INE PAN Working Paper Series

Paper number 50

Measuring inequality in the distribution of socioeconomic status and health

Martyna Kobus, Radosław Kurek

Warsaw, 25.02.2022

wp@inepan.waw.pl

Measuring inequality in the distribution of socioeconomic status and health

Martyna Kobus[*], Radosław Kurek[†]

Abstract

The measurement of socioeconomic inequalities in health receives a lot of attention in economic literature. Measurement problems that arise because of qualitative nature of health and status indicators are widely acknowledged. Methods based on polarization orderings have been developed to address this. We further extend the set of available tools by proposing dominance ordering and measures that are sensitive not only to between-group heterogeneity (i.e. increased *spread*), but also to within-group homogeneity (i.e. increased *bipolarity*). Using data from the Survey of Health, Ageing and Retirement in Europe we show that accounting for increased bipolarity significantly increases the value of inequality, although it does not affect the ranking of countries.

Keywords: socioeconomic inequalities in health; ordinal data; polarization
JEL classification: I14; I31; D63

1 Introduction

The measurement of health inequalities and socioeconomic inequalities in health is a fast growing area of economic research. Problems arise because of often qualitative nature of both health and socioeconomic indicators, which makes them difficult to use with standard measures of inequality. Such measures depend on the mean, which in case of ordinal

[*]Institute of Economics, Polish Academy of Sciences, 00-330 Warsaw, Nowy Swiat 72, Poland, mkobus@inepan.waw.pl.

[†]Institute of Economics, Polish Academy of Sciences, radek.kurek@inepan.waw.pl.

variables changes with the form of cardinalization. This may lead to the reversals of conclusions depending on the form of cardinalization, which is of course undesirable, e.g. Abul Naga and Yalcin (2008), Lazar and Silber (2013), Kobus (2015), Bond and Lang (2019). The concentration index (van Doorslaer et al. (1997), van Doorslaer and Koolman (2004)) typically used to assess socioeconomic inequalities in health is also inappropriate for similar reasons (see e.g. Makdisi and Yazbeck (2014), Fleurbaey and Schokkaert (2012)).

Following the work of Allison and Foster (2004) (henceforth, AF) a large body of literature has emerged to solve the problem of measuring inequality with ordered response data, e.g. Apouey (2007), Abul Naga and Yalcin (2008), Apouey and Silber (2013), Abul Naga and Stapenhurst (2015), Gravel, Magdalou and Moyes (2021), Lv, Wang and Xu (2015), Kobus (2015), and Cowell and Flachaire (2017). The methodology is non-parametric and avoids scaling problems. Allison and Foster (2004) proposed a bi-polarization dominance relation to compare distributions of health in terms of inequality. The most unequal distribution according to this relation is one for which half of the population occupies the lowest ordinal category and half the highest i.e. the most bi-polarized distribution. Recently, Kobus and Kurek (2018) extend it to two dimensions, which is a framework necessary to capture socioeconomic inequality in health. The proposed dominance relation and measures, however, care only about how far the mass is *away* from the median; the so called *increased spread*. They do not take into account the distribution of the mass below and above the median, namely, how concentrated the mass is around particular categories; the so called *increased bipolarity*. As noted by Esteban and Ray (1994) in their classic paper, full definition of polarization takes into account both concepts. In this short article, we show how already existing methods can be used to construct a dominance ordering and measures that capture both high between–group heterogeneity (spread) and high within–group homogeneity (bipolarity) and thus can be used to better measure socioeconomic inequality in health.

We draw on the work of Chakravarty and Maharaj (2015) (CM) that received much less attention than Allison and Foster (2004), but in fact improves AF dominance relation that considers increased spread only by adding increased bipolarity. In essence, to compare two distributions according to CM relation means to compare partial sums of cumulative distribution functions below the median, and partial sums of survival functions above the median. The summation of both cumulative distribution functions (cdfs) and survival functions starts from the median. A distribution which has higher such sums is considered more unequal. Our proposed extension of CM to two dimensions is a dominance relation that compares sums of cumulative distribution function and survival function of the *joint* distribution. In the context of health and socioeconomic status, such relation can be seen as the aggregation of health inequality, status inequality and the socioeconomic gradient in health (i.e. dependence between health and status indicator). We offer a simple index that is consistent with this relation and has some standard properties. We develop estimation and inference procedures. In order to increase the conclusiveness of our dominance ordering we remove the assumption of identical medians embedded in both AF and CM definitions, utilizing recent results by Sarkar and Santra (2020). Furthermore, we allow for the concentration of probability mass around different quantiles, not necessarily the median. This idea dates back to Mendelson (1987), long before Allison and Foster (2004) work.

We apply the developed methods to the data from Wave 7 of the Survey of Health, Ageing and Retirement in Europe (SHARE) (27 countries). We proxy socioeconomic status by 7–category educational attainment, and health by 5 – category self–reported health status. We find dominances in both health and education in around 8% of all pairwise comparisons, e.g. Kobus, Polchlopek and Yalonetzky (2018) get 4% for OECD countries. By comparing measures developed here with the measures proposed by Kobus and Kurek (2018), we observe that accounting for increased bipolarity increases inequality in each case. The increases are significant, e.g. inequality goes up from around 0.3 to around 0.5. However, countries' ranking is unaffected.

The paper is organized as follows. Section 2 introduces basic framework and definitions. In Section 3 we develop a new polarization ordering, the associated class of measures and show that it has desired properties. In Section 4 we develop inference and estimation procedures for the proposed class of measures. In Section 5, using the proposed tools, we analyze educational inequalities in health in Europe.

2 Basic definitions

We define $\mathbb{I} := \{1, \ldots, n_1\} \times \{1, \ldots, n_2\} \times \ldots \times \{1, \ldots, n_k\}$ which is endowed with the usual partial order: $(i_1, \ldots, i_k) \preceq (i'_1, \ldots, i'_k)$ if, and only, if $i_j \leq i'_j$ for all $j \in \{1, \ldots, k\}$. $\mathbb{I}$ gives the labeling of ordinal categories; the results are the same if such labeling is transformed monotonically. Let $\mathbb{i} = (i_1, \ldots, i_k)$ denote the element of $\mathbb{I}$. Throughout the article $\mathbb{I}, k, n_i$ are fixed unless we explicitly state otherwise. Let $\mathbb{p}$ be a probability distribution on the set $\mathbb{I}$.[1]

Obviously we require $\sum_{\mathbb{i}\in\mathbb{I}} \mathbb{p}(\mathbb{i}) = 1$ and $\mathbb{p}(\mathbb{i}) \geq 0$ $\forall_{\mathbb{i}\in\mathbb{I}}$. Let $\mathbb{p}$ be a probability distribution on $\mathbb{I}$ as above. For $j \in \{1, 2, \ldots, k\}$ we define $p^j(i) :=\sum_{\mathbb{i}\in\mathbb{I} \text{ such that } i_j=i} \mathbb{p}(\mathbb{i})$, $l \in \{1, 2, \ldots, n_j\}$. We notice that p^j is a unidimensional distribution for which we define the cumulative distribution function $P^j(i) = \sum_{h\leq i} p^j(h)$, $j \in \{1, 2, \ldots, k\}$. Let $\bar{P^j}(i) = \sum_{h>i} p^j(h)$, $j \in \{1, 2, \ldots, k\}$ denote the survival function for dimension j-th. In a similar manner we define a multidimensional cumulative distribution function by $\mathbb{P}(\mathbb{i}) = \sum_{\mathbb{h}\preceq\mathbb{i}} \mathbb{p}(\mathbb{h})$ and a multidimensional survival function by $\bar{\mathbb{P}}(\mathbb{i}) = \sum_{\mathbb{h}\succ\mathbb{i}} \mathbb{p}(\mathbb{h})$. Let λ, Λ denote, respectively, the set of all probability distributions and cumulative distribution functions.

For each dimension j we define q_j which is the number of the category for which $P^j(q_j - 1) < \tau$ and $P^j(q_j) \geq \tau$. That is, q_j is τ-quantile in dimension j. In particular, for $\tau = \frac{1}{2}$, we denote m_j as the median. Let $\mathbb{m} = (m_1, \ldots, m_k)$ denote the vector of unidimensional medians and $\mathbb{q} = (q_1, \ldots, q_k)$ similarly for any quantile. Finally, let a multidimensional polarization index be denoted by $\mathcal{P} : \Lambda \to \mathbb{R}$.

Allison and Foster (2004) propose the following bi-polarization ordering to compare distributions of ordinal variables in terms of inequality.

Definition 1. ***Allison and Foster (2004)(AF)***
Let p_1, p_2 be two distributions and let m denote the median. We write $p_1 \precsim_{AF} p_2$ if, and only, if the following conditions hold

(AF1) *p_1, p_2 have a unique and common median m,*

(AF2) *$P_1(i) \leq P_2(i)$ for any $i < m$,*

(AF3) *$P_1(i) \geq P_2(i)$ for any $i \geq m$.*

The interpretation of the AF ordering is intuitive: p_1 is more concentrated (i.e. when there is more probability mass) around the median than p_2. The most equal distribution has

[1]By focusing on probability distributions instead of actual individuals (e.g. see Apouey (2007)), the usual Anonymity and Population Principle axioms are assumed.

all probability mass in one category. It is a partial ordering, that is, when two distributions cross, no ranking of distributions can be obtained.

As already mentioned, AF focus on how far the probability mass is from the median. This is known as increased spread and is equivalent to a median-preserving spread i.e. a transfer of mass from someone below the median to someone above the median that does not change the median itself (Kobus (2015)). Such a transfer moves groups below and above the median further apart from each other. It does not, however, take into account how homogeneous these groups are. In other words, it does not pay attention to the distribution of mass below and above the median. The condition that does so is known as increased bipolarity (Apouey (2007)). It is implemented by the ordering that compares partial sums of cumulative distribution functions below the median and partial sums of survival functions above the median, as developed by Chakravarty and Maharaj (2015). It is implied by AF, but not vice versa.

Definition 2. ***Chakravarty and Maharaj (2015)(CM)***
Let $CS(P, i, m)$ denote the cumulative sum of P up to i-th category, starting from the median m, i.e. $CS(P, i, m) = \Sigma_{i \leq \mathbb{h} \prec m} P(h)$ for $i < m$ and $CS(\bar{P}, i, m) = \Sigma_{m \leq h \leq i} \bar{P}(h)$ for $i \geq m$. Let us note that CS includes values in the median category for $i \geq m$ but not for $i < m$. We write $p_1 \precsim_{CM} p_2$ if, and only, if the following conditions hold

(CM1) *p_1, p_2 have a unique and common median m,*

(CM2) *$CS(P_1, i) \leq CS(P_2, i)$ for any $i < m$,*

(CM3) *$CS(\bar{P}_1, i) \leq CS(\bar{P}_2, i)$ for any $i \geq m$.*

Here distribution p_2 is more concentrated around given categories below and/or above the median than p_1. That is, p_2 is characterized by more homogeneous groups on both sides of the median than p_1.

3 New bidimensional polarization ordering

We now propose a natural multidimensional extension of CM.

Definition 3. ***Multidimensional CM (mCM)***

Let $\mathbb{p}_1, \mathbb{p}_2$ be two probability distributions with a unique and common τ-quantile $\mathbb{q}$. We say that $\mathbb{p}_1 \precsim_{mCM} \mathbb{p}_2$ if, and only, if the following two conditions hold

(1) $CS(\mathbb{P}_1, \mathbb{i}, \mathbb{q}) \leq CS(\mathbb{P}_2, \mathbb{i}, \mathbb{q})$ *for* $\mathbb{i} \prec \mathbb{q}$

(2) $CS(\bar{\mathbb{P}}_1, \mathbb{i}, \mathbb{q}) \leq CS(\bar{\mathbb{P}}_2, \mathbb{i}, \mathbb{q})$ *for* $\mathbb{i} \succeq \mathbb{q}$ *where* $\bar{\mathbb{P}}_1, \bar{\mathbb{P}}_2$ *denote survival functions of* $\mathbb{p}_1$ *and* $\mathbb{p}_2$*, respectively.*

The mCM ordering compares sums of cumulative distribution and survival functions below and above a given quantile. Distribution $\mathbb{p}_2$ is more concentrated around bidimensional categories, for example, around a given education-health category, than distribution $\mathbb{p}_1$. The following family of indices is consistent with this ordering.

Definition 4.

$$\mathcal{P}_{a,b}^{mCM}(\mathbb{p}, \mathbb{q}) = \frac{\Sigma_{\mathbb{i} \prec \mathbb{q}} a_{\mathbb{i}} \mathbb{P}(\mathbb{i}) + \Sigma_{\mathbb{i} \succeq \mathbb{q}} b_{\mathbb{i}} \bar{\mathbb{P}}(\mathbb{i})}{C} \tag{1}$$

where $a_{\mathbb{i}} = \Pi_{j=1}^{k}(n_j - q_j + i_j)$*,* $b_{\mathbb{i}} = \Pi_{j=1}^{k}(n_j + q_j - i_j)$*,* $C = \sup_{\mathbb{q}}\{\min \tau_j \Sigma_{\mathbb{i} \prec \mathbb{q}} \Pi_{j=1}^{k}(n_j - q_j + i_j) + (1 - \max \tau_j)(\Sigma_{\mathbb{q} \preceq \mathbb{i} \prec \mathbb{n}} \Pi_{j=1}^{k}(n_j + q_j - i_j) - 1)\}$.

Weights a, b allow for differential treatment of inequality below and above the chosen quantile. In particular, when $a > b$, then more weight is attached to inequality in the lower tail of the distribution and the reverse holds when $a < b$. Furthermore, the weights and the denominator C have been chosen to normalize the index (please see NORM axiom below). In particular, weights at $\mathbb{P}$ are increasing and at $\bar{\mathbb{P}}$ are decreasing to ensure consistency with mCM. Finally, these multidimensional weights are a multiplication of weights that would hold on each dimension if the measure reduced to a unidimensional index consistent with CM.

We will now show that $\mathcal{P}_{a,b}^{mCM}$ has a number of properties that are desired for a polarization measure. Before we do this, we will define those properties.

CON $\mathcal{P} : \lambda \to \mathbb{R}$ is a continuous function.

NORM The range of $\mathcal{P}$ ($Ran(\mathcal{P})$) is the closed interval $[0, 1]$.

DECOMP There exists $f : Ran(\mathcal{P}) \times Ran(\mathcal{P}) \times (0,1) \to \mathbb{R}$ continuous and strictly increasing with respect to the first two coordinates such that for any $\mathbb{p}_1, \mathbb{p}_2 \in \lambda$, $\alpha \in (0,1)$

$$\mathcal{P}(\alpha\mathbb{p}_1 + (1-\alpha)\mathbb{p}_2) = f(\mathcal{P}(\mathbb{p}_1), \mathcal{P}(\mathbb{p}_2), \alpha),$$

where $\alpha\mathbb{p}_1 + (1-\alpha)\mathbb{p}_2$ is a weighted sum of probability distributions, i.e. if $\mathbb{p}_1$ assigns mass $\mathbb{p}_1(\mathbb{i})$ to category $\mathbb{i}$ and $\mathbb{p}_2$ assigns mass $\mathbb{p}_2(\mathbb{i})$, then the probability mass attributed to $\mathbb{i}$ in $\alpha\mathbb{p}_1 + (1-\alpha)\mathbb{p}_2$ is $\alpha\mathbb{p}_1(\mathbb{i}) + (1-\alpha)\mathbb{p}_2(\mathbb{i})$.

ADDSEP $\mathcal{P}(\mathbb{p}) = f(\mathcal{P}(p^1), \mathcal{P}(p^2), \dots, \mathcal{P}(p^k))$, where $f(x) = \sum_{j=1}^{k} f_j(x_j)$.

CATADD Let $\mathbb{p}_2, \mathbb{p}_3$ be such distributions that $p_2^j(1) = 0$, $\mathbb{p}_2((i_1, \dots, i_j + 1, \dots, i_k)) = \mathbb{p}_1((i_1, \dots, i_j, \dots, i_k))$ for $1 \le i_j \le n_j$, $p_3^j(n_j + 1) = 0$, $\mathbb{p}_3((i_1, \dots, i_j, \dots, i_k)) = \mathbb{p}_1((i_1, \dots, i_j, \dots, i_k))$ for $1 \le i_j \le n_j$ and let $\mathbb{q}_1, \mathbb{q}_2, \mathbb{q}_3$ be obtained in the same way then $\mathcal{P}(\mathbb{p}_1) \le \mathcal{P}(\mathbb{q}_1) \iff \mathcal{P}(\mathbb{p}_2) \le \mathcal{P}(\mathbb{q}_2) \iff \mathcal{P}(\mathbb{p}_3) \le \mathcal{P}(\mathbb{q}_3)$.

SLIDE Let $\mathbb{p}_1$ be such distribution that $p_1^j(1) = 0$, let $\mathbb{p}_2((i_1, \dots, i_j, \dots, i_k)) = \mathbb{p}_1((i_1, \dots, i_j + 1, \dots, i_k))$ for $i_j < n_j$ and $p_2^j(n_j) = 0$, then $\mathcal{P}(\mathbb{p}_1) \le (\ge)\mathcal{P}(\mathbb{q}) \iff \mathcal{P}(\mathbb{p}_2) \le (\ge)\mathcal{P}(\mathbb{q})$.

CON is a natural technical assumption. NORM means that the index achieves the lowest value equal to zero for the most equal distribution, namely, the distribution such that all mass is in one category. The index admits the highest value equal to one for the most unequal distribution, in other words, bi-polarized distribution. DECOMP means that the index is decomposable by population subgroups, namely, that it is a function of the weighted mean of the indices' values in the subgroups, with weights corresponding to population size of the subgroups (Shorrocks (1984), Kobus and Milos (2012)). In addition, ADDSEP states that the index is an additive function of the unidimensional indices. CATADD involves an operation which adds an empty category to one dimension either below the lowest category or above the highest category, whereas SLIDE moves the probability mass to empty categories so the chosen quantiles of two distributions agree (provided that there are enough empty categories). These two axioms allow for comparisons of distributions with different quantiles (Sarkar and Santra (2020)).

Theorem 1. $\mathcal{P}_{a,b}^{mCM}$ *fulfills CON, NORM, DECOMP, SLIDE. It also fulfills CATADD (with respect to attribute j) if* $n_j = \inf\{i|\bar{P}^j(i) = 0\} - \sup\{i \geq 0|P^j(i) = 0\}$.

4 Statistical inference

In this section we present how to estimate $\mathcal{P}_{a,b}^{mCM}$ and we establish its large sample distribution. We consider the following setting. Each individual chooses one and only one answer in every category, which gives a unique response in the set of all states $\mathbb{I}$. We will assume that there are N independent responses $(N_{1\ldots 11}, N_{1\ldots 12}, \ldots, N_{n_1\ldots n_{k-1}n_k})$ defining the vector of frequencies $\hat{\mathbb{p}}$. In such a model, $\mathbb{I}$ has joint multinomial distribution with parameters $p = (p_{1\ldots 11}, p_{1\ldots 12}, \ldots, p_{n_1\ldots n_{k-1}n_k})$ and N. Empirical probability distribution is the following

$$\hat{p}_{i_1 i_2\ldots i_k} = \frac{1}{N}\Sigma_{t=1}^{N}\mathbf{1}(I_1^t = i_1, I_2^t = i_2, \ldots, I_k^t = i_k) = \frac{N_{i_1 i_2 \ldots i_k}}{N},$$

In Formby et al. (2004) it is shown that multinomial distributions are asymptotically distributed as normal with mean p and covariance matrix Σ, where (for $\mathbb{I} = \{1, 2, \ldots, n_1\} \times \{1, 2, \ldots, n_2\} \times \cdots \times \{1, 2, \ldots, n_k\}$) we assume the following notation

$$p = \begin{pmatrix} p_{1\ldots 1} \\ p_{1\ldots 12} \\ \vdots \\ p_{n_1\ldots n_{k-1}n_k} \end{pmatrix},$$

$$\Sigma = \begin{pmatrix} p_{1\ldots 11}(1 - p_{1\ldots 11}) & -p_{1\ldots 11}p_{1\ldots 12} & \cdots & -p_{1\ldots 11}p_{n_1\ldots n_{k-1}n_k} \\ -p_{1\ldots 12}p_{1\ldots 11} & p_{1\ldots 12}(1 - p_{1\ldots 12}) & \cdots & -p_{1\ldots 12}p_{n_1\ldots n_{k-1}n_k} \\ \vdots & \vdots & \ddots & \vdots \\ -p_{n_1\ldots n_{k-1}n_k}p_{1\ldots 11} & -p_{n_1\ldots n_{k-1}n_k}p_{1\ldots 12} & \cdots & p_{n_1\ldots n_{k-1}n_k}(1 - p_{n_1\ldots n_{k-1}n_k}) \end{pmatrix}.$$

That is, lth row/column corresponds to $(i_1, i_2, \ldots i_k)$ category, where $l = i_k + \Sigma_{j=1}^{k-1}(i_j - 1)n_{j+1}n_{j+2}\ldots n_k$. We have $i_k = l \bmod n_k + n_k\mathbf{1}_{n_k|l}$, $l_k = \frac{l-i_k}{n_k}$ and $i_j - 1 = l_{j+1} \bmod n_j$, $l_j = \frac{l_{j+1}-(i_j-1)}{n_j}$ for $1 \leq j < k$. For example, for $k = 3$ and $(n_1, n_2, n_3) = (6, 5, 8)$, the 28th category is $28 = 4 + 3 \times 8$ i.e. $(0 + 1, 3 + 1, 4) = (1, 4, 4)$. The multivariate, multinomial distributions of two samples can be compared by stacking the probabilities in respective probability vectors $\hat{\mathcal{E}} = \hat{\mathbb{p}} - \hat{\mathbb{q}}$. Under the null hypothesis of homogenous distributions,

$$\hat{\mathcal{E}} \xrightarrow{d} \mathcal{N}\left(0, \frac{N_{\mathbb{p}} + N_{\mathbb{q}}}{N_{\mathbb{p}} N_{\mathbb{q}}} \Sigma\right). \tag{2}$$

Using Continuous Mapping Theorem we get that under appropriate assumptions $F(\hat{\mathbb{P}})$ will be a consistent estimator of $F(\mathbb{P})$. Let us define $n_1 n_2 \dots n_k$-dimensional Jacobian (horizontal) vector of the transformation F,

$$J = \left(\frac{\partial F}{\partial \mathbb{P}_{1\dots 11}}, \frac{\partial F}{\partial \mathbb{P}_{1\dots 12}}, \dots \frac{\partial F}{\partial \mathbb{P}_{n_1 \dots n_{k-1} n_k}}\right).$$

N independent responses $(N_{1\dots 11}, N_{1\dots 12}, \dots, N_{n_1 \dots n_{k-1} n_k})$ are jointly distributed from a multinomial distribution with parameters N and p and such that $\mathrm{cov}(\sqrt{n}\hat{\mathbb{p}}) = \Sigma(\mathbb{p})$, where $\Sigma(\mathbb{p})$ is a finite positive semi-definite matrix.

Theorem 2. *For multinomial distribution $\mathbb{P}$ and function F which is continuously differentiable at $\mathbb{P}$ and does not involve total number of observations N, we have that*

$$\sqrt{n}(F(\hat{\mathbb{P}}) - F(\mathbb{P})) \xrightarrow{d} \mathcal{N}(0, JL\Sigma L^T J^T)$$

where $J = J(\mathbb{P})$ is Jacobian vector evaluated at $\mathbb{P}$, and $L = (l_{ij})_{1 \le i,j \le n_1 n_2 \dots n_k}$ is such that $l_{ij} = 1$ if $j \equiv (j_1, j_2 \dots j_k) \preceq (i_1, i_2, \dots i_k) \equiv i$ and 0 otherwise where $\preceq$ is lexicographical order.[2]

A consistent estimator for Σ is

$$\hat{\sigma}_{(i_1 i_2 \dots i_k)(i_1 i_2 \dots i_k)} = \frac{N_{i_1 i_2 \dots i_k}}{N}(1 - \frac{N_{i_1 i_2 \dots i_k}}{N})$$

and

$$\hat{\sigma}_{(i_1 i_2 \dots i_k)(j_1 j_2 \dots j_k)} = -\frac{N_{i_1 i_2 \dots i_k} N_{j_1 j_2 \dots j_k}}{N^2} \text{ if } i_t \neq j_t \text{ for some } 1 \le t \le k.$$

It follows that standard error σ is equal to $\sigma = \sqrt{\frac{J(\hat{\mathbb{P}}) L \hat{\Sigma} L^T J(\hat{\mathbb{P}})^T}{n}}$.. The Jacobian vector for $F = \mathcal{P}^{mCM}_{a,b}$ is given by $J_i = \frac{a_{\mathbb{i}}}{C}$ for $i \equiv \mathbb{i} \prec \mathbb{m}$, $J_i = \frac{b_{\mathbb{i}}}{C}$ for $i - n_1 n_2 \dots n_k \equiv \mathbb{h} \succeq \mathbb{m}$ and $J_i = 0$ otherwise, where $a_{\mathbb{i}}, b_{\mathbb{i}}, C$ are defined as in Definition 1.

[2] The result is unchanged when $\mathbb{m}$ is replaced by τ.

5 Empirical application

We use data from Survey of Health, Ageing and Retirement in Europe (SHARE), Wave 7, which was collected in 2017. SHARE is a multi-country representative survey of non-institutionalized individuals aged 50. It contains rich data on health and socioeconomic status. We measure socioeconomic inequalities in health. Socioeconomic status is proxied by educational attainment and health by self-reported health status. The specific indicators are presented in Table 1.

Table 1: Ordinal indicators of health and education

Dimension	Indicator	Level	Construction
Education	Highest educational level attained	1	Pre-primary education
		2	Primary education
		3	Lower secondary education
		4	(Upper) secondary education
		5	Post-secondary non-tertiary education
		6	First stage of tertiary education
		7	Second stage of tertiary education (research)
Health	Self-reported health status	1	Poor
		2	Fair
		3	Good
		4	Very good
		5	Excellent

Source: Wave 7 of SHARE.

For each pair of 27 countries that took part in the survey we search for a bidimensional dominance cCM at either $q = .25$, $q = .5$ or $q = .75$. Dominance holds in around 8% of all pairwise comparisons. In fact, there is more dominance because in case there was dominance for, say, $q = .25$ and $q = .5$ we choose a lower quantile. For these dominances, we report the values of indices and test statistics related to $\mathcal{P}_{a,b}^{mCM}$ and a measure proposed by Kobus and Kurek (2018). $\mathcal{P}_{a,b}^{mAF2}$ uses cumulative distribution and survival function to compare distributions below and above a given quantiles, but not their sums, so it does not take into account increased bipolarity. Therefore, the comparison of two measures allows us to detect the impact of increased bipolarity on the observed inequality.

In Table 2 we report the values of two measures for the list of countries where dominance holds. The first and third column report values for the first country in the comparison, and the second and fourth column report values for the second country. We can see that the values of $\mathcal{P}_{a,b}^{mCM}$ are in each case significantly higher than the values of $\mathcal{P}_{1,1}^{mAF2}$ indicating that accounting for group homogeneity significantly increases inequality score, e.g. 0.34 vs. 0.45 for AUT (first row of Table 2). The joint distribution of education and health becomes much more polarized when increased bipolarity is considered too.

Table 2: The values of polarization indices

Dominance	$\mathcal{P}_{1,1}^{mAF2}$		$\mathcal{P}_{1,1}^{mCM}$	
AUT-FIN	0.34(0.0077)	0.39(0.0101)	0.45(0.0093)	0.50(0.0114)
ISR-AUT	0.25(0.0052)	0.18(0.0034)	0.32(0.0061)	0.24(0.0042)
ISR-BGR	0.25(0.0052)	0.20(0.0047)	0.32(0.0061)	0.25(0.0054)
ISR-BEL	0.32(0.0069)	0.27(0.0042)	0.43(0.0086)	0.38(0.0052)
CYP-CZE	0.26(0.0067)	0.15(0.0026)	0.34(0.0080)	0.19(0.0031)
ISR-CYP	0.32(0.0069)	0.31(0.0092)	0.43(0.0086)	0.45(0.0120)
CYP-MLT	0.46(0.0139)	0.29(0.0096)	0.60(0.0169)	0.41(0.0124)
ESP-CZE	0.34(0.0048)	0.19(0.0034)	0.43(0.0055)	0.25(0.0041)
FRA-CZE	0.23(0.0041)	0.15(0.0026)	0.28(0.0046)	0.19(0.0031)
FIN-CZE	0.29(0.0067)	0.17(0.0031)	0.39(0.0082)	0.25(0.0040)
ISR-CZE	0.32(0.0069)	0.17(0.0031)	0.43(0.0086)	0.25(0.0040)
LUX-CZE	0.23(0.0060)	0.15(0.0026)	0.29(0.0071)	0.19(0.0031)
SWE-CHE	0.28(0.0053)	0.20(0.0046)	0.38(0.0063)	0.27(0.0056)
ISR-DEU	0.28(0.0057)	0.18(0.0031)	0.40(0.0074)	0.26(0.0041)
DNK-PRT	0.36(0.0054)	0.16(0.0110)	0.47(0.0066)	0.27(0.0153)
DNK-SVK	0.36(0.0054)	0.20(0.0046)	0.47(0.0066)	0.28(0.0057)
ISR-EST	0.25(0.0052)	0.16(0.0024)	0.32(0.0061)	0.23(0.0032)
SWE-EST	0.24(0.0039)	0.16(0.0024)	0.31(0.0047)	0.23(0.0032)
FRA-ITA	0.26(0.0044)	0.19(0.0030)	0.41(0.0058)	0.31(0.0042)
ISR-ITA	0.28(0.0057)	0.19(0.0030)	0.40(0.0074)	0.31(0.0042)
ISR-LTU	0.34(0.0071)	0.23(0.0053)	0.47(0.0090)	0.34(0.0071)
ISR-LVA	0.42(0.0109)	0.28(0.0090)	0.52(0.0124)	0.36(0.0101)
ISR-MLT	0.28(0.0057)	0.19(0.0055)	0.40(0.0074)	0.33(0.0081)
ISR-SVN	0.34(0.0071)	0.21(0.0042)	0.47(0.0090)	0.31(0.0054)
SWE-MLT	0.31(0.0049)	0.19(0.0055)	0.43(0.0061)	0.33(0.0081)
ROU-PRT	0.15(0.0032)	0.11(0.0075)	0.25(0.0046)	0.19(0.0105)

Notes: Own calculations based on SHARE data (Wave 7).

While increased bipolarity strongly affects the assessment of inequality, it does not, however, change the countries' ranking, as can be seen in Table 3, where we report test statistics from comparisons of indices $\mathcal{P}_{a,b}^{mAF2}$ and $\mathcal{P}_{a,b}^{mCM}$. The sign of a test statistic is the same in $\mathcal{P}_{a,b}^{mCM}$ as in $\mathcal{P}_{a,b}^{mAF2}$ for each pairwise comparison. This holds for a various weights: (1,1), (10,1) and (1,10).

Table 3: The values of test statistic

Dominance	$\mathcal{P}_{a,b}^{mAF2}$			$\mathcal{P}_{a,b}^{mCM}$		
A-B	1,1	10,1	1,10	1,1	10,1	1,10
AUT-FIN	-3.83	[-1.52]	-6.41	-3.64	[-0.70]	-6.16
ISR-AUT	10.45	8.06	5.67	10.21	7.66	5.59
ISR-BGR	7.03	2.84	6.97	7.58	2.74	7.11
ISR-BEL	5.92	7.89	2.50	5.58	7.69	2.21
CYP-CZE	16.22	9.08	13.61	17.56	8.38	14.35
ISR-CYP	[0.88]	[-1.75]	[1.27]	[-1.15]	-2.41	[-0.18]
CYP-MLT	10.00	5.07	8.30	9.32	4.76	7.51
ESP-CZE	25.14	27.29	5.33	24.91	25.96	5.56
FRA-CZE	16.64	11.19	12.06	16.42	10.52	11.46
FIN-CZE	16.00	9.62	12.80	15.38	9.67	11.73
ISR-CZE	20.15	11.64	15.91	19.25	11.47	14.74
LUX-CZE	12.36	7.28	9.52	12.66	7.26	9.19
SWE-CHE	12.59	10.26	7.71	12.28	10.09	6.19
ISR-DEU	15.78	16.62	13.01	16.71	15.78	13.36
DNK-PRT	16.24	[-1.30]	20.30	12.08	-2.34	15.99
DNK-SVK	22.76	7.74	21.23	22.26	7.52	20.43
ISR-EST	15.29	10.51	10.51	12.16	10.40	7.05
SWE-EST	16.51	6.31	16.04	12.69	6.70	11.61
FRA-ITA	13.62	11.52	13.34	13.81	10.33	13.37
ISR-ITA	14.63	12.33	13.38	11.21	9.43	10.19
ISR-LTU	12.66	6.95	10.52	11.33	6.54	9.07
ISR-LVA	9.64	2.25	14.50	10.17	2.26	13.28
ISR-MLT	10.91	10.60	10.04	6.90	7.52	6.29
ISR-SVN	15.79	5.53	14.55	15.85	5.32	14.39
SWE-MLT	15.08	8.32	15.49	10.03	3.50	10.83
ROU-PRT	4.85	2.12	4.29	5.10	[0.93]	4.88

Notes: Own calculations based on SHARE data (Wave 7). [..] denotes insignificant values of a test statistic.

Acknowledgement

This research was funded by National Science Centre, Poland, grant number 2017/26/M/HS4/00905.

References

Abul Naga R H, Stapenhurst Ch. Estimation of inequality indices of the cumulative distribution function. *Econ Lett* **2008**, *130*, 109-112.

Abul Naga R H, Yalcin T. Inequality measurement for ordered response health data. *JHE* **2008**, *27(6)*, 1614-1625.

Allison R A, Foster J E. Measuring health inequality using qualitative data. *JHE* **2004**, *23(3)*, 505-524.

Apouey B. Measuring health polarization with self-assessed health data. *HE* **2007**, *16*, 875-894.

Apouey B, Silber J. Inequality and bi-polarization in socioeconomic status and health: ordinal approaches. In *Health and Inequality*; P.R.Dias, O. O'Donnell, Eds.; Emerald Group Publishing Limited: Bingley, UK, 2013; Vol. 21, Research on Economic Inequality.

Bond T, Lang K. 2019. The Sad Truth About Happiness Scales. *JPE* **2019**, *127(4)*., :1629–1640.

Chakravarty S R, Maharaj B. Generalized Gini polarization indices for an ordinal dimension of human well-being. *IJET* **2015**, *11(1)*, 231-246.

Cowell F, Flachaire E. Inequality with ordinal data. *ECMA* **2017**, *84(334)*, 290-321.

Esteban J M, Ray D. On the measurement of polarization. *ECTA* **1994**, *62(4)*, 819-851.

Fleurbaey M, Schokkaert E. Equity in Health and Health Care. In *Handbook of Health Economics*: M. V. Pauly, T. G. McGuire, P. P. Barros, Eds.; Elsevier, North-Holland; Amsterdam and Oxford; Vol. 2, 1003–1092.

Formby J, Smith J, Zheng B. Mobility measurement, transition matrices and statistical inference. *J. Econom.* **2004**, *120*, 181–205.

Gravel N, Magdalou B, Moyes P. Ranking distributions of an ordinal attribute. *Econ. Theory* **2021**, *71(1)*, 33-80.

Kobus M, Miłoś P. Inequality decomposition by population subgroups for ordinal data. *JHE* **2012**, *31(1)*, 15-21.

Kobus M. Polarization measurement with ordinal data. *JoEI* **2015**, *13(2)*, 275-297.

Kobus M, Kurek R. Multidimensional polarization for ordinal data. *JoEI* **2018**, doi: 10.1007/s10888-018-9402-1.

Kobus M, Polchlopek O, Yalonetzky G. Inequality and welfare in quality of life among OECD countries: non-parametric treatment of ordinal data. *Soc. Indic. Res.* **2018**, doi:10.1007/s11205-018-1962-8.

Lazar A, Silber J. On the cardinal measurement of health inequality when only ordinal information is available on individual health status. *HE* **2013**, *22(1)*, 106-113.

Lv G, Wang Y, Xu Y. On a new class of measures for health inequality based on ordinal data. *JoEI* **2015**, *13(3)*, 465-477.

Madden D. Ordinal and cardinal measures of health inequality: an empirical comparison. *HE* **2010**, *19(2)*, 243–25.

Makdisi P, Yazbeck M. Measuring socioeconomic health inequalities in the presence of multiple categorical information. *JHE* **2014**, *34*, 84-95.

Mendelson H. Quintile-preserving spread. *JET* **1987**, *42(2)*, 334-351.

Sarkar S, Santra S. Extending the approaches to polarization ordering of ordinal variables. *JoEI* **2020**, *18*, 421–440.

Shorrocks A F. Inequality decomposition by population subgroups. *Econometrica* **1984**, *52(6)*,1369-85.

van Doorslaer E, Koolman X. Explaining the differences in income-related health inequalities across European countries. *HE* **2004**, *13(7)*, 609-628.

van Doorslaer E, Wagstaff A, Bleichrodt H, Calonge S, Gerdtham U, Gerfin M, Geurts J, Gross L, Hakkinen U, Leu R. Income-related inequalities in health: some international comparisons. *JHE* **1997**, *16(1)*, 93–112.

Appendix

Proof of Theorem 1. First of all, as $\mathcal{P}^{mCM}_{a,b}$ is consistent with mCM ordering, because it is increasing with $CS(\mathbb{P})$, which follows from the fact that weights at $\mathbb{P}(\mathbb{i})$ are increasing and at $\bar{\mathbb{P}}(\mathbb{i})$. The fact that $\mathcal{P}^{mCM}_{a,b}$ fulfills CON and DECOMP is obvious since it is a linear function. Let us check the NORM axiom. Let $\mathbb{p}$ and $\mathbb{q}$ be the best and worst distributions according to AF; $\mathbb{q}$ has all probability mass in the median category, while $\mathbb{p}$ has half the mass minus infinitesimal mass in the first category, and infinitesimal mass in the median and half of the mass in the last category. We have

$$\mathcal{P}^{mCM}_{a,b}(\mathbb{q}) = \frac{0}{C} = 0$$

and

$$\sup_{\mathbb{m}} \mathcal{P}^{mCM}_{a,b}(\mathbb{p}) = \frac{\sup_{\mathbb{m}}\{\frac{1}{2}(a\Sigma_{\mathbb{i}\prec\mathbb{m}}\Pi^k_{j=1}(n_j - m_j + i_j) + b\Sigma_{\mathbb{m}\preceq\mathbb{i}\prec\mathbb{n}}\Pi^k_{j=1}(n_j + m_j - i_j))\}}{C} = \frac{C}{C} = 1.$$

We note that $\mathbb{i} + 1 = (i_1, \dots, i_j + 1, \dots, i_k)$. Let us now take distributions $\mathbb{p}, \mathbb{q}$ such that $p^j(1) = 0$, $\mathbb{q}_{\mathbb{i}} = \mathbb{p}_{\mathbb{i}+1}$ and $q^j(n_j) = 0$. We know that

$$a^{\mathbb{p}}_{\mathbb{i}+1} = a(n_j - m_j + i_j + 1)\Pi^k_{l\neq j}(n_l - m_l + i_l) = a(n_j - (m_j - 1) + i_j)\Pi^k_{l\neq j}(n_l - m_l + i_l) = a^{\mathbb{q}}_{\mathbb{i}},$$

$$b^{\mathbb{p}}_{\mathbb{i}+1} = b(n_j + m_j - i_j - 1)\Pi^k_{l\neq j}(n_l + m_l - i_l) = b(n_j + (m_j - 1) - i_j)\Pi^k_{l\neq j}(n_l + m_l - i_l) = b^{\mathbb{q}}_{\mathbb{i}}.$$

Finally

$$
\begin{aligned}
\mathcal{P}^{mCM}_{a,b}(\mathbb{p}) &= \frac{\Sigma_{\mathbb{i}\prec\mathbb{m}} a^{\mathbb{p}}_{\mathbb{i}}\mathbb{P}(\mathbb{i}) + \Sigma_{\mathbb{i}\succeq\mathbb{m}} b^{\mathbb{p}}_{\mathbb{i}}\bar{\mathbb{P}}(\mathbb{i})}{C} = \\
&= \frac{\Sigma_{\mathbb{i}+1\prec\mathbb{m}} a^{\mathbb{p}}_{\mathbb{i}+1}\mathbb{P}(\mathbb{i}+1) + \Sigma_{\mathbb{i}+1\succeq\mathbb{m}} b^{\mathbb{p}}_{\mathbb{i}+1}\bar{\mathbb{P}}(\mathbb{i}+1)}{C} = \\
&= \frac{\Sigma_{\mathbb{i}\prec\mathbb{m}-1} a^{\mathbb{q}}_{\mathbb{i}}\mathbb{Q}(\mathbb{i}) + \Sigma_{\mathbb{i}\succeq\mathbb{m}-1} b^{\mathbb{q}}_{\mathbb{i}}\bar{\mathbb{Q}}(\mathbb{i})}{C} = \mathcal{P}^{mCM}_{a,b}(\mathbb{q})
\end{aligned}
$$

□

Proof of Theorem 2. For $k = 2$ and $(n_1, n_2) = (3, 3)$, L is the following

$$
L = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{pmatrix}.
$$

In general, let L^t be the square matrix of dimension t satisfying $l^t_{ij} = 1$ if $i \geq j$ and $l^t_{ij} = 0$ otherwise. We define L recursively by setting $L^{n_{k-t}}_{t+1} = L^{n_{k-t}} \otimes L^{n_{k-t+1}}_t$, where $\otimes$ denotes Kronecker product (for any $a \times b$ matrix A and $c \times d$ matrix B, $A \otimes B$ is $ac \times bd$ matrix C such that $c_{(b(i-1)+k)(d(j-1)+l)} = a_{ij}b_{kl}$). Next, let $L^{n_k}_1 = L^{n_k}$ and finally, $L = L^{n_1}_k$.

From Central Limit Theorem, we know that $\sqrt{n}(\hat{\mathbb{p}} - \mathbb{p}) \xrightarrow{d} \mathcal{N}(0, \Sigma)$. It is straightforward to check that for

$$
\hat{\mathbb{P}} = L\hat{\mathbb{p}}, \mathbb{P} = L\mathbb{p}
$$

we have $\sqrt{n}(\hat{\mathbb{P}} - \mathbb{P}) \xrightarrow{d} \mathcal{N}(0, L\Sigma L^T)$. From the Law of Large Numbers $\hat{\mathbb{P}}$ converges in probability to $\mathbb{P}$ and from Continuous Mapping Theorem we see that $J(\hat{\mathbb{P}})$ converges in probability to $J(\mathbb{P})$ (due to continuous differentiability). Finally, from the delta method we obtain

$$
\sqrt{n}(F(\hat{\mathbb{P}}) - F(\mathbb{P})) \xrightarrow{d} \mathcal{N}(0, JL\Sigma L^T J^T).
$$

□