Multidimensional polarization for ordinal data

Martyna Kobus[a,*]

[a] *Institute of Economics, Polish Academy of Sciences, Nowy Swiat 72, 00-330 Warszawa, Poland*

Abstract

Western governments increasingly place more emphasis on non-income dimensions in measuring national well-being (e.g. the UK, France). Not only averages, but the characteristics of the whole distribution (e.g. inequalities) are taken into consideration. Commonly used data such as life satisfaction, declared health status or level of education, however, are ordinal in nature and the fundamental problem of measuring inequality with ordinal variables exists. Here, a class of multidimensional inequality indices for ordinal data is characterized by inequality axioms and based on the characterization theorem an inequality measure is proposed. The method ensures that the index is also attribute decomposable, that is, we can estimate the contribution to overall inequality from inequality in dimensions and from their association. It was found for the period 1972-2010 in the US, excluding 1985 that inequality in perceived happiness contributed more to overall inequality than health inequality. Joint inequality in health and happiness was significantly higher in the first half of the study period (0.3 vs. 0.2). In the 1970s and 1980s most healthy people were also happier and this positive association increased inequality by around 20 percent. This trend was reversed in the late 1980s when the contribution of association became negative. This trend for the healthiest to no longer be the happiest persisted with the exception of three years.

Keywords: Multidimensional inequality, Ordinal data, Copula function
JEL: D3, D6

Introduction

Governments are increasingly interested in measuring social well-being. Indeed, several countries, including France, Germany, Italy, Japan, Korea, Spain and the UK, are now searching beyond GDP to measure progress. In November 2010 the British Prime Minister launched his happiness index, announcing that in evaluating quality of life the Government would rely not only on GDP growth but also on non-income indicators such as education, health and environment. The report on well-being was published by the Office for National Statistics in July last year. Cameron earlier

☆Phone + 48225549149; Fax + 48228312846

☆☆The project was supported by National Science Centre in Poland and by a grant from the CERGE-EI Foundation under a program of the Global Development Network. All opinions expressed are those of the author and have not been endorsed by CERGE-EI or the GDN.

*Corresponding author

Email address: mkobus@wne.uw.edu.pl (Martyna Kobus)

described monitoring people's well-being as one of the central political issues of our time. The Canadian government has been working with the UK since 2010 on the problem of measuring socio-economic progress. In May 2011, in its 50th Anniversary Week, the OECD launched the Better Life Index which allows comparison of life in 34 countries, based on 11 dimensions - housing, income, jobs, community, education, environment, governance, health, life satisfaction, safety, work-life balance - applying their own weight to each dimension. It appears that politicians have responded to economists' urges for a multidimensional analysis of well-being (Atkinson and Bourguignon 1982; Sen 1973, 1987; Maasoumi 1986; Tsui 1995; Gajdos and Weymark 2005; Duclos, Sahn and Younger 2011). In 2008 the French government created the Stiglitz-Sen-Fitousi commission to resolve the issues involved with measuring social and economic performance. The commission's final report (Stiglitz, Sen and Fitoussi, 2009) underlines the multidimensionality of the concept of well-being, its relation to non-income dimensions and the importance of including both objective and subjective measures. The authors write "To define what well-being means a multidimensional definition has to be used"(p.14), whereas in Recommendation 7 they stress "Quality-of-life indicators in all the dimensions covered should assess inequalities in a comprehensive way." Yet widely used non-income dimensions such as self-reported health status (Allison and Foster, 2004; Apouey, 2007; Zheng, 2010) or declared happiness (Di Tella and McCulloch, 2006; Diener et al., 1999; Frey and Stutzer, 2002; Kahneman and Krueger, 2006; Layard, 2005; Oswald, 1997) are ordinal. This is critical as there has been no published account of the construction of multidimensional indices to measure inequality in ordinal data. This paper proposes a methodology offering an opportunity to extend the reach of measurement and relieve this setback.

The standard procedure to avoid the problem of ordinal variables is to assign numerical values to categories in a manner that is consistent with the order of preference - a *scale*. Any increasing transformation of a scale reflects the same ordering of categories. Then, standard indices such as the Gini coefficient, Atkinson's index or the Theil index can be applied. However, as the following example indicates this procedure is fundamentally flawed. The distributions of self-reported health status among men π and women ω are, respectively, $\pi = (0.2, 0.2, 0.2, 0.2, 0.2)$ and $\omega = (0.3, 0.2, 0.1, 0.1, 0.3)$, that is, there are twenty percent men in each health category and thirty percent women in the first category. By assumption, higher category number indicates better health status. We consider two scales: $c = (1, 2, 3, 4, 5)$ and $\tilde{c} = (1, 2, 3, 4, 100)$; please note that both correspond to the same order of health categories. Then, under scale c the Gini index for the men's distribution is[1] $GINI(\pi, c) = 0.26$ whereas for women's distribution we get $GINI(\omega, c) = 0.31$, hence health inequality is lower among men than women. However, under scale $\tilde{c}$ the ranking is reversed; $GINI(\pi, \tilde{c}) = 0.72 > GINI(\omega, \tilde{c}) = 0.66$. Sensitivity to rescaling is clearly an undesirable property of standard inequality measures. In other words, conventional inequality measures should not be used with ordinal data as they are dependent on the mean which is sensitive to rescaling. Therefore, inequality measurement theory for ordered response data is based on the distribution of a variable

[1]We calculated the Gini index assuming there are two men in each health category, three women in the first health category, two women in the second health category and the like. This is valid since the Gini index is replication invariant.

rather than its values. The scatter around the median is the key concept involving qualitative variables. To avoid the problem with scale changes it is often postulated that inequality measures for ordinal data should be invariant with respect to rescaling or scale independent.

In this paper inequality indices are proposed which are both multidimensional and invariant to change of scale and so can be used with several ordinal variables. An inequality index is not an arbitrary function but a function that meets certain criteria (axioms), considered reasonable in the context of inequality measurement. Multidimensional inequality concerns two main ideas: the spread of the distribution (or, equivalently, we call it majorization) and association between variables. Methods to represent majorization of ordinal data in the light of inequality measurement and association are described. The two can be combined further to form an inequality relation, and finally, a class of inequality measures, consistent with the inequality relation and characterisable using the other standard axioms (continuity and scale independence). The additive medial correlation index is defined in the last section as one such measure.

Considering majorization first, an inequality index is sensitive to distribution changes that influence spread. In a unidimensional framework it is expressed by Allison and Foster (AF) as the partial ordering (Allison and Foster, 2004) of two distributions, as in the following case: Fixing $n \geq 1$ and allowing $\mathbf{p}_1, \mathbf{p}_2$ to be two probability distributions on $\{1, \dots, n\}$. $\mathbf{p}_1 \leqslant_{AF} \mathbf{p}_2$ if and only if the following three conditions are met:

(AF1) $\mathbf{p}_1, \mathbf{p}_2$ have identical median states m,

(AF2) $\mathbf{P}_1(j) \leq \mathbf{P}_2(j)$ for any $j < m$,

(AF3) $\mathbf{P}_1(j) \geq \mathbf{P}_2(j)$ for any $j \geq m$,

where $\mathbf{P}_1$ and $\mathbf{P}_2$ are the cdf's corresponding to $\mathbf{p}_1$ and $\mathbf{p}_2$ respectively. Interpretation of this ordering is intuitive, in particular, $\mathbf{p}_1 \leqslant_{AF} \mathbf{p}_2$ when $\mathbf{p}_1$ is more concentrated around the median state than $\mathbf{p}_2$. Considering a natural multidimensional extension of the AF relation, the multidimensional distribution $\mathbb{p}_1$ is less scattered than $\mathbb{p}_2$ if the AF relation is true for every dimension.

Now association between variables which cannot apply to unidimensional but which is a salient feature of multidimensional measures is considered. According to the multidimensional inequality literature (Tsui, 1999) it is asserted that for two distributions with the same margins, the distribution with the higher correlation between variables exhibits more inequality. In the case of categorical data a correlation coefficient cannot be used; yet, it appears that other coefficients such as Kendall's tau, Spearman's rho or the medial correlation coefficient (later defined) are better suited for the use with ordinal variables. They are based on a copula which is the most general measure of association between qualitative variables. As Schweizer and Wolff (1981) note, "it is precisely the copula which captures those properties of the joint distribution which are invariant under (...) strictly increasing transformations." Copulas are well-known in mathematics and statistics due to the celebrated Sklar's theorem (Sklar, 1959) which states that copula and marginal distributions characterize joint distribution fully. In this paper we assert that distributions with the same margins can be evaluated according to an ordering on copulas; such ordering can be consistent with Kendall's tau or Spearman's rho or another measure of dependence. The medial correlation coefficient is the measure considered in this work.[2] For example, if the ordering on copulas is given by the medial correlation coefficient then the distribution with a higher value of medial correlation coefficient is more unequal between two distributions with the same margins.

Combining majorization (expressed by multidimensional AF ordering) with association (expressed by the ordering on copulas) an inequality ordering (called CAF), with which the index needs to be consistent, is obtained. More precisely, two distributions are ordered with respect to CAF if they are either ordered with respect to multidimensional AF and they have identical copulas or, they are ordered with respect to ordering on copulas and they have identical marginals. It is demonstrated as a main result that indices which fulfil continuity, normalization, scale independence and are consistent with inequality ordering (CAF) are continuous increasing functions of copulas and marginals. With respect to copulas these functions rise according to the order on copulas (such as, for example, the medial correlation coefficient) whereas with respect to marginals they rise according to multidimensional AF. Based on the theorem a specific inequality index, the so-called additive medial correlation index is proposed. Moreover, this framework (separating majorization and association) allows the introduction of indices that are attribute decomposable. Contributions from dimensional inequality and association to overall inequality can then be isolated.

Attribute decomposability is a recently introduced topic (first described by Abul Naga and Geoffard, 2006), however its applications appear far reaching, for example, if in two regions of a country there is a similar level of multidimensional inequality arising from different source causes and if in one region educational inequality dominates over health, policy makers seeking to reduce inequality in both regions need to apply different policy measures. Attribute decomposability reveals the detail needed to devise and implement suitable policy. Without attribute decomposability such insight is absent, since the two regions are indistinguishable in terms of overall inequality score.

Here the methodology is applied to data on happiness and declared health status taken from General Social Survey in the United States between 1972-2010. The sample consists of not institutionalized randomly-selected adults (18+). Throughout 1972-2010 inequality in declared happiness and health status in the US ranges from 0.217 to 0.381. In general, it is higher in the 1970s and 1980s than in the 1990s and 2000s (0.3 versus 0.2). To a major extent (around 80 percent) inequality is determined by both health and happiness inequality, what remains is attributed to association. Health and happiness inequality follow the same trend, with happiness being dominant except for in 1985. Association is more variable than inequality in the dimensions considered. Its contribution to overall inequality is reasonably stable and significantly positive in the first twenty years , but from the late 80s it decreases dramatically to negative or near zero levels. With a few exceptional years (1998, 2002, 2004) this suggests that the predominant trend in the first twenty years (healthier are happier) disappears in the last twenty years. This general picture does not apply to

[2] We explain why we choose this coefficient in in Section 5.

2004 when there was a remarkable rise in the association level when its contribution equalled 0.41, but after which it returned to levels close to zero.

The paper is organized as follows. In Section 1 notation and definitions are introduced. In Section 2 majorization (multidimensional AF) and then association (copula ordering) are considered. In Section 3 the focus is on the most equal and most unequal distributions. In Section 4 axioms are presented and a characterization theorem is given. This is the main result of the paper. Based on the theorem in Section 5 an inequality index is presented and proved to fulfil the axioms postulated by the characterization theorem. Section 6 which is empirical is followed by conclusions. Proofs are collected in the Appendix.

1. Basic definitions and notation

A relation $\precsim$ is a partial ordering if and only if it is reflexive ($\mathbb{p} \precsim \mathbb{p}$), antisymmetric ($\mathbb{p}_1 \precsim \mathbb{p}_2$ and $\mathbb{p}_2 \precsim \mathbb{p}_1$ implies $\mathbb{p}_1 = \mathbb{p}$) and transitive ($\mathbb{p}_1 \precsim \mathbb{p}_2$ and $\mathbb{p}_2 \precsim \mathbb{p}_3$ implies $\mathbb{p}_1 \precsim \mathbb{p}_3$). In what follows relations which are only reflexive and transitive will be used, these are quasi-orderings. An ordering is a partial ordering in which all elements are comparable (a chain). Each such ordering has an associated indifference relation (being the equivalence relation) defined as $\mathbb{p}_1 \sim \mathbb{p}_2$ if and only if $\mathbb{p}_1 \precsim \mathbb{p}_2$ and $\mathbb{p}_2 \precsim \mathbb{p}_1$. An element $\mathbb{p}$ is considered maximal (minimal) in $\precsim$ if there exists no element $\tilde{\mathbb{p}}$ such that $\mathbb{p} \precsim \tilde{\mathbb{p}}$ ($\tilde{\mathbb{p}} \precsim \mathbb{p}$) and $\mathbb{p} \nsim \tilde{\mathbb{p}}$. For demonstration purposes two well-being dimensions are analysed, however, this can be extended to an arbitrary number of dimensions.[3] In the notation unidimensional objects are typed in bold.

Let us consider self-declared health status and education level as two dimensions of well-being. We call a vector $\mathbf{c} = (\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_n)$ a unidimensional scale whenever $\mathbf{l}_1 < \mathbf{l}_2 < \ldots < \mathbf{l}_n$, n is the number of categories. For instance, we have ordered responses to a well-being dimension such as health and $\mathbf{c}_1 = (1, 2, 3, 4, 5)$ means that the first health category is assigned number 1 and the second health category is assigned number 2. For education level the scale is $\mathbf{c}_2 = (2, 4, 8)$. The set of bi-dimensional scales is denoted by $\mathbb{C}$. Continuing the example, $\mathbb{c} := (\mathbf{c}_1, \mathbf{c}_2) = ((1, 2, 3, 4, 5), (2, 4, 8)) \in \mathbb{C}$. We define $\mathbb{I} := \{1, \ldots, n_1\} \times \{1, \ldots, n_2\}$, n_j is the number of categories in the $j-$th dimension; $j = 1, 2$. In our example $\mathbb{I} := \{1, 2, 3, 4, 5\} \times \{1, 2, 3\}$. Throughout our article $\mathbb{I}, n_j$ are fixed unless we explicitly state otherwise.

Now let $\mathbb{p}$ be a probability distribution on the set $\mathbb{I}$. Given a scale $\mathbb{c} \in \mathbb{C}$ one may also consider $\mathbb{p}$ as a distribution on $\mathbb{c}$. By defining probability distribution $\mathbb{p}$ on $\mathbb{I}$ we make it independent of scale; that is, if there are two different scales with the same number of categories on each dimension ($\mathbb{I}$ does not change) then on both scales the same $\mathbb{p}$ can be considered a probability distribution. Obviously we require $\sum_{\mathbb{i} \in \mathbb{I}} \mathbb{p}(\mathbb{i}) = 1$ and for all $\mathbb{i} \in \mathbb{I}$, $\mathbb{p}(\mathbb{i}) \geq 0$. We now define marginal distributions. Let $\mathbb{p}$ be a probability distribution on $\mathbb{I}$ as above. For $j \in \{1, 2\}$ we put

$$\mathbf{p}^j(\mathbf{i}) := \sum_{\mathbb{i} \in \mathbb{I} \text{ such that } \mathbb{i}_j = \mathbf{i}} \mathbb{p}(\mathbb{i}), \quad \mathbf{i} \in \{1, 2, \ldots, n_j\}. \tag{1}$$

For example, for educational level there are three categories hence $\mathbf{i} \in \{1, 2, 3\}$ and the distribution can be the following $\mathbf{p}^2(1) = 0.15, \mathbf{p}^2(2) = 0.55, \mathbf{p}^2(3) = 0.30$, that

[3] See *coin.wne.uw.edu.pl/mkobus/Multi1extended.pdf*.

is, fifteen percent of the society is in the first category of education level and fifty five percent is in the second category. We notice that $\mathbf{p}^j$ is a unidimensional distribution for which we can define the cumulative distribution function

$$\mathbf{P}^j(k) = \sum_{h \leq k} \mathbf{p}^j(k), \quad j \in \{1,2\}.$$

In our previous example, we obtain $\mathbf{P}^2(2) = 0.15 + 0.55 = 0.70$. In a similar manner we define a multidimensional cumulative distribution function by

$$\mathbb{P}(\mathbb{i}) = \sum_{\mathbb{h} \in \{1,2,\ldots,\mathbb{i}_1\} \times \{1,2,\ldots,\mathbb{i}_2\}} \mathbb{p}(\mathbb{h}).$$

Continuing the example we could have $\mathbb{P}(2,4) = 0.40$. For each dimension j we define a median m_j which is the number for which $\mathbf{P}^j(m_j - 1) \leq 1/2$ and $\mathbf{P}^j(m_j) \geq 1/2$. Let Λ denote a set of probability distributions with given marginals on all dimensions. Finally, let inequality index be denoted by $I : \Lambda \times \mathbb{C} \mapsto \mathbb{R}$.

2. Majorization, association and inequality ordering

Having begun defining ordering that reflects one of two distributions as being less scattered (majorization), we next offer a suggestion of how association can be represented using copulas. We follow with an example of association ordering. The two ideas are then combined by constructing a general inequality relation.

2.1. Majorization

When matrix A is less scattered than matrix B we say that it is majorized by B. In the literature on unidimensional inequality indices for continuous data the notion of less spread is expressed by Pigou-Dalton transfer axiom which states that transfer of resources from the richer to the poorer in such a way not to change their relative positions, reduces inequality. Hence, individuals are moved towards the mean of the distribution. Similarly, Allison and Foster (2004) postulate that inequality in categorical data increases when probability mass is moved away from the median. They introduce a particular relation on the space of distributions which we term unidimensional AF ordering $\leqslant_{AF}$ and define formally in the Introduction. This relation embodies the notion of one distribution being more equal than the other. As a bi-dimensional counterpart of $\leqslant_{AF}$ we focus on its natural extension.

Definition 1. *Multidimensional AF*
Fixing m_1, m_2, let $\mathbb{p}_1, \mathbb{p}_2$ be two probability measures on $\mathbb{I}$ respectively. $\mathbb{p}_1 \precsim_{AF} \mathbb{p}_2$ if and only if

$$\mathbf{p}_1^1 \leqslant_{AF} \mathbf{p}_2^1 \ \textit{and} \ \mathbf{p}_1^2 \leqslant_{AF} \mathbf{p}_2^2$$

where $\leqslant_{AF}$ is defined by (AF1) - (AF3) and the median states for dimensions 1 and 2 are, respectively, m_1 and m_2.

In the above definition $\mathbf{p}_1^j, \mathbf{p}_2^j$ are marginals of $\mathbb{p}_1, \mathbb{p}_2$ given by (1). This definition is a straightforward extension of a unidimensional relation $\leqslant_{AF}$ to a multivariate framework. In particular, one bivariate distribution is less scattered, if it is so for each dimension. Further, if marginals are concentrated at one point then the joint distribution shows the same. The converse also holds true, that is, if a joint distribution is concentrated at one point then so are the marginals. The tendency is, therefore, for a better concentration of the marginal distributions to imply a better concentration of the joint distribution.

2.2. Association

As already mentioned association is a salient feature of multidimensional distributions. Generally, the greater the association the greater the tendency to inequality. In this section we show that in the case of ordinal data this concept is well-captured by the ordering on copulas which have the advantage of being scale-invariant. Copulas were popularized by Sklar (1959) to study the dependence structure between random variables. We know from Sklar's theorem that the copula of a distribution $\mathbb{p}$ is the only information necessary to recover $\mathbb{p}$ from its marginal distributions $\mathbf{p}^1, \mathbf{p}^2$. Formally, a bi-dimensional copula $cop : [0,1]^2 \mapsto [0,1]$ is a function such that

$$\mathbb{P}((j_1, j_2)) = cop\left(\mathbf{P}^1(j_1), \mathbf{P}^2(j_2)\right). \tag{2}$$

Let Cop denote a function which given a multidimensional distribution returns its copula.[4]

In order to understand copulas and how they differ from cdf better it is easiest to consider an example. In this setting it is impossible to show the invariance of copulas with respect to increasing transformations since probability distribution on $\mathbb{I}$ has already been defined, irrespective of scale. Therefore, we need to move away from the framework and assume that $\mathbb{p}$ is a probability distribution on a given scale. In the original distribution (Distribution (a) in Figure 1) there are two dimensions: health and education level with respectively $(1, 2)$ and $(3, 5)$ as categories. The copula is constructed as follows: we calculate $cop(0.5, 0.5)$; here the first 0.5 is the probability value in, say, the health dimension, so in Distribution (a) it corresponds to health category 1; second 0.5 is the probability value in education distribution, so in Distribution (a) it corresponds to education category 3; since $\mathbb{P}(1, 3) = 0.5$ we obtain $cop(0.5, 0.5) = 0.5$. Now if we change the scale in which we measure health from 1 and 2 into, respectively, 10 and 20 (Distribution (b)) the cdf changes from $\mathbb{P}(1, 2) = 0.5$ into $\mathbb{P}(1, 2) = 0$, whereas the copula does not change, namely, the copula calculated at a point $(0.5, 1)$ is still 0.5. This is what is meant when it is stated that the copula is invariant to increasing transformations of variables (cdf is not). Therefore, the copula is a particularly useful concept when applied to measure association between variables for which only order matters.

Let association be captured by $\ll$ a total ordering on the set of copulas; $\sim$ is its indifference relation. The specific $\ll$ we will use extensively later on is the ordering on copulas induced by the medial correlation coefficient denoted by $\ll_{medial}$. It was proposed by Blomqvist (1950). In our setting it has the following form.

[4] Technical problems related to the fact that in the discrete setting such as ours there are infinitely many copulas associated with a given distribution are dealt with in the Appendix.

	Health 1	Health 2	
Edu 2	$\frac{1}{2}$	0	$\frac{1}{2}$
Edu 5	0	$\frac{1}{2}$	$\frac{1}{2}$
	$\frac{1}{2}$	$\frac{1}{2}$	

(a) Distribution (a)

	$\frac{1}{2}$	1
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
1	$\frac{1}{2}$	1

(b) Copula of (a)

	Health 10	Health 20	
Edu 2	$\frac{1}{2}$	0	$\frac{1}{2}$
Edu 5	0	$\frac{1}{2}$	$\frac{1}{2}$
	$\frac{1}{2}$	$\frac{1}{2}$	

(c) Distribution (b): transformed health categories

Figure 1: The difference between copula and cdf.

Definition 2. *Medial correlation coefficient*

$$\beta(\mathbb{p}) = \sum_{i=1}^{m_1-1} \sum_{j=1}^{m_2-1} \mathbb{p}((i,j)) + \sum_{i=m_1+1}^{n_1} \sum_{j=m_2+1}^{n_2} \mathbb{p}((i,j)) - \sum_{i=1}^{m_1-1} \sum_{j=m_2+1}^{n_2} \mathbb{p}((i,j)) - \sum_{i=m_1+1}^{n_1} \sum_{j=1}^{m_2-1} \mathbb{p}((i,j)).$$

The first two expressions indicate a situation in which two variables values fall on the same side of their respective medians, namely, either below the median or above the median. The higher the probability mass associated with these two cases the higher the association is between two variables. In contrast, the last two expressions rise when two variables values fall on opposite sides of their medians. Association is then lowered. Obviously, higher coefficient value implies more association between two dimensions.

2.3. Inequality ordering

We combine majorization and association by defining the relation $<_{CAF}$ as follows.

Definition 3. $\mathbb{p}_1 <_{CAF} \mathbb{p}_2$ *if either*

(CAF1) $\mathbb{p}_1 \precsim_{AF} \mathbb{p}_2$ *and* $Cop(\mathbb{p}_1) \sim Cop(\mathbb{p}_2)$.

or

(CAF2) $Cop(\mathbb{p}_1) \ll Cop(\mathbb{p}_2)$ *and* $\mathbb{p}_1, \mathbb{p}_2$ *have the same marginal distributions.*

There are two cases in which distribution $\mathbb{p}_1$ is more equal than $\mathbb{p}_2$ in the sense of $<_{CAF}$. When the degree of dimensional association is identical for two distributions, distribution $\mathbb{p}_1$ is more equal than $\mathbb{p}_2$ if it has less spread. On the other hand, if both distributions are equally scattered then the more equal is the distribution that has lower association between dimensions. In other words, with association fixed, equality is markedly based on majorization and also, leaving majorization unaltered it is association that determines which distribution is the more equal.

$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{3}$
$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{3}$
$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{3}$
$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	

(a) Distribution $\mathbb{p}_1$

$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{4}$
$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$
$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{4}$
$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	

(b) Distribution $\mathbb{p}_2$

0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$
$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$
$\frac{1}{8}$	$\frac{1}{8}$	0	$\frac{1}{4}$
$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	

(c) Distribution $\mathbb{p}_3$

Figure 2: Violation of transitivity.

Relation $<_{CAF}$ is reflexive because of reflexivity of multidimensional AF relation, but typically, $<_{CAF}$ is not transitive. To see why, we consider a simple intuitive example (Figure 2) of three distributions. Since $Cop(\mathbb{p}_1) \sim_{medial} Cop(\mathbb{p}_2)$[5] by (CAF1) $\mathbb{p}_2 <_{CAF} \mathbb{p}_1$. On the other hand, $\mathbb{p}_2$ and $\mathbb{p}_3$ have the same marginals and $Cop(\mathbb{p}_3) \ll_{medial} Cop(\mathbb{p}_2)$[6], hence by (CAF2) $\mathbb{p}_3 <_{CAF} \mathbb{p}_2$. There is no reason however to have $\mathbb{p}_3 <_{CAF} \mathbb{p}_1$; neither (CAF1) nor (CAF2) applies. In order to remove non-transitivity we introduce the following definition.

Definition 4. *$\precsim_{CAF}$ is a transitive relation induced by $<_{CAF}$, that is, $\precsim_{CAF}$ is the intersection of all quasi-orderings containing $<_{CAF}$.*

In other words $\precsim_{CAF}$ is "the smallest order" compatible with $<_{CAF}$.[7]

3. The most equal distribution and the most unequal distributions

We will now consider two distinctive distributions, namely, the most equal and the most unequal distribution. We start with the most equal distribution. We argue that the best distribution in the set Λ with respect to inequality is $\hat{\mathbb{p}}$ such that

$$\hat{\mathbf{p}}^1(m_1) = 1 \text{ and } \hat{\mathbf{p}}^2(m_2) = 1. \tag{3}$$

Namely, the best distribution occurs when all probability mass is concentrated in the median state on each marginal. Let us note that it is trivial to confirm that such a distribution is unique and concentrated in one point: $\mathbb{i} = (m_1, m_2)$. If mass assigned to a point were different than $\mathbb{i}$ then, clearly, some marginal would vary from that prescribed by (3). This distribution is optimal with respect to $\precsim_{AF}$ as each marginal is minimum with respect to unidimensional AF ordering. For example, we might consider $\mathbb{I} = \{1, 2, 3, 4, 5\} \times \{1, 2, 3\}$, setting $m_1 = 4, m_2 = 2$. According to (3) the most equal distribution among distributions with prescribed medians is $\hat{\mathbb{p}}((4, 2)) = 1$ and is otherwise zero. We note that $\hat{\mathbb{p}}$ is the only minimal element of $\precsim_{AF}$ and $\precsim_{CAF}$.[8]

[5] The medians of $\mathbb{p}_1$ and $\mathbb{p}_2$ are second row and second column; $\beta(\mathbb{p}_1) = \frac{1}{9} + \frac{1}{9} - \frac{1}{9} - \frac{1}{9} = 0$ and $\beta(\mathbb{p}_2) = \frac{1}{16} + \frac{1}{16} - \frac{1}{16} - \frac{1}{16} = 0$.

[6] The medians of $\mathbb{p}_3$ and $\mathbb{p}_2$ are second row and second column; $\beta(\mathbb{p}_3) = 0 + 0 - \frac{1}{8} - \frac{1}{8} = -\frac{1}{4}$.

[7] Such transitive relation can be easily constructed. Please see *coin.wne.uw.edu.pl/mkobus/Multi1extended.pdf*.

[8] Please note that the medians are fixed.

Let us now study the most unequal distribution. In a unidimensional setting the distribution which assigns half of the probability mass to "the worst" category and the other half of mass to "the best category" is typically considered as the most unequal.[9] We argue that the worst distribution in the set Λ is $\check{\mathbb{p}}$ such that

$$\check{\mathbf{p}}^1(1) = \check{\mathbf{p}}^1(n_1) = \frac{1}{2} \text{ and } \check{\mathbf{p}}^2(1) = \check{\mathbf{p}}^2(n_2) = \frac{1}{2} \text{ and } \beta(\check{\mathbb{p}}) = 1. \tag{4}$$

That is, we require $\check{\mathbb{p}}$ to be such that each of its marginal distributions has "maximal unidimensional spread" and the lowest association. For example, let the number of categories on both dimensions be three, that is, $n_1 = n_2 = 3$. Then all such 3×3 distributions can be parametrized by k as in Figure 3, where $k \in [-\frac{1}{4}, \frac{1}{4}]$. Thus in the set of 3×3 distributions, the most unequal distribution is the one for which $k = -\frac{1}{4}$ because then medial correlation coefficient equals 1. That is, the higher association the higher inequality.

Figure 3: A parametrization of all 3×3 distributions.

$\frac{1}{4}+k$	0	$\frac{1}{4}-k$	$\frac{1}{2}$
0	0	0	0
$\frac{1}{4}-k$	0	$\frac{1}{4}+k$	$\frac{1}{2}$
$\frac{1}{2}$	0	$\frac{1}{2}$	

4. Characterization theorem

In this section the multidimensional inequality axioms are set out. These axioms parallel standard axioms used in inequality measurement.

CON $I : \Lambda \times \mathbb{C} \mapsto \mathbb{R}$ is a continuous function.

SCALINDEP $I(\mathbb{p}, \mathbb{c}_1) = I(\mathbb{p}, \mathbb{c}_2)$ for any $\mathbb{c}_1, \mathbb{c}_2 \in \mathbb{C}$.

NORM1 $I(\mathbb{p}, \mathbb{c}) \geq 0$ and $I(\hat{\mathbb{p}}, \mathbb{c}) = 0$ for any $\mathbb{c} \in \mathbb{C}$.

NORM2 $I(\mathbb{p}, \mathbb{c}) \leq 1$ and $I(\check{\mathbb{p}}, \mathbb{c}) = 1$ for any $\mathbb{c} \in \mathbb{C}$.

EQUAL $(\mathbb{p}_1 \precsim_{CAF} \mathbb{p}_2) \Rightarrow (I(\mathbb{p}_1, \mathbb{c}) \leq I(\mathbb{p}_2, \mathbb{c}))$ for any $\mathbb{c} \in \mathbb{C}$.

(CON) states that an index is continuous. The (SCALEINDEP) postulates that an index is independent of scale.[10] Thus the problem of scale changes influencing the

[9] This is distribution $\check{\Pi}$ in Abul Naga and Yalcin (2008).

[10] The analysis could also be conducted for a weaker requirement, namely, than an index is invariant with scale changes (scale invariance). For further details please refer to *coin.wne.uw.edu.pl/mkobus/Multi1extended.pdf*.

ranking of two distributions no longer exists. The (NORM1) normalizes the possible range of values an index admits by assigning lowest value (zero) to the most equal distribution. In addition, the (NORM2) requires that the index admits the highest value for the most unequal distribution. The (EQUAL) says that an index must rank distributions according to equality ordering $\precsim_{CAF}$. We are now ready to derive the following theorem.

Theorem 1. $I : \Lambda \times \mathbb{C} \mapsto \mathbb{R}$ *fulfils CON, NORM1, NORM2, EQUAL, SCALINDEP if and only if* I *is of the form*

$$I(\mathbb{p}) = f(Cop(\mathbb{p}), \mathbf{p}^1, \mathbf{p}^2), \tag{5}$$

where f *is a continuous increasing function. Along the first coordinate it increases with respect to* $\ll$ *and on the coordinates* $2, 3$ *with respect to (unidimensional)* $\leqslant_{AF}$. *Moreover,* $f(\hat{\mathbb{p}}) = 0$ *and* $f(\check{\mathbb{p}}) = 1$.

Theorem 1 states that indices satisfying the postulated axioms admit a particular functional form. This result facilitates the choice of an inequality measure, namely, a given measure needs to be a member of the (5) family of measures. Otherwise one of the axioms does not hold. A particularly attractive property of the proposed index (and in fact, of all the indices characterized in Theorem 1) is attribute decomposability, in so far as we can distinguish between inequality arising from association and from marginal distributions. In fact, equation (5) can be viewed as the special case of the definition of the attribute decomposability of inequality indices, namely, an index is attribute decomposable if it can be represented as a function of inequality in marginals and some measure of association between marginals.

5. Additive medial correlation index

In this section we present a concrete index that is characterized by Theorem 1.

Definition 5. *Additive medial correlation index*

$$I(\mathbb{p}) = A1\beta(\mathbb{p}) + A2I_1(\mathbf{p}_1) + A2I_2(\mathbf{p}_2), \tag{6}$$

where

$$A1 = \frac{n_1 + n_2 - 2}{n_1 n_2 - 1}$$

$$A2 = \frac{(n_1 - 1)(n_2 - 1)}{2(n_1 n_2 - 1)}$$

$$I_1(\mathbf{p}_1) = \frac{\sum_{j<m_1} \mathbf{P}^1(j) - \sum_{j \geq m_1} \mathbf{P}^1(j) + n_1 - m_1 + 1}{\frac{1}{2}(n_1 - 1)}$$

$$I_2(\mathbf{p}_2) = \frac{\sum_{j<m_2} \mathbf{P}^2(j) - \sum_{j \geq m_2} \mathbf{P}^2(j) + n_2 - m_2 + 1}{\frac{1}{2}(n_2 - 1)}.$$

The unidimensional indices $I_1(\mathbf{p}_1), I_2(\mathbf{p}_2)$ both denote an absolute value index defined in Naga and Yalcin (2008). Kobus and Miłoś (2011) proved that this index is the only inequality index for ordinal data known in the literature that is decomposable by population subgroups. Obviously, the additive medial correlation index has the form given by (5). It is continuous and also scale independent. What is less obvious is that it satisfies the normalization axioms. Although clearly $I(\hat{\mathbb{p}}) = 0$ and $I(\bar{\mathbb{p}}) = 1$, indeed, we also need to show that these are the minimum and maximum values of the index. The weights in the index were chosen to ensure that the following condition is satisfied.

Theorem 2. *Index defined in* (6) *fulfils (NORM1) and (NORM2).*

Normalization proved to be the most difficult with other association measures such as Kendall's tau or Spearman's rho. Once bi-dimensional scale is fixed, it is not possible to find a distribution for which the value of one is admitted. This is caused by the substantial chance of obtaining a pair which is neither concordant nor discordant. Such a pair is the result of the discrete nature of distributions, it is well-known that such a situation cannot happen for any continuous distribution. On the other hand, medial correlation coefficient does not have this unfavourable property.

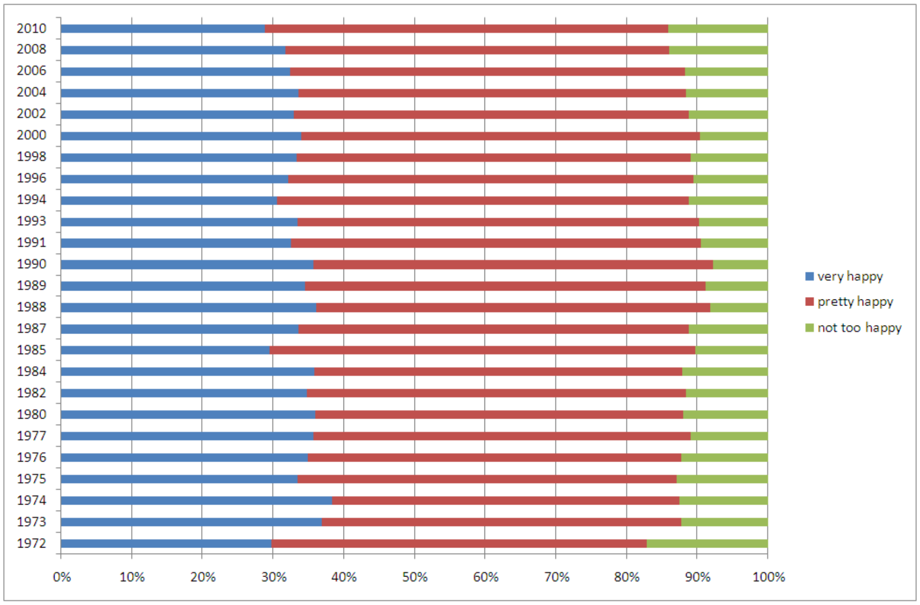
6. Empirical application

The methodology was applied to the data on happiness and declared health status taken from General Social Survey in the United States between years 1972 and 2010. The General Social Survey is a data bank of demographic characteristics and opinions of residents of the United States ranging from race and gender issues to religion. The survey is conducted via face-to-face interview and administered by the National Opinion Research Center at the University of Chicago. The sample is of randomly-selected adults (18+) who are not institutionalized. The survey started in 1972 and was conducted every year until 1994 (except in 1979, 1981, and 1992). Since 1994, it has been conducted every other year. As of 2010 twenty-eight national samples with 55,087 respondents and 5,417 variables had been collected. The variables were weighted appropriately to account for the black oversample in some years, problems with randomization procedures and the number of adults in the household.[11] The variables to point are as follows: happy (Taken all together, how would you say things are these days - would you say that you are very happy, pretty happy, or not too happy?) and health (Would you say your own health, in general, is excellent, good, fair, or poor?). The time trends for these two variables are presented in Figures 4 and 5. Through the whole period the median category of the happiness distribution was 'pretty happy', whereas the median of the health distribution was 'good'. The weights were the following: $A1 = \frac{5}{11}, A2 = \frac{6}{22}$.

Figure 4 indicates that the proportion of not too happy increased in the last decade in comparison with the 90s and 80s and returned to the levels observed in the 70s. Indeed, in the 1970s and 2000s, on average, 12 percent of the sample was not

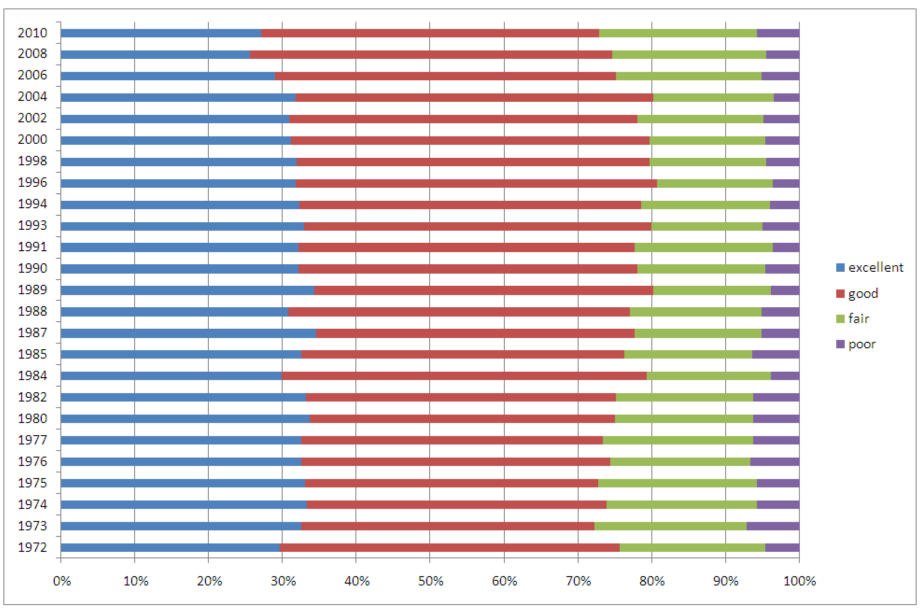
[11] Strictly speaking, the weight variables we use are the following: oversamp, formwt and wtssall. More on sampling design and weighting can be found in Appendix A of Cumulative Codebook of General Social Survey available at http://www.norc.uchicago.edu/GSS+Website/Documentation/.

Figure 4: General happiness by year in the US (1972-2010)



Source: General Social Survey database.

Figure 5: Condition of health by year in the US (1972-2010)



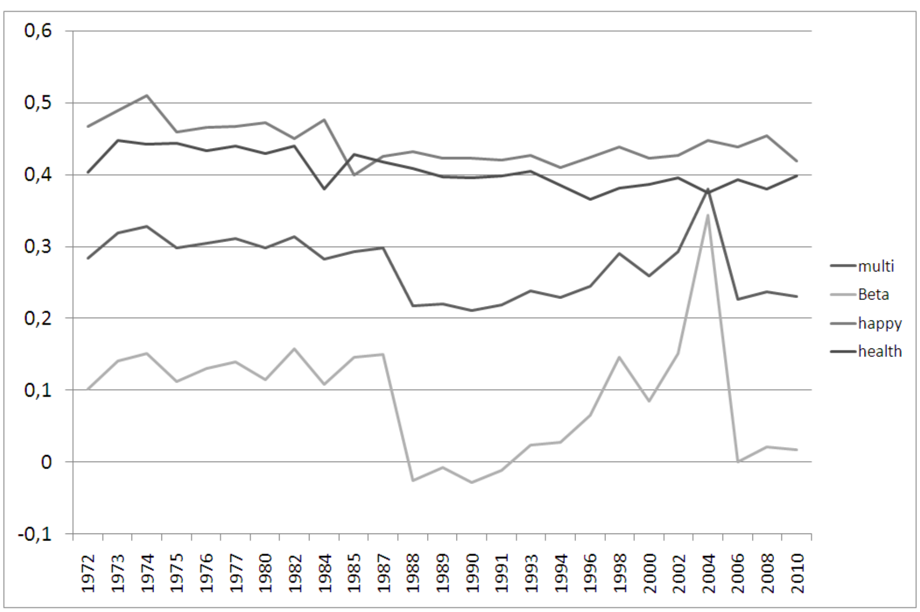
Source: General Social Survey database.

too happy, whereas in the 80s and 90s the average was around 10 percent and some years it was as low as 7.7 percent. On the other hand, a different declining trend with individuals who declare themselves as very happy is observed; in the 1970s very happy individuals constituted around 35 percent of the population, whereas in the first decade of the new millennium this number was close to 32 percent. With respect to health distribution (Figure 5) the greatest variability concerns those with fair health, the percentage ranges from 7 to more than 21. The percentage of individuals with poor health was slightly reduced, from 6 percent on average in the 1970s to around

5 percent between 2000 and 2010. There was a significant rise (by around 6 percent) in the proportion of those with good health for the last 40 years.

The values of the multidimensional index, the happiness and the health inequalities and the medial correlation coefficient are presented in Figure 6. The value of the additive medial correlation index varies between 0.217 (1988) and 0.381 (2004). Inequality was fairly stable until the late 80s when it decreased significantly. Since then, excluding the years 1998, 2002 and 2004, joint inequality in happiness and health is lower than it used to be throughout most of the 1970s and 1980s. Except for the year 1985 inequality in happiness is higher than health inequality throughout. Also health and happiness inequality were much less variable than the association which may explain why the time trend of the additive medial correlation index mimicked the behaviour of the association measure. On the other hand overall inequality score is for the whole period significantly higher than the association coefficient, which implies that inequality in marginals to a large extent determines the behaviour of the former. Indeed, this is what we observe when analyzing the contributions of happiness, health and medial correlation coefficient presented in Figure 7: the contribution of association is the highest only in 2004. The contribution of variable happy evolves between 0.321 (year 2004) and 0.547 (year 1990), whereas the contribution of variable health changes between 0.268 (year 2004) and 0.511 (year 1990). Except for 1985, inequality in perceived happiness contributed more to overall inequality than inequality in declared health status. The contribution of medial correlation coefficient varied between minus 0.06 (year 1990) and plus 0.41 (year 2004). Here minus means that association lowers inequality.

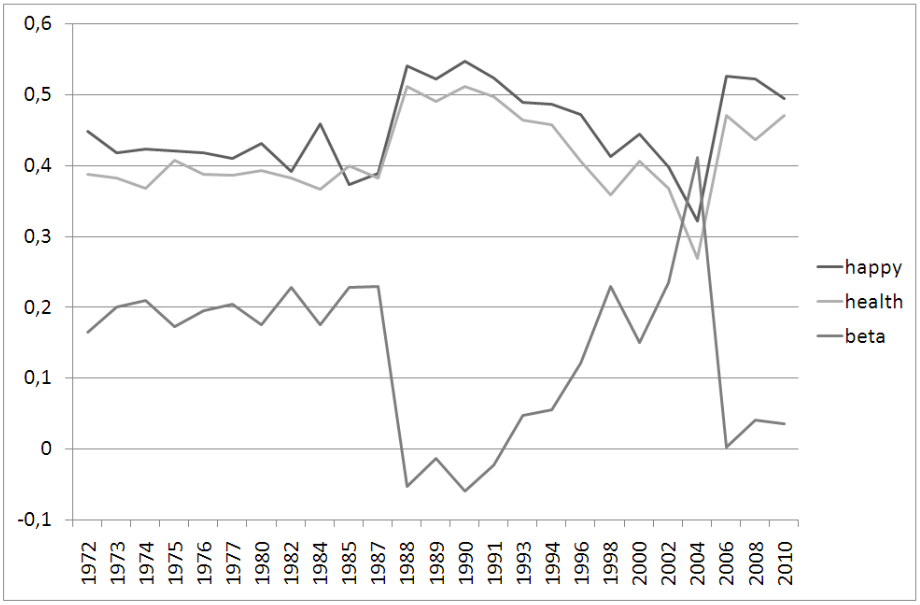
Figure 6: Inequality in happiness and health in the US in years 1972-2010



Source: General Social Survey database.

The Figures above show that until 1987 the contribution of association remained relatively stable at around plus 0.17 - 0.23. It fell in 1984, following which different trends of health and happiness can be observed (Figure 5); the percentage of those with excellent health reduced as did the percentage of individuals declaring poor

Figure 7: Contribution of happiness, health and association to overall inequality in the US in years 1972-2010



Source: General Social Survey database.

health, whereas as Figure 4 showed that number of both "very happy" and "not too happy" rise. The following year 1985 brought a reversal of these tendencies and was the only year in which health inequality prevailed over happiness inequality. Until 1988 both health and happiness were associated positively; those with poor health did not feel very happy and those with excellent health were happier. In 1988 the association dropped markedly and became negative with its contribution. It then grew to became as large as 0.41 in 2004 and this, as already stated, was the only year when association contributed more to inequality than dimensions. Afterwards, again, the association returned to near previous values.

In the year 2004 we observe an exceptionally high value of medial correlation coefficient. This looks almost like an outlier, therefore we studied this year more closely. In 2004 a new approach to demographic sampling frame construction was introduced, but this is taken into account by the weighting system. Also the approach was continued and we do not observe anything unusual (comparing to the 1990s) in years 2006-2010. We analyzed the raw data using an indicator variable to indicate whether observations lay on the same or on the opposite sides of the medians. In particular, the indicator variable assigned one to observations for which values of both happiness and health lie on the same side of the medians (are both lower or higher than its medians), zero when either health value or happiness value is the median value and minus one if the two values were on opposite sides of the median (i.e. health value was higher than its median value whereas happiness was lower). If we look at the percentile distribution of this indicator variable then in most years value one is admitted at around eightieth percentile, whereas in the year 2004 value one appears at the fiftieth percentile, so this really means there are substantially more cases in 2004 in which both happiness and health values lie on the same sides of their medians, that is, when they are positively correlated.

To sum up the observations, throughout 1972-2010 inequality in declared happiness and health status in the US is between 0.217 and 0.381. It was the lowest in 1988 and the highest in 2004. Yet on average it was greater in the first two decades of the considered period than in the last twenty years. The contributions of health inequality and happiness inequality follow the same time pattern with happiness inequality contributing more to overall inequality except for year 1985. The degree of association between happiness and health was less stable than both inequality in happiness and inequality in health and similar to its contribution. Until the late 1980s association was mildly positive and then, with the exclusion of years 1998, 2002, 2004 it became either negative or zero meaning that the healthiest were not necessarily the happiest. This was a reversal of the tendency that was prevalent in the 1970s and most of the 1980s.

Concluding remarks

In the introduction several initiatives undertaken by Western countries to improve measurement of social progress were mentioned. These initiatives prove that politicians are finally responding to economists' and social scientists' calls to look beyond GDP in order to estimate a nation's state of well-being. These pleas are documented in the extensive literature that has evolved over recent decades. As noted by McGillivray and Shorrocks (2005), twenty years ago, the comparison of living standards across countries was achieved by comparison of average incomes, whereas now it almost always includes non-income dimensions of well-being. Empirical research reveals factors that determine income inequality (social security provision, minimum wage access) that do not necessarily coincide with determinants of educational inequality, namely public provision of schools and child labour legislation (Jensen and Skyt Nielsen, 1997). Unlike income however, which is a continuous variable, many welfare dimensions are ordinal. Such is the case with happiness, health, occupational status and educational attainment. Indicators of living conditions such as drinking water sources and types of sanitation are evaluated differently for levels of quality; spring water is higher quality than collected rainwater, but is lower quality than public tap; latrines are classed as inferior to 'flush to piped sewer systems'. In measurement exercises, these categories are typically assigned numerical values consistent with the order of preference, to enable researchers to use standard inequality indices, yet such practice is inherently flawed. In this article we show how inequality in several ordinal variables can be measured.

Many research questions remain open. Firstly, there may be alternative definitions of multidimensional majorization ordering. Secondly, multidimensional AF ordering might be extended to allow for distribution comparison with different median states. In the unidimensional case Naga and Yalcin (2010) have already proposed methodology for comparison of distributions irrespective of their medians. Finally, dependence on association measure does not necessarily have to be explicit, nevertheless this may compromise attribute decomposability.

References

Abul Naga R H, Geoffard P Y. 2006. Decomposition of bivariate inequality indices

by attributes, Economics Letters, Elsevier, 90(3):362-367.
Abul Naga R H, Yalcin T.2010. Median independent inequality orderings. University of Aberdeen Business School Working Paper 03,University of Aberdeen Business School.
Abul Naga R H, Yalcin T. 2008. Inequality measurement for ordered response health data, Journal of Health Economics; 27(6):1614-1625.
Allison R A, Foster J E. 2004. Measuring health inequality using qualitative data, Journal of Health Economics; 23(3):505-524.
Apouey B. 2007. Measuring health polarization with self-assessed health data, Health Economics, 16:875-894.
Atkinson A B. 1970. On the measurement of inequality, Journal of Economic Theory, 2(3):244-263.
Atkinson A B, Bourguignon F. 1982. The Comparison of MultiDimensioned Distributions of Economic Status, Review of Economic Studies, 49(2):183-201.
Blackorby C, Primont D, Russell, R. 1978. Duality, Separability and Functional Structure: Theory and Economic Applications. North-Holland, New York.
Blomqvist N. 1950. On a measure of dependence between two random variables, Annals of Mathematical Statistics, 21:593–600.
Diener E, Lucas, R E. 1999. Personality and subjective well-being. In Kahneman, D., Diener E. and Schwarz, N. (Eds), Foundations of Hedonic Psychology: Scientific Perspectives on Enjoyment and Suffering, Chapter 11. New York: Russell Sage Foundation.
Di Tella R, McCulloch R. 2006. Some Uses of Happiness Data in Economics, Journal of Economic Perspectives, 20(1):25-46.
Duclos J Y, Sahn D E, Younger S D. 2011. Partial Multidimensional Inequality Orderings, Journal of Public Economics, 95(3-4):225-238.
Frey B S, Stutzer, A. 2002. Happiness and Economics. Princeton University Press, Princeton, NJ.
Gajdos T, Weymark J A. 2005. Multidimensional Generalized Gini Indices, Economic Theory, 26(3):471-496.
Jensen P, Skyt Nielsen, H. 1997. Child Labour or School Attendance? Evidence from Zambia, Journal of Population Economics 10:407-424.
Kahneman D, Krueger A B. 2006. Developments in the measurement of subjective wellbeing. Journal of Economic Perspectives, 22:3-24.
Kobus M, Miłoś P. 2012. Inequality decomposition by population subgroups for ordinal data, Journal of Health Economics 31:15-21. Layard R. 2005. Happiness. Lessons from a New Science, London: Allen Lane.
Maasoumi E. 1986. The Measurement and Decomposition of Multi-dimensional Inequality, Econometrica, 54(4):991-997.
Marshall A W, Olkin I. 1979. Inequalities: theory of majorization and its applications. Academic Press, San Diego.
Meng X, Miller W P. 1995. Occupational segregation and its impact on gender wage discrimination in China's rural industrial sector, Oxford Economic Papers, 47.
Nelsen R B. 1999. An Introduction to Copulas. New York: Springer-Verlag.
Oswald A J. 1997. Happiness and economic performance. Economic Journal, 107:1815-1831.
Tsui K Y. 1995. Multidimensional Generalizations of the Relative and Absolute In-

equality Indices: The Atkinson-Kolm-Sen Approach, Journal of Economic Theory, 67 (1):251-265.
Tsui K Y. 1999. Multidimensional inequality and multidimensional generalized entropy measures: An axiomatic derivation, Social Choice and Welfare, Springer; 16(1):145-157.
Schweizer B, Wolff E F. 1981. On nonparametric measures of dependence for random variables, Annals of Statistics, 9:870–885.
Sen A K. 1973. On economic inequality. Oxford, UK: Clarendon Press.
Sen A.K. 1987. The standard of living. Cambridge: Cambridge University Press.
Sklar A. 1959. Fonctions de repartition an dimensions et leurs marges, Publ. Inst. Statist. Univ. Paris, 8:229–231.
Stiglitz J, Sen, A, Fitoussi, JP. 2009. Report by the Commission on the Measurement of Economic Performance and Social Progress, www.stiglitz-sen-fitoussi.fr
Zheng B. 2010. A new approach to measure socioeconomic inequality in health, Journal of Economic Inequality.

Appendix

Theorem 1

Proof. Let us first observe that if I is of the form (5) axioms are obviously satisfied. We assume that the axioms are satisfied. SCALEINDEP ensures that $I(\mathbb{p}, \mathbb{c}) = \Phi(\mathbb{p})$ for some function Φ. Let us first take a function fulfilling the conditions above. To check that it is increasing with respect to $\precsim_{CAF}$ it is enough to check that it is increasing with respect to cases listed in Definition 3. We start with (CAF1); let $\mathbb{p}_1, \mathbb{p}_2$ be such that $\mathbb{p}_1 \precsim_{AF} \mathbb{p}_2$ and $Cop(\mathbb{p}_1) \sim Cop(\mathbb{p}_2)$. From the second property we know that $f(Cop(\mathbb{p}_2), \mathbf{p}_2^1, \mathbf{p}_2^2) = f(Cop(\mathbb{p}_1), \mathbf{p}_2^1, \mathbf{p}_2^2)$ while the first property gives us $f(Cop(\mathbb{p}_1), \mathbf{p}_2^1, \mathbf{p}_2^2) \geq f(Cop(\mathbb{p}_1), \mathbf{p}_1^1, \mathbf{p}_1^2)$ and finally $f(Cop(\mathbb{p}_2), \mathbf{p}_2^1, \mathbf{p}_2^2) \geq f(Cop(\mathbb{p}_1), \mathbf{p}_1^1, \mathbf{p}_1^2)$. (CAF2) is even simpler as it is basically equivalent to the fact that f is increasing with respect to the first coordinate. In this way we proved that the ordering induced by f contains $<_{CAF}$ and also $\precsim_{CAF}$ (Definition 4). By $Uncop(c, \mathbf{p}^1, \mathbf{p}^2)$ we denote a function which produces a distribution $\mathbb{p}$ such that $Cop(\mathbb{p}) = c$ and the marginal distributions are given by $(\mathbf{p}^1, \mathbf{p}^2)$. Thus $f = h \circ Uncop$ gives us the required representation. □

Theorem 2

Proof. First we will prove that minimum of the additive medial correlation index is attained at the distribution concentrated at (m_1, m_2): $\hat{\mathbb{p}}$. To this end let us compute a derivative of I with respect to $\mathbb{p}((i,j))$. To ease the notation we work with $\tilde{I} := (n_1 n_2 - 1)I$. We have the following cases

- $i < m_1, j < m_2$; $\frac{\partial \tilde{I}}{\partial \mathbb{p}((i,j))} = n_1 + (n_1 - 1)(-j + 2m_2 - n_2 - 1) + (-i + 2m_1 - n_1 - 1)(n_2 - 1) + n_2 - 2.$
- $i = m_1, j < m_2$; $\frac{\partial \tilde{I}}{\partial \mathbb{p}((i,j))} = (n_1 - 1)(-j + 2m_2 - n_2 - 1) + (m_1 - n_1 - 1)(n_2 - 1).$

- $i > m_1, j < m_2$; $\frac{\partial \tilde{I}}{\partial \mathbb{p}((i,j))} = -n_1 + (n_1 - 1)(-j + 2m_2 - n_2 - 1) + (i - n_1 - 1)(n_2 - 1) - n_2 + 2.$
- $i < m_1, j = m_2$; $\frac{\partial \tilde{I}}{\partial \mathbb{p}((i,j))} = (n_1 - 1)(m_2 - n_2 - 1) + (-i + 2m_1 - n_1 - 1)(n_2 - 1).$
- $i = m_1, j = m_2$; $\frac{\partial \tilde{I}}{\partial \mathbb{p}((i,j))} = (n_1 - 1)(m_2 - n_2 - 1) + (m_1 - n_1 - 1)(n_2 - 1).$
- $i > m_1, j = m_2$; $\frac{\partial \tilde{I}}{\partial \mathbb{p}((i,j))} = (n_1 - 1)(m_2 - n_2 - 1) + (i - n_1 - 1)(n_2 - 1).$
- $i < m_1, j > m_2$; $\frac{\partial \tilde{I}}{\partial \mathbb{p}((i,j))} = -n_1 + (n_1 - 1)(j - n_2 - 1) + (-i + 2m_1 - n_1 - 1)(n_2 - 1) - n_2 + 2.$
- $i = m_1, j > m_2$; $\frac{\partial \tilde{I}}{\partial \mathbb{p}((i,j))} = (n_1 - 1)(j - n_2 - 1) + (m_1 - n_1 - 1)(n_2 - 1).$
- $i > m_1, j > m_2$; $\frac{\partial \tilde{I}}{\partial \mathbb{p}((i,j))} = n_1 + (n_1 - 1)(j - n_2 - 1) + (i - n_1 - 1)(n_2 - 1) + n_2 - 2.$

Let $\mathbb{p}$ be a distribution with medians $(i,j) \neq (m_1, m_2)$. We consider the probability distribution such that $\mathbb{p}_((i,j)) := \epsilon$, $\mathbb{p}_((m_1, m_2)) := a - \epsilon$ and arbitrary values in other points. We define $f(\epsilon) = \tilde{I}(\mathbb{p})$ and calculate its derivative. We have $f'(\epsilon) = \frac{\partial \tilde{I}}{\partial \mathbb{p}((i,j))} - \frac{\partial \tilde{I}}{\partial \mathbb{p}((m_1,m_2))}$. Hence we have the following cases:

- $i < m_1, j < m_2$; $f'(\epsilon) = n_1 - (n_1 - 1)(m_2 - n_2 - 1) + (n_1 - 1)(-j + 2m_2 - n_2 - 1) - {} - (m_1 - n_1 - 1)(n_2 - 1) + (-i + 2m_1 - n_1 - 1)(n_2 - 1) + n_2 - 2 \geq 0.$
- $i = m_1, j < m_2$; $f'(\epsilon) = (n_1 - 1)(-j + 2m_2 - n_2 - 1) - (n_1 - 1)(m_2 - n_2 - 1) \geq 0.$
- $i > m_1, j < m_2$; $f'(\epsilon) = -n_1 - (n_1 - 1)(m_2 - n_2 - 1) + (n_1 - 1)(-j + 2m_2 - n_2 - 1) + (i - n_1 - 1)(n_2 - 1) - (m_1 - n_1 - 1)(n_2 - 1) - n_2 + 2 \geq 0.$
- $i < m_1, j = m_2$; $f'(\epsilon) = (-i + 2m_1 - n_1 - 1)(n_2 - 1) - (m_1 - n_1 - 1)(n_2 - 1) \geq 0.$
- $i = m_1, j = m_2$; $f'(\epsilon) = 0.$
- $i > m_1, j = m_2$; $f'(\epsilon) = (i - n_1 - 1)(n_2 - 1) - (m_1 - n_1 - 1)(n_2 - 1) \geq 0.$
- $i < m_1, j > m_2$; $f'(\epsilon) = -n_1 + (n_1 - 1)(j - n_2 - 1) - (n_1 - 1)(m_2 - n_2 - 1) - (m_1 - n_1 - 1)(n_2 - 1) + (-i + 2m_1 - n_1 - 1)(n_2 - 1) - n_2 + 2 \geq 0.$
- $i = m_1, j > m_2$; $f'(\epsilon) = (n_1 - 1)(j - n_2 - 1) - (n_1 - 1)(m_2 - n_2 - 1) \geq 0.$
- $i > m_1, j > m_2$; $f'(\epsilon) = n_1 + (n_1 - 1)(j - n_2 - 1) - (n_1 - 1)(m_2 - n_2 - 1) + (i - n_1 - 1)(n_2 - 1) - (m_1 - n_1 - 1)(n_2 - 1) + n_2 - 2 \geq 0.$

Therefore we proved that regardless of the choice of (i,j) the derivative of $f'(\epsilon) > 0$ hence the function is increasing. One easily checks that $\mathbb{p}_0 = \hat{\mathbb{p}}$. Let us now consider any $\mathbb{p} \neq \hat{\mathbb{p}}$. We notice that $\mathbb{p}$ can be obtained from $\hat{\mathbb{p}}$ by means of sequence of transfers of the probability mass from (m_1, m_2). As we have just proved each such transfer increases the value of $\tilde{I}$ therefore we have $\tilde{I}(\hat{\mathbb{p}}) < \tilde{I}(\mathbb{p})$. This concludes the proof of (NORM1).

Now we take any distribution $\mathbb{p}$ and show a sequence of transfers which transform this distribution into $\check{\mathbb{p}}$ and such that the index increases following each transfer. Let (i_0, j_0) be such that $1 < i_0 < m_1$, let us consider the probability distribution $\mathbb{p}$ such that $\mathbb{p}((i_0, j_0)) > 0$. We construct a new distribution $\mathbb{p}_\epsilon$ by setting $\mathbb{p}_\epsilon((i_0, j_0)) := \mathbb{p}((i_0, j_0)) - \epsilon$, $\mathbb{p}_\epsilon((1, j_0)) = \mathbb{p}((1, j_0)) + \epsilon$ and $\mathbb{p}_\epsilon$ agrees with $\mathbb{p}$ otherwise. The transformation moves the mass towards the edge of the first coordinate without moving the other. We denote $g(\epsilon) := \tilde{I}(\mathbb{p}_\epsilon)$ and calculate the derivative $g'(\epsilon) = \frac{\partial \tilde{I}}{\partial \mathbb{p}((1,j_0))} - \frac{\partial \tilde{I}}{\partial \mathbb{p}((i_0,j_0))}$. We have only three cases this time $i_0 < m_1, j_0 < m_2$, $i_0 < m_1, j_0 = m_2$ and $i_0 < m_1, j_0 > m_2$. In each case $g'(\epsilon) = (i_0 - 1)(n_2 - 1) > 0$. Such transfers on both sides of the median ensure that the whole mass is concentrated at a point for which $i \in \{1, m_1, n_1\}$. When transporting the mass from points (m_1, j) to $(1, j)$ or (n_1, j) the situation is more delicate since the transfers are required to be such that m_1 is median, hence the total mass gathered in the points of the form $(1, j)$ must be equal to $1/2$ and the same holds for the points of the form (n_1, j). The final step is to transfer all the mass from points of the form $(1, j)$ to $(1, 1)$ and all from (n_1, j) to (n_1, n_2). Following such transfer (6) medial correlation coefficient β increases and so does inequality in the second dimension as the spread rises. At each step of our procedure we increase the index hence the value attained at $\check{\mathbb{p}}$ is the highest possible. □

The problem of infinitely many copulas associated with any distribution

In our context there are infinitely many copulas associated with any distribution. This results from the fact that our distributions are discrete. Indeed, one can see that to calculate $\mathbb{P}$ in (2) one needs only the values of *cop* attained on a range of $\mathbf{P}^1$ and $\mathbf{P}^2$, which as a range of values of cumulative functions of discrete distributions is finite. The remaining points of $[0, 1]^2$ can be assigned arbitrary values compatible with the properties of the copula. However, this does not seem to be problematic since Sklar's theorem ensures that copula is unique on the range of cdf of marginal distributions (which we call copula's significant points) and this is indeed the only relevant information.[12] Further, we will assume that given a multidimensional distribution we always choose the smallest possible copula, where smallest means that a chosen copula is the smallest as a function, namely for every argument it admits smaller value than any other copula function. We also note that this copula can be recovered from its values at significant points, hence sometimes its values are only presented at these points (as in the example in Section 2.2). Technically, without this assertion Definition 3 would be incomplete, namely, conditions $Cop(\mathbb{p}_1) \sim Cop(\mathbb{p}_2)$ would not be well-defined. For instance, if $cop, \tilde{cop}$ are two copulas associated with $\mathbb{p}_1$ and additionally $\hat{cop}$ is associated with $\mathbb{p}_2$ and we have an order $\ll$ such that $cop \ll \hat{cop}$ and $\hat{cop} \ll \tilde{cop}$, then it is unclear whether we should have $\mathbb{p}_1 <_{CAF} \mathbb{p}_2$ or $\mathbb{p}_2 <_{CAF} \mathbb{p}_1$ or declare indifference. Therefore we needed to decide on choosing a single copula associated with a given multidimensional distribution $\mathbb{p}$. This solution

[12] Please refer further to Nelsen (1999): Theorem 2.3.3.

is equivalent to working with copulas restricted to their significant points.