This draft chapter that has been published by Edward Elgar Publishing in *Handbook of Research on Economic and Social Well-Being* by D'Ambrosio, C. (Ed.), published in 2018 [https://www.elgaronline.com/view/edcoll/9781781953709/9781781953709.00024.xml]

Author's pre-print:

Permanyer, Iñaki (2018) "Income and Social Polarization: Theoretical Approaches". D'Ambrosio, C. (Ed.) Handbook of Research on Economic and Social Well-Being. Cheltenham (UK)-Northampton, MA (USA): Edward Elgar Publishing Limited, Chapter 19, pages 434-459. ISBN 9781781953709.

Theoretical Approaches to the Measurement of Income and Social Polarization

1. Introduction

Over the last two decades, we have witnessed a growing interest in the conceptualization and measurement of polarization. This interest can be partly attributable to the existing relationship between polarization, socio-economic stability and economic growth. Loosely speaking, polarization has to do with the clustering of individuals forming groups in different parts of a given distribution – particularly in its extremes. In politics, a society can be said to become more polarized if their members fly away from a moderate 'center position' and become more radicalized towards opposing and irreconcilable poles. In the case of income distributions, societies becomes more polarized when the middle class is hollowed-out and greater proportions of individuals fall in the 'poor' and 'rich' categories. It has been persuasively argued elsewhere that the disappearance of the middle class hinders economic mobility and economic growth. Moreover, highly polarized societies are more prone to experience episodes of social unrest, tension, revolt and even armed conflict.

For many years, it has been common to relate the origins of social conflictivity with high levels of inequality. In this respect, the empirical literature offers – at best – mixed and inconclusive evidence. As will be shown in a companion chapter of this volume, polarization measures have typically performed better than its inequality counterparts when explaining episodes of social tension or conflict. This chapter, however, is of a more theoretical and technical nature. Because of space limitations, the measures presented in this paper will not be discussed in great depth (the interested reader can check the corresponding references for greater detail). Rather, our main goal is to provide a wide overview of the different approaches that have been proposed so far in the conceptualization and measurement of polarization.

As it happens with many other terms related to socio-economic phenomena (e.g.: 'inequality', 'poverty' or 'well-being'), the term 'polarization' means different things to different people. The many and diverse ways in which researchers have tried to formalize the intuitions underlying the notion of 'polarization' has generated a myriad of measures – each of which attempting to approach the same object from different angles – that will be explored in this chapter. During the last twenty years or so when this literature has been blossoming, it has not been uncommon to see some confusion regarding the notions of polarization and inequality because of the close relationship existing among them. This is why many contributions to the polarization literature – particularly the original ones in the mid 90s – have made strong efforts to establish a clear distinction and avoid misunderstandings. As will be seen in this chapter, even if the notion of inequality is intimately embedded in the conceptualization of polarization and it has been particularly influential in the developments of the later, the two ideas are fundamentally different and very often lead to opposing views and results when evaluating alternative states of affairs.

Roughly speaking, most of the contributions to the measurement of polarization can be classified under the headings of 'income polarization' and 'social polarization' measures. Income polarization indices measure the extent to which individuals are clustered around local and antagonistic poles in the income distribution: they will be reviewed in section 2. In that section, we will examine the so-called 'bipolarization' indices and the 'multipolar' ones. Clearly, when a particular income polarization index is chosen, alternative income distributions will be ranked completely – that is: given any couple of distributions the index will be able to establish which one of them is more polarized than the other. However, if we use different indices, there may be different rankings of the distributions. In this context, it is common to seek for classes of indices that produce the same orderings of alternative sets of income distributions. The incipient literature on polarization orderings – which is very much related to the literature on inequality and poverty orderings – will also be explored in section 2.

It has been argued that social tension or conflict may be articulated along certain factors other than income distributions, like culture or biology. These factors typically include ethnicity, race or religion. The measures attempting to capture the notion of polarization when the salient characteristic that divides a given society is of that nature can be classified under the label of 'social polarization'. They will be analyzed in section 3. Given the fact that social tension can take place along economic and non-economic lines, it is also desirable to have polarization measures able to capture both factors simultaneously. These are the so-called 'hybrid polarization' measures, which will be described in section 4. Lastly, section 5 is devoted to future research lines that need to be further investigated.

Throughout this chapter, we will use the following general notation. A polarization index will be denoted with the letter 'P' *and* the corresponding sub-index and super-index. The super-index will typically be an acronym with the first letter of the author(s)' last name and the sub-index will indicate the kind of index we will be referring to. More specifically, P_I will denote an *income* polarization index, P_S a *social* polarization index, P_H a *hybrid* polarization index and P_M a *multidimensional* polarization index.

2. Income Polarization Measures

As many other subfields in welfare analysis, the first polarization measures introduced in the literature were primarily interested in the distribution of income across the population. This section will be devoted to explore the different income polarization measures that have been proposed so far. For that purpose we need to introduce some notation. For a population of size N, an income distribution will be given by a pair

(n, y), where $y = (y_1,\ldots, y_k)$ is the vector of different income levels and $n = (n_1,\ldots, n_k)$ is the vector of corresponding population sizes (i.e.: n_i is the number of individuals with income exactly equal to y_i). Clearly, $\sum_i n_i = N$. Defining $\pi_i := n_i/N$, $\pi = (\pi_1,\ldots, \pi_k)$ is the vector of corresponding population shares. It is assumed that each y_i belongs to a right-open interval of the real line R^1: $[a,+\infty)$ with $a \geq 0$. The set of income distributions for such population will be denoted by D. For any $(n, y) \in D$, the mean and the median will be denoted by $\mu(n, y)$ and $m(n, y)$ respectively (or μ, m for short). We will denote by 1^k the k-coordinated vector of ones. Assuming that the y_i's are ordered non-decreasingly, we denote by y^U (resp. y^L) the vector of such y_i's above (resp. below) the median m. The expressions μ^U and μ^L will be used to denote the mean values of the numbers in y^U and y^L respectively.

An *income polarization index* is defined as a real-valued function $P_I: D \rightarrow R^1$. For all $(n, y) \in D$, the value $P_I(n, y)$ indicates the level of polarization corresponding to the distribution (n, y).

Income polarization indices can be classified in two subgroups: the so-called 'bipolarization' indices (the ones that basically measure the extent to which an income distribution is clustered around the 'poor' and 'rich' poles) and 'multipolar' indices (the ones that measure the extent to which an income distribution has an arbitrary number of antagonistic poles). Before exploring the corresponding indices that have been proposed in the literature (see below) we will briefly examine their normative foundations, that is: the axiomatic properties upon which they are based. Some of these axioms apply for both 'bipolarization' indices and 'multipolar' ones, but some of them are group-specific. The most basic axioms are the following ones:

Normalization: If $(n, y) \in D$ is such that there exists $j \in \{1,\ldots,k\}$ with $n_j = N$, then $P_I(n, y) = 0$.

Scale invariance: For all $(n, y) \in D$ and all scalars $\lambda > 0$, $P_I(n, y) = P_I(n, \lambda y)$.

Translation invariance: For all $(n, y) \in D$ and all scalars $\lambda > 0$, $P_I(n, y) = P_I(n, y+\lambda 1^k)$.

Symmetry: For all $(n, y) \in D$, $P_I(n, y) = P_I(n\Pi, y\Pi)$, where Π is any $k \times k$ permutation matrix.

Population Principle: For all $(n, y) \in D$, $P_I(n, y) = P_I(\lambda n, y)$, where $\lambda > 0$.

Continuity: P_I is a continuous function in its arguments.

These axioms are so mild that they are not able to pin down a specific polarization index. As a matter of fact they are so general that they have also been applied with minor modifications when characterizing inequality or poverty indices, so they will not be discussed here. The following two axioms are the cornerstones upon which income bipolarization measures are based.

Increased spread: Consider any $(n, y) \in D$ and $(n, x) \in D$ such that $m(n, y) = m(n, x) = m$. Consider the following scenarios: (i) There exists $j \in \{1,\ldots,k\}$ such that $x_j<y_j<m$ and $x_i = y_i$ for all $i \neq j$; (ii) There exists $l \in \{1,\ldots,k\}$ such that $m <y_l<x_l$ and $x_i = y_i$ for all $i \neq l$. If either (i), (ii) or both (i) and (ii) are true, then $P_I (n, x) > P_I (n, y)$.

This axiom basically states that greater distancing between the groups below and above the median should make the distribution more polarized (see Figure 1).

[[[Figure 1 around here]]]

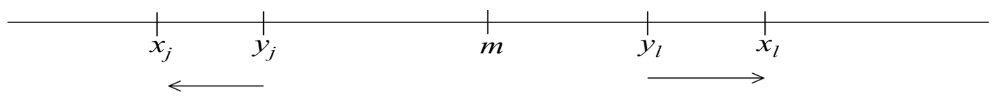

Figure 1. Illustration of 'Increasing Spread'.

Increased bipolarity: Consider any $(n, y) \in D$ and $(n, x) \in D$ such that $m(n, y) = m(n, x) = m$. Consider the following scenarios: (i) x has been obtained from y by a progressive transfer of income from richer person '*b*' to poorer person '*a*' with $y_b<m$; (ii) x has been obtained from y by a progressive transfer of income from richer person '*d*' to poorer person '*c*' with $y_c>m$. If either (i), (ii) or both (i) and (ii) are true, then $P_I (n, x) > P_I (n, y)$.

Increased bipolarity is a clustering or bunching principle. It basically states that when egalitarian transfers between individuals on the same side of the median take place, polarization should increase (see Figure 2). This axiom is what essentially distinguishes polarization from inequality. While all inequality measures satisfying the Pigou-Dalton transfers principle would decrease after the transfer, Increased Bipolarity states that polarization should increase.

[[[Figure 2 around here]]]

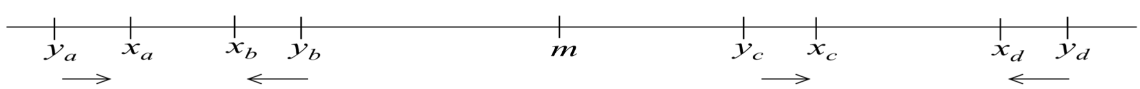

Figure 2. Illustration of 'Increased Bipolarity'.

With these and other similar axioms, different authors have characterized some of the income bipolarization measures we will now review.

2.1. Income bipolarization measures

The first bipolarization measures were proposed in Foster and Wolfson (1992) and later used in Wolfson (1994, 1997). While originally derived from a polarization ordering approach (see section 3), the authors proposed the following bipolarization index:

$$P_I^{FW} = (T - G)\frac{\mu}{m} \qquad [1]$$

where $T = \left(\mu^U - \mu^L\right)/\mu$ and G is the relative Gini index. The value of T is known as the *relative median deviation*. When the distribution is symmetric, $\mu = m$, so bipolarization simply equals the difference between T and G, which is always non-negative. Foster and Wolfson (1992) also showed that P_I^{FW} can be written as

$$P_I^{FW} = (G_B - G_W)\frac{\mu}{m} \qquad [2]$$

where G_B and G_W are the between- and within- group inequality values (as measured with the Gini index) in a partitioning of the population in two groups: those above the median and those below it[1]. Equation [2] clearly shows that in some cases, inequality and polarization can go in opposite directions: *ceteris paribus*, a greater level of within group inequality raises overall inequality but lowers polarization. The bipolarization index P_I^{FW} is simple, intuitive and – as will be shown below – it has been generalized in different directions. To be sure, it should be pointed out that different authors speak about the Wolfson index when referring to P_I^{FW}. The indices shown in equations [1] and [2] are written in *relative* terms (i.e.: the expressions between parentheses are divided by the mean μ). It is straightforward to obtain their absolute counterparts substituting G and T with the absolute Gini and the absolute median deviation respectively.

Another original approach to measure bipolarization was proposed by Wang and Tsui (2000), who suggested the following indices.

$$P_I^{WTa} = \frac{1}{N}\sum_{i=1}^{k} n_i \left|y_i - m\right|^r \qquad [3]$$

$$P_I^{WTr} = \frac{1}{N}\sum_{i=1}^{k} n_i \left|\frac{y_i - m}{m}\right|^r \qquad [4]$$

where $r \in (0,1)$. The Wang and Tsui indices simply aggregate the deviations of the individual incomes from the median when those are measured in absolute or relative terms (equations [3] and [4] respectively). Larger values of parameter r give more importance to very large deviations from the median.

Extensions

---

[1]Since these two subgroups are non-overlapping, one has that overall inequality as measured with the Gini index (G) can be written as $G = G_B + G_W$.

The indices presented in the previous section gave way to a number of extensions and generalizations aiming to improve some of the former deficiencies. As noted in Chakravarty and Majumder (2001) and Chakravarty et al (2007), the indices shown in equations [1]-[4] are purely descriptive regarding the distribution of incomes around the median but they were derived without making any use of welfare concepts. For this reason, these authors suggested using relative and absolute indices of bipolarization using explicit forms of social welfare functions. The relative index proposed by Chakravarty and Majumder (2001) can be written as follows:

$$P_I^{CM} = \frac{\mu^U(1-I(y^U)) + 2\mu^U + \mu^L(1-I(y^L)) - B(m)\mu^L}{2m} + H(m) \qquad [5]$$

where $B(m)$ and $H(m)$ are chosen so that the bipolarization axioms are satisfied and $I(y^U)$ (resp. $I(y^L)$) is the value of an inequality index for the numbers in the vector y^U (resp. y^L). Equation [5] generalizes the Foster-Wolfson index by using other indices of inequality rather than the original Gini (in fact, whenever I is the Gini index, equation [5] essentially reduces to the Foster-Wolfson bipolarization index shown in equations [1], [2]). When alternative inequality formulations like the Atkinson or the Theil indices are chosen, the corresponding polarization index satisfies other ethical principles the decision-maker might be interested in. In a very similar fashion, Chakravarty et al (2007) proposed the following index:

$$P_I^{CMR} = \frac{\mu^U(1-I(y^U)) + 2\mu^U + \mu^L(1-I(y^L)) - \widetilde{B}(m)\mu^L}{2} + \widetilde{H}(m) \qquad [6]$$

where $\widetilde{B}(m)$ and $\widetilde{H}(m)$ serve the same purpose as the $B(m)$ and $H(m)$ in equation [5]. Clearly, [6] is the absolute version of [5]: it can also be interpreted as a generalization of the absolute version of the Foster-Wolfson index.

Also inspired in the Foster-Wolfson index, Rodriguez and Salas (2003) proposed the so-called *extended Wolfson bi-polarization measure* as

$$P_I^{RS} = G_B(v) - G_W(v) \qquad [7]$$

where $G_B(v)$ (resp. $G_W(v)$) is the between group (resp. within group) component associated with the Donaldson and Weymark (1980) extended Gini inequality index[2] assuming that the population is bi-partitioned across the median. In order to satisfy the Increased Bipolarity axiom, Rodriguez and Salas (2003) show that parameter v must lie somewhere in the interval [2, 3]. The idea of defining bipolarization as the difference of between-group and within-group inequality when the population is split in two across the median income was also present in the index proposed by Silber et al. (2007):

[2]In its continuous version, the extended Gini inequality index is defined as $G(v) = 1 - v(v-1)\int_0^1 (1-q)^{v-2} L(q)dq$, where v is an inequality aversion parameter such that $v>1$ and L is the Lorenz curve.

$$P_I^{SDH} = (G_B - G_W)/G \qquad [8]$$

Following the same basic idea, bipolarization has also been defined as

$$P_I^{ZK_G} = G_B / G_W \qquad [9]$$

Clearly, equation [9] can be seen as a 'multiplicative' version of equations [7] and [8] (i.e.: rather than using the *difference* in between-group and within-group inequalities one uses their *ratio*). Equation [9] is very similar to the hybrid polarization measure proposed by Zhang and Kanbur (2001) and presented in equation [41]. The similarity between equations [2], [7], [8] and [9] is clear. As is clear from equation [9], such polarization index can be problematic when there is no within group inequality. In this respect, it is easy to show that $P_I^{SDH} = \left(P_I^{ZK_G} - 1\right) / \left(P_I^{ZK_G} + 1\right)$, so this measure avoids the eventual problem of zero denominators. An axiomatic characterization of a broad class of bipolarization indices can be found in Bossert and Schworm (2008).

Recently, Chakravarty (2009:117) proposed the following bipolarization indices:

$$P_I^{C\varepsilon} = \frac{\left(\frac{1}{N}\sum_{i=1}^{k} n_i \, | m - y_i |^{\varepsilon}\right)^{1/\varepsilon}}{m}, \qquad 0 < \varepsilon < 1 \qquad [10]$$

$P_I^{C\varepsilon}$ is the ratio between the generalized mean of order ε of deviations of individual incomes from the median and the median. The absolute version of equation [10] is given by

$$P_I^{C_a\varepsilon} = \left(\frac{1}{N}\sum_{i=1}^{k} n_i \, | m - y_i |^{\varepsilon}\right)^{1/\varepsilon} = m P_I^{C\varepsilon}, \qquad 0 < \varepsilon < 1 \qquad [11]$$

While not being exactly the same, the indices $P_I^{C\varepsilon}$ and $P_I^{C_a\varepsilon}$ are reminiscent of the bipolarization indices proposed by Wang and Tsui (2000) in equations [3] and [4].

2.2. Polarization orderings

As is well-known, different bipolarization indices might rank alternative distributions in different directions. In some cases, one might be interested in having a more robust procedure to rank distributions that remains unchanged when choosing all possible indices belonging to a particular class. This partial ordering approach will be reviewed in this section, in which we assume that the population weights n_i=1 for all i to simplify notation and define $\overline{N} = (N+1)/2$. Similarly to the Lorenz curve in the inequality framework, different authors have defined polarization curves to rank distributions in terms of polarization in a robust manner.

The first authors to speak about polarization curves were Foster and Wolfson (1992), who defined the so-called *relative polarization curve*. That curve shows the extent to which a given distribution is different from the hypothetical situation in which everybody has an income equal to the median. The ordinate corresponding to the population proportion q/N equals

$$FWC_R(y,q) = \begin{cases} \dfrac{1}{N}\displaystyle\sum_{q\leq i\leq \overline{N}} \dfrac{(m(y)-y_i)}{m(y)} & \text{if} \quad 1\leq q\leq \overline{N} \\ \dfrac{1}{N}\displaystyle\sum_{\overline{N}\leq i\leq q} \dfrac{(y_i-m(y))}{m(y)} & \text{if} \quad \overline{N}\leq q\leq N \end{cases} \qquad [12]$$

Note that the ordinate at $\overline{N}/N$ involves the income level $m(y)$. Whenever q is not an integer, the corresponding ordinate is obtained via linear interpolation (see Chakravarty 2009 for details). For a typical income distribution, FWC_R is decreasing up to $\overline{N}/N$, at $\overline{N}/N$ the curve coincides with the horizontal axis and then it increases monotonically (see Figure 3). In case of an equal distribution ($y_i = m$ for all i), the FWC_R curve corresponds to a horizontal line passing through the origin. Interestingly, the area under FWC_R corresponds to the Foster and Wolfson (1992) bipolarization index shown in equations [1] and [2][3].

[[[Figure 3 around here]]]

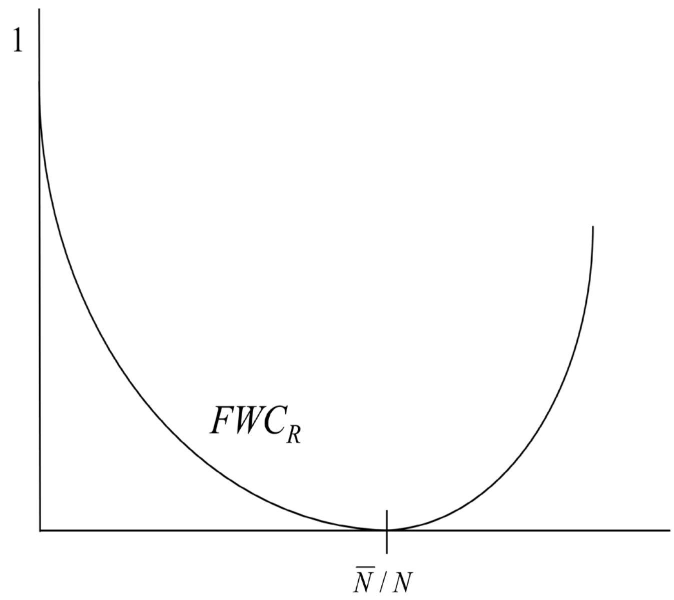

Figure 3. An illustration of a relative polarization curve.

[3] In that paper, the polarization curve was introduced first and the Foster-Wolfson polarization index was defined after it.

The FWC_R curve allows introducing the definition of dominance criteria. For any two income distributions *x*, *y*, we say that '*x* relative *FW* dominates *y*' (written as $x \geq_{FW_R} y$) if and only if FWC_R (*x, q*) ≥ FWC_R (*y, q*) for all $q \in \{1,\ldots,N\}$ and the strict inequality holds at least once. The next one is the kind of result a polarization ordering approach aims to establish:

Theorem 1 (Chakravarty 2009): Let *x*, *y* be a couple of arbitrary income distributions. Then, the following conditions are equivalent: (i) $x \geq_{FW_R} y$; (ii) $P_I(x)>P_I(y)$ for all relative bipolarization indices that satisfy Increased Spread, Increased Bipolarity and Symmetry.

Extensions

The notion of relative polarization curve (FWC_R) can be extended in several directions. The easiest one is to consider an *absolute polarization curve*, which is simply obtained from FWC_R scaling up by the median. This is the approach followed by Chakravarty et al (2007). According to that paper, one has that: (i) the area under the absolute polarization curve corresponds to the absolute version of the Foster and Wolfson index P_I^{FW}; (ii) Theorem 1 can be restated simply using absolute bipolarization indices.

As is known, relative polarization curves (and the corresponding indices) remain unchanged with proportional changes in all incomes and absolute polarization curves (and the corresponding indices) remain unchanged if the same amount is added to all individual incomes. However, there are intermediate concepts of invariance which have been explored in the field of income inequality (see, for instance, Pfingsten 1986). In an interesting contribution, Chakravarty and D'Ambrosio (2010) define an *intermediate polarization curve* as

$$IPC_1(y,q,\lambda)=\begin{cases}\dfrac{1}{N}\displaystyle\sum_{q\leq i\leq \overline{N}}\dfrac{(m(y)-y_i)}{\lambda m(y)+1-\lambda} & if \quad 1\leq q\leq \overline{N}\\ \dfrac{1}{N}\displaystyle\sum_{\overline{N}\leq i\leq q}\dfrac{(y_i-m(y))}{\lambda m(y)+1-\lambda} & if \quad \overline{N}\leq q\leq N\end{cases} \qquad [13]$$

where λ is a real parameter between 0 and 1. Clearly, when $\lambda = 1$, IPC_1 corresponds to the relative polarization curve (FWC_R) and when $\lambda = 0$, IPC_1 corresponds to the absolute polarization curve. Chakravarty and D'Ambrosio (2010) show that: (i) the area under IPC_1 is another bipolarization index (called intermediate polarization index); (ii) Theorem 1 can be generalized for the case of intermediate polarization indices.

In this line, Lasso et al (2010) propose other notions of intermediateness. In particular, they propose the following generalization of the intermediate polarization curve shown in equation [13]:

$$IPC_2(y,q,\lambda,\varepsilon)=\begin{cases}\dfrac{1}{N}\displaystyle\sum_{q\le i\le \overline{N}}\dfrac{(m(y)-y_i)}{(\lambda m(y)+1-\lambda)^{\varepsilon}} & if \quad 1\le q\le \overline{N}\\ \dfrac{1}{N}\displaystyle\sum_{\overline{N}\le i\le q}\dfrac{(y_i-m(y))}{(\lambda m(y)+1-\lambda)^{\varepsilon}} & if \quad \overline{N}\le q\le N\end{cases} \qquad [14]$$

Clearly, when $\varepsilon = 1$, IPC_2 corresponds to IPC_1. This family of curves comes from the adoption of the Krtscha-type notion of intermediateness (Krtscha 1994). In turn, Lasso et al (2010) also show that the area under IPC_2 corresponds to a Krtscha-type intermediate polarization index and elaborate the corresponding version of Theorem 1 for Krtscha-type intermediate bipolarization indices. Interestingly, Lasso et al (2010) also show the conditions under which their intermediate bipolarization indices satisfy the 'polarization version' of the *Unit Consistency*[4] axiom proposed by Zheng in the context of inequality and poverty measurement. That axiom is *not* satisfied by the intermediate polarization indices proposed by Chakravarty and D'Ambrosio (2010).

Lastly, Duclos and Echevin (2005) introduce simple dominance tests for income distributions. Defining $d_y(i)=\left|1-y_i/m\right|$ and $Q_y(\lambda)=N^{-1}\sum_{i=1}^{N}I(d_y(i)\ge\lambda)$–where $I(.)$ is an indicator function that takes the value of 1 if its argument is true and 0 otherwise–the authors suggest that a reasonable way to test bipolarization dominance between any two income distributions x, y is to compare the relative position of the curves $Q_x(\lambda)$ and $Q_y(\lambda)$ for all $\lambda>0$ (see Duclos and Echevin 2005 for details). Recall that $Q_y(\lambda)$ gives the proportion of the population whose proportional distance from the median exceeds λ.

2.3. Income multipolar indices

The different bipolarization measures presented in the previous sections were constructed under the assumption that a society is split into two-equal sized groups: the 'poor' and the 'rich' (i.e.: those below and above the median income). However, this is just one possible partition and one might wonder what happens when alternative grouping of the population are proposed. This intuition has lead to the creation of other polarization indices that will be investigated in this section.

In a fundamental contribution, Esteban and Ray (1994) proposed a polarization measure (henceforth ER) that attempted to measure the extent to which an income distribution is clustered around an arbitrary number of poles. For that purpose, the authors presented the so-called Identification-Alienation approach (IA). According to IA, polarization can be assumed to be equal to the sum of all possible effective antagonisms existing in a given society, which in turn depend on individuals' sense of identification and alienation. On the one hand, individuals are assumed to feel identified with other individuals that are 'similar' to themselves. On the other hand, individuals are assumed

[4] A bipolarization measure P_I is *unit consistent* if for any distributions x, y such that $P_I(x) < P_I(y)$, then $P_I(\theta x) < P_I(\theta y)$ for any $\theta > 0$.

to feel alienated vis-à-vis other individuals that are 'very different'. More specifically, Esteban and Ray (1994) posit that the interpersonal antagonism $T(i,a)$ of a person with an income level x with respect to one with an income level y is the result of her own sense of identity i – which depends on the group size n_i – and of the interpersonal alienation a – which is assumed to depend on the income distance $|y-x|$. Moreover, it is assumed that T is some function increasing in its second argument with $T(0,a) = T(i,0) = 0$. Therefore, the IA approach is summarized in the following hypothesis.

Hypothesis: Polarization in a given society is postulated to be the sum of all effective antagonisms:

$$P_I(n,y) = \sum_i \sum_j n_i n_j T\left(n_i, |y_i - y_j|\right) \qquad [15]$$

This assumption is a bit of a black box and some extra work is necessary to derive it from other – much weaker – axioms. Be that as it may, Esteban and Ray (1994) proposed the following three axioms to pin down an explicit functional form for equation [15] so that it could be implemented empirically[5]. All the three axioms are based on an income distribution constituted by three different values $y_1 = 0 < y_2 < y_3$ and the corresponding population masses n_1, n_2 and n_3.

Axiom ER1: Let $n_1 > n_2 = n_3 > 0$. Fix $n_1 > 0$ and $y_2 > 0$. There exists $c_1 > 0$ and $c_2 > 0$ (possibly depending on n_1 and y_2) such that if $|y_2 - y_3| < c_1$ and $n_2 < c_2 n_1$, then joining of the masses n_2 and n_3 at their mid-point $(y_2 + y_3)/2$ increases polarization.

Axiom ER2: Let $n_1, n_2, n_3 > 0$, $n_1 > n_3$ and $|y_2 - y_3| < y_2$. There exists $c_3 > 0$ such that if n_2 is moved to the right towards n_3 by an amount not exceeding c_3, polarization increases.

Axiom ER3: Let $n_1, n_2, n_3 > 0$, $n_1 = n_3$ and $y_2 = y_3 - y_2 = c_4$. Any new distribution formed by shifting population mass from the central mass n_2 equally to the two lateral masses n_1 and n_3, each c_4 units of distance away, must increase polarization.

Imposing axioms ER1, ER2 and ER3 to the functional form of equation [15], Esteban and Ray (1994) derived the following polarization index:

$$P_I^{ER}(n,y) = c \sum_i \sum_j n_i^{1+\alpha} n_j |y_i - y_j| \qquad [16]$$

where $c > 0$ is a proportionality constant and $\alpha \in (0, 1.6]$. Recall that when $\alpha = 0$, P_I^{ER} corresponds to the (absolute) Gini index. Therefore, α is usually interpreted as a polarization sensitivity parameter: the greater the value of α, the greater the difference between inequality and polarization. Interestingly, the maximum possible value of P_I^{ER} is attained when the population is split in two equal sized income groups, and is

[5] The formulation of these axioms is taken from Chakravarty (2009), who in turn adapted them from Esteban and Ray (1994).

minimal when all population is concentrated on a single income group. In another contribution, Esteban and Ray (1999) show that polarization indices like the one shown in equation [16] arise naturally in behavioral models that link the level and pattern of social conflict to the society-wide distribution of individual characteristics.

Variants and extensions

The polarization index suggested by Esteban and Ray (1994) has been generalized in a number of directions that attempt to overcome some of its shortcomings. One of the problems with the ER framework is that in most real world cases, income distributions are modeled using continuous distributions rather than with a discrete list of income values $\{y_1,\ldots,y_k\}$. In this line, Duclos, Esteban and Ray (2004) proposed an axiomatically characterized continuous version of the *ER* index, which will be referred to as *DER*. In order to characterize the new index, Duclos, Esteban and Ray (2004) introduced the notion of *basic densities*, that is: un-normalized, symmetric, unimodal densities with compact support. Those basic densities are used to model income distributions. The axioms used in that paper are listed below.

Axiom DER1: If a distribution is composed of a single basic density, then a squeeze[6] of that density cannot increase polarization.

After squeezing a basic density, this one is more concentrated towards the mean and the distribution is more homogeneous (see Figure 4), so it seems natural to expect that polarization should reduce in such case.

[[[Figure 4 around here]]]

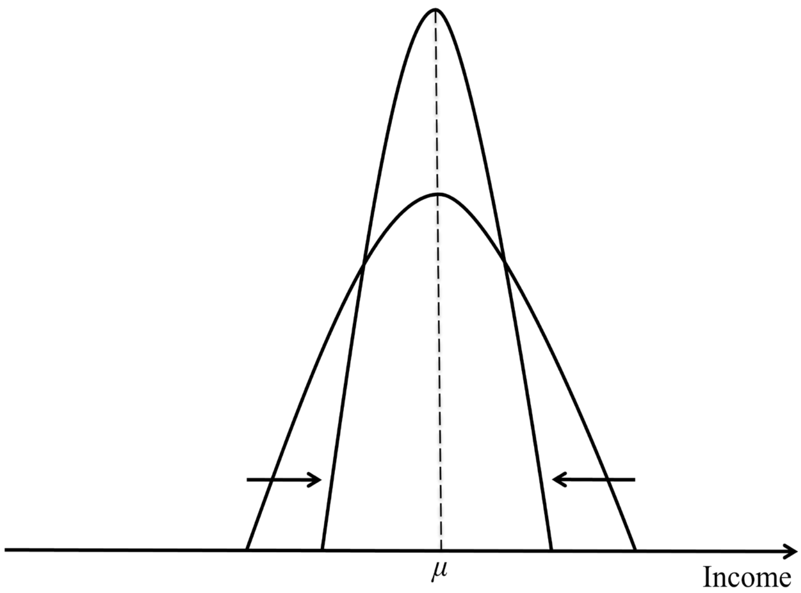

Figure 4. An illustration of a single squeeze.

[6] Technically speaking, a λ-squeeze of a basic density *f* is the following mean-preserving reduction in the spread of *f*: $f^{\lambda}(x) \equiv (1/\lambda) f\big(x-(1-\lambda)\mu/\lambda\big)$, where μ is the mean of *f*.

Axiom DER2: If a symmetric distribution is composed of three basic densities with mutually disjoint supports, then a symmetric squeeze of the side densities cannot reduce polarization.

This is the basic axiom that makes polarization essentially different from inequality measurement. Since the squeezes of the side densities are accomplished via Pigou-Dalton progressive transfers, virtually all inequality measures would reduce, while polarization goes in the opposite direction because more cohesive antagonistic groups are formed after the transfers (see Figure 5).

[[[Figure 5 around here]]]

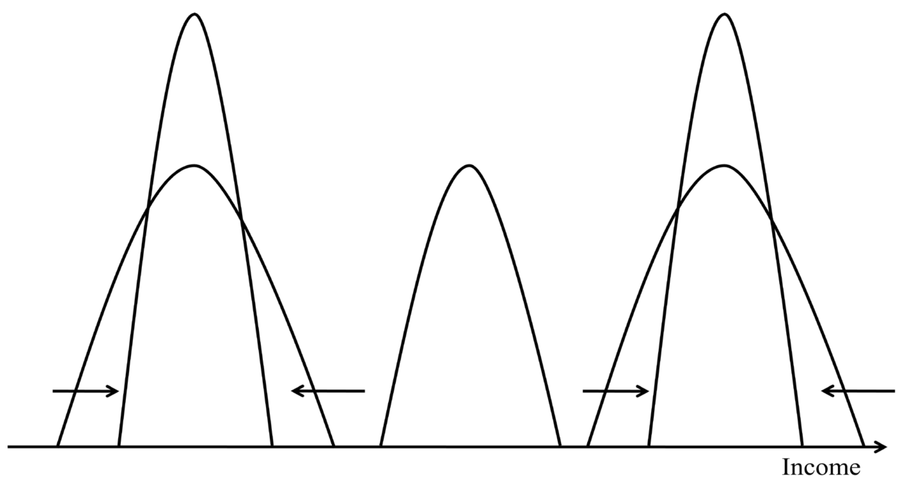

Figure 5. An illustration of a double squeeze.

Axiom DER3: Consider a symmetric distribution composed of four basic densities with mutually disjoint supports, as in Figure 6. Slide the two middle densities to the side as shown (keeping all supports disjoint). Then polarization must increase.

After the slides of the two middle densities to the corresponding sides, it looks as if two more cohesive and antagonistic groups emerged from the distribution: the first corresponding to the two densities to the right and the second to the two densities to the left. Since these changes depart from a uniform-like distribution and approach the bipolar case, it is expected that polarization should increase.

[[[Figure 6 around here]]]

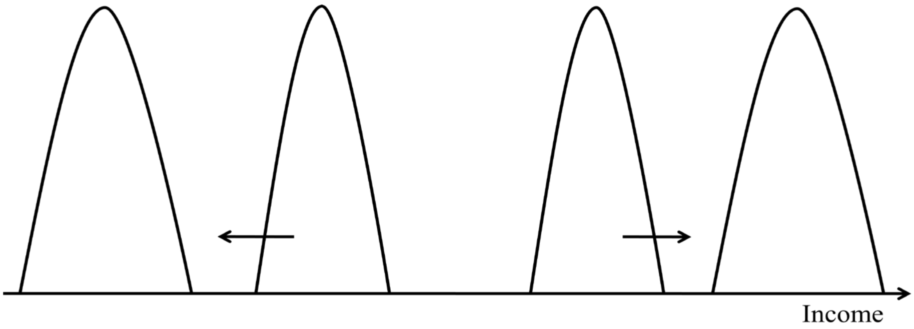

Figure 6. An illustration of a symmetric outward slide.

In a well-known result, Duclos, Esteban and Ray (2004:1744) prove that a polarization index defined in the IA framework satisfies axioms DER1, DER2, DER3 and a continuous version of the population principle if and only if it is proportional to

$$P_I^{DER} \equiv \int\int f^{1+\alpha}(x) f(y) \,|\, x - y \,|\, dxdy \qquad [17]$$

In this formula, $f(x)$ is the density function corresponding to the original income distribution and $\alpha \in [0.25,1]$. As before, when $\alpha = 0$, P_I^{DER} corresponds to the continuous version of the Gini index, so α can be treated as a polarization sensitivity parameter: the larger its values, the greater the difference between polarization and inequality. In order to facilitate the empirical implementation of P_I^{DER}, Abdelkrim and Duclos (2007) have introduced a polarization module in their distributive analysis software DASP ('Distributive Analysis Stata Package') which can be freely downloaded from http://dasp.ecn.ulaval.ca. In an interesting contribution, Esteban and Ray (2012) discuss the axiomatic characterization of *DER* and compare it with other characterizations presented in the literature for other well-known measures like the Foster and Wolfson (1992) and the Wang and Tsui (2000) indices shown in equations [1] to [4].

Another criticism directed against the *ER* and *DER* indices is that individuals are assumed to feel identified only with other individuals with *exactly* the same income level. However, it has been argued that such identification might also exist when the differences in individuals' incomes are relatively 'small'. In order to put into practice this idea, Esteban et al (2007) proposed a statistical approach that can be summarized as follows. Given an income distribution modeled by a density function $f(x)$, an n-spike representation of f is a collection ρ of numbers ($y_0, y_1,\ldots,y_n; \pi_1,\ldots, \pi_n; \mu_1,\ldots, \mu_n$) such that $y_0<\ldots<y_n$ and

$$\pi_i = \int_{y_{i-1}}^{y_i} f(y)dy \qquad [18]$$

$$\mu_i = \frac{1}{\pi_i}\int_{y_{i-1}}^{y_i} yf(y)dy \qquad [19]$$

for all i = 1,…,n. Each n-spike representation ρ of f induces an approximation error, which in Esteban et al (2007) is defined as

$$\varepsilon(f,\rho)=\frac{1}{2}\sum_{i}\int_{y_{i-1}}^{y_i}\int_{y_{i-1}}^{y_i}|x-z|\,f(x)f(z)dxdz \qquad [20]$$

Esteban et al (2007) propose to work with the n-spike representation that minimizes this error, which is denoted by ρ^*. With that representation, the polarization index they propose can be written as follows:

$$P_I^{EGR}(f,\alpha,\beta)=ER(\alpha,\rho^*)-\beta\varepsilon(f,\rho^*) \qquad [21]$$

where $ER(\alpha,\rho^*)$ and $\varepsilon(f,\rho^*)$ are the ER index and error terms associated to ρ^*; α is the polarization sensitivity parameter and β is a free parameter which measures the weight we attach to the 'measurement error'. The income polarization index shown in equation [21] is commonly referred to as the Esteban-Gradín-Ray (EGR) index. Clearly, when β = 0 EGR reduces to ER. Interestingly, for the special case where there are only two groups (n = 2) and $\alpha = \beta = 1$, one has that

$$P_I^{EGR}(f,1,1)=\frac{m}{2}P_I^{FW} \qquad [22]$$

That is: the Foster and Wolfson (1992) index of bipolarization can be seen as a particular case of the EGR index. This approach has the virtue of casting both measures – each of which derived from a completely different perspective – in the context of a statistically unified framework. Again, the EGR index can be implemented empirically using the software package DASP implemented by Abdelkrim and Duclos (2007).

Assuming that the income distribution is partitioned into different groups, Lasso de la Vega and Urrutia (2006) have proposed a variant of the EGR index that can be written as follows:

$$P_I^{LU}=\sum_{i}\sum_{j}\pi_i^{1+\alpha}\pi_j(1-G_i)^{\theta}|y_i-y_j| \qquad [23]$$

In this equation, G_i stands out as the Gini index for group i and $\theta \geq 0$ is a constant representing the sensitivity towards group cohesion. Introducing the term $(1-G_i)^{\theta}$, the P_I^{LU} index is also sensitive to within group dispersion.

3. Social Polarization

It has been argued that social tension might arise not only because of particular characteristics of the income distribution, but by other salient characteristics like ethnicity or religiosity. In this respect, several researchers have attempted to measure polarization on the basis of alternative groupings of the population (typically based on ethnic or religious lines) that do not depend exclusively on the distribution of income or wealth. These constructs will be referred to as 'social polarization measures'. One of the most well-known examples of a social polarization index was introduced by Reynal-Querol (2002) and is defined as

$$RQ = 1 - \sum_{i=1}^{k} \left(\frac{1/2 - \pi_i}{1/2} \right)^2 \pi_i = 4 \sum_{i=1}^{k} \pi_i^2 (1 - \pi_i) \qquad [24]$$

where we assume there are k exogenously given groups with population shares π_i. According to Montalvo and Reynal-Querol (2002), the *RQ* index satisfies the following basic properties:

Property 1: If there are three groups with shares π_1, π_2 and π_3, and $\pi_1 > \pi_2 \geq \pi_3$, then if we merge the two smallest groups into a new group, the new distribution is more polarized than the original one.

Property 2: Suppose there are two groups with shares π_1, π_2. Take one of the groups and split it into $g \geq 2$ groups in such a way that $\pi_1 = \tilde{\pi}_1 \geq \tilde{\pi}_i \quad \forall i = 2 \ldots, g+1$, with strict inequality for at least one i, where $\tilde{\pi}$ is the new vector of population shares. Then polarization under $\tilde{\pi}$ is smaller than under π.

Property 3: Assume there are three groups with shares π_1, π_2 and $\pi_3 = \pi_1$. Then, if we shift mass from the second group equally to the other two groups, polarization increases.

These properties are reminiscent of the different axioms used to characterize the income polarization index P_I^{ER}. It should be noted that *RQ* takes its maximal value of 1 when a population is split in two equal-sized groups (the bipolar case) and it takes its minimal value of 0 when there is only a single group. As a matter of fact, the original purpose of the index was to capture how far the distribution of the different groups is from the (1/2,0,…,0,1/2) bipolar distribution, so *RQ* can be interpreted as an index measuring 'how bipolar' a given population distribution is. In this respect, *RQ* differs substantially from other well-known heterogeneity measures that have been widely used in the literature, like the Fractionalization index shown below.

$$FRAC = 1 - \sum_{i=1}^{k} \pi_i^2 = \sum_{i=1}^{k} \pi_i (1 - \pi_i) \qquad [25]$$

As is known, *FRAC* should be interpreted as the probability that two randomly selected individuals do not belong to the same group. Keeping all else constant, when we increase arbitrarily the number of groups (k), *RQ* will decrease but *FRAC* will increase, so the two measures are fundamentally different.

Axiomatic characterization of RQ

In its original formulation, the *RQ* index was not axiomatically characterized (observe that *RQ* satisfies properties 1, 2 and 3, but these do not characterize the index univocally). In an attempt to fill this gap, Chakravarty and Maharaj (2011) characterized axiomatically *RQ*, a useful exercise that is important to fully understand the normative foundations upon which an index is based. In that paper, the authors posit that a social polarization index has to be of the following form:

$$P_S^{CM}(\pi_1,\ldots,\pi_k)=\sum_{i=1}^{k}\psi(\pi_i) \qquad [26]$$

where ψ:[0,1]→R is a continuous function called 'influence function' and $\psi(\pi_i)$ is assumed to represent the impact of group i on overall polarization. In this context, Chakravarty and Maharaj (2011) propose the following axioms.

Axiom S1: For all k and all $\pi=(\pi_1,\ldots,\pi_k)$, $0\le P_S^{CM}\le 1$.

Axiom S2: For all k, $P_S^{CM}(\pi)=0$ if π is some permutation of (1,0,…,0).

Axiom S3: For all k, $P_S^{CM}(\pi)=1$ if π is some permutation of (1/2,1/2,0,…,0).

Axiom S4: Consider the distribution $\pi=(\pi_1,\ldots,\pi_k)$ and the distribution $\pi'=((\pi_1+\pi_2)/2, (\pi_1+\pi_2)/2, \pi_3,\ldots, \pi_k)$. Then the polarization difference $P_S^{CM}(\pi)-P_S^{CM}(\pi')$ can be expressed as

$$P_S^{CM}(\pi)-P_S^{CM}(\pi')=(\pi_1-\pi_2)^2 f(\pi_1+\pi_2) \qquad [27]$$

where f:[0,1]→R is a continuous function.

Axiom S5: $P_S^{CM}(1/k,\ldots,1/k)$ becomes arbitrarily small for sufficiently large k.

All these axioms – which are named with the letter 'S' standing for 'social' – are quite clear and simple (except perhaps for axiom S4, which might not look as intuitive as the other ones). With them, the following result can be presented.

Theorem 2 (Chakravarty and Maharaj 2011): Let P_S^{CM} be a function as in [26] for which the influence function ψ is twice continuously differentiable. Then the following statements are equivalent:

(*i*) P_S^{CM} satisfies axioms S1, S2, S3, S4 and Property 1.

(*ii*) P_S^{CM} satisfies axioms S1, S2, S3, S4 and Property 2.

(*iii*) P_S^{CM} satisfies axioms S2, S3, S4 and Property 3.

(*iv*) P_S^{CM} satisfies axioms S2, S3, S4 and S5.

(v) P_S^{CM} is precisely the index RQ.

Moreover, Chakravarty and Maharaj (2011) also show that the different sets of axioms are independent (see their Theorem 7).

Variants and Extensions

One of the greatest advantages of the RQ index is its simplicity: its values can be easily calculated simply knowing the population shares of the subgroups $\pi_1,\ldots,\pi_k$. In turn, this leaves room for a significant number of extensions in alternative directions. For instance, Chakravarty and Maharaj (2012) have recently proposed the so-called 'Generalized RQ-index of order θ', which is defined as:

$$RQ_\theta = 4\sum_{i=1}^{k}\pi_i^2(1-\pi_i)+\theta\sum_{1\le i_1<i_2<i_3\le k}\pi_{i_1}\pi_{i_2}\pi_{i_3} \qquad [28]$$

where $\theta\in[0,3]$ and $k\geq 3$. The first term appearing in equation [28] gives a multiple of the probability that out of three randomly selected individuals, two will belong to a single group and the third to another one. The second term gives a multiple of the probability that all three individuals belong to three different groups. Since RQ_θ is a linear combination of both probabilities, its values can be interpreted as an extent of heterogeneity of a population partitioned in k groups. Clearly, when $\theta = 0$, RQ_θ reduces to RQ.

In Montalvo and Reynal-Querol (2002), the RQ index is presented as a particular case of the so-called 'discrete polarization index'

$$P_S^{DP}(\alpha,\theta)=\theta\sum_{i=1}^{k}\sum_{j\neq i}\pi_i^{1+\alpha}\pi_j \qquad [29]$$

It is trivial to show that $P_S^{DP}(1,4)=RQ$. This equation bears a lot of resemblance to the ER income polarization index shown in equation [16]. Rather than using the term $\left|y_i-y_j\right|$ for measuring distance between groups, equation [29] uses a discrete metric that takes a value of 1 when two individuals belong to different groups and 0 otherwise.

It can be argued that the discrete distance measure might be too crude to capture appropriately the existing distances between alternative social groups. In this respect, Permanyer (2012) proposes to greatly enrich such distance measure in the following way. Firstly, it is postulated that social polarization is greatly influenced by the extent to which different individuals feel identified with the group to which they belong. The intensity with which individuals feel to belong to their particular group is called *radicalism degree*, and it is measured with a real number x ($x \in R^+$). For each population group i there is an un-normalized density function $f_i(x)$ that measures the way in which radicalism degrees are distributed therein. We denote by $f=(f_1(x),\ldots,f_k(x))$ the collection of the k density functions. In order to operationalize the IA approach, it is postulated that an individual belonging to group i with radicalism degree x has a feeling of identification proportional to $f_i(x)$. Regarding alienation, two different assumptions are made: one of them posits that alienation can only be felt between members of different groups[7] (assumption A1) and the second one states that alienation can *also* be felt between members of the same group[8] (assumption A2). Under assumption A1, Permanyer (2012) axiomatically characterizes the following social polarization index:

$$P_S^{P1}(f) = \sum_{i=1}^{k} \sum_{j \neq i} \pi_i^{1+\alpha} \pi_j (\mu_i + \mu_j) \qquad [30]$$

where μ_i is the mean of radicalism degrees' density function $f_i(x)$ and $\alpha \in (0,1]$. As usual, α is interpreted as a polarization sensitivity parameter. Recall this measure can be seen as a generalization of the discrete social polarization measure $P_S^{DP}(\alpha,\theta)$ shown in equation [29]: the absence of an alienation component has been substituted by a much richer structure that is sensitive to individuals' radicalism distributions in equation [30]. Clearly, if $\mu_i = \mu_j \forall i \neq j, P_S^{P1} \equiv P_S^{DP}(\alpha,\theta)$. Interestingly, [30] also bears some resemblance with equation [42] (see below), the difference being that in equation [30] one deals with radicalism degrees distributions while in equation [42] one deals with income distributions.

Since one could argue that alienation might operate not only between individuals of different groups but also between individuals of the same group (assumption A2), Permanyer (2012) also characterizes axiomatically the following index:

$$P_S^{P2}(f) = \sum_{i=1}^{k} \int \int f_i^{1+\alpha}(x) f_i(y) \,|\, x - y \,|\, dxdy + \sum_{i=1}^{k} \sum_{j \neq i} \int \int f_i^{1+\alpha}(x) f_j(y)(x+y) dxdy \quad [31]$$

where the polarization sensitivity parameter α is bounded between 1/2 and 1. This formulation bears some resemblance with the income polarization index *DER* shown in equation [17]. As before, however, equation [31] is based on radicalism degrees distributions but equation [17] is based on income distributions.

The main challenge when attempting to implement empirically the polarization indices shown in equations [30] and [31] is to measure radicalism degrees and their distribution $f_i(x)$. In the empirical application shown in Permanyer (2012), religious radicalism degree is proxied applying Principal Components Analysis to a set of attitudinal questions regarding religiosity. In order to facilitate the computation of equations [30] and [31], Abdelkrim and Duclos (2007) have implemented a specific module in the software package DASP.

Polarization measures on the basis of ordinal information

[7] In case radicalism degrees of different individuals from different groups equal x and y, their alienation is assumed to be $x+y$.

[8] In case radicalism degrees of different individuals from the same group equal x and y, their alienation is assumed to be $|x-y|$.

So far, the polarization indices shown in this chapter have been computed on the basis of cardinal or categorical data. In the former category we can find the income polarization indices of section 2 plus the social polarization indices shown in equations [30]-[31]. In the latter category we can find the other social polarization indices. However, in many circumstances important dimensions might be coded with ordinal variables (a typical example is that of health, which in many circumstances is measured via respondents' self-assessments in an ordinal scale). Interestingly, other socio-economic indices have recently made room for ordinal variables as well, for instance in the field of inequality or multidimensional poverty measurement (e.g.: Allison and Foster 2004, Alkire and Foster 2011 among many others).

In this subsection we will show a couple of polarization measures that are meant to be used with ordinal information. Before that, we introduce some basic notation. The number of categories will be denoted by *k*, and the number of individuals in category '*i*' by n_i. As before, let $\pi_i = n_i \Big/ \sum_{i=1}^{k} n_i$. For each category $c \in \{1,\ldots,k\}$ we define $N_c = \sum_{i=1}^{c} n_i$ and $\Pi_c = \sum_{i=1}^{c} \pi_i$. In an original contribution, Apouey (2007) presented the following polarization indices on the basis of ordinal data:

$$P_S^{A1} = \theta \left[\left(\frac{N_k}{2} \right)^{\beta} - \frac{1}{k-1} \sum_{c=1}^{k-1} \left| N_k \left(\frac{N_c}{N_k} - \frac{1}{2} \right) \right|^{\beta} \right] \qquad [32]$$

$$P_S^{A2} = 1 - \frac{2^{\beta}}{k-1} \sum_{c=1}^{k-1} \left| \Pi_c - \frac{1}{2} \right|^{\beta} \qquad [33]$$

where θ is a strictly positive constant and $\beta \in (0,1)$. In equation [32], θ is a proportionality constant that can be used to normalize the values of the index into the desired range. In equations [32] and [33], parameter β reflects the importance that is given to the median category. When $\beta \to 0$ (resp. $\beta \to 1$), the relative contribution of the median category increases (resp. decreases) whereas the relative contribution of the other categories decrease (resp. increase).

The ordinal polarization indices shown in equations [32] and [33] are inspired in Wang and Tsui (2000) cardinal bipolarization measures (see equations [3] and [4]). That is, it is assumed that the population is divided in two subgroups: those below the median and those above it. In this respect, the polarization indices P_S^{A1}, P_S^{A2} are meant to satisfy the ordinal version of the axioms of Increased Spread and Increased Bipolarity (see Apouey 2007 for details). Essentially, both P_S^{A1} and P_S^{A2} measure how far a given distribution is from the extreme bipolar case where half the population is concentrated in the lower end of the distribution and the other half in the upper extreme. Therefore, both measures are maximized in the bipolar case and minimized in the case where all population is concentrated in a single category. This property is shared with all income bipolarization measures shown in section 2.1 and with the *RQ* index.

In a recent contribution, D'Ambrosio and Permanyer (2012) presented a new class of social polarization indices meant for those cases where an attribute of interest (e.g.: health, happiness or degree of satisfaction) is measured in an ordinal scale *and* the population is assumed to be partitioned in exogenously given groups $g_1,\ldots,g_k$ (e.g.: along ethnic, religious or other lines). In that paper, the authors use the IA approach, so polarization is assumed to be proportional to the sum of all effective antagonisms in the population. As is usual, individuals' identification is assumed to be proportional to the size of the group to which they belong. Regarding alienation, the authors propose two different alternatives: symmetric and asymmetric alienation. For symmetric alienation between groups g_i, g_j, the 'discrete overlap coefficient' has been proposed:

$$\theta_{ij} = \sum_{c=1}^{k} \min\{p_{g_i,c}, p_{g_j,c}\} \qquad [34]$$

Where $p_{g_i,c}$ is the share of group g_i in category c. This coefficient lies between 0 (disjoint groups) and 1 (perfectly-overlapping groups). Alienation is then defined as $1-\theta_{ij}$, taking the value 0 when the groups overlap completely and 1 when the groups are completely disjoint. The greater the degree of overlap, the more similar the groups are, and hence the less the degree of alienation between them. Alternatively, it has been argued that feelings of alienation between groups should not necessarily be reciprocal. Consider, say, a comparison between a poor and a rich individual: while the poor person might have reason to feel animosity towards the rich person, the opposite might not hold necessarily. In this context, alienation between groups '*i*' and '*j*' can be defined as a function of:

$$A_{ij} = \frac{\sum_{s=1}^{N_i}\sum_{t=1}^{N_j}\delta_{st}}{N_i N_j} \qquad [35]$$

where δ_{st} equals 1 if individual '*i*' from group *i* is ranked below individual '*t*' from group *j* and 0 otherwise. This procedure yields an asymmetric function ($A_{ij} \neq A_{ji}$), consistent with the alienation felt from underprivileged towards more privileged groups not necessarily being reciprocated (this contrasts with traditional income-polarization measures, where alienation is always symmetric). The value of A_{ij} measures the extent to which group g_i is underprivileged with respect to group g_j. When $A_{ij} = 1$, all of the members of group g_i are ranked below any member of group g_j with respect to the ordinal attribute we take into consideration: this is the case of maximal alienation. At the other extreme, $A_{ij} = 0$ when no member of group g_i is ranked below any member of group g_j, which refers to minimal alienation.

With these symmetric and asymmetric alienation functions, D'Ambrosio and Permanyer (2012) axiomatically characterized the following social polarization measures:

$$P_S^{AP1} \equiv \sum_{s=1}^{k}\sum_{t=1}^{k} \pi_s^{1+\alpha}\pi_t(1-\theta_{st}) \qquad [36]$$

$$P_S^{AP2} \equiv \sum_{s=1}^{k}\sum_{t=1}^{k} \pi_s^{1+\alpha}\pi_t A_{st} \qquad [37]$$

where $\alpha \in [\alpha^*,1],\, with\; \alpha^* = \frac{2-\log_2 3}{\log_2 3 - 1} \approx 0.71$ is the polarization sensitivity parameter. Once again, these polarization indices can be seen as generalizations of the polarization index $P_S^{DP}(\alpha,\theta)$ shown in equation [29] where the discrete distance function has been substituted by a richer structure that is sensitive to the way in which a given attribute of interest is distributed across individuals and groups.

4. Hybrid Polarization measures

The income polarization measures shown in section 2 describe the extent to which the distribution of income is clustered around certain poles. Following a different approach, the social polarization indices shown in section 3 basically measure the extent to which a population is clustered around certain poles defined on the basis of non-income characteristics like race, ethnicity or religion. In this section, we describe other polarization indices that lie somewhere in between of the aforementioned intuitions, that is: the distribution of a given attribute – typically income – *as well as* other qualitative characteristics play a non-trivial role when defining the groups among which polarization is going to be measured. For this reason, this kind of indices will be referred to as 'hybrid'. In this respect, the polarization indices P_S^{AP1} and P_S^{AP2} shown in equations [36] and [37] could also be considered as hybrid because the distribution of an (ordinal) attribute *and* the existence of exogenously given groups are relevant when defining the alienation component.

Zhang and Kanbur (2001) defined a simple and intuitive hybrid index of polarization on the basis of the Generalized Entropy (*GE*) index of inequality (Shorrocks 1980). If we assume that the population has *n* members, the index is written as

$$I(y) = \begin{cases} \sum_{i=1}^{n} f(y_i)\left[\left(\frac{y_i}{\mu}\right)^c - 1\right] & if \quad c \neq 0,1 \\ \sum_{i=1}^{n} f(y_i)\left(\frac{y_i}{\mu}\right)\log\left(\frac{y_i}{\mu}\right) & if \quad c = 1 \\ \sum_{i=1}^{n} f(y_i)\log\left(\frac{y_i}{\mu}\right) & if \quad c = 0 \end{cases} \qquad [38]$$

Recall that *GE* is additively decomposable, that is: if we assume that the population is split in *k* groups, total inequality can be written as the sum of withing group and between group inequality. This is generally written as:

$$I(y)=\sum_{g=1}^{k} w_g I_g + I(\mu_1 e_1,\ldots,\mu_k e_k) \qquad [39]$$

where e_g is a vector of ones of length n_g, μ_g is the mean of group g, I_g is the internal inequality within group g and

$$w_g = \begin{cases} \frac{n_g}{N}\left(\frac{\mu_g}{\mu}\right)^c & if \quad c \neq 0,1 \\ \frac{n_g}{N}\left(\frac{\mu_g}{\mu}\right) & if \quad c = 1 \\ \frac{n_g}{N} & if \quad c = 0 \end{cases} \qquad [40]$$

Taking advantage of this well-known inequality index, Zhang and Kanbur (2001) proposed the following hybrid polarization measure:

$$P_H^{ZK} = \frac{I(\mu_1 e_1,\ldots,\mu_k e_k)}{\sum_{g=1}^{k} w_g I_g} \qquad [41]$$

This simple index satisfies the most basic intuitions regarding a polarization index, namely: it increases with larger between group inequality and it decreases with larger within group inequality. It must be emphasized that the income polarization index shown in equation [9] is very similar to the index shown in [41]: in the former, it is assumed that the population is split in two non-overlapping groups – those above the median income and those below it – and the inequality measure that is used is the Gini index rather than GE[9].

In an interesting contribution, D'Ambrosio (2001) proposed other hybrid polarization measures inspired in the IA approach. Essentially, it is assumed that the population is split in k relevant groups, each of which with the corresponding income distribution $f_i(x)$. With this notation, the D'Ambrosio (2001) hybrid polarization measure can be written as

$$P_H^A = \sum_i \sum_j \pi_i^{1+\alpha} \pi_j Kov_{ij} \qquad [42]$$

where $Kov_{ij} = \frac{1}{2}\int |f_i(y) - f_j(y)| dy$ is the Kolmogorov measure of variation distance. This is a measure of the lack of overlapping between income distributions $f_i(x)$ and $f_j(x)$. $Kov_{ij} = 0$ if $f_i(x) = f_j(x)$ for all x and Kov_{ij} reaches its maximum of 1 when $f_i(x)$ and $f_j(x)$ do not overlap. There is a clear relationship between the Kolmogorov measure of

[9] The Gini index is not additively decomposable (i.e.: it does not satisfy equation [39]) in general. However, in the particular case where the different groups are non-overlapping, it is possible to write the Gini as the sum of between group and within group inequalities.

variation distance and the overlap coefficient shown in equation [34], which can be rewritten in its continuous version as

$$\theta_{ij} = \int_{-\infty}^{\infty} \min\{f_i(x), f_j(x)\} dx \qquad [43]$$

The statistical properties of the overlap coefficient θ_{ij} and some of its multivariate generalizations have been explored in detail by Anderson et al (2009). Other interesting hybrid polarization measures have been proposed – albeit not axiomatically characterized – in Duclos Esteban and Ray (2004). Following the IA approach, the authors consider the case in which notions of identification are mediated not just by group membership but also by income similarities as well, while the antagonism equation remains untouched. Then they get what one might call social polarization with income-mediated identification, which can be written as

$$P_H^{DER1} = \sum_{j=1}^{k} (1 - n_j) \int_x f_j(x)^{\alpha} dF_j(x) \qquad [44]$$

Duclos Esteban and Ray (2004) also make the assumption that alienation can also be income-mediated (for alienation, two individuals must belong to different groups and have different incomes). According to the authors, "*groups have only a demarcating role – they are necessary (but not sufficient) for identity, and they are necessary (but not sufficient) for alienation*" (Duclos Esteban and Ray 2004, p. 1760). The corresponding index would look as follows

$$P_H^{DER2} = \sum_{j=1}^{k} \sum_{k \neq j} \int_x \int_y f_j(x)^{\alpha} |x - y| dF_j(x) dF_k(y) \qquad [45]$$

5. Future research lines

As has been shown in this literature review, the measurement of polarization has expanded considerably in the last two decades since the seminal contributions of Foster and Wolfson (1994) and Esteban and Ray (1994). However, much works remains to be done and we aim to finish this chapter indicating some research lines in polarization measurement that deserve to be further explored.

Among the most promising lines of research in the field of polarization measurement we find that of multidimensional polarization. As has been the case with inequality or poverty measurement, it is possible to conceive the notion of polarization in a multidimensional setting where attributes other than income are taken into account when trying to identify clusters and distances between them. In this respect, it is worth highlighting that there have already been some initial attempts to produce such measures. In an innovative contribution, Gigliarano and Mosler (2009) introduce a polarization index defined on the basis of many attributes at the same time. If we denote by X an $n \times k$ matrix with non-negative entries (n being the number of individuals in the society and k the number of attributes that are being taken into account), the authors posit that multidimensional polarization should be written as follows:

$$P_M^{GM}(X) = \varsigma(B(X), W(X), S(X)) \qquad [46]$$

where $B(X)$ and $W(X)$ are indices that measure between and within group inequality respectively, $S(X)$ is an index of relative group size and ς is a function that increases on B and S and decreases on W. In other words, the authors construct multivariate polarization indices using the between and within decomposition by subgroups of certain indices of multivariate inequality. Clearly, this formulation is an attempt to generalize the functional forms of the polarization indices shown in equations [2], [7], [8], [9] and [41]. In order to be able to obtain meaningful between and within group inequality decompositions, Gigliarano and Mosler (2009) use different inequality decompositions via Generalized Entropy indices, multivariate extensions of Atkinson's inequality index and Gini decompositions where the supports of the different groups have no geometric overlap. To simplify matters, the authors further assume that equation [46] must take one of the following forms:

$$P_M^{GM1}(X) = \phi\left(\frac{B(X)}{W(X)+c}\right)S(X) \qquad [47]$$

$$P_M^{GM2}(X) = \psi\big(B(X) - W(X)\big)S(X) \qquad [48]$$

$$P_M^{GM3}(X) = \tau\left(\frac{B(X)}{B(X)+W(X)+c}\right)S(X) \qquad [49]$$

Recall these are quite general functional forms that include relatively similar versions of the polarization indices of Foster and Wolfson (1994), Zheng and Kanbur (2001) or Silber et al (2007) as particular cases. Interestingly, Gigliarano and Mosler (2009) also propose a list of desirable axioms a multidimensional polarization index should satisfy. Among these, we highlight a couple of them that are specifically multidimensional: the 'between groups correlation increasing majorization property' and the 'within groups correlation increasing majorization property'. Both of them describe the way in which a polarization index should respond under correlation increasing switches among attributes[10]. Despite the interest that these measures have, it is fair to say that multidimensional polarization measurement is still in its infancy and much work remains to be done. Among other things, the polarization indices shown in [46]-[49] are not axiomatically characterized and their structure is somewhat arbitrary. It would be interesting to derive other – more general – functional forms from weaker assumptions than those imposed in the aforementioned equations.

---

[10] In the multi-attribute welfare literature, a *correlation increasing switch* is defined as follows. Assume we are comparing two individuals i and j in a two-dimensional achievement space associated with attributes a and b. Assume also that i has more of a but less of b than j does. If we interchange the achievements in attribute b between the two persons, now individual i has more of a and more of b, so there has been an increase in the correlation of the attributes. Such change is called a 'correlation increasing switch'.

In this line, another topic that to our knowledge has never been explored before and which might be interesting to investigate in the future is the possibility of defining multidimensional polarization orderings. The literature of inequality and poverty measurement has already introduced the notions of multidimensional inequality and multidimensional poverty orderings (e.g.: Duclos, Sahn and Younger 2011, Bourguignon and Chakravarty 2008). One might just wonder whether it might be possible to generalize the existing ideas on polarization orderings to a multivariate setting.

After reading the existing list of polarization indices, it seems clear that there is room for much further improvement in the measurement of polarization on the basis of ordinal information. In many cases the variables of interest are not cardinal, particularly when they refer to non-income dimensions (e.g.: health or educational attainment). This issue has also been highlighted by Alkire and Foster (2011), who argue that in multidimensional poverty measurement it is very useful to define indices admitting ordinal structures. To date, there are only a couple of measures defined on an ordinal basis (see above) and these measures can be expanded in many different directions. In particular, it might also be of interest to follow Alkire and Foster's approach and define multidimensional polarization indices that can be constructed with non-cardinal variables.

An important issue that definitely needs to be further explored is the definition of appropriate socio-economic distance functions between groups when these are defined along a variety of economic or social lines (e.g.: wealth, income, class, race, ethnicity or religion). As we have seen in different sections of this chapter, the literature has already proposed different ways of approaching this issue but none of them has gained widespread adherence. In this respect, we anticipate that the methodology used to measure distances in trees (for instance: linguistic trees, see Greenberg 1956, Fearon and Laitin 2000, Laitin 2000, Desmet, Ortuño-Ortin and Weber 2009, and Desmet, Ortuño-Ortin and Wacziarg 2012) can be useful to generate distance/alienation functions between groups. Similarly, we speculate that the ideas behind the construction of diversity theories (e.g.: Witzman 1992, 1998, Nehring and Puppe 2002, 2004) can be fruitful when attempting to define compelling inter-group distance functions.

References

Abdelkrim, A. and J-Y. Duclos (2007), *DASP: Distributive Analysis Stata Package*, PEP, World Bank, UNDP and Université Laval.

Alkire, S. and J. Foster (2011), 'Counting and multidimensional poverty measurement', *Journal of Public Economics* 95 (7-8), 476-487.

Allison, R. and J. Foster (2004), 'Measuring Health Inequality Using Qualitative Data', *Journal of Health Economics*, 23, 505-524.

Anderson, G., Y. Ge and L. Wah (2009), 'Distributional Overlap: Simple, Multivariate, Parametric and Non-Parametric Tests for Alienation, Convergence and general distributional difference issues', *Econometric Reviews*, 29 (3), 247-275.

Apouey, B. (2007), 'Measuring health polarization with self-assessed health data', *Health Economics,* 16, 875-894.

Bossert, W. and W. Schworm (2008), 'A class of two-group polarization measures', *Journal of Public Economic Theory*, 10, 1169-1187.

Bourguignon, François and Satya Chakravarty (2008), 'Multi-dimensional poverty orderings: Theory and applications', in Basu, K. and Kanbur, R. (Eds.), *Essays for a better world: Essays in honor of Amartya Sen*, Volume I, chapter 18, Oxford University Press.

Chakravarty, Satya (2009), *Inequality, polarization and poverty. Advances in distributional analysis*, Springer, Berlin.

Chakravarty, S. and A. Majumder (2001), 'Inequality, polarization and welfare: theory and applications', *Australian Economic Papers*, 40.

Chakravarty S., A. Majumder and R. Roy (2007), 'A treatment of absolute indices of polarization', *The Japanese Economic Review,* 58, 273–293.

Chakravarty, S. and C. D'Ambrosio (2010), 'Polarization orderings of income distributions', *Review of Income and Wealth,* 56 (1), 47-64.

Chakravarty, S. and B. Maharaj (2011), 'Measuring ethnic polarization', *Social Choice and Welfare,* 37, 431-452.

Chakravarty, S. and B. Maharaj (2012), 'Ethnic polarization orderings and indices', *Journal of Economic Interaction and Coordination,* 7, 99-123.

D'Ambrosio, C. (2001), 'Household characteristics and the distribution of income in Italy: An application of social distance measures', *Review of Income and Wealth,* 47(1), 43-64.

D'Ambrosio, C. and I. Permanyer (2012), 'Measuring social polarization with ordinal and categorical data', Mimeo (forthcoming in the *Journal of Public Economic Theory*).

Desmet, K., I. Ortuño-Ortín, and S. Weber (2009), 'Linguistic Diversity and Redistribution', *Journal of the European Economic Association,* 7 (6), 1291–318.

Desmet, K., I. Ortuño-Ortín and R. Wacziarg (2012), 'The Political Economy of Ethnolinguistic Cleavages', *Journal of Development Economics,* 97 (1), 322–32.

Donaldson, D., and J. Weymark (1980), 'A single-parameter generalization of the Gini indices of inequality', *Journal of Economic Theory,* 22, 67-86.

Duclos, J.Y. and D. Echevin (2005), 'Bipolarization comparisons', *Economics Letters,* 87, 249-258.

Duclos, J-Y., J. Esteban and D. Ray (2004), 'Polarization: Concepts, Measurement, Estimation', *Econometrica,* 72, 1737-1772.

Duclos, J-Y., D. Sahn and S. Younger (2011), 'Partial multidimensional inequality orderings', *Journal of Public Economics,* 95(3-4), 225-238.

Esteban, J., C. Gradín and D. Ray (2007), 'An Extension of a Measure of Polarization, with an application to the income distribution of five OECD countries', *Journal of Economic Inequality,* 5, 1-19.

Esteban, J. and D. Ray (1994), 'On the Measurement of Polarization', *Econometrica,* 62, 819–852.

Esteban, J. and D. Ray (1999), 'Conflict and distribution', *Journal of Economic Theory,* 87, 379-415.

Esteban, Joan and Debraj Ray (2012), 'Comparing Polarization Measures", in M. Garfinkel and S. Skaperdas, Oxford Handbook of the Economics of Peace and Conflict, Oxford University Press: Oxford, 2012, 127-151.

Fearon, J. and D. Laitin (2000), 'Violence and the Social Construction of Ethnic Identity: Review Essay', *International Organization,* 54 (4), 845–77.

Foster, J. and M. Wolfson (1992), 'Polarization and the decline of the middle class: Canada and the US', Mimeo, Vanderbilt University (now published in the *Journal of Economic Inequality* 2010, 8:247-273).

Gigliarano, C. and K. Mosler (2009), 'Constructing indices of multivariate polarization', *Journal of Economic Inequality,* 7, 435-460.

Greenberg, J. (1956), 'The Measurement of Linguistic Diversity', *Language* 32 (1), 109–15.

Krtscha, M. (1994), 'A new compromise measure of inequality', in W. Eichhorn (ed.), *Models and measurement of welfare and inequality*, Springer, Heidelberg.

Laitin, D. (2000), 'What Is a Language Community?', *American Journal of Political Science,* 44 (1), 142–55.

Lasso, C., A. Urrutia and H. Diez (2010), 'Unit consistency and bipolarization of income distributions', *Review of Income and Wealth,* 56(1), 65-83.

Lasso, C. and A. Urrutia (2006), 'An alternative formulation of the Esteban-Gradin-Ray extended measure of bipolarization', *Journal of Income Distribution,* 15, 42–54.

Montalvo, J., and M. Reynal-Querol (2002), 'Why ethnic fractionalization? Polarization, ethnic conflict and growth', Universitat Pompeu Fabra, Economics Working Paper No. 660.

Nehring, K. and C. Puppe (2002), 'A Theory of Diversity', *Econometrica,* 70(3), 1155-1198.

Nehring, K. and C. Puppe, (2004), 'Modelling phylogenetic diversity', *Resource and Energy Economics,* 26, 205-235.

Permanyer, I. (2012), 'The Conceptualization and Measurement of Social Polarization', *Journal of Economic Inequality,* 10(1), 45-74.

Pfingsten, A. (1986), 'Distributionally-neutral tax changes for different inequality concepts', *Journal of Public Economics,* 30, 385-393.

Reynal-Querol, M. (2002), 'Ethnicity, political systems and civil wars', *Journal of Conflict Resolution,* 46(1), 29-54.

Rodriguez J. and R. Salas (2003), 'Extended bipolarization and inequality measures', *Research on Economic Inequality*, 9, 69–83.

Shorrocks, A. (1980), 'The class of additively decomposable inequality measures', *Econometrica,* 48, 613-625.

Silber, J., J. Deutsch and M. Hanoka (2007), 'On the link between the concepts of kurtosis and bipolarization', *Economics Bulletin,* 4, 1–6.

Wang, Y. and K. Tsui (2000), 'Polarization Orderings and New Classes of Polarization Indices', *Journal of Public Economic Theory,* 2, 349–363.

Weitzman, M. (1992), 'On Diversity', *The Quarterly Journal of Economics,* 107 (2), 363-405.

Weitzman, M. (1998) ‘The Noah’s Ark Problem’, *Econometrica,* 66 (6), 1279-1298.

Wolfson, M. (1994), ‘When Inequalities Diverge’, *American Economic Review P&P*, 94, 353-358.

Wolfson, M. (1997), ‘Divergent Inequalities: theory and empirical results’, *Review of Income and Wealth,* 43, 401-421.

Zhang, X. and R. Kanbur (2001), ‘What differences do polarization measures make? An application to China’, *Journal of Development Studies*, 37, 85-98.