A MULTIRESOLUTION ANALYSIS OF INCOME POLARIZATION

Palacios-González, Federico fpalacio@ugr.es
García-Fernández, Rosa María rosamgf@ugr.es

Department of Quantitative Methods in Economics
University of Granada
Campus de Cartuja s/n.
Granada, 18071, Spain

Extended Abstract

A review of the literature related to the income distribution shows the increasing interest in the measurement of income polarization since the end of the last century. This interest is motivated by the evolution in income distribution of different countries, partly as a consequence of social and technological changes. The advantages of the use of new technologies tend to fall on a part of society solely showing bimodalities in the income distribution which imply a diminishing weight of the middle class. This fact has important political repercussions. On the one hand affects to the design of the redistributive policy increasing the size of the social transfers. On the other hand the existence of differentiated poles in society could generate conflict between classes and consequently political instability.

Following Esteban and Ray (1994) the polarization is focused on the extent to which population is grouped around a small number of poles. According to these authors if there is a high degree of homogeneity within each group and a high degree of heterogeneity across groups society is polarized. This idea shows the difference between polarization and inequality. An increase in the degree of homogeneity within each group should be considered as diminishing in inequality while the polarization should increase. Indeed inequality measures the concentration around the global mean whereas polarization focuses on the concentration around different poles. For this reason inequality fits better with unimodal distributions and polarization with multimodal distributions.

The previous comments emphasize the importance of providing an answer of how polarized a distribution is. This answer requires firstly to define a concept of polarization and secondly to propose a class of measures to quantify the polarization of a distribution. In line

with these ideas different measures of polarization have been proposed. The measures defined by Esteban and Ray (1994), Esteban, Gradín and Ray (1999), Gradín (2000) and Duclos, Esteban and Ray (2005) are developed following what they called an identity-alienation framework. Esteban and Ray (1994) through an axiomatic approach defined a measure of polarization based on the sum of antagonisms between individuals that belong to different groups. Duclos, Esteban and Ray (2005) developed the axiomatic of the polarization and provided a measure derived from the sum of antagonisms for the case in which income distributions are described by density functions. Esteban, Gradín and Ray (1999) proposed an extension of the Esteban and Ray measure which corrects the error that appears when the distribution is arranged into groups. The measure of polarization proposed by Wolson (1994) is derived from the Lorenz curve and it shows how distant the distribution is from the symmetric bimodal. It is appropriate to measure the bipolarization which can be obtained as a particular case of the measure proposed by Esteban, Gradín and Ray. Wang and Tsui (1998) defined a class of polarization indices using two partial ordering axioms of "increased bipolarity" and "increased spread". These measures will be explained in detail in a later section.

An important issue when measuring polarization is how to determine the size of the group, to which each individual belongs, and its location. In the papers of Esteban and Ray (1994), Esteban, Gradín, and Ray (1999) and Gradín (2000) the number of poles is chosen exogenously according to standard economic categories and their locations are determined minimizing the error which is expressed as the average of income distances within all groups. Duclos, Esteban and Ray (2005) estimated group size non-parametrically making use of kernel density procedures.

The aim of this paper is to develop an approach to estimate the size of the groups and their locations at different levels of resolution, and to provide a measure of polarization following the definition as well as the identification-alienation framework proposed by Esteban and Ray.

The tool proposed is a family of probability density functions which are obtained by mixing dilations and translation (over a regular partition on a closed interval [a, b]) of a box spline of degree three (Mallat, S. 1999). This is a flexible family of density functions which

is able to model a great variety of income distributions. Moreover as we will show, this family is by its own construction appropriate to describe polarization.

Assume that the income distribution is built over a closed interval $[a,b]$ that contains the sample data and which is partitioned at *m* regular segments. Let $\theta(x)$ be a box spline of degree three. The family of densities f is defined by the expression:

$$f(x) = \sum_{k=0}^{m} a_{mk}\theta_{mk}(x) \qquad (1)$$

where $a_{mk} > 0 \quad \forall k = 0,1,...,m; \sum_{k=0}^{n} a_{mk} = 1$ and $\theta_{mk}(x) = s\theta\big(s(x-a)-k\big)$ with $s = \frac{m}{b-a}$.

Observe that $\theta_{mk}(x)$ is a translation of $\theta(x)$ toward the point $a + \frac{k}{s}$ and dilation by the factor $s^{-1} = \frac{b-a}{m}$.

The scale parameter *m* or *s* determines the level of resolution. For each value of *m*, the coefficients of the model are estimated by maximum likelihood.

Each $\theta_{mk}(x)$ is an "atomic" density located at level of income $a + \frac{k}{s}$. The coefficients a_{mk} are interpreted as the share of population for a given value of *m*.

Several coefficients located together and distinct from zero capture a sub-population and may generate a not necessarily symmetric unimodal income distribution. The sum of such coefficients is equal, in proportion, to the size of the sub-population.

Different groups of coefficients isolated by coefficients equal to zero or with a lower relative importance than the rest of the coefficients may define different sub-populations, showing the polarization of the distribution. Several levels of resolution will be used to show how population is clustered around determinate levels of income. An optimum value of *m* will be obtained making use of cross-validation procedure which will be applied to locate the poles. Furthermore the optimum resolution will define the border between population group and random sample grouping.

The features of the model defined make it especially useful to study polarization. It allows us to identify, by its own method of estimation, those sub-populations whose incomes are concentrated around poles. As a consequence the number of poles can be

established from the income distribution. This is especially useful when the number of poles cannot be determined exogenously according to standard economic categories.

The model proposed is more advantageous than the nonparametric estimation of the density using kernel estimation procedures. For instance, one advantage is the coefficients of this model are estimated by maximum likelihood, with the qualities of this procedure. A further advantage is that this density model can be as flexible and adaptable as the nonparametric method and it has capacities for local analysis that are not found in the kernel procedures.

The potential of the proposed procedure has been proved empirically, by means of simulated and sample data, and demonstrates how population is grouped around different levels of income as well as the estimated size of the groups.

Two samples have been simulated whose sizes are 400 and 200 respectively. The former comes from a beta distribution with parameters $(a = 6; b = 24; p = 2; q = 3)$ and mean 13.2, that is: $f(x) \propto K(x-6)(24-x)^2$. The latter comes from a beta distribution with parameters $(a' = 25; b' = 50; p' = 3; q' = 2)$ and mean equal to 40. This is equivalent to simulate a stratified sample over the density obtained as a convex linear combination of two beta distributions with coefficients equal to 2/3 and 1/3 respectively. These beta distributions were combined to obtain the density of the sample. Model (1) was estimated for this sample considering a wider range of the level of resolution. The estimation of the size and location of the groups allow us to observe the two groups that form the sample as well as their importance. For instance, for $m = 10$, 66.7% of the population is grouped around an estimated mean income of 12.93 and 33.3% is grouped around an estimated mean income of 40.07 which demonstrates that the model defined fits well to the data.

In addition the approach proposed facilitates the calculation of the inter-group and intra-group variance. They will be used to provide a measure of polarization following an identification-alienation framework. This measure is defined as a function which is increasing with the variance inter-group and decreasing with respect to the variance intra-group, the number of groups and their importance.

Keywords: Box-spine of degree three, level of resolution, polarization.