than association measures on $\mathcal{A}(m,n)$. In fact, this is worth doing because some association orderings cannot be represented by an association measure.

The suitable reformulation of Class-Size Invariance, Row-Size Invariance and Monotonicity for association orderings is obvious. The reformulation of Theorem 1 is equally straightforward: *an association ordering $\succsim$ on $\mathcal{A}(m,n)$ is class-size invariant if and only if there exists an ordering $\succsim_0$ on $\mathcal{A}(m-1, n-1)$ such that, for all $A, B \in \mathcal{A}(m,n)$, $A \succsim B$ if and only if $r(A) \succsim_0 r(B)$.* As for Theorem 2, call an ordering $\succsim$ on $\mathcal{A}(m,n)$ continuous if, for every $A \in \mathcal{A}(m,n)$, the sets $\{B \in \mathcal{A}(m,n) \mid B \succsim A\}$ and $\{B \in \mathcal{A}(m,n) \mid A \succsim B\}$ are closed in the standard (Euclidean) topology of $\mathcal{A}(m,n)$. By a classic result of Debreu (1954), every continuous association ordering is representable by a continuous association measure. It therefore follows from

Theorem 2 that, *if $m \geq 4$, no continuous association ordering $\succsim$ on $\mathcal{A}(m,n)$ satisfies Monotonicity and Row-Size Invariance.*

We do not know whether monotonic and row-size invariant (necessarily discontinuous) association orderings exist on $\mathcal{A}(m,n)$ when m and n are arbitrary. Continuity plays an important role in the proof of Theorem 2.

5　Some possibility results

This section discusses a few special cases in which possibility results can be established.

1) Monotonicity, Class-Size Invariance and Continuity are compatible in the dichotomous case, i.e., when $m = 2$ and $n = 2$. In fact, it follows easily from Theorem 1 that the first two axioms single out a unique ordering, which happens to be continuous. This ordering on $\mathcal{A}(2,2)$ is given by

$$A \succsim B \Leftrightarrow \max\left(\frac{a_1^1 a_2^2}{a_1^2 a_2^1}, \frac{a_1^2 a_2^1}{a_1^1 a_2^2}\right) \geq \max\left(\frac{b_1^1 b_2^2}{b_1^2 b_2^1}, \frac{b_1^2 b_2^1}{b_1^1 b_2^2}\right).$$

2) Monotonic and class-size invariant orderings do exist when $m = 2$ *or* $n = 2$. This is of some interest because dichotomous characteristics are common in practice: a good deal of the segregation measurement literature focuses on that case. Here is an example when, say, $m = 2$ and n is arbitrary. For every $A \in \mathcal{A}(2,n)$ and every each pair $\{j,l\} \in 2^N$ define

$$\rho^{\{j,l\}}(A) = \max\left(\frac{a_1^j a_2^l}{a_2^j a_1^l}, \frac{a_2^j a_1^l}{a_1^j a_2^l}\right). \tag{26}$$

Let $\rho(A) \in \mathbb{R}^{\frac{(n-1)n}{2}}$ be the vector $\rho(A) = (\rho^{\{j,l\}}(A))_{\{j,l\} \in 2^N}$. Let $\succsim^L$ denote the *leximax ordering*[2] on $\mathbb{R}^{\frac{(n-1)n}{2}}$ and define the ordering $\succsim$ on $\mathcal{A}(2,n)$ by

$$B \succsim A \Leftrightarrow \rho(B) \succsim^L \rho(A). \tag{27}$$

---

[2]For any $x \in \mathbb{R}^{\frac{(n-1)n}{2}}$, let $x^* \in \mathbb{R}^{\frac{(n-1)n}{2}}$ denote the vector obtained by reordering the coordinates of x in nondecreasing order. The leximax ordering $\succsim^L$ on $\mathbb{R}^{\frac{(n-1)n}{2}}$ is defined by letting $x \succsim^L y$ if and only if either there exists $j \in \left\{1, ..., \frac{(n-1)n}{2}\right\}$ such that $x_i^* = y_i^*$ for all $i > j$ and $x_j^* > y_j^*$ (in which case we write $x \succ^L y$) or $x_i^* = y_i^*$ for all $i \in \left\{1, ..., \frac{(n-1)n}{2}\right\}$ (in which case we write $x \sim^L y$).

We prove in the Appendix that $\succsim$ is class-size invariant and monotonic. Of course, $\succsim$ is not continuous.

3) Continuous, monotonic and class-size invariant measures do exist when $m = 2$ and $n = 3$ (or $m = 3$ and $n = 2$). Two examples, as the reader may check, are $F(A) = \max\left(\rho^{\{1,2\}}(A), \rho^{\{1,3\}}(A), \rho^{\{2,3\}}(A)\right)$ and $F(A) = \rho^{\{1,2\}}(A)\rho^{\{1,3\}}(A)\rho^{\{2,3\}}(A)$. Both generate the same ordering.

4) A matrix $A \in \mathcal{A}(m,n)$ is *isotropic* (Yule and Kendall (1950)) if $r_i^j(A) \geq 1$ for all $i \in M\backslash\{m\}$ and all $j \in N\backslash\{n\}$. Because of (7), this implies $r_{i,k}^{j,l}(A) \geq 1$ for all $i, k \in M$ such that $i < k$ and all $j, l \in N$ such that $j < l$. Moreover,

$$r_{i,k}^{j,l}(A) \leq r_{i',k'}^{j',l'}(A) \text{ if } i' \leq i < k \leq k' \text{ and } j' \leq j < l \leq l'. \tag{28}$$

A *rearrangement* of a matrix $A \in \mathcal{A}(m,n)$ is any matrix $\widetilde{A} \in \mathcal{A}(m,n)$ obtained by permuting the rows and columns of A. We let $\mathcal{A}^{iso}(m,n)$ denote the subset of matrices in $\mathcal{A}(m,n)$ that possess an isotropic rearrangement. This corresponds to the case where, although the characteristics are only categorical, their values can be "endogenously" ordered in a way that reveals a positive correlation between them.

Recall our notation $\overline{r}(A) = (r_{i,k}^{j,l}(A))_{1\leq i<k\leq m}^{1\leq j<l\leq n} \in \mathbb{R}_{++}^{\mu}$, where μ is the number of 2×2 submatrices of an $m\times n$ matrix and let $\succsim^L$ be the leximax ordering on $\mathbb{R}_{++}^{\mu}$. If $\widetilde{A}, \widehat{A}$ are two isotropic rearrangements of a matrix $A \in \mathcal{A}^{iso}(m,n)$, it is easy to see that $\overline{r}(\widetilde{A}) \sim^L \overline{r}(\widehat{A})$. We may therefore define the cpr-leximax ordering $\succsim$ on $\mathcal{A}^{iso}(m,n)$ by

$$B \succsim A \Leftrightarrow \overline{r}(\widetilde{B}) \succsim^L \overline{r}(\widetilde{A}) \text{ for any isotropic rearrangements } \widetilde{A}, \widetilde{B} \text{ of } A, B. \tag{29}$$

Proposition. *The cpr-leximax ordering $\succsim$ on $\mathcal{A}^{iso}(m,n)$ is class-size invariant and monotonic.*

Proof. Class-size invariance is obvious. To establish Monotonicity, let $A, B \in \mathcal{A}^{iso}(m,n)$, let $i, k \in M$ and $j, l \in N$ be such that $i \neq k$ and $j \neq l$, and suppose that i, j and k, l are positively associated in A whereas i, l and k, j are negatively associated, i.e., $a_i^j a > a_i a^j$, $a_k^l a > a_k a^l$, $a_i^l a > a_i a^l$, $a_k^j a > a_k a^j$. These inequalities imply $\frac{a_i^j a_k^l}{a_i^l a_k^j} > 1$. Let $\widetilde{A}, \widetilde{B}$ be isotropic rearrangements of A, B. To simplify notation, suppose $\widetilde{A} = A$ and $\widetilde{B} = B$. Since A is isotropic,

22

it follows that either (i) $i < k$ and $j < l$ or (ii) $i > k$ and $j > l$. Without loss of generality, assume $i < k$ and $j < l$.

Suppose there is $\delta > 0$ such that such that $b_i^j = a_i^j + \delta$, $b_i^l = a_i^l + \delta$, $b_k^j = a_k^j - \delta$, $b_k^l = a_k^l - \delta$, and all other entries of A and B coincide. We must show that $\bar{r}(B) \succ^L \bar{r}(A)$.

Without loss of generality, assume that δ is small enough to guarantee that for all $i', i'', k', k'' \in M$ and $j', j'', l', l'' \in N$ such that $i' < k', i'' < k'', j' < l', j'' < l''$,

$$r_{i',k'}^{j',l'}(A) < r_{i'',k''}^{j'',l''}(A) \Rightarrow r_{i',k'}^{j',l'}(B) \leq r_{i'',k''}^{j'',l''}(B)$$

and

$$r_{i',k'}^{j',l'}(B) < r_{i'',k''}^{j'',l''}(B) \Rightarrow r_{i',k'}^{j',l'}(A) \leq r_{i'',k''}^{j'',l''}(A).$$

If δ violates this assumption, we can write it as a sum of increments that do satisfy the assumption and apply the argument below to each of these increments.

In order to prove that $\bar{r}(B) \succ^L \bar{r}(A)$, it suffices to show that if there exist $i', k' \in M$ and $j', l' \in N$ such that $i' < k'$, $j' < l'$, and

$$r_{i',k'}^{j',l'}(B) < r_{i',k'}^{j',l'}(A), \tag{30}$$

then there exist $i'', k'' \in M$ and $j'', l'' \in N$ such that $i'' < k''$, $j'' < l''$, and

$$r_{i'',k''}^{j'',l''}(B) > r_{i',k'}^{j',l'}(B), \ r_{i'',k''}^{j'',l''}(A). \tag{31}$$

Because A is isotropic, inequality (30) only holds in the following cases: (i) $(i,j) = (i',l')$, (ii) $(i,j) = (k',j')$, (iii) $(i,l) = (i',j')$, (iv) $(i,l) = (k',l')$, (v) $(k,j) = (k',l')$, (vi) $(k,j) = (i',j')$, (vii) $(k,l) = (i',l')$, (viii) $(k,l) = (k',j')$.

Let us assume (i); all other cases are treated similarly. If $k' \neq k$, we have

$$r_{i',k'}^{j',l}(B) = \frac{a_i^{j'} a_{k'}^l}{a_{k'}^{j'}(a_i^l - \delta)} > \frac{a_i^{j'} a_{k'}^l}{a_{k'}^{j'} a_i^l} = r_{i',k'}^{j',l}(A) > \frac{a_i^{j'} a_{k'}^j}{a_{k'}^{j'} a_i^j} > \frac{a_i^{j'} a_{k'}^j}{a_{k'}^{j'}(a_i^j + \delta)} = r_{i',k'}^{j',l'}(B),$$

where the second inequality is an application of property (28). If $k' = k$, we have

$$r_{i',k'}^{j',l}(B) = \frac{a_i^{j'}(a_k^l + \delta)}{a_k^{j'}(a_i^l - \delta)} > \frac{a_i^{j'} a_k^l}{a_k^{j'} a_i^l} = r_{i',k}^{j',l}(A) > \frac{a_i^{j'} a_k^j}{a_k^{j'} a_i^j} > \frac{a_i^{j'}(a_k^j - \delta)}{a_k^{j'}(a_i^j + \delta)} = r_{i',k'}^{j',l'}(B),$$

In both cases, (31) holds for $i'' = i', k'' = k', j'' = j', l'' = l$.∎

6 Appendix

Let $\succsim$ be the ordering on $\mathcal{A}(2,n)$ defined in (27). We prove that $\succsim$ is class-size invariant and monotonic.

Class-Size Invariance follows immediately from the observation that if B is obtained by multiplying a row or a column of A, then $\rho(A) = \rho(B)$, hence, $A \sim B$.

As for Monotonicity, let $A, B \in \mathcal{A}(2,n)$. Note that by definition $\rho^{\{j,l\}}(A) \geq 1$ for every $\{j,l\} \in 2^N$. Assume, without loss of generality, that $i = 1$, $j = 1$ as well as $i = 2$, $j = 2$ are positively associated whereas $i = 1$, $j = 2$ as well as $i = 2$, $j = 1$ are negatively associated in A. These assumptions imply that $a_1^1 a_2^2 > a_2^1 a_1^2$, hence by (26),

$$\rho^{\{1,2\}}(A) = \frac{a_1^1 a_2^2}{a_2^1 a_1^2} > 1. \tag{32}$$

Suppose there is $\delta > 0$ such that $b_1^1 = a_1^1 + \delta$, $b_1^2 = a_1^2 - \delta$, $b_2^1 = a_2^1 - \delta$, $b_2^2 = a_2^2 + \delta$ and $b_i^j = a_i^j$ for $i = 1, 2$ and all $j \in N \setminus \{1, 2\}$. We must check that $B \succ A$.

Assume that δ is small enough to guarantee that for all $\{j,l\}, \{j',l'\}$,

$$\rho^{\{j,l\}}(A) < \rho^{\{j',l'\}}(A) \Rightarrow \rho^{\{j,l\}}(B) \leq \rho^{\{j',l'\}}(B)$$

and

$$\rho^{\{j,l\}}(B) < \rho^{\{j',l'\}}(B) \Rightarrow \rho^{\{j,l\}}(A) \leq \rho^{\{j',l'\}}(A).$$

This is without loss of generality because if δ violates this assumption, we can write it as a sum of increments that do satisfy the assumption and apply the argument below to each of these increments.

By definition of $\succsim$, it is necessary and sufficient to prove that $\rho(B) \succ^L \rho(A)$. Note first that $\rho^{\{j,l\}}(B) = \rho^{\{j,l\}}(A)$ whenever $j, l \notin \{1, 2\}$. Moreover, $\rho^{\{1,2\}}(B) = \frac{(a_1^1 + \delta)(a_2^2 + \delta)}{(a_2^1 - \delta)(a_1^2 - \delta)} > \frac{a_1^1 a_2^2}{a_2^1 a_1^2} = \rho^{\{1,2\}}(A)$. This means that $\rho^{\{j,l\}}(B) < \rho^{\{j,l\}}(A)$ only if (i) $\{j,l\} = \{1, l\}$ and $l \neq 2$ or (ii) $\{j,l\} = \{2, l\}$ and $l \neq 1$.

Suppose first that there exists $l \neq 2$ such that

$$\rho^{\{1,l\}}(B) < \rho^{\{1,l\}}(A). \tag{33}$$

This can only occur if $\rho^{\{1,l\}}(A) = \frac{a_2^1 a_1^l}{a_1^1 a_2^l}$ and $\rho^{\{1,l\}}(B) = \frac{(a_2^1 - \delta)a_1^l}{(a_1^1 + \delta)a_2^l}$. In that case, however, we have

$$\frac{a_1^l a_2^2}{a_2^l a_1^2} = \frac{a_2^1 a_1^l}{a_1^1 a_2^l} \frac{a_1^1 a_2^2}{a_2^1 a_1^2} = \rho^{\{1,l\}}(A)\rho^{\{1,2\}}(A).$$

24

Since $\rho^{\{1,l\}}(A) \geq 1$ and $\rho^{\{1,2\}}(A) > 1$, it follows that $\frac{a_1^l a_2^2}{a_2^l a_1^2} > 1$, hence, $\rho^{\{2,l\}}(A) = \frac{a_1^l a_2^2}{a_2^l a_1^2} > \rho^{\{1,l\}}(A)$. Furthermore, $\rho^{\{2,l\}}(B) = \frac{a_1^l(a_2^2+\delta)}{a_2^l(a_1^2-\delta)} > \frac{a_1^l a_2^2}{a_2^l a_1^2} = \rho^{\{2,l\}}(A)$. Hence,

$$\rho^{\{1,l\}}(B) < \rho^{\{1,l\}}(A) < \rho^{\{2,l\}}(A) < \rho^{\{2,l\}}(B). \tag{34}$$

Suppose next that there exists $l \neq 1$ such that

$$\rho^{\{2,l\}}(B) < \rho^{\{2,l\}}(A). \tag{35}$$

This can only occur if $\rho^{\{2,l\}}(A) = \frac{a_1^2 a_2^l}{a_2^2 a_1^l}$ and $\rho^{\{2,l\}}(B) = \frac{(a_1^2-\delta)a_2^l}{(a_2^2+\delta)a_1^l}$. In that case, however, we have

$$\frac{a_2^l a_1^1}{a_1^l a_2^1} = \frac{a_1^2 a_2^l}{a_2^2 a_1^l}\frac{a_1^1 a_2^2}{a_2^1 a_1^2} = \rho^{\{2,l\}}(A)\rho^{\{1,2\}}(A).$$

Since $\rho^{\{2,l\}}(A) \geq 1$ and $\rho^{\{1,2\}}(A) > 1$, it follows that $\frac{a_2^l a_1^1}{a_1^l a_2^1} > 1$, hence, $\rho^{\{1,l\}}(A) = \frac{a_2^l a_1^1}{a_1^l a_2^1} > \rho^{\{2,l\}}(A)$. Furthermore, $\rho^{\{1,l\}}(B) = \frac{a_2^l(a_1^1+\delta)}{a_1^l(a_2^1-\delta)} > \frac{a_2^l a_1^1}{a_1^l a_2^1} = \rho^{\{1,l\}}(A)$. Hence,

$$\rho^{\{2,l\}}(B) < \rho^{\{2,l\}}(A) < \rho^{\{1,l\}}(A) < \rho^{\{1,l\}}(B). \tag{36}$$

Equations (34) and (36) say that if a coordinate of the $\rho(X)$ vector decreases from $X = A$ to $X = B$, some other coordinate whose value is greater than the value of the decreasing coordinate must increase. By definition of the leximax ordering, this implies that $\rho(B) \succ^L \rho(A)$.

7   References

Debreu, G. (1954), "Representation of a Preference Ordering by a Numerical Function," chapter XL in R. M. Thrall, C. H. Coombs, and R. L. Davis (Eds.), Decision Processes, New York, 159-166.

Edwards, A. W. F. (1963), "The Measure of Association in a 2×2 Table," *Journal of the Royal Statistical Society* **126**, 109-114.

Goodman, L. A. (1965), "On the Multivariate Analysis of Three Dichotomous Variables," *American Journal of Sociology* **71**, 290–301.

Goodman, L. A. (1969), "How to Ransack Social Mobility Tables and Other Kinds of Cross-Classification Tables," *American Journal of Sociology* **75**, 1-39.

Goodman, L. A. and Kruskal, W. H. (1954), "Measures of Association for Cross Classification," *Journal of the American Statistical Association* **49**, 732-764.

Holland, P. W. and Wang, Y. J. (1987), "Dependence function for continuous bivariate densities," *Communications in Statistics – Theory and Methods* **16**, 863-876.

James, D. R. and Taeuber, K. E. (1985), "Measures of Segregation," *Sociological Methodology* **14**, 1-32.

Mosteller, F. (1968), "Association and Estimation in Contingency Tables," *Journal of the American Statistical Association* **63**, 1-28.

Plackett, R. L. (1965), "A Class of Bivariate Distributions," *Journal of the American Statistical Association* **60**, 516–522.

Reardon, S. and Firebaugh, G. (2002), "Measures of Multigroup Segregation," *Sociological Methodology* **32**, 33-67.

Theil, H. and Finizza, A. J. (1971), "A Note on the Measurement of Racial Integration of Schools by Means of Informational Concepts," *Journal of Mathematical Sociology* **1**, 187-194.

Yule, G. U. (1912), "On The Methods of Measuring Association Between Two Attributes," *Journal of the Royal Statistical Society* **75**, 579-652.

Yule, G. U. and Kendall, M. G. (1950), An Introduction to the Theory of Statistics, New York: Hafner.