An index of multivariate polarization*

Sorawoot Srisuma† Alberto Vesperoni‡

March 2015

Abstract

We define and axiomatically characterize a multivariate index of polarization which can be interpreted as a predictor of social conflict. We consider a distribution of multiple attributes in a population, such as individual income, age, education, etc. Assuming that there is some degree of antagonism between each pair of individuals that is determined by the distance of their attributes, we restrict our attention to a class of measures which are expressions of the expected antagonism of a randomly matched pair. We characterize our polarization index uniquely from this class via two axioms, which respectively require social conflict to increase when individuals become more extremist, and when homogeneous groups are formed. We show that our index shares several properties with well-known polarization measures in the univariate setting, while it differs fundamentally from well-known multivariate inequality measures. A natural estimator for our index takes the form of a second order U-statistic and has well-behaved statistical properties. We illustrate the distinction between our polarization measure and inequality measures by analyzing data from the British Household Panel Survey, focusing on income and highest level of education across England, Wales, Scotland and Northern Ireland.

Keywords: conflict; polarization; index; multivariate analysis.
JEL classification: C43; D63; D74.

1 Introduction

The idea of polarization is pervasive in economics and other social sciences. When one thinks of a highly polarized distribution, a picture of two large and homogeneous groups opposite to each other comes to mind. This situation should typically be problematic, inducing tension and incompatibilities within the population. Although powerful and evocative, to our knowledge there is no general consensus of what exactly polarization means beyond this idea.

*PRELIMINARY DRAFT, please do not circulate. We are grateful to Ronny Razin for useful advice and suggestions, and to Nattavudt Powdthavee for help with the data.

†University of Surrey E-mail: s.srisuma@surrey.ac.uk
‡University of Siegen E-mail: alberto.vesperoni@gmail.com

Polarization is inevitably related to inequality, but there are some fundamental differences between the two concepts. Inequality measurement has a long tradition and it is typically based on concepts of normative nature, which are often derived from a social welfare function. (See citations below.) In contrast, the formalization of polarization is relatively recent and has a strong empirical orientation, which necessarily creates more ambiguities with respect to a normative approach. In their seminal work Esteban and Ray (1994) put forward a class of measures that represent a certain average of inter-group differences in a population. They impose axioms based on empirically motivated conditions of what a polarization measure should satisfy, and characterize a class of univariate measures that fulfill these axioms. The goal of this paper is to develop a multivariate index of polarization within this tradition.

A crucial aim of polarization measurement is to predict social conflict. While pure economic inequality between individuals is a poor predictor of social tension and institutional failure, there is ample evidence that economic divergences between ethno-cultural groups are a leading cause of inefficiencies, leading to violent crime (Blau and Blau (1982)), low public good provision (Baldwin and Huber (2010)) and civil war (Cederman et al. (2011)). In these studies individuals are partitioned into homogeneous ethno-cultural groups, and measures of inequality across aggregate incomes are provided. These measures are known as indices of horizontal inequality. Within this approach ethno-cultural traits are necessarily categorical, as they are used to partition the population into groups. Then, by construction, horizontal inequality overlooks the magnitude of ethno-cultural distances between groups, which is also relevant as argued in Fearon (2003), Desmet et al. (2012) and Esteban et al. (2012). Moreover, while Fearon (2003) and Desmet et al. (2012) focus on measures of ethno-cultural inequality, Esteban et al. (2012) show that the univariate index of polarization of Esteban and Ray (1994) calculated with ethno-cultural data systematically outperforms inequality measures as predictor of social conflict.[1] The polarization measure of Esteban and Ray (1994) is built as a function of the expected antagonism between two randomly matched groups. Although similar to the Gini index, it crucially differs from inequality measurement as it fulfills the typical empirically-motivated properties of a polarization measure.

We now summarize three crucial lessons from the evidence above. To develop a better predictor of social conflict one should: (*i*) define a multivariate measure which focuses on the correlation across economic and ethno-cultural attributes; (*ii*) quantify economic as well as ethno-cultural distances and build the measure so that it can take all these distances into account; (*iii*) define the measure so that it is represents the expected antagonism in the population and it satisfies the typical properties of a polarization measure. We motivate our goal of defining an index of multivariate polarization following the three lessons above.

We consider a distribution of multiple attributes in a population, such as individual income, age, education, etc. We assume that there is some degree of antagonism between each pair of individuals, which is determined by the distance of their attributes. Given this, we restrict our attention to a class of measures which are expressions of the *expected antagonism* between a randomly matched pair. Despite the strong similarity with the framework of Esteban and

[1]The measure of Esteban and Ray (1994) is extended to accommodate for continuous density functions in Duclos et al. (2004).

Ray (1994), we crucially differ from them in that we rule out the effects of group size. As we extensively explain in Section 2, we do so because groups may encounter collective action problems. Moreover, taking a neutral attitude towards group size guarantees the uniqueness of our index. In contrast, Esteban and Ray (1994) characterize a continuum of parametrized measures which can lead to ambiguous findings in empirical studies.[2]

In Section 3 we focus on a specific index from our class of measures of expected antagonism, showing that it is the only index from this class to fulfill two properties (or *axioms*) which are strongly linked to the idea of social conflict. The first axiom requires social conflict to increase when equal amounts of individuals are shifted from the center to each side of a distribution. In other words conflict should be higher when moderate individuals become more extremist by "taking sides". Conversely, the second axiom demands that social conflict should be higher when groups whose members have perfectly homogeneous attributes are formed, even if some distances decrease in this process. The intuition is that the antagonism between two individuals is high if all their attributes are different, while it is low whenever some of their attributes are alike. From a purely technical view point, the first axiom imposes our index to be weakly convex and the second axiom to be weakly concave and null at the origin. Then, they jointly require the index to be linear.

Our axioms are in line with well-established approaches in the literature. We validate our index as polarization measure in the tradition of Esteban and Ray (1994) in Section 4, where we show that the index is exclusively maximized by distributions which divide the population into two homogeneous and equally sized groups with opposite attributes, i.e., by *most polarized distributions*. Discussing other properties, we argue that our index is among the conservative ones in the literature, in the sense that it tends to give a "lower bound" of polarization. At the same time our framework differs fundamentally from inequality measurement. Multivariate inequality indices typically satisfy some majorization criteria, such as uniform majorization and correlation increasing majorization, which are defined to generalize the Pigou-Dalton Transfer Principle (we refer the reader to Tsui (1999) for further discussion on multivariate inequality measures). To distinguish our approach from inequality measurement, we show that our polarization index does not satisfy either of these majorization criteria.

We show in Section 5 that our index has a simple geometrical interpretation. For the case with two attributes only, the index is proportional to the inner product of the two vectors of distances of attributes of all pairs of individuals. More generally, the index can be seen as the length of a vector whose elements are the average distances of all pairs of individuals for each attribute. In Section 5 we also rewrite our index in terms of density functions, which is useful for comparisons with related models in the literature. We discuss the univariate polarization measures in Esteban and Ray (1994), Duclos et al. (2004) and Anderson et al. (2012).[3] Then, we focus on the index in Gigliarano and Mosler (2009), which to our knowledge is the only

[2]For instance Duclos et al. (2004) rank countries by their extension of the univariate index of Esteban and Ray (1994) calculated with income data and find that the ranking changes with different permissible parameter values.

[3]There are other measures of univariate polarization in the literature, see for instance Wang and Tsui (2000) and Foster and Wolfson (2010). As argued in Section 5, we focus on Esteban and Ray (1994), Duclos et al. (2004) and Anderson et al. (2012) as they are the closest to our framework.

multivariate measure of polarization in the literature. Lastly, we consider the multivariate inequality index in Koshevoy and Mosler (1997), as it belongs to the same class of measures of expected antagonism as ours.

We develop an estimator of our index in Section 6 and characterize its statistical properties. Any index from the class of measures of expected antagonism can be understood as a mathematical expectation of some function of differences in attributes of a randomly matched pair of individuals. If the individuals are drawn from the same distribution, then it is natural to think of the index as a feature of such distribution. We formalize this idea by modeling the inference of the index from a random sample of any distribution. We propose to estimate our polarization measure using a second order U-statistic. We show that our estimator is consistent and has asymptotic normal distribution under very weak conditions.

In Section 7 we illustrate the distinctions between our polarization index and two well-known multivariate inequality measures empirically. We use a 7-year dataset from the British Household Panel Survey, and we compute all the three measures with data on individual income and highest level of education as the relevant attributes. We show that, throughout the time period, there are systematic differences between the measurement of polarization and inequality across England, Wales, Scotland and Northern Ireland.

To summarize, the paper develops as follows. In Section 2 we define our class of indices of expected antagonism, while in Section 3 we take an axiomatic approach to uniquely characterize an index from this class. In Section 4 we discuss properties of this index which validate it as a measure of polarization, they set it apart from inequality measurement, and show that it is a conservative measure of polarization. We consider different representations of this index in Section 5, focusing on an intuitive geometrical interpretation based on vectors of distances of attributes. In the same Section we also compare our index with related measures in the literature. We develop the statistical theory to perform inference from a random sample in Section 6, while in Section 7 we provide an empirical illustration to highlight the significant differences between our index and some measures of inequality. Section 8 concludes. All proofs are in Appendix.

2 Model

Consider a population $N = \{1, \ldots, n\}$ with $n \geq 4$. Each individual $i \in N$ is associated with attributes of various *types*. Let $T = \{1, \ldots, \tau\}$ be the set of all types, where $\tau \geq 2$. For any individual $i \in N$ denote by $x_i^t \in [0,1]$ its attribute of type $t \in T$ and $x_i = (x_i^1, \ldots, x_i^\tau) \in [0,1]^\tau$ its *attribute profile*. Let $x^t = (x_1^t, \ldots, x_n^t) \in [0,1]^n$ be the distribution of attributes of type $t \in T$, and $x \in [0,1]^{\tau \times n}$ the *joint distribution* of attributes of all types. Then, an index is a mapping $I : [0,1]^{\tau \times n} \to [0,1]$ where $I(x)$ measures some property of the joint distribution $x \in [0,1]^{\tau \times n}$. We assume that there is a certain degree of antagonism between each pair of individuals. The antagonism between a pair $i, j \in N$ is an increasing function of the distance of their attributes $|x_i^t - x_j^t|$ of each type $t \in T$. The mapping $p : [0,1]^\tau \to [0,1]$ defines its degree, where $p(|x_i^1 - x_j^1|, \ldots, |x_i^\tau - x_j^\tau|)$ is continuous, symmetric, $p(0, \ldots, 0) = 0$ and $p(1, \ldots, 1) = 1$.

Given this, we require our index to be a measure of *expected antagonism*, in the sense that it represents the degree of antagonism between two randomly matched individuals. Then, for any $x \in [0,1]^{\tau \times n}$, the index takes the form

$$I(x) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} p(|x_i^1 - x_j^1|, \ldots, |x_i^\tau - x_j^\tau|) \text{ for some } p. \tag{1}$$

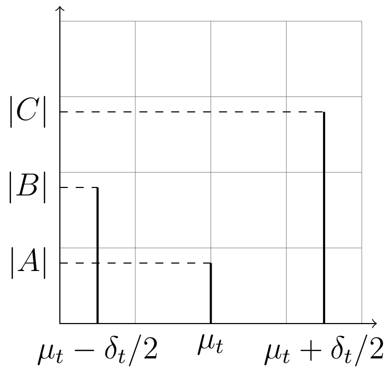
Our definition of expected antagonism is analogous to the well-known framework of Esteban and Ray (1994) for the univariate case. See Section 5 for a representation of their index. However, while we assume that the antagonism between two individuals is exclusively determined by the distance of their attributes, they additionally require this antagonism to depend on how many individuals share the same attributes with the two individuals. Their assumption is that, the larger the group of individuals who share some attribute, the more each member identifies with the attribute, and therefore the higher the antagonism of each member towards any outsider. This idea is certainly appealing. However, it is well-known that groups may encounter severe collective action problems when they mobilize, hence cooperation may be more difficult for larger groups than for smaller ones. Olson (1965) and Isaac and Walker (1988) are two of the many contributions which highlight the issue of free-riding in large groups. Given these contrasting forces, in this work we restrict our measure to take a neutral stance on the subject: we rule out a priori effects of group size to exclusively focus on antagonism between individuals. As discussed in Section 4, this minimalistic approach is sufficient to validate our index as a measure of polarization.

3 Characterization

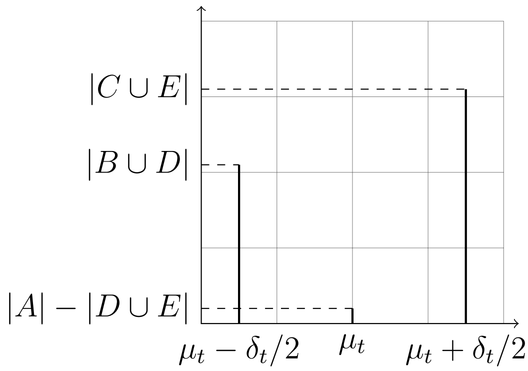
The restriction in (1) defines a class of multivariate indices which are intuitively linked to social conflict. In the rest of the article we focus on a particular index from this class, arguing that it is the only one to fulfill some properties that we motivate to be essential for the notion of polarization. We shall call these properties *axioms*.

The following axiom imposes a very natural property of a predictor of social conflict. It requires the index to increase when equal amounts of individuals are shifted from the center to each side of a distribution of some type, everything else equal. In other words, social conflict should increase when some moderate individuals take sides, becoming more extremist.

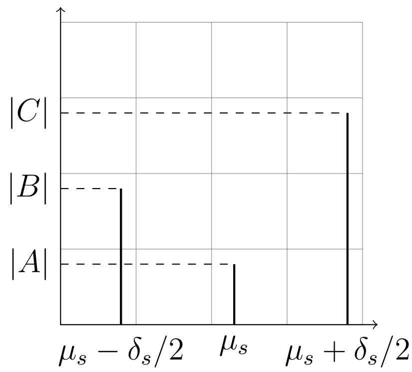
Axiom 1 Data: *Consider $x \in [0,1]^{\tau \times n}$ which partitions the population into $A, B, C \subseteq N$ such that $|A| \geq 2$ and $|B|, |C| \geq 1$. For any $i \in N$ and $t \in T$ we have $x_i^t = \mu_t$ if $i \in A$, $x_i^t = \mu_t - \delta_t/2$ if $i \in B$, $x_i^t = \mu_t + \delta_t/2$ if $i \in C$, where $\mu_t, \delta_t \in (0,1)$. For any $t \in T$, let $y^t \in [0,1]^n$ be such that there are $D, E \subseteq A$ with $|D|, |E| \geq 1$ such that for all $i \in N$ we have $y_i^t = \mu_t - \delta_t/2$ if $i \in B \cup D$, $y_i^t = \mu_t + \delta_t/2$ if $i \in C \cup E$, $y_i^t = x_i^t$ if $i \in A \setminus \{D \cup E\}$.* Statement: *For any $t \in T$, we have $I(x^1, \ldots, y^t, \ldots, x^\tau) \geq I(x)$ if $|D| = |E| \leq |B| + |C|$.*



Distribution x^t for some $t \in T$



Distribution y^t for some $t \in T$

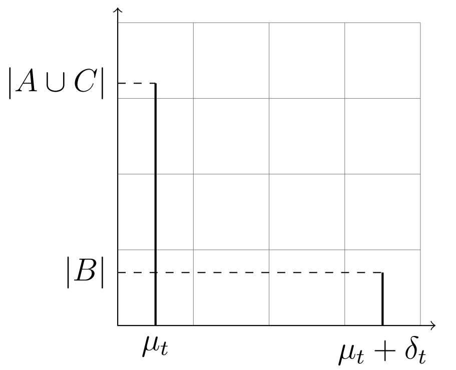


Distribution x^s for all $s \neq t$

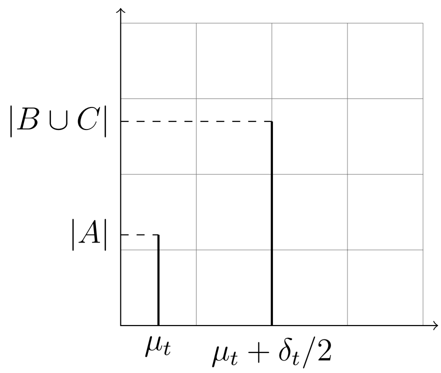
Axiom 1 demands the index to increase when the distribution of some type is changed, taking all other distributions as given. It is straightforward that, by induction, $I(x)$ should increase also when all distributions are simultaneously changed, i.e., $I(y) \geq I(x)$. Given this, one can see that analogous properties to Axiom 1 are fulfilled by most univariate indices of polarization, see for instance Esteban and Ray (1994), Duclos et al. (2004) and Anderson et al. (2012). In this sense, this is perhaps the most characterizing property of a polarization measure.

The next axiom requires that, under very specific circumstances, shifting attributes of some type closer to each other can increase social conflict. This should be the case when, by doing so, individuals are partitioned into two groups whose members have perfectly homogeneous attributes. Then, two contrasting forces are at play, and the gain from the formation of homogeneous groups should be worth the cost of reducing some distances.

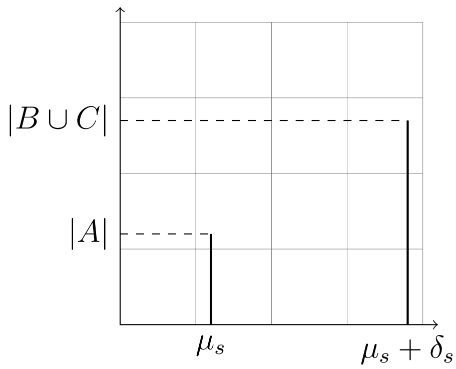
Axiom 2 Data: *Consider $x \in [0,1]^{\tau \times n}$ which partitions the population into $A, B, C \subseteq N$ such that $|A|, |B|, |C| \geq 1$. For any $i \in N$, some $t \in T$ and all $s \neq t$ we have $x_i^t = \mu_t$ and $x_i^s = \mu_s$ if $i \in A$, $x_i^t = \mu_t + \delta_t$ and $x_i^s = \mu_s + \delta_s$ if $i \in B$, $x_i^t = \mu_t$ and $x_i^s = \mu_s + \delta_s$ if $i \in C$, where $\mu_t, \delta_t, \mu_s, \delta_s \in (0,1)$. For the same $t \in T$, let $y^t \in [0,1]^n$ be such that for all $i \in N$ we have $y_i^t = \mu_t$ if $i \in A$ and $y_i^t = \mu_t + \delta_t/2$ if $i \in B \cup C$.* Statement: *For any $t \in T$, we have $I(x^1, \ldots, y^t, \ldots, x^\tau) \geq I(x)$ if $|C| \geq |B|$.*



Distribution x^t for some $t \in T$



Distribution y^t for some $t \in T$



Distribution x^s for all $s \neq t$

Axiom 2 is somewhat related to correlation increasing majorization (CIM), which is a well-known dominance criterion in the literature on multidimensional inequality. See Section 4 for further details. In short, CIM requires an index to increase when the attributes of two individuals are rearranged such that they are perfectly positively correlated. While CIM applies locally, i.e. independently of the rest of the distribution, Axiom 2 has bite only under specific circumstances which depend on the global properties of the distribution. Then, although both properties concern the correlation of attributes across types, Axiom 2 is of much weaker nature than CIM. This is in line with the standard approach in the literature on univariate polarization, where indices typically fail to fulfill local majorization criteria such as the Pigou-Dalton Transfer Principle and related concepts. See Esteban and Ray (1994) for a discussion.

To summarize, our axioms focus on two crucial aspects of social conflict. Axiom 1 requires the index to increase when individuals take sides, in the sense that their attributes take extreme values rather than central ones. Conversely, Axiom 2 demands social conflict to be related to the formation of groups which are homogeneous with respect to all types of attributes. The theorem below characterizes a specific index from our family of measures of expected antagonism via these two axioms. Note that the functional form of this index is unique, therefore it always provides unambiguous ranks of distributions by their degree of polarization.

Theorem 1 *A measure of expected antagonism satisfies Axioms 1-2 if and only if for any* $x \in [0,1]^{\tau \times n}$ *it takes the form*

$$I(x) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \prod_{t \in T} |x_i^t - x_j^t|. \tag{2}$$

The structure of the proof of Theorem 1, which is in Appendix, is straightforward. We briefly sketch here the role of the axioms in the proof. Given that $I(x)$ must take the form in (1) for some function p, we show that Axiom 1 requires p to be weakly convex in each of its arguments. Conversely, Axiom 2 implies that p is weakly concave. Then, as p is both weakly concave and weakly convex, p is linear. Moreover the axioms jointly require also that $I(x)$ is null whenever one of its arguments is null, i.e., $p(\cdot, 0, \cdot) = 0$. Together with linearity this implies $p(|x_i^1 - x_j^1|, \ldots, |x_i^\tau - x_j^\tau|) = \prod_{t \in T} |x_i^t - x_j^t|$, hence $I(x)$ must take the form in (2).

4 Majorization properties

We now discuss a series of properties which, together with Axioms 1-2, should help us to better understand the nature of our index. Recall that our purpose is to develop a tool to predict social conflict. Then, one may wonder which distributions maximize the index. Intuitively, these distributions are "highly polarized". To give an idea of what a highly polarized distribution should look like, we quote a passage from Esteban and Ray (1994): *"The polarization of a distribution of individual attributes must exhibit the following basic features. (i) There must be a high degree of homogeneity within each group. (ii) There must be a high degree of heterogeneity across groups. (iii) There must be a small number of significantly sized groups."* We follow their approach closely in our definition below.

A distribution $x \in [0,1]^{\tau \times n}$ is *most polarized* if there is a partition of N denoted by $\{L_x, R_x\}$ which satisfies the following three conditions.

- $|x_i^t - x_j^t| = 0$ for any $t \in T$ and pair $i, j \in N$ such that $i, j \in L_x$ or $i, j \in R_x$.
- $|x_i^t - x_j^t| = 1$ for any $t \in T$ and pair $i, j \in N$ such that $i \in L_x$ and $j \in R_x$.
- $|L_x| = |R_x|$ if n is even, and $|L_x| = |R_x| + 1$ if n is odd.

Let $P \subseteq [0,1]^{\tau \times n}$ be the set of all most polarized distributions. Then, we have $x \in P$ if and only if x partitions the population in two equally sized groups whose members have perfectly homogeneous attributes, and the distance between the attributes of the two groups is maximal. The following proposition argues that P is the set of all maximizers of our index of polarization.

Proposition 1 *A joint distribution* $x^* \in [0,1]^{\tau \times n}$ *maximizes the index in (2) if and only if* $x^* \in P$.

Note that, if distributions were univariate, P would coincide with the set of maximizers of the polarization measure in Esteban and Ray (1994). This is somewhat surprising. As discussed in Section 2, Esteban and Ray (1994) assume that the antagonism between two individuals increases in the size of their groups, while our framework rules out a priori effects of group size. Then, Proposition 1 shows that our minimalistic framework is sufficient to capture the crucial feature of a polarization measure: to be exclusively maximized at most polarized distributions.

While validating our index as a measure of polarization, Proposition 1 sets us apart from inequality measurement, as inequality is typically maximized by distributions where a single individual detains the highest attributes and all others the lowest.[4] To further appreciate this, we now argue that our index does not satisfy two well-known majorization criteria of multivariate inequality. These criteria, named *uniform majorization* (UM) and *correlation increasing majorization* (CIM), are alternative generalizations of the Pigou-Dalton Transfer Principle for the univariate case. UM requires that, for any $x, y \in [0,1]^{\tau \times n}$ and any bistochastic matrix B, $y = xB$ implies $I(y) < I(x)$ if and only if y cannot be derived from x by permutation. To define CIM, we need first to define the concept of correlation increasing transfer. For any $x, y \in [0,1]^{\tau \times n}$, y is derived from x by a correlation increasing transfer if for some $i, j \in N$ we have $y_k = x_k$ for any $k \neq i, j$ and

$$y_i = \left(\min\left\{x_i^1, x_j^1\right\}, \ldots, \min\left\{x_i^\tau, x_j^\tau\right\}\right),$$
$$y_j = \left(\max\left\{x_i^1, x_j^1\right\}, \ldots, \max\left\{x_i^\tau, x_j^\tau\right\}\right).$$

This transfer is said to be strict whenever $y \neq x$. Given this, CIM demands that, for any $x, y \in [0,1]^{\tau \times n}$, $I(y) > I(x)$ whenever y is derived from x by permutation and a finite sequence of correlation increasing transfers at least one of which is strict. See Tsui (1999) for further details. We are now ready to state our result.

Proposition 2 *The index in (2) does not satisfy either UM or CIM.*

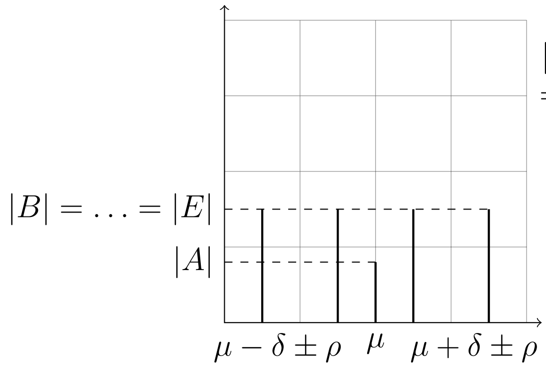
We have shown that our index is maximized only at most polarized distributions, and that it does not behave like a standard inequality measure. Then, one may wonder to which extent it is in line with polarization and the related literature. We now argue that our index does not fulfill a property which is satisfied by the measure of Duclos et al. (2004), but it is violated by other polarization indices, see for instance Esteban and Ray (1994) and Anderson et al. (2012).[5] Roughly speaking, this property requires the index to increase when all distributions are symmetrically clustered both on the left and right side of the global median, on the respective local medians.

Property 1 Data: *Consider $x \in [0,1]^{\tau \times n}$ which partitions the population into $A, B, C, D, E \subseteq N$ such that $|A| \geq 0$ and $|B| = |C| = |D| = |E| \geq 1$. For any $i \in N$ and $t \in T$, let $x_i^t = \mu$ if*

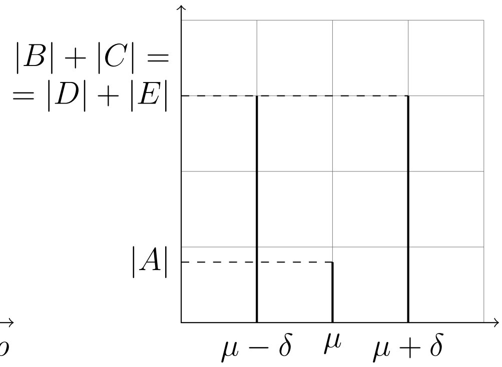
[4]See for instance the broad family of inequality measures defined in Tsui (1995) (see Section 7). The univariate Gini index is an exception, as one can show that it is exclusively maximized at most polarized distributions. However, this is not generally true for its multivariate versions. See for instance the average of univariate Gini indices across all types of attributes, $\frac{1}{n^2\tau}\sum_{t \in T}\sum_{i \in N}\sum_{j \in N}|x_i^t - x_j^t|$.

[5]These three indices are defined in Section 5. Esteban and Ray (2010) show that an analogous version of Property 1 is fulfilled by the index in Duclos et al. (2004). Conversely, the index in Esteban and Ray (1994) does not fulfill Property 1 when the parameter α is small. This easily follows from the arguments in the proof of Proposition 3 and continuity. Lastly, by construction, it is obvious that the index in Anderson et al. (2012) violates Property 1.

$i \in A$, $x_i^t = \mu - \delta - \rho$ if $i \in B$, $x_i^t = \mu - \delta + \rho$ if $i \in C$, $x_i^t = \mu + \delta - \rho$ if $i \in D$ and $x_i^t = \mu + \delta + \rho$ if $i \in E$, where $\mu, \delta, \rho \in (0,1)$ and $\delta > \rho$. Let $y \in [0,1]^{\tau \times n}$ be such that for any $i \in N$ and $t \in T$ we have $x_i^t = \mu$ if $i \in A$, $x_i^t = \mu - \delta$ if $i \in B \cup C$ and $x_i^t = \mu + \delta$ if $i \in D \cup E$. Statement: *$I(y) \geq I(x)$.*



Distribution x^t for all $t \in T$

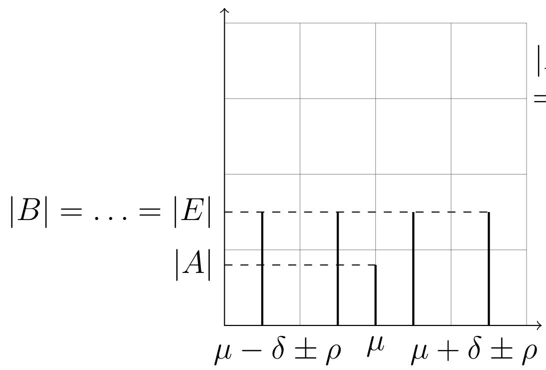


Distribution y^t for all $t \in T$

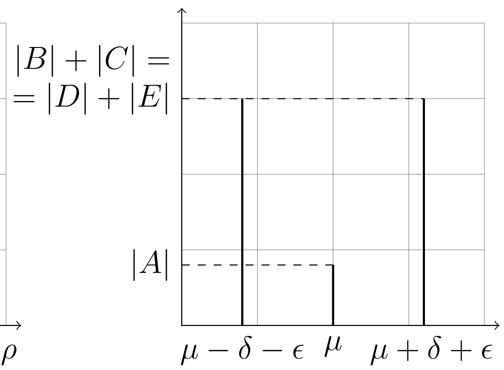
Proposition 3 *The index in (2) does not satisfy Property 1.*

One may wonder to which extent our framework violates Property 1. It turns out that a slightly weaker version is always satisfied by our index. This other property is essentially the same as Property 1, the only difference being that clustering does not take place on the local medians, but on marginally more extreme attributes.

Property 2 Data: *Let $x \in [0,1]^{\tau \times n}$ be as in Property 1. Consider $y \in [0,1]^{\tau \times n}$ such that for any $i \in N$ and $t \in T$ we have $x_i^t = \mu$ if $i \in A$, $x_i^t = \mu - \delta - \epsilon$ if $i \in B \cup C$ and $x_i^t = \mu + \delta + \epsilon$ if $i \in D \cup E$, where $\epsilon \in [0, \rho]$.* Statement: *There exists $\epsilon^* \in (0, \rho)$ such that for all $\epsilon \in [\epsilon^*, \rho]$ we have $I(y) \geq I(x)$.*



Distribution x^t for all $t \in T$



Distribution y^t for all $t \in T$

Proposition 4 *The index in (2) satisfies Property 2.*

Propositions 3 and 4 show that, when clustering on both sides of the global median takes place, additional shifts towards the extremes are necessary to increase polarization. This suggests that our index is among the conservative ones in the literature, as it is less sensitive to the formation of homogeneous groups than other approaches. In this sense, our index tends to give a "lower bound" of polarization. This idea is further confirmed by looking at the distributions which minimize our index. By (2), the antagonism between two individuals is null whenever their attributes are equal for at least one type, i.e., $p(|x_i^1 - x_j^1|, \ldots, |x_i^\tau - x_j^\tau|) = 0$ if and only if $x_i^t = x_j^t$ for some $t \in T$. Then, $I(x) = 0$ for all $x \in X$ where for each pair $i, j \in N$ there is a pair-specific type $t_{i,j} \in T$ such that $x_i^{t_{i,j}} = x_j^{t_{i,j}}$. Clearly, this is a rather broad set of distributions.

Let us briefly summarize this Section. We have shown in Proposition 1 that our index is exclusively maximized by most polarized distributions, which is the crucial property of a polarization measure. On the other hand Proposition 2 shows that the index does not satisfy two well-known majorization criteria of inequality measurement. Then, our index is a measure of polarization. The results in Propositions 3 and 4 are useful to locate our framework within the polarization literature. Together, these two findings suggest that our index is a conservative measure which gives a lower bound of polarization. This intuition is confirmed by considering the broad set of minimizers of our index.

5 Alternative representations

The index in (2) is intuitive. As it is stated, it clearly shows its close relation with social conflict. We now give two alternative representations. The first states the index in geometrical terms, helping us to better understand its mathematical properties. The second rewrites it with density functions, which is useful to compare the index with other measures in the literature.

Geometrical representation

For any $x \in [0,1]^{\tau \times n}$ and $t \in T$, let $\Delta_x^t = (|x_1^t - x_2^t|, \ldots, |x_{n-1}^t - x_n^t|)$ be the vector of distances of attributes of all unordered pairs of different individuals, where

$$||\Delta_x^t||_\tau = \left[\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (|x_i^t - x_j^t|)^\tau\right]^{1/\tau} \quad \text{is its } \tau\text{-norm.} \tag{3}$$

It is easy to verify that, for the bivariate case ($\tau = 2$), the index in (2) is a function of the dot product of the two vectors Δ_x^1 and Δ_x^2, as it can be written as $I(x) = \frac{2}{n^2}(\Delta_x^1 \cdot \Delta_x^2)$. Then, by definition of dot product, for $\tau = 2$ our index takes the form

$$I(x) = \frac{2}{n^2} ||\Delta_x^1||_2 ||\Delta_x^2||_2 \cos[\theta_x],$$

where $||\Delta_x^1||_2$ and $||\Delta_x^2||_2$ are the Euclidean norms of the vectors Δ_x^1 and Δ_x^2, while θ_x is the angle between these two vectors (see Figure 1).

We can write the index as a dot product only for $\tau = 2$. However, an alternative geometrical representation can be given for all $\tau \geq 2$. Let $\gamma_x^{i,j} = \prod_{t \in T} |x_i^t - x_j^t|^{1/\tau}$ be the geometric mean of all distances of attributes between a pair $i, j \in N$, and $\Gamma_x = (\gamma_x^{1,2}, \ldots, \gamma_x^{n-1,n})$ be the vector of mean distances of all pairs. Then, for any $\tau \geq 2$, our index can be written as

$$I(x) = \frac{2}{n^2} \left(||\Gamma_x||_\tau\right)^\tau .$$

Then, $I(x)$ is proportional to the length (the $\tau-norm$) of the vector of mean distances Γ_x. Given this, it is straightforward that the upper bound of the index is defined by Hölder inequality (see Finner (1992))

$$\frac{n^2 I(x)}{2} = \left(||\Gamma_x||_\tau\right)^\tau \leq \prod_{t \in T} ||\Delta_x^t||_\tau, \tag{4}$$

which is strict if and only if all vectors Δ_x^t are linearly independent.

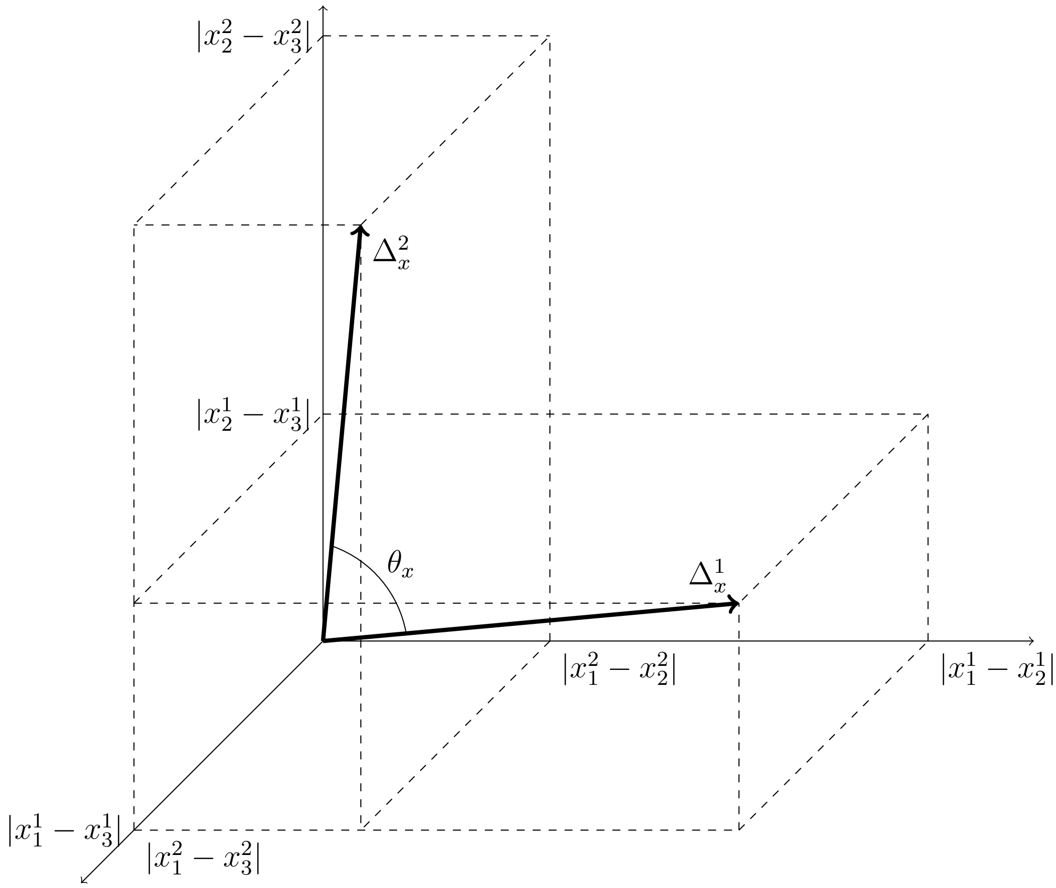


Figure 1. Graphical representation of the vectors Δ_x^1 and Δ_x^2 for the bivariate case. Given $n = 3$, $x^1 = (.1, .9, .6)$ and $x^2 = (.6, .9, .1)$, we have $\Delta_x^1 = (.8, .5, .3)$ and $\Delta_x^2 = (.3, .5, .8)$. Note that there is an axis for each pair $i, j \in N$, and on such axis we have both $|x_i^1 - x_j^1|$ and $|x_i^2 - x_j^2|$.

Density functions

For any $x \in [0,1]^{\tau \times n}$, let $f_x : [0,1]^\tau \rightarrow [0,1]$ assign a fraction of population to each attribute profile $x_i \in [0,1]^\tau$ according to x, i.e., f_x is the *density function* of individual attributes defined by x. For each attribute profile $u \in [0,1]^\tau$, the respective fraction of population is denoted by $f_x(u)$. Then, for any $x \in [0,1]^{\tau \times n}$, the index in (2) can be written as

$$I(x) = \int_u \int_v f_x(u) f_x(v) \prod_{t \in T} |u^t - v^t| \partial u \partial v, \tag{5}$$

where $u, v \in [0,1]^\tau$ are two generic attribute profiles and $u^t, v^t \in [0,1]$ are their respective attributes of type $t \in T$. Given this, we can easily compare our index with other polarization measures.

We start with the indices in Esteban and Ray (1994) and Duclos et al. (2004) due to the similarity of their framework to ours. Although they exclusively focus on the univariate case ($\tau = 1$) which is ruled out in our framework, they can be written in our notation as

$$\tilde{I}(x) = \int_u \int_v f_x(u)^{1+\alpha} f_x(v) |u - v| \partial u \partial v,$$

where $u \in [0,1]$ is a generic attribute and $\alpha > 0$.[6] Note that the mathematical expressions of $\tilde{I}(x)$ and (5) coincide if and only if $\tau = 1$ and $\alpha = 0$, which are ruled out. For these parameter values, both $\tilde{I}(x)$ and (5) would coincide with the Gini index of inequality.

Among other univariate polarization measures, we wish to focus on the index of Anderson et al. (2012), as it can be seen as a bivariate measure under some circumstances. To see this, let $\tau = 2$ and assume that, while attributes of the first type can take any value, attributes of second type are binary, i.e., $x_i^1 \in [0,1]$ and $x_i^2 \in \{0,1\}$ for all $i \in N$. Then, the index of Anderson et al. (2012) can be written as

$$\hat{I}(x) = 1 - \int_u \min \{f_x(u,0), f_x(u,1)\} \partial u,$$

where $u \in [0,1]$ is a generic attribute of the first type. Conversely, for this particular case (5) takes the form

$$I(x) = \int_u \int_v f_x(u,0) f_x(v,1) |u - v| \partial u \partial v,$$

where $u, v \in [0,1]$ are two generic attributes of first type. One can clearly see that (5) and $\hat{I}(x)$ are very different, as $\hat{I}(x)$ is not a function of the distances of attributes but only of the density function.

To our knowledge the only multivariate measure of polarization in the literature is in Gigliarano and Mosler (2009), which follow a completely different approach than ours. They define a class of indices which are separable in within-group inequality $W(x)$, between-group inequality $B(x)$ and homogeneity of group sizes $S(x)$, where the functional forms of $W(x)$, $B(x)$ and $S(x)$ are

[6] The mathematical expressions of the two indices in Esteban and Ray (1994) and Duclos et al. (2004) are equivalent, although they fundamentally differ for the nature of the density function f_x, which is respectively discrete and continuous. Moreover, the two indices present different ranges of admissible values of the parameter α, which are respectively $(0, 1.6]$ and $[.25, 1]$.

taken from well-known models in the literature. Then, their multivariate polarization index has the structure

$$\breve{I}(x) = \phi(W(x), B(x), S(x)),$$

where the function $\phi : \mathbb{R}^3_+ \to [0, 1]$ is decreasing in $W(x)$, increasing in $B(x)$ and $S(x)$ and typically non-linear. A crucial aspect of this framework is that if attributes of type $t \in T$ are used to partition of the population into groups, then $B(x)$ and $W(x)$ must be independent of attributes of type t. This leads to the same limitations of the measures of horizontal inequality discussed in the introduction. Another noteworthy feature is that, due to the monotonicity in $S(x)$, their index is always maximized when all groups have equal size, no matter the number of groups. Then, this index is not maximized at most polarized distributions.

Lastly, we consider the multivariate inequality index developed in Koshevoy and Mosler (1997), which can be written with our notation as

$$\breve{I}(x) = \int_u \int_v f_x(u) f_x(v) \left[\frac{1}{\tau^2} \sum_{t \in T} (u^t - v^t)^2 \right]^{1/2} \partial u \partial v, \tag{6}$$

where $u, v \in [0, 1]^\tau$ are two generic attribute profiles and $u^t, v^t \in [0, 1]$ are their respective attributes of type $t \in T$. The index in (6) belongs to the class of measures of expected antagonism defined in (1), although it does not satisfy our axioms and it is not maximized at most polarized distributions.[7] However, there is a further link with our approach. If we renormalize $\breve{I}(x)$ by the multiplying constant $\sqrt{\tau}$, we can write the function p as the power mean

$$p(|x_i^1 - x_j^1|, \ldots, |x_i^\tau - x_j^\tau|) = \left[\frac{1}{\tau} \sum_{t \in T} (|u^t - v^t|)^r \right]^{1/r} \text{ evaluated at } r = 2.$$

Then, taking the limit $r \to 0$ the power mean becomes

$$p(|x_i^1 - x_j^1|, \ldots, |x_i^\tau - x_j^\tau|) = \left[\prod_{t \in T} |u^t - v^t| \right]^{1/\tau},$$

which is the τ-root of the functional form that p takes in (2).

6 Inference

Let the n individuals described in the previous section to be independent from one another, for example such as those collected from a survey. Given real data, it will be useful to think of them as a random sample, $\{X_i\}_{i=1}^n$, drawn from some population distribution of a $\tau-$dimensional

[7]The index (6) cannot satisfy our axioms due to the uniqueness of our characterization within the class. To see that (6) is not maximized at most polarized distributions, consider the simplest case with $n = 4$ and $\tau = 2$. Let $x \in X$ be such that $x_1^t = x_2^t = 0$ for all $t \in T$, and $x_3^1 = x_3^2 = x_4^1 = 1$. Then, x is most polarized if and only if $x_4^2 = 0$. However, it is easy to show that, given the other attributes are fixed, $\breve{I}(x)$ is always maximized at $x_4^2 = 1$.

random vector X. Then it is natural to think of polarization as a feature related to the distribution of X.

For any $p : [0,1]^{\tau} \to [0,1]$, we can analogously interpret the class of polarization measures in Section 2 to be the expected antagonism between any two randomly matched individuals, namely:

$$\theta = E\left[p\left(\left|X^1 - Y^1\right|, \ldots, \left|X^{\tau} - Y^{\tau}\right|\right)\right], \tag{7}$$

where Y is another random vector that has the same distribution as X, but is independent of X. This expectation representation of the polarization measure is useful since it encompasses all types of random variables; continuous, discrete or a mixture. Indeed, from taking the law of iterated expectation, our definition of polarization from equation (1) equals the quantity above when the expectation operator is replaced by the empirical expectation operator. Specifically, let E_n denote an expectation operator that integrates X and Y by putting a constant measure of n^{-1} on $\{x_i\}_{i=1}^n$ and zero everywhere else, then we have

$$I(x) = E_n\left[E_n\left[p\left(\left|X^1 - Y^1\right|, \ldots, \left|X^{\tau} - Y^{\tau}\right|\right)|Y\right]\right].$$

We now describe how one can perform inference on θ given a random sample $\{X_i\}_{i=1}^n$. We propose a *leave-one-out* estimator based on the sample counterpart of θ. Let

$$f_n(X_i) = \frac{1}{(n-1)} \sum_{j \neq i}^{n} p\left(\left|X_i^1 - X_j^1\right|, \ldots, \left|X_i^{\tau} - X_j^{\tau}\right|\right).$$

We define our estimator as:

$$\begin{aligned}
\theta_n &= \frac{1}{n} \sum_{i=1}^{n} f_n(X_i) && (8)\\
&= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} p\left(\left|X_i^1 - X_j^1\right|, \ldots, \left|X_i^{\tau} - X_j^{\tau}\right|\right)\\
&= \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} p\left(\left|X_i^1 - X_j^1\right|, \ldots, \left|X_i^{\tau} - X_j^{\tau}\right|\right).
\end{aligned}$$

Since p is symmetric in its arguments, the leave-one-out estimator is a standard *second order U-statistic.* Our motivation for defining our estimator in the form of a U-statistic is due to the fact that the latter object has well-established statistical properties (Hoeffding (1948)), and the omission of the $i-th$ observation in the summand of $f_n(X_i)$ makes no difference to the large sample properties of our estimator asymptotically. The following propositions summarize the large sample properties of our estimator.

Proposition 5 $\theta_n = \theta + o_p(1)$.

Proposition 6 $\sqrt{n}(\theta_n - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ *where*

$$\sigma^2 = 4Var\left(E\left[p\left(\left|X_i^1 - X_j^1\right|, \ldots, \left|X_i^{\tau} - X_j^{\tau}\right|\right)|X_i\right]\right).$$

Propositions 5 and 6 say that our estimator of the polarization index is consistent, and it converges to the true at the parametric rate of root-n with a limiting normal distribution. Since we have the explicit form for the asymptotic variance of θ_n, this can be estimated by using its sample counterpart. Note that, for any $i \neq j$, we have

$$\begin{aligned} & Var\left(E\left[p\left(\left|X_i^1 - X_j^1\right|, \ldots, \left|X_i^\tau - X_j^\tau\right|\right)|X_i\right]\right) \\ = \; & E\left[E\left[p\left(\left|X_i^1 - X_j^1\right|, \ldots, \left|X_i^\tau - X_j^\tau\right|\right)|X_i\right]^2\right] - \left(E\left[p\left(\left|X_i^1 - X_j^1\right|, \ldots, \left|X_i^\tau - X_j^\tau\right|\right)\right]\right)^2. \end{aligned}$$

Let σ_n^2 denote the estimator of σ^2. One natural candidate of σ_n^2 is the following,

$$\sigma_n^2 = \frac{4}{n}\sum_{i=1}^{n}\left(\frac{1}{n-1}\sum_{j\neq i}^{n} p\left(\left|X_i^1 - X_j^1\right|, \ldots, \left|X_i^\tau - X_j^\tau\right|\right)\right)^2 - 4\theta_n^2. \tag{9}$$

Our next proposition confirms that σ_n^2 is a consistent estimator of σ^2.

Proposition 7 $\sigma_n^2 = \sigma^2 + o_p(1)$.

Given this, we can construct confidence intervals and perform hypothesis tests on θ based on normal approximation. Alternatively the distribution of θ_n can also be approximated using a nonparametric bootstrap. See Arcones and Giné (1992) for general results on bootstrapping U-statistics.

7 Empirical Illustration

In this section we illustrate how polarization and inequality differ across the four countries of the United Kingdom (England, Wales, Scotland and Northern Ireland) and how they evolve over time. We use a 7-year annual dataset from the British Household Panel Survey (BHPS) constituted of individuals interviewed in the years 2002-2009.[8] We compute our polarization measure based on two individual attributes, income and highest education level, and compare it with multivariate inequality measures based on the same data.

We begin by normalizing our data to take value between zero and one for each country and year. We construct various indices based on the samples that omit individuals with extreme reported value of income. The latter is defined as income earning that is more than 20 times higher or lower than the median income for each country and year. (It seems reasonable to exclude income values that are zero or negative, as well as individuals reporting very high inccome since income distributions are known to have thick tails.) The income can then be trivially translated into a unit interval. We use 7 categories of the highest levels of education recorded in the survey, consisting of: "*Higher Degree*", "*1st Degree*", "*HND, HNC, Teaching*", "*A-Level*", "*O-Level*", "*CSE*" and "*None of these*".[9,10] Clearly it not possible to have a universally accepted ordinal

[8] The dataset is publicly available from https://www.iser.essex.ac.uk/bhps/.

[9] Higher National Certificates (HNCs), Higher National Diplomas (HNDs) and Teaching qualifications are work-related, or vocational, higher education qualifications in the United Kingdom.

[10] The Certificate of Secondary Education (CSE) was an academic qualification awarded between 1965 and 1987 in England, Wales, and Northern Ireland that is comparable to O-Levels.

ranking of all these qualifications. We correspondingly rank them in the same order recorded on the survey, and assign value 1 to "*Higher Degree*" and 0 to "*None of these*", with all other levels of qualification having equidistance with their immediate neighbors.

We construct and compare three indices. We use the leave-one-out estimator to construct the polarization measure:

$$\theta_{n,SV} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \prod_{t=1}^{2} \left| x_i^t - x_j^t \right|.$$

Given our discussions in Sections 5 and 4, for the purpose of comparison, it seems most natural to consider the multivariate inequality measures of Tsui (1995) and Koshevoy and Mosler (1997). The former index takes the form:

$$\theta_{n,T} = 1 - \frac{1}{n} \sum_{i=1}^{n} \prod_{t=1}^{2} \left(\frac{x_i^t}{\mu(x^t)} \right)^{1/2},$$

where $\mu(x^t)$ denotes $\frac{1}{n}\sum_{i=1}^{n} x_i^t$ for $t = 1, 2$.[11] The leave-one-out version of the latter index takes the form:

$$\theta_{n,KM} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left[\frac{1}{2^2} \sum_{t=1}^{2} (x_i^t - x_j^t)^2 \right]^{1/2}$$

Table 1 gives the estimates and the standard errors for each index for different countries and years. The rankings of indices for each year are also reported in the table. All standard errors are constructed using nonparametric bootstrap. We also computed the standard errors for our estimators based on the sample counterparts, as suggested in the previous section, and find they give almost identical values. Table 2 gives the sample sizes used to construct the indices in Table 1. The percentages of individuals omitted for reporting extreme income levels are also provided.

[Tables 1 and 2 here]

It is difficult to compare different indices cardinally. Our illustration shall only focus on the rankings of polarization and inequality indices to highlight the differences across measure concepts (e.g. as has been done in related work, see Duclos et al. (2004)).

Apart from Year 2002/3, our index and Tsui's measure agree in all years that Northern Ireland and Wales are the most polarized and the most unequal countries, and that England and Scotland are the least polarized and the least unequal countries. However, our orderings within the two groups often do not coincide with Tsui's. In addition we see high variation in the ranking of polarization over time. In particular our measure suggests that Wales is more

[11] A general form for the inequality indices proposed in Tsui (1995) takes the form: $1 - \left[\frac{1}{n}\sum_{i=1}^{n}\prod_{t=1}^{2}\left(\frac{x_i^t}{\mu(x^t)}\right)^{1/r_t}\right]^{1/(r_1+r_2)}$. The choice of r_1, r_2 allows users to put different weight on different attributes. We compute the neutral inequality index where $r_1 = r_2 = 1/2$.

polarized than Northern Ireland from the samples taken from 2002/3 - 2004/5 and in 2008/9, whereas the measure of Tsui consistently ranks Northern Ireland as the most unequal country in the United Kingdom. Assuming that attributes of individuals from different countries are independent within each year, it is easy to see that the ranking of Northern Ireland and Wales based on our index statistically differs from the one of Tsui for many of the years. The other measure, of Koshevoy and Mosler, shows high variation of its ranking over time and it generally induces different rankings than ours and Tsui's. For instance, the measure of Koshevoy and Mosler ranks Scotland and Northern Ireland as the two most unequal countries for all years apart from 2008/9. Moreover, while our indices and Tsui's rank Wales as more polarized and more unequal than Scotland (statistically) for almost every year, the measure of Koshevoy and Mosler suggests that Scotland is at least as unequal as Wales in all years.

We have also performed analogous comparisons within different subsamples (e.g. splitting the data by gender) and with other censoring schemes (excluding individuals with more/less extreme income levels), always finding that the three indices generally induce different rankings.

8 Conclusion

We define an index of polarization for multivariate attributes which is an expression of the expected antagonism of two randomly matched individuals. We argue our index to be within the tradition of the seminal work of Esteban and Ray (1994), although we crucially differ from them as we rule out a priori effects of group size. Ruling out these effects guarantees the uniqueness of our index. As theirs, our index is always maximized when the population is divided into two equally sized and homogeneous groups with opposite attributes, which is the crucial property of polarization measures. We discuss other properties of our index that set it apart from inequality measurement and close to the polarization literature. We characterize the large sample properties of the estimator of our index and provide an empirical illustration to show the statistically significant differences between polarization and inequality.

In this paper we take a neutral stance on the effects of group size. However, a natural question is how to generalize Esteban and Ray (1994) to a multivariate setting. We believe this approach to be complementary to ours. Moreover, there are some common issues to this and our approach which should be tackled. The index in Esteban and Ray (1994) is defined under the assumption that individual attributes take value in $\mathbb{R}$, as well as our measure and all other indices in the literature to our knowledge. However, in empirical work it is not always possible to order attributes from lowest to highest. In such circumstances, as a natural ordering does not exist, attributes must take value in some broader space than $\mathbb{R}$, and the polarization measure should be defined accordingly. For instance this issue arises with the linguistic data employed in Esteban et al. (2012).[12] We leave this and other issues to future work.

[12] To see this, suppose that there are three individuals 1, 2 and 3 which speak languages which are equally different form each other, so that $|x_i^t - x_j^t| = \delta$ for each pair $i, j \in \{1, 2, 3\}$ and some $\delta > 0$. If we impose $x_i^t \in \mathbb{R}$, then $|x_1^t - x_2^t| = |x_2^t - x_3^t| = \delta$ necessarily implies $|x_1^t - x_3^t| \in \{0, 2\delta\}$, which is a contradiction. Then, to allow for $|x_1^t - x_3^t| = \delta$, x_i^t should take value in $\mathbb{R}^k$ with $k \geq 2$.

Appendix

Proof of Theorem 1

We want to show that the index in (1) fulfills Axioms 1-2 if and only if it takes the form in (2), i.e., if and only if $p(|x_i^1 - x_j^1|, \dots, |x_i^\tau - x_j^\tau|) = \prod_{t \in T} |x_i^t - x_j^t|$. It is easy to verify that, if the function p takes this form, the index in (1) fulfills Axioms 1-2. Let us show the converse: if Axioms 1-2 are satisfied by the index in (1), then p must take such form. Let $t \in T$ be any type. Consider the data of Axiom 2, and let $|A| = a$, $|B| = b$ and $|C| = c$ for some $a, b, c \geq 1$. Then

$$n(n-1)I(x)/2 = abp\left(\delta_1, \dots, \delta_\tau\right) + acp(\delta_1, \dots, 0, \dots, \delta_\tau) + bcp(0, \dots, \delta_t, \dots, 0) \text{ and}$$

$$n(n-1)I(x_1, \dots, y_t, \dots, x_\tau)/2 = a(b+c)p\left(\delta_1, \dots, \delta_t/2, \dots, \delta_\tau\right).$$

Axiom 2 demands $I(x_1, \dots, y_t, \dots, x_\tau) \geq I(x)$ if $c \geq b$. Assuming $c = b$ this implies

$$2p\left(\delta_1, \dots, \delta_t/2, \dots, \delta_\tau\right) - p\left(\delta_1, \dots, \delta_\tau\right) \geq p(\delta_1, \dots, 0, \dots, \delta_\tau) + \frac{c}{a}p(0, \dots, \delta_t, \dots, 0). \tag{10}$$

As the RHS is weakly positive, the LHS must be so. Then, by Jensen inequality p is always weakly concave in δ_t. Consider the data of Axiom 1 for the same t. Let $|A| = a'$, $|B| = b'$, $|C| = c'$, $|D| = d'$ and $|E| = e'$ for some $a', b', c', d', e' \geq 1$ such that $a' \geq d' + e'$. Then

$$n(n-1)I(x)/2 = b'c'p\left(\delta_1', \dots, \delta_\tau'\right) + a'(b'+c')p\left(\delta_1'/2, \dots, \delta_\tau'/2\right),$$

$$\text{while } n(n-1)I(x_1, \dots, y_t, \dots, x_\tau)/2 =$$

$$= b'c'p\left(\delta_1', \dots, \delta_\tau'\right) + (a'-d'-e')(b'+c')p\left(\delta_1'/2, \dots, \delta_\tau'/2\right) + (b'e'+c'd')p\left(\delta_1'/2, \dots, \delta_t', \dots, \delta_\tau'/2\right) +$$

$$+ e'd'p(0, \dots, \delta_t', \dots, 0) + (a'-d'-e')(d'+e')p(0 \dots, \delta_t'/2, \dots, 0).$$

By Axiom 1, $I(x_1, \dots, y_t, \dots, x_\tau) \geq I(x)$ if $e' = d' \leq b' + c'$. Assuming $d' + e' = a'$ this implies

$$\frac{d'}{(b'+c')}p(0, \dots, \delta_t', \dots, 0) \geq 2p\left(\delta_1'/2, \dots, \delta_\tau'/2\right) - p\left(\delta_1'/2, \dots, \delta_t', \dots, \delta_\tau'/2\right). \tag{11}$$

Let $\delta_t' = \delta_t$ and $\delta_s' = 2\delta_s$ for all $s \neq t$. As $d'/(b'+c') \leq 1$, both (10) and (11) hold only if

$$p(0, \dots, \delta_t, \dots, 0) \geq p(\delta_1, \dots, 0, \dots, \delta_\tau) + \frac{c}{a}p(0, \dots, \delta_t, \dots, 0). \tag{12}$$

If $c \geq a$, (12) must always hold with equality, which implies $p(|x_i^1 - x_j^1|, \dots, |x_i^\tau - x_j^\tau|) = 0$ whenever $|x_i^t - x_j^t| = 0$ for some $t \in T$. Given this, it follows by (10) and (11) that

$$2p\left(\delta_1, \dots, \delta_t/2, \dots, \delta_\tau\right) - p\left(\delta_1, \dots, \delta_\tau\right) = 0,$$

hence $p(|x_i^1 - x_j^1|, \dots, |x_i^\tau - x_j^\tau|)$ is linear in $|x_i^t - x_j^t|$ for any $t \in T$ by Jensen inequality. Then, these results jointly imply $p(|x_i^1 - x_j^1|, \dots, |x_i^\tau - x_j^\tau|) = \prod_{t \in T} |x_i^t - x_j^t|$.

□

Proof of Proposition 1

By (4), it can be shown that $y \in [0,1]^{\tau\times n}$ maximizes (2) if and only if the following two conditions are fulfilled: (a) $y^t \in \arg\max[||\Delta_x^t||]$ for any $t \in T$, (b) for some $s \in T$ and all $t \neq s$ there is $\alpha_t > 0$ such that $|y_i^t - y_j^t| = \alpha_t|y_i^s - y_j^s|$ for any $i, j \in N$. It is easy to verify that conditions (a) and (b) are always fulfilled if $y \in P$, therefore I is maximized at $y \in P$. Let us show the converse: if $y \in [0,1]^{\tau\times n}$ satisfies these conditions, which means that I is maximized at y, then $y \in P$.

By (3), condition (a) can be written as

$$y^t \in \arg\max\left[\left(\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(|x_i^t - x_j^t|)^\tau\right)^{1/\tau}\right] \text{ for any } t \in T.$$

One can show that, for any $i \in N$, $\partial[||\Delta_x||^2]/\partial x_i^t \geq 0$ if and only if $x_i^t \geq \sum_{j=1}^n x_j^t/n$. Then, the function $(\sum_{i=1}^{n-1}\sum_{j=i+1}^n(|x_i^t - x_j^t|)^\tau)^{1/\tau}$ is minimized at $x^t = (1/2, \ldots, 1/2)$, and it is strictly increasing (decreasing) in x_i^t if $x_i^t > 1/2$ ($x_i^t < 1/2$). It follows that, if $y \in \arg\max[||\Delta_x||]$, we must have $y_i^t \in \{0,1\}$ for any $i \in N$ and $t \in T$, which implies $|y_i^t - y_j^t| \in \{0,1\}$ for any $i, j \in N$.

We have shown that condition (a) implies $|y_i^t - y_j^t| \in \{0,1\}$ for any $i, j \in N$ and $t \in T$. Condition (b) requires all vectors Δ_x^t to be linearly dependent, therefore (a) and (b) jointly imply $|y_i^t - y_j^t| = |y_i^s - y_j^s| \in \{0,1\}$ for all $i, j \in N$ and $t, s \in T$. It follows that y must induce a partition of the population in two groups $G, H \subseteq N$ such that

(d) $|y_i^t - y_j^t| = 0$ for any $t \in T$ and pair $i, j \in N$ such that $i, j \in G$ or $i, j \in H$.

(e) $|y_i^t - y_j^t| = 1$ for any $t \in T$ and pair $i, j \in N$ such that $i \in G$ and $j \in H$.

In other words y maximizes I only if it divides the population in two perfectly homogeneous groups whose attributes are the most distant from each other. Given this, the index must take value $I(y) = 2g(n-g)/[n^2]$, where $g = |G|$. It is straightforward that I is maximized if and only if

(f) $g = n/2$ if n is even and $g \in \{n - 1/2, n + 1/2\}$ if n is odd.

Condition (f) essentially requires G, H to be equally sized. Then, conditions (d), (e) and (f) are respectively equivalent to (i), (ii) and (iii), which are the three defining properties of a most polarized distribution. It follows that $y \in [0,1]^{\tau\times n}$ maximizes (2) if and only if $y \in P$.

$\square$

Proof of Proposition 2

To see that our index does not fulfill either UM or CIM, it is sufficient to focus on cases with $n = 4$ and $\tau = 2$.[13] Let $x \in [0,1]^{\tau\times n}$ be such that $x_1 = x_2 = x_3 = x_4 = (1/2, 1/2)$. Consider

[13] This can be easily shown by "cloning" players 3 and/or 4 and distributions of types 1 and/or 2 in an opportune fashion.

another distribution $y \in [0,1]^{\tau\times n}$ where $y_1 = (1/2, 0)$, $y_2 = (1/2, 1)$ and $y_3 = y_4 = (1/2, 1/2)$. Note that $I(y) = I(x) = 0$. It is easy to show that, for the bistochastic matrix

$$B = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \text{ we have } x = yB.$$

It follows that the index in (2) does not satisfy UM. Let $x \in [0,1]^{\tau\times n}$ be such that $x_1 = (0,1)$, $x_2 = (1,0)$, $x_3 = (1,0)$ and $x_3 = (0,1)$. Consider another distribution $y \in [0,1]^{\tau\times n}$ where $y_1 = (0,0)$, $y_2 = (1,1)$, $y_3 = (1,0)$ and $y_4 = (0,1)$. Note that, while x is a most polarized distribution, y is not, therefore $I(y) < I(x)$. Moreover, as $y_1 = (\min\{0,1\}, \min\{0,1\})$, $y_2 = (\max\{0,1\}, \max\{0,1\})$ $y_3 = x_3$ and $y_4 = x_4$, y is derived from x by a single correlation increasing transfer, which is strict. It follows that the index in (2) does not satisfy CIM.

□

Proof of Proposition 3

Consider the data of Property 1, and let $|A| = a \geq 0$ and $|B| = \ldots = |E| = b \geq 1$. It is straightforward that $n^2 I(y)/2 = (2b)^2(2\delta)^\tau + 4ab(\delta)^\tau$, while

$$n^2 I(x)/2 = b^2 \left[2(2\rho)^\tau + (2\delta + 2\rho)^\tau + (2\delta - 2\rho)^\tau + 2(2\delta)^\tau\right] + 2ab\left[(\delta+\rho)^\tau + (\delta-\rho)^\tau\right].$$

By Property 1 we have $I(y) \geq I(x)$, which requires

$$\left[(2\delta)^\tau - \frac{(2\delta+2\rho)^\tau + (2\delta-2\rho)^\tau}{2}\right] + \frac{2a}{b}\left[(\delta)^\tau - \frac{(\delta+\rho)^\tau + (\delta-\rho)^\tau}{2}\right] \geq \frac{(2\rho)^\tau}{2}.$$

Note that for any $\tau \geq 2$ the functions $(\delta)^\tau$ and $(2\delta)^\tau$ are strictly convex in δ, therefore the LHS is strictly negative for any $\rho < \delta$. As the RHS is always positive, our index never fulfills Property 1.

□

Proof of Proposition 4

Consider the data of Property 2, and let $|A| = a \geq 0$ and $|B| = \ldots = |E| = b \geq 1$. It is easy to see that, for some $\epsilon \in [0, \rho]$, $n^2 I(y)/2 = (2b)^2(2\delta + 2\epsilon)^\tau + 4ab(\delta + \epsilon)^\tau$, while

$$n^2 I(x)/2 = b^2 \left[2(2\rho)^\tau + (2\delta + 2\rho)^\tau + (2\delta - 2\rho)^\tau + 2(2\delta)^\tau\right] + 2ab\left[(\delta+\rho)^\tau + (\delta-\rho)^\tau\right].$$

Given $\rho < \delta$, it can be shown that, for $\epsilon = \rho$, a sufficient condition for $I(y) > I(x)$ is

$$(2\delta + 2\rho)^\tau - (2\delta)^\tau - (2\rho)^\tau > 0.$$

As $\tau \geq 2$ the function $(2\delta + 2\rho)^\tau$ is strictly convex in $2\delta + 2\rho$, therefore the LHS is strictly positive. It follows that, for $\epsilon = \rho$, we must have $I(y) > I(x)$. By Proposition 3, for $\epsilon = 0$ we have $I(y) < I(x)$. Then, by continuity there exists $\epsilon^* \in (0, \rho)$ such that for all $\epsilon \in [\epsilon^*, \rho]$ we have $I(y) \geq I(x)$, hence our index fulfills Property 2.

$\square$

Proof of Propositions 5 and 6

Our proof follows from some standard results from the literature of U-statistics (for example see Chapter 5.3 in Serfling (1980)). The first step is to define the projection of the U-statistic. Let $r(X_i) = E\left[p\left(\left|X_i^1 - X_j^1\right|, \ldots, \left|X_i^\tau - X_j^\tau\right|\right)|X_i\right]$, then we denote the *projection* of θ_n by:

$$\begin{aligned}\widehat{\theta}_n &= \sum_{i=1}^{n} E\left[\theta_n|X_i\right] - (n-1)\, E\left[r(X_i)\right] \\ &= E\left[r(X_i)\right] + \frac{2}{n}\sum_{i=1}^{n}\left(r(X_i) - E\left[r(X_i)\right]\right).\end{aligned}$$

Since $\theta = E\left[r(X_i)\right]$, we have

$$\widehat{\theta}_n - \theta = \frac{2}{n}\sum_{i=1}^{n}\left(r(X_i) - E\left[r(X_i)\right]\right).$$

Then following Hoeffding (1948), $\theta_n - \theta$ is asymptotically equivalent to $\widehat{\theta}_n - \theta$ to the order $O_p\left(n^{-1/2}\right)$ when $E\left[\left|p\left(\left|X_i^1 - X_j^1\right|, \ldots, \left|X_i^\tau - X_j^\tau\right|\right)\right|^2\right] < \infty$. The latter condition trivially holds since p is bounded above by 1, so that $\theta_n = \widehat{\theta}_n + O_p\left(n^{-1/2}\right)$. Therefore $\theta_n - \theta$ can be approximated by a sum of i.i.d. zero mean variables. Propositions 5 and 6 then follow immediately from a standard law of large numbers and central limit theorem for i.i.d. variables respectively.

$\square$

Proof of Proposition 7

Let r_n denote the sample counterpart of r, which is defined in the previous proof, so that $r_n(X_i) = \frac{1}{n-1}\sum_{j\neq i}^{n} p\left(\left|X_i^1 - X_j^1\right|, \ldots, \left|X_i^\tau - X_j^\tau\right|\right)$. We can now write (9) as,

$$\sigma_n^2 = \frac{4}{n}\sum_{i=1}^{n} r_n(X_i)^2 - 4\theta_n^2,$$

and σ^2, defined in Proposition 6, can be written as

$$\sigma_n^2 = 4E\left[r(X_i)^2\right] - 4\theta^2.$$

Since θ_n is consistent, it suffices to show

$$\frac{1}{n}\sum_{i=1}^{n} r_n(X_i)^2 = E\left[r(X_i)^2\right] + o_p(1).$$

To this end,

$$\begin{aligned} E\left[\left|r_n\left(X_i\right)-r\left(X_i\right)\right|^2\right] &= E\left[Var\left(r_n\left(X_i\right)|X_i\right)\right] \\ &= \frac{1}{n-1}E\left[Var\left(p\left(\left|X_i^1-X_j^1\right|,\ldots,\left|X_i^\tau-X_j^\tau\right|\right)|X_i\right)\right] \\ &\leq \frac{1}{n-1}E\left[\left|p\left(\left|X_i^1-X_j^1\right|,\ldots,\left|X_i^\tau-X_j^\tau\right|\right)\right|^2\right] \\ &= O\left(n^{-1}\right). \end{aligned}$$

Therefore $E\left[\left|r_n\left(X_i\right)-r\left(X_i\right)\right|^2\right]=o\left(1\right)$, which implies $E\left[\left|r_n\left(X_i\right)^2-r\left(X_i\right)^2\right|\right]=o\left(1\right)$, and the required result follows from Markov inequality. The proof then follows from applications of the continuous mapping theorem.

$\square$

References

Anderson, G., O. Linton, and Y.J. Whang, "Nonparametric estimation and inference about the overlap of two distributions," *Journal of Econometrics*, 2012, *171* (1), 1–23.

Arcones, M.A. and E. Giné, "On the bootstrap of U and V statistics," *The Annals of Statistics*, 1992, pp. 655–674.

Baldwin, K. and J.D. Huber, "Economic versus cultural differences: forms of ethnic diversity and public goods provision," *American Political Science Review*, 2010, *104* (04), 644–662.

Blau, J.R. and P.M. Blau, "The cost of inequality: metropolitan structure and violent crime," *American Sociological Review*, 1982, pp. 114–129.

Cederman, L.E., N.B. Weidmann, and K.S. Gleditsch, "Horizontal inequalities and ethnonationalist civil war: a global comparison," *American Political Science Review*, 2011, *105* (03), 478–495.

Desmet, K., I. Ortuño-Ortín, and R. Wacziarg, "The political economy of linguistic cleavages," *Journal of Development Economics*, 2012, *97* (2), 322–338.

Duclos, J.Y., J. Esteban, and D. Ray, "Polarization: concepts, measurement, estimation," *Econometrica*, 2004, *72* (6), 1737–1772.

Esteban, J. and D. Ray, "On the measurement of polarization," *Econometrica*, 1994, pp. 819–851.

_ **and** _ , "Comparing polarization measures," *Working paper*, 2010.

_ **, L. Mayoral, and D. Ray**, "Ethnicity and conflict: an empirical study," *The American Economic Review*, 2012, *102* (4), 1310–1342.

Fearon, J.D., "Ethnic and cultural diversity by country," *Journal of Economic Growth*, 2003, *8* (2), 195–222.

Finner, H., "A generalization of Hölder's inequality and some probability inequalities," *The Annals of Probability*, 1992, *20* (4), 1893–1901.

Foster, J.E. and M.C. Wolfson, "Polarization and the decline of the middle class: Canada and the US," *The Journal of Economic Inequality*, 2010, *8* (2), 247–273.

Gigliarano, C. and K. Mosler, "Constructing indices of multivariate polarization," *The Journal of Economic Inequality*, 2009, *7* (4), 435–460.

Hoeffding, W., "A class of statistics with asymptotically normal distribution," *The Annals of Mathematical Statistics*, 1948, pp. 293–325.

Isaac, R.M and J.M Walker, "Group size effects in public goods provision: the voluntary contributions mechanism," *The Quarterly Journal of Economics*, 1988, pp. 179–199.

Koshevoy, G.A. and K. Mosler, "Multivariate gini indices," *Journal of Multivariate Analysis*, 1997, *60* (2), 252–276.

Olson, M., *The logic of collective action; public goods and the theory of groups*, Harvard University Press, 1965.

Serfling, R.J., *Asymptotic theorems of mathematical statistics*, New York: Wiley, 1980.

Tsui, K.Y., "Multidimensional generalizations of the relative and absolute inequality indices: the Atkinson-Kolm-Sen approach," *Journal of Economic Theory*, 1995, *67* (1), 251–265.

_ , "Multidimensional inequality and multidimensional generalized entropy measures: an axiomatic derivation," *Social Choice and Welfare*, 1999, *16* (1), 145–157.

Wang, Y.Q. and K.Y. Tsui, "Polarization orderings and new classes of polarization indices," *Journal of Public Economic Theory*, 2000, *2* (3), 349–363.

Index	Cntry	2002/3	*Rk*	2003/4	*Rk*	2004/5	*Rk*	2005/6	*Rk*	2006/7	*Rk*	2007/8	*Rk*	2008/9	*Rk*
$\theta_{n,SV}$	E	0.0230	*3*	0.0238	*3*	0.0214	*4*	0.0263	*3*	0.0208	*4*	0.0215	*3*	0.0270	*3*
		(0.0005)		(0.0005)		(0.0004)		(0.0006)		(0.0004)		(0.0005)		(0.0006)	
	W	0.0318	*1*	0.0336	*1*	0.0370	*1*	0.0290	*2*	0.0277	*2*	0.0262	*2*	0.0413	*1*
		(0.0012)		(0.0012)		(0.0012)		(0.0011)		(0.0009)		(0.0010)		(0.0015)	
	S	0.0316	*2*	0.0225	*4*	0.0344	*3*	0.0231	*4*	0.0235	*3*	0.0211	*4*	0.0268	*4*
		(0.0011)		(0.0008)		(0.0011)		(0.0008)		(0.0009)		(0.0009)		(0.0010)	
	NI	0.0228	*4*	0.0310	*2*	0.0346	*2*	0.0316	*1*	0.0344	*1*	0.0390	*1*	0.0320	*2*
		(0.0009)		(0.0011)		(0.0014)		(0.0011)		(0.0013)		(0.0017)		(0.0015)	
$\theta_{n,T}$	E	0.1636	*4*	0.1591	*4*	0.1556	*4*	0.1523	*4*	0.1552	*4*	0.1519	*4*	0.1473	*4*
		(0.0032)		(0.0030)		(0.0029)		(0.0028)		(0.0028)		(0.0030)		(0.0029)	
	W	0.2016	*2*	0.1910	*2*	0.1836	*2*	0.1831	*2*	0.1865	*2*	0.1802	*2*	0.1744	*2*
		(0.0065)		(0.0057)		(0.0058)		(0.0055)		(0.0061)		(0.0057)		(0.0056)	
	S	0.1666	*3*	0.1647	*3*	0.1576	*3*	0.1565	*3*	0.1642	*3*	0.1607	*3*	0.1510	*3*
		(0.0053)		(0.0058)		(0.0049)		(0.0053)		(0.0054)		(0.0051)		(0.0047)	
	NI	0.2160	*1*	0.2132	*1*	0.1992	*1*	0.1960	*1*	0.1994	*1*	0.1979	*1*	0.1774	*1*
		(0.0065)		(0.0071)		(0.0070)		(0.0071)		(0.0079)		(0.0076)		(0.0077)	
$\theta_{n,KM}$	E	0.1725	*4*	0.1733	*4*	0.1732	*4*	0.1754	*3*	0.1725	*4*	0.1736	*4*	0.1766	*4*
		(0.0011)		(0.0011)		(0.0012)		(0.0012)		(0.0012)		(0.0012)		(0.0012)	
	W	0.1749	*3*	0.1764	*3*	0.1792	*3*	0.1750	*4*	0.1738	*3*	0.1743	*3*	0.1851	*1*
		(0.0020)		(0.0020)		(0.0020)		(0.0019)		(0.0020)		(0.0021)		(0.0022)	
	S	0.1836	*1*	0.1793	*2*	0.1856	*1*	0.1797	*2*	0.1803	*2*	0.1796	*2*	0.1829	*2*
		(0.0018)		(0.0020)		(0.0018)		(0.0021)		(0.0018)		(0.0018)		(0.0022)	
	NI	0.1762	*2*	0.1813	*1*	0.1848	*2*	0.1816	*1*	0.1818	*1*	0.1853	*1*	0.1822	*3*
		(0.0022)		(0.0021)		(0.0023)		(0.0024)		(0.0026)		(0.0026)		(0.0028)	

Table 1: *This table consists of multivariate polarization and other inequality indices and their rankings between England, Wales, Scotland and Northern Ireland across a 7-year period from 2002 to 2009. Each index is computed using annual dataset of attributes based on individuals' normalized annual income and highest education level. Standard errors are in parentheses.* $\theta_{n,SV}$ *denotes the multivariate polarization index proposed in this paper.* $\theta_{n,T}$ *denotes the multivariate inequality index proposed in Tsui (1995).* $\theta_{n,KM}$ *denotes the multivariate inequality index proposed in Koshevoy and Mosler (1997).*

Cntry/Yr	2002/3	2003/4	2004/5	2005/6	2006/7	2007/8	2008/9
E	6617	6451	6304	6393	6326	6161	5967
	(5.25)	(5.52)	(5.67)	(5.74)	(5.57)	(5.67)	(5.53)
W	2348	2285	2204	2232	2244	2150	2092
	(7.16)	(7.08)	(6.13)	(5.86)	(5.44)	(6.07)	(6.77)
S	2565	2472	2359	2332	2259	2151	2072
	(5.25)	(5.65)	(4.11)	(4.93)	(4.80)	(4.91)	(4.69)
NI	2425	2333	2145	2108	1854	1769	1491
	(7.05)	(7.46)	(6.50)	(7.30)	(6.88)	(6.85)	(7.74)

Table 2: *This table displays the sample sizes for each country across years we use to construct the indices in Table 2. Each figure in parentheses denotes the percentage of individuals omitted from the original sample due to his/her income taking values either more than 20 times above or below the median value.*