Utilization of mixed distributions in the calculation of polarization.

The case of Spain.

Authors:

Ana Karina ALFARO* (University of Alcalá)

José Javier NÚÑEZ VELAZQUEZ (University of Alcalá)

* Corresponding author

E-mails: anakarinaalfaro@hotmail.com

Tel: (0034) 630 428 202

Fax: (0034) 918 854 239

Address: Universidad de Alcalá
Facultad de Ciencias Económicas, Empresariales y Turismo
Departamento de Economía
Plaza de la Victoria, 2
(28802) Alcalá de Henares
Madrid

The author acknowledges financial support from Madrid's regional Government for her research activities (Project S2015/HUM-3416-DEPOPOR-CM)

Keywords: Mixed distributions; polarization; unemployment.

JEL classification: D63, D31, C10 and J60

Abstract: This study analyses the evolution of polarization in the Spanish labor market in the period 2004-2012, to examine the consequences of the crisis on one of the European markets where the rate of unemployment has had greatest impact. The proposal of the paper is to use mixed distributions, with the objective of analyzing separately the population of the unemployed and their contribution in the calculations performed. Thus, in the present article the polarization indexes traditionally used in the literature are calculated taking into account the unemployed, in distinction to analogous studies in which the unemployed are not considered.

1. INTRODUCTION

Polarization is a concept which emerged in the 1980s, due to the impossibility of measuring the disappearance of the middle class with the measurements of inequality familiar until then. At its outset, the disappearance of the middle class was related to the concept of a greater inequality of income and inequality indexes were used for its measurement. However, these indicators proved inadequate to measure this phenomenon because the results obtained differed according to the measurements used (Levy and Murnane, 1992; Blackburn and Bloom, 1986; Horrigan and Haugen, 1988, among others). Furthermore, it was clear that greater polarization could be produced while inequality was reduced, or vice versa. On the basis of these facts there began to emerge a new branch of research in the area of the analysis of income distribution, dedicated to the identification and calculation of polarization. From among the pioneering studies in this field can be cited Wolfson (1994) or Esteban and Ray (1994) who, in addition to formalizing the measurement of polarization, demonstrated the differences between this concept and inequality and gave examples in which the two measurements presented opposing behavior. An extreme case of inequality would be, for example, a society in which one person possessed all the income of a region and the rest of the population had none. Such a society is unequal, but nevertheless is not polarized given that the measurements of polarization take into account the size of the groups (the person possessing all the income does not have a significant size). Similarly, the measurements of inequality are not capable of distinguishing whether the population is concentrated around a pole (the average), or instead around two or more poles. Thus, the interest of these two concepts has distinct origins:

- Inequality, the measurement of equity. The measurement of inequality is based on the Pigou-Dalton transfer principle, which establishes that any transfer of income from one individual to another with fewer resources reduces inequality as long as the initial order between the two is preserved (Núñez, 2006).

- Polarization, the concern for social cohesion. Polarization measures whether a society is divided into groups of significant size. A society is more polarized and more prone to social conflict when the groups into which it is divided are more similar internally and less equal among them (Gradín, 2000; Alfaro, 2005, among others).

The term polarization has expanded along two lines:

- As a particular form of inequality, in which the extremes of the distribution increase at the cost of the intermediate levels (Autor et al., 2006).
- In the formation of small groups, internally homogenous and different among them, within a distribution (Bernat, 2006; Ezcurra, 2009, among others).

Nevertheless, it is the second definition which best permits differentiation of the concepts of inequality and polarization. Greater internal homogeneity in a group signifies greater polarization and, in turn, less inequality. Thus, briefly:

- Polarization: measures concentration around different poles (multimodal distributions)
- Inequality: concentration around the average (unimodal distributions).

Taking into account that the origin of the concept of polarization resides in the preoccupation for social cohesion, and the worrying situation in Spain due to its high unemployment rates, this study is intended to find answers to the following questions: How has salary polarization evolved in Spain in the period 2004-2012? Should specific consideration be made of the fact of having a significant quantity of unemployed (that is to say, with a salary of zero)? If the unemployed are taken into account in the calculation of polarization, should the group of unemployed be separated in the calculation? In order to answer these questions, and principally the last two, this article proposes a methodology which attempts to adapt the measure of polarization in order to

disaggregate the unemployed through the use of mixed distributions, permitting the separation of the contribution associated with this particular group. Furthermore, analysis will be made of whether this methodology contributes correctly to the treatment of mixed distributions in comparison with the results obtained if complete distributions are considered.

With the aim of performing the research detailed above, this study is structured in the following way. The next section presents the measurements used for the calculation of polarization. Similarly, it presents the theoretical framework of mixed distributions and the adaptation performed of the measurements of polarization in order to consider the inclusion of these distributions in their calculation. The third section is divided into two parts: the first offers the results of the analysis of polarization obtained for Spain, using both complete and mixed distributions, and the second presents the contribution of mixed distributions to the calculation of polarization. Finally, the concluding comments are included in the fourth section.

2. THE MEASUREMENT OF POLARIZATION

The important changes observed in the distribution of income in diverse countries in the 1980s generated great controversy. The principal debate was motivated by the definition and quantification of the middle class. In studies of labor economics in that decade, the analysis of the disappearance of the middle class was based on the use of indicators of inequality (Levy and Murnane, 1992; Blackburn and Bloom, 1986, for example). However, in many such studies the conclusions were insufficiently robust, as they compared the sensitivity of the results with alternative specifications of the middle class or employed varying techniques to measure its size (Haugen, 1988, among others).

It was Esteban and Ray (1994) who presented a more general notion and formally expounded the phenomenon of polarization, presenting their concept and the derivation from its measurement (for the discrete case). Thus, the polarization index they propose is:

$$(1) \quad ER(\alpha, y) = \sum_{i=1}^k \sum_{j=1}^k \pi_i^{1+\alpha} \pi_j |y_i - y_j|$$

where π_i , π_j , y_i and y_j are the proportions of the groups i and j , and their respective income, while $0 < \alpha \leq 1,6$ is a sensitivity parameter with regard to polarization. The fact that the population weights are raised to a value greater than unity is what differentiates the polarization measurement from that of inequality, since the greater its value the greater will be the difference of behavior between these measurements. Thus, for $\alpha=0$ this measure equals the Gini index. The fact that α must be greater than 0 imposes a minimum distancing from the Gini index.

2.1. The measurement proposed by Esteban, Gradín and Ray

However, Esteban and Ray (1994) measurement poses certain problems which limit its practical application. With the aim of solving these problems, Esteban, Gradín and Ray (2007) present an extension for the notion of polarization which makes its calculation more operative, as it does not depend on the form in which the data are presented. Furthermore, they demonstrate that the measurements of Wolfson and of Esteban and Ray are particular cases of this approach.

The measurement proposed by Esteban and Ray (*ER* hereinafter) does not take into account the loss of information occurring when intra-group information is not contemplated. For this reason, Esteban, Gradín and Ray (*EGR* hereinafter) propose a criterion for grouping individuals and thereby succeed in incorporating the information concerning intra-group distribution. Initially,

they define a simplified representation of the original distribution in a series of mutually exclusive groups.

It is assumed that $f(x)$ represents the density function of the variable to be studied, transformed in such a way that its average is 1 and its variation interval is $[a, b]$. Thus, ϱ is the simplified representation of $f(x)$ in k groups, such that its characterization would be: $\rho = (z_0, z_1, \dots, z_k; \pi_1, \pi_2, \dots, \pi_k; \mu_1, \mu_2, \dots, \mu_k)$, donde $a = z_0 < z_1 \dots < z_k = b$ y para $i=1, 2, \dots, k$.

That is to say, z are the cut-off points, π the proportions of persons in each group, μ the average income of each group and k the quantity of groups formed.

On using ϱ in place of f , the representation induces an approximation error $\varepsilon(f, \varrho)$, due to the degree of internal dispersion of the groups (internal heterogeneity), which may be interpreted as the lack of their identification. Esteban, Gradín and Ray define this error in terms of the average of the distances among all the components of each group. Thus, this can be rewritten as the intra-group component of the decomposition into groups of the Gini index, if the subgroup income range does not overlap (Lambert and Aronson, 1993).

Consequently, it can be deduced that the error of approximation is identified with the inequality of distribution f minus the inequality among groups given by ϱ , both measures using the Gini index. In this way:

$$(2) \quad \varepsilon(f; \rho) = G(f) - G(\rho)$$

where G denotes the Gini index.

Expressing the index in the case of the optimal representation ϱ^* , the measurement of the extended (or generalized) measurement of polarization proposed by Esteban, Gradín and Ray is obtained:

$$(3) \quad \mathbf{EGR}(\mathbf{f}; \alpha, \beta, \rho^*) = \mathbf{ER}(\alpha, \rho^*) - \beta[\mathbf{G}(\mathbf{f}) - \mathbf{G}(\rho^*)]$$

Measurement is first made of the simplified polarization, which is the ER calculated in the representation ϱ , and then the degree of internal heterogeneity is discounted. β is a parameter, greater than or equal to zero, which measures the weight given to the lack of identification produced when a simplified distribution is used in place of the original.

Thus, when implementing the EGR approach:

- Exogenous choice is made of the number k of groups, according to whether we are interested in studying a bimodal, trimodal, etc. distribution, and
- The partition which minimizes the aggregate error induced (heterogeneity among groups) is obtained endogenously.

Esteban, Gradín and Ray consider as exogenous the selection of the number of groups permitting the application of the method proposed by Aghevli and Mehran (1981) and Davies and Shorrocks (1989). It is thereby intended to obtain the optimal partition which minimizes the loss of information resulting from grouping. The solution of this problem coincides with the minimization of the area existing between the Lorenz curves of the original distribution and the simplified distribution. Thus, the expression which determines the cut-off point can be written as:

$$(4) \quad z_i^* = \frac{\pi_i^* \mu_i^* + \pi_{i+1}^* \mu_{i+1}^*}{\pi_i^* + \pi_{i+1}^*}$$

That is to say, the cut-off point between two adjacent groups is equal to the combined average of the two groups. This means that on the basis of this expression it can be observed that for the case of bipolarization, z must equal μ , μ being average income.

2.2. Mixed distributions and the adaptation of polarization measurements

One of the principal questions when it comes to calculating a measure of polarization is the determination of the number of groups to be formed and the cut-off points which determine them. In the period under study, Spain has passed from a stage of economic growth to an important recession, with serious consequences for unemployment. Thus, according to the Spanish Statistical Office (INE), it moved from having the lowest unemployment rate in its entire democratic history, approximately 8%, in 2007, to surmount 24% in 2012.

With this reality in mind, the principal interest of this study is centered on the analysis of salary polarization, considering the unemployed and their zero salaries due to their being an important part of the social tension which the polarization measure attempts to quantify. In such situations, the study population responds to a mixture of two distributions. One of these is a discrete probabilistic mass taking the value of zero and another is a continuous distribution of positive values (Aitchison, 1955). The literature denominates these cases a non-standard mixture of two populations. Thus, it considers a random variable X , with the following distribution of probability:

- ✓ The probability of X taking the value of 0 is $1-p$ and, consequently, the probability of X taking a value distinct from 0 is p .

- ✓ The distribution of X is conditioned on $X > 0$ proceeding from a continuous random variable with positive values, designated by Z, with a density function $f(z)$.

In general, it can be expressed as:

$$X = \begin{cases} 0 & 1 - p \\ Z > 0 & p \end{cases}$$

Consequently, its density function takes the form:

$$f_X(x) = (1 - p) \cdot I_0(x) + p \cdot f_Z(x), \quad x \geq 0$$

$I_0(\cdot)$ being the indicator function at point 0. The distribution function is determined by:

$$F_X(x) = 1 - p + F_Z(x), \quad x > 0$$

$$F_X(0) = P(X \leq 0) = 1 - p$$

Let μ_z and σ_z^2 be the average and the variance of the distribution of Z, and let, μ_x and σ_x^2 be the corresponding parameters of the variable X. Consequently:

$$\mu_x = p \cdot E[Z] = p \cdot \mu_z$$

$$\sigma_x^2 = E(x^2) - [E(x)]^2 = p \cdot (\sigma_z^2 + \mu_z^2) - p^2 \mu_z^2 = p \sigma_z^2 + p \cdot (1 - p) \mu_z^2$$

Thus, for the calculation of the partitions of the groups, this study uses both the optimum partitions already mentioned (Aghevli and Mehran, 1981 and Davies and Shorrocks, 1989)

and that generated by taking into account the mixed distributions. The objective of applying these two approximations is to compare the results of the polarization obtained in the two cases.

3. EMPIRICAL EVIDENCE: POLARIZATION IN SPAIN

A study is now made of polarization in Spain during the period for which data are available from the Living Conditions Survey (ECV), that is to say between 2004 and 2012, utilizing both complete and mixed distributions, since that corresponding to the unemployed who receive no salary can be considered to be degenerated in 0.

3.1. Database and analysis of the distribution of salaries

The database used is the ECV between the years 2004 and 2012, which corresponds to the data for Spain from the Statistics on Income and Living Conditions survey of the European Union (EU-SILC) developed by the Statistical Office of the European Union (Eurostat) and harmonized for all the countries of the EU, making it possible to effect international comparisons in future extensions of the present study.

In Spain, in a context such as the present one, high rates of unemployment are a clear sign of social tension. Accordingly, this study has obtained the indexes of polarization proposed by Esteban and Ray (1994) and by Esteban, Gradín and Ray (2007), on the basis of complete distributions of salaries. All these measurements are computed with normalized net monthly salaries. That is to say, salaries are normalized with regard to their average, giving a result of $\mu=1$. One of the contributions of this study is to consider in the distribution of salaries those which are zero, that is to say the entire active population (workers plus the unemployed) is studied, with the

exception of the self-employed. It is considered that the unemployed do not perceive any salary, given that salary is understood to be a continuous flow and not a sum received sporadically. Thus, another of the contributions of this study consists of considering mixed distributions in the calculation of polarization. The objective of this contribution consists of analyzing how the indexes obtained are calculated if the group of unemployed is considered separately in the calculation of these indicators.

As a first step, it is interesting to present the density functions of the normalized net monthly salaries. A Kernel function was used to obtain an estimation of the density of each type of income, with an Epanechnikov nucleus and the size of the window determined endogenously (**Figure 1**). Since standard Kernel methods have issues regarding the boundaries, mixed distribution are used in these graphics too. Therefore, zero salaries are excluded from Kernel distribution shown in **Figure 1** and the probability of having a zero salary is shown at the bottom of each graph. Thus, these graphs present the densities of the normalized net monthly salaries, where the line found in value one is the average.

Figure 1

At first sight, it can be appreciated how the distribution of salaries in Spain is asymmetric to the right and that the early years present a marked pointedness in the mean zone of salaries' distribution, which progressively increases until 2007 but decreases in 2008-2012. The opposite happens with probabilities of zeros (unemployed workers), which diminishes until 2007 and increases since 2008. Consequently, it can be deduced that with the passage of time, two well differentiated groups have been formed, the unemployed and salaried workers. For this reason, it is desirable to study whether the formation of these two groups has provoked greater polarization in Spain. By proposing the use of mixed distributions it is possible to separate zero

salaries from the rest of the distribution of salaries. This type of distributions is used so that, on determining the optimal partitions, the zero salaries do not contaminate the formation of the groups.

An important question in the calculation of polarization is the number of poles to consider. Although this decision remains at the criterion of the analyst, two factors must be taken into account. On the one hand, beyond three or four groups the improvements obtained in the precision of the representation are very scanty. On the other hand, although the number of poles to be considered is an exogenous decision, the form of the distribution or the nature of the problem it is wished to study condition this choice. In this article, the population is divided into up to four groups, in an attempt to represent the distinct social classes there may be in society. That is to say, on dividing the population into two groups two social classes are assumed, one low and the other high. The representation of society in the low, middle and high classes is sought by dividing the population into three. And, finally, the purpose of dividing society into four is to compare four social classes: the working lower class, which is divided into two, lower and middle-lower and the upper class, which is separated into middle-upper and upper. Having determined the number of poles, their placement is calculated endogenously, using the method proposed by Davies and Shorrocks (1983).

3.1.1. Results obtained from the complete distribution of salaries

The polarization indexes of *EGR* (2007), obtained from the complete distribution of salaries, are presented in

Table 1 (**Figure 2** in the appendix).

Table 1

Before beginning the analysis of the results obtained, it should be emphasized that the period studied can be divided into two stages, the first being that taking place from the year 2004 until 2007, a period of economic growth in Spain in which the lowest unemployment rate of the entire democratic period (8.2%) was achieved, and the second stage, from 2008 to 2012, marked by a sharp economic crisis which caused the unemployment rate to increase considerably in Spain, reaching 24.3% in 2012 according to the data from the Active Population Survey supplied by the INE.

This said, it is not surprising that the indexes calculated from *EGR* decrease until 2007 and in 2008 began to increase for all the groups and for all the values considered from α and β . The choices of the values of α and β are for purposes of comparison with related studies. The greater level of polarization considering the *EGR* index is obtained by dividing the population into three groups.

On analyzing the internal heterogeneity of the groups, that is to say the error terms, it is observed that, considering two groups, this fell until 2007 and from 2008 on began to increase. When considering three and four groups the tendency, at a general level, is to slightly decrease in the period analyzed, such that the members of the groups feel more identified with each other with the passage of time. As was to be expected, the groups are more heterogeneous when the population is divided into two groups and more homogenous when divided into four.

Table 2

The optimal partitions of the distribution of salaries are presented in **Table 2**. When there are two groups in question, it can be observed that the proportions of persons remained stable until 2007; however, from 2008 onwards, the persons with lower salaries have progressively decreased in proportion, while those with larger salaries have gained in the proportion of the population. This change of trend has meant that in the last two years analyzed there is a greater proportion of persons with higher comparative salaries. It should be remembered that the salaries used are relative to the average of the population and thus salaries are not discussed in nominal terms. Consequently, with regard to the average relative salaries of these groups, it can be observed that those of the persons with higher salaries diminished until 2007 and increased since 2008. The opposite has occurred with the average salaries of the persons with lower salaries. From this analysis it can be deduced that the crisis has widened the breach between the groups of relative salaries considered.

3.1.2. Results obtained from a mixed distribution for salaries

Table 3 presents the polarization indexes EGR , considering that the distribution of salaries is mixed (**Figure 3** in the appendix). In general, the trends continue to be the same as when considering a complete distribution, that is to say polarization decreased until 2007 and increased from 2008 onwards; three principal differences are perceived. On the one hand the values obtained in the indexes are lower. On the other, the measurement of EGR , considering two groups, presents, in some cases, negative values. These values are due to the fact that on considering a mixed distribution for two groups the approximation error is high. Finally, the tendencies are more pronounced in the case of two groups. This indicates that consideration of the persons with zero salaries in a separate group in the calculation of polarization leads to distinct results in the indicators.

Table 3

The heterogeneity present in each group (a correction used to calculate the *EGR* index) decreases in the entire period analyzed, particularly so for the case of two groups, which also presents greater values. Once more, the lowest heterogeneity is produced when the population is divided into four groups. The highest levels of polarization are reached when the population is divided into three groups for all the parameters examined.

Table 4

Table 4 presents the partitions obtained when a mixed distribution is considered. In all cases, the first group is comprised of unemployed individuals, that is to say persons with a zero salary. Thus, the proportion of persons with zero salaries decreases until 2007 and increases from 2008 onwards. Logically, for the case of two groups, the remaining salaried persons present the opposite trend, increasing until 2007 and decreasing from 2008 on. If three groups are considered, the proportions of the first and second groups present the tendencies described above, while the third group increases its proportion until 2008 and in 2009 begins to reduce in size. If four population groups are considered, the second and third groups, in general, reduce their percentages in the period analyzed, while the fourth increases its proportion until 2008 and in 2009 begins to lose population weight. At the same time, the group of unemployed presents the opposite trend, making the increasing tendency of this last period move from being the group of lowest proportion in 2004 to be the second population group in size in 2012.

Consequently, the introduction of mixed distributions appears to add a greater clarity to the results derived from the study of polarization when considering the group of the unemployed and to do so separately from the rest.

3.2. The contribution of mixed distributions in the calculation of polarization

So far the polarization indexes have been calculated using the complete distribution of salaries and the mixed distribution. In this section, the idea is to show the contribution of using mixed distributions in the calculation of the polarization.

Table 5

Since the partition using the complete distribution looks for the greater inequality between the groups, it is not surprising that its index gives higher polarization values than using the mixed distribution. In this paper, the use of this last distribution seeks an exogenous group classification and the most outstanding results are presented when considering bipolarization. Thus, the annual changes that occur in the indices calculated from the mixed distribution are much more pronounced as can be observed in **Figure 2** and **Figure 3** and whose calculations are presented in the **Table 5**.

Also, the greater the value of the indices, the greater the difference with respect to the considered alpha. Therefore, in this case the sensitivity becomes more important with respect to polarization, as can be seen in **Figure 2** and **Figure 3**.

Assuming that the level of social polarization is captured by the complete distribution, and in order to show what part of the social tension is due to the exogenous disaggregation of the

unemployed group (ie, it is the result of unemployment), a comparative between both indexes is done. Thus, the first step is to make them comparable, so that these indices are normalized linearly for each alpha (Table 6). Then, the proportions of the values obtained with mixed distribution are compared with respect to the others to analyze how much polarization is collected by the two groups chosen here exogenously (corresponding to unemployed and employed). These results are shown in Table 7, where it can be seen that the lowest percentages are found in 2007 and the highest in 2012, that is, they depend directly on the percentage of unemployment. Thus, in the last year analyzed this percentage represents values close to 50%.

4. FINAL COMMENTS

The principal objective of this article has been to analyze the evolution of polarization in Spain between 2004 and 2012, taking into account the unemployed. To do this, distinct polarization indexes have been used, with different values in their parameters, to evaluate the robustness of the results. Similarly, for the calculations two types of distributions have been used, complete and mixed. The latter has been proposed with the aim of differentiating the group of zero salaries, that is to say those of the unemployed, thereby permitting the study of the evolution of their impact on the global measurement of polarization, to identify signals of social tension. Furthermore, a percentage decomposition of the contribution to polarization has been proposed, taking into account the group of the unemployed via mixed distributions.

As commented throughout this study, the period studied is comprised of two differentiated stages. In the first (2004-2007), Spain experienced a period of economic growth, while in the second (2008-2012) the effects of the worldwide economic crisis deeply affected the Spanish economy, principally in drastic increases in the unemployment rates. It is unsurprising, as a result, the clear downward tendency of the polarization indexes calculated until 2007 and their change

of trend in 2008, for the distinct groups considered, the distinct distributions employed and the distinct parameters assumed. In this way, the incorporation of unemployment people in the calculation of the index is proven to be important to show social tension.

The fact of considering a mixed distribution, in place of a complete one, separates the extreme (zero) salaries, meaning that on considering two groups, the tendencies are more pronounced.

Given the way to separate the groups, the highest levels of polarization are obtained by using the complete distribution. However, in the case of bipolarization, with mixed distribution, the tendencies are more pronounced. That is, annual changes in indices are more sensitive to exogenous separation into two proposed groups (employed and unemployed). Since, the most outstanding results are presented when considering bipolarization and the decomposition proposed has proved useful to tackle the importance of the groups of employed and unemployed in the result obtained for the existing polarization.

Finally, in order to analyze how much polarization is collected by the mixed distribution, that is to say, it is due to the separation in two exogenous groups (employed and unemployed), both indexes are compared. Unsurprisingly, in the years of higher unemployment, this proportion is larger, reaching almost 50% in 2012.

References

1. Aitchison, J.: On the distribution of a positive random variable having a discrete probability mass at the origin. *Journal of the American Statistical Association* 50 (271), 901-08 (1955)
2. Alfaro, A.K.: “Estudio de la pobreza, la desigualdad del ingreso y la polarización en Argentina en el período 1993-2003”, mimeo, University of Alcalá - Complutense University (2005)

3. Aghevli, B. and Mehran, F.: Optimal grouping of income distribution data. *Journal of the American Statistical Association* 76 (373), 22-26 (1981)
4. Autor, D.H.; Katz, L.F. and Kearny, M.S.: The polarization of the US labor market. *American Economic Review* 96, 189-94 (2006)
5. Blackburn, M.L. and Bloom, D.: Family Income Inequality in the United States, 1967-84. *Proceedings of the 39th Annual Meetings. Industrial Relations Research Association*, 349-58 (1986)
6. Davies, J.B. and Shorrocks, A.F.: Optimal grouping of income and wealth data. *Journal of Econometrics* 42, 97-108 (1989)
7. Esteban, J.M. and Ray, D.: On the measurement of polarization. *Econometrica* 62 (4), 819-51 (1994)
8. Esteban J.M., Gradín C and Ray, D.: An Extension of a measure of polarization with an application to the income distribution of five OECD countries. *Journal of Income Inequality* 5 (1), 1-19 (2007)
9. Ezcurra, R.: Does income polarization affect economic growth? The case of the European regions. *Regional Studies* 43 (2), 267-85 (2009)
10. Gradín, C.: Polarization by sub-populations in Spain, 1973-91. *Review of Income and Wealth* 46 (4), 457-74 (2000)
11. Horrigan, M.W. and Haugen S. E.: The declining middle-class thesis: a sensitivity analysis. *Monthly Labor Review* 111 (5), 3-13 (1988)
12. Lambert, P.J. and Aronson, J.R. : Inequality Decomposition Analysis and the Gini Coefficient Revisited. *The Economic Journal* 103 (420), 1221-27 (1993)
13. Levy, F. and Murnane, R.J.: U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations. *Journal of Economic Literature* 30 (3), 1333-81 (1992)

14. Núñez Velázquez, J.J.: La desigualdad económica medida a través de las curvas de Lorenz.
Revista de Métodos Cuantitativos para la Economía y la Empresa 2, 67-108 (2006)
15. Wolfson, M.C.: When Inequalities Diverge. American Economic Review 84, 353-58 (1994)

Table 1: Measures of *EGR* polarization.
Complete distribution of normalized net monthly salaries, 2004-2012.

Two groups									
Year	<i>ER</i>			$G(f)$	$G(\rho^*)$	ε	<i>EGR</i> ($\beta=1$)		
	$\alpha=1$	$\alpha=1,3$	$\alpha=1,6$				$\alpha=1$	$\alpha=1,3$	$\alpha=1,6$
2004	0,267	0,217	0,177	0,396	0,270	0,126	0,142	0,092	0,052
2005	0,251	0,204	0,166	0,374	0,254	0,120	0,133	0,086	0,048
2006	0,236	0,193	0,157	0,354	0,241	0,113	0,124	0,080	0,045
2007	0,230	0,187	0,153	0,351	0,238	0,113	0,120	0,077	0,043
2008	0,238	0,194	0,158	0,357	0,243	0,113	0,126	0,082	0,046
2009	0,279	0,226	0,184	0,411	0,284	0,126	0,153	0,101	0,059
2010	0,295	0,240	0,195	0,425	0,297	0,128	0,168	0,113	0,068
2011	0,309	0,251	0,204	0,430	0,300	0,129	0,180	0,122	0,075
2012	0,329	0,268	0,217	0,460	0,329	0,132	0,201	0,139	0,089
Three groups									
Year	<i>ER</i>			$G(f)$	$G(\rho^*)$	ε	<i>EGR</i> ($\beta=1$)		
	$\alpha=1$	$\alpha=1,3$	$\alpha=1,6$				$\alpha=1$	$\alpha=1,3$	$\alpha=1,6$
2004	0,223	0,162	0,119	0,396	0,342	0,054	0,170	0,109	0,065
2005	0,210	0,153	0,112	0,374	0,322	0,051	0,158	0,101	0,060
2006	0,196	0,142	0,103	0,354	0,304	0,050	0,145	0,091	0,053
2007	0,191	0,139	0,101	0,351	0,301	0,049	0,142	0,090	0,052
2008	0,198	0,144	0,105	0,357	0,308	0,049	0,149	0,095	0,056
2009	0,233	0,168	0,122	0,411	0,360	0,050	0,182	0,118	0,072
2010	0,245	0,177	0,128	0,425	0,375	0,050	0,196	0,127	0,079
2011	0,257	0,185	0,134	0,430	0,381	0,049	0,208	0,137	0,086
2012	0,271	0,196	0,142	0,460	0,412	0,049	0,223	0,148	0,094
Four groups									
Year	<i>ER</i>			$G(f)$	$G(\rho^*)$	ε	<i>EGR</i> ($\beta=1$)		
	$\alpha=1$	$\alpha=1,3$	$\alpha=1,6$				$\alpha=1$	$\alpha=1,3$	$\alpha=1,6$
2004	0,178	0,118	0,079	0,396	0,368	0,028	0,149	0,090	0,050
2005	0,168	0,112	0,075	0,374	0,346	0,028	0,140	0,084	0,047
2006	0,157	0,105	0,070	0,354	0,327	0,027	0,129	0,077	0,042
2007	0,153	0,102	0,068	0,351	0,324	0,027	0,126	0,075	0,041
2008	0,159	0,106	0,071	0,357	0,330	0,027	0,132	0,079	0,044
2009	0,185	0,123	0,081	0,411	0,384	0,027	0,158	0,096	0,054
2010	0,195	0,129	0,086	0,425	0,399	0,026	0,169	0,103	0,059
2011	0,204	0,135	0,090	0,430	0,404	0,026	0,179	0,110	0,065
2012	0,218	0,146	0,097	0,460	0,435	0,025	0,194	0,121	0,073

Table 2: Group population shares, mean salaries by group and inter-group cut-off salaries.
Complete distribution of normalized net monthly salaries, 2004-2012.

Two groups					
Year	Group population shares		Inter-group cut-off salaries		
	π_1	π_2	Mean salaries by group	y_1	y_2
2004	0,548	0,452	0,512	1,591	1,000
2005	0,538	0,462	0,533	1,544	1,000
2006	0,556	0,444	0,575	1,533	1,000
2007	0,560	0,440	0,590	1,522	1,000
2008	0,556	0,444	0,572	1,537	1,000
2009	0,515	0,485	0,459	1,574	1,000
2010	0,498	0,502	0,407	1,588	1,000
2011	0,490	0,510	0,370	1,607	1,000
2012	0,489	0,511	0,326	1,645	1,000

Three groups								
Year	Group population shares			Mean salaries by group			Inter-group cut-off salaries	
	π_1	π_2	π_3	y_1	y_2	y_3	z_1	z_2
2004	0,306	0,454	0,240	0,238	1,000	1,969	0,693	1,336
2005	0,281	0,464	0,255	0,252	0,969	1,881	0,698	1,293
2006	0,312	0,444	0,244	0,349	0,985	1,861	0,723	1,296
2007	0,289	0,454	0,257	0,337	0,961	1,811	0,719	1,269
2008	0,283	0,452	0,265	0,295	0,966	1,812	0,708	1,279
2009	0,284	0,441	0,275	0,146	0,987	1,903	0,658	1,338
2010	0,306	0,429	0,265	0,134	1,018	1,968	0,650	1,381
2011	0,319	0,426	0,255	0,113	1,050	2,026	0,648	1,416
2012	0,337	0,409	0,254	0,088	1,084	2,074	0,634	1,463

Four groups											
Year	Group population shares				Mean salaries by group				Inter-group cut-off salaries		
	π_1	π_2	π_3	π_4	y_1	y_2	y_3	y_4	z_1	z_2	z_3
2004	0,223	0,326	0,286	0,166	0,099	0,796	1,236	2,207	0,513	1,002	1,592
2005	0,213	0,325	0,308	0,154	0,136	0,794	1,228	2,179	0,533	1,005	1,545
2006	0,207	0,337	0,289	0,167	0,193	0,794	1,194	2,077	0,565	0,979	1,518
2007	0,192	0,331	0,305	0,172	0,180	0,783	1,169	2,030	0,562	0,968	1,480
2008	0,203	0,349	0,281	0,167	0,160	0,806	1,222	2,050	0,568	0,992	1,531
2009	0,239	0,312	0,276	0,173	0,067	0,824	1,282	2,161	0,495	1,039	1,620
2010	0,255	0,296	0,274	0,175	0,050	0,827	1,305	2,207	0,467	1,056	1,656
2011	0,268	0,283	0,277	0,172	0,028	0,836	1,326	2,257	0,443	1,078	1,684
2012	0,294	0,254	0,282	0,170	0,022	0,848	1,365	2,315	0,405	1,120	1,722

Table 3: Measures of *EGR* polarization.
Mixed distribution of normalized net monthly salaries, 2004-2012.

Two groups									
<i>ER</i>							<i>EGR</i> ($\beta=1$)		
Year	$\alpha=1$	$\alpha=1,3$	$\alpha=1,6$	$G(f)$	$G(\rho^*)$	ε	$\alpha=1$	$\alpha=1,3$	$\alpha=1,6$
2004	0,134	0,119	0,109	0,392	0,160	0,233	-0,099	-0,114	-0,124
2005	0,117	0,105	0,097	0,370	0,135	0,235	-0,118	-0,129	-0,137
2006	0,100	0,091	0,085	0,349	0,112	0,236	-0,137	-0,145	-0,151
2007	0,095	0,087	0,082	0,339	0,106	0,234	-0,139	-0,147	-0,152
2008	0,106	0,097	0,090	0,350	0,121	0,229	-0,122	-0,132	-0,138
2009	0,157	0,137	0,122	0,404	0,195	0,209	-0,052	-0,072	-0,086
2010	0,171	0,148	0,130	0,422	0,219	0,203	-0,033	-0,056	-0,073
2011	0,185	0,158	0,137	0,438	0,244	0,193	-0,009	-0,036	-0,056
2012	0,198	0,168	0,144	0,458	0,273	0,185	0,013	-0,017	-0,041
Three groups									
<i>ER</i>							<i>EGR</i> ($\beta=1$)		
Year	$\alpha=1$	$\alpha=1,3$	$\alpha=1,6$	$G(f)$	$G(\rho^*)$	ε	$\alpha=1$	$\alpha=1,3$	$\alpha=1,6$
2004	0,193	0,145	0,111	0,392	0,323	0,070	0,123	0,076	0,041
2005	0,191	0,145	0,112	0,370	0,300	0,070	0,121	0,075	0,042
2006	0,186	0,144	0,112	0,349	0,279	0,070	0,116	0,074	0,042
2007	0,186	0,144	0,113	0,339	0,271	0,069	0,117	0,075	0,044
2008	0,185	0,142	0,110	0,350	0,282	0,067	0,117	0,074	0,042
2009	0,189	0,140	0,105	0,404	0,343	0,061	0,129	0,079	0,044
2010	0,192	0,141	0,104	0,422	0,363	0,059	0,133	0,081	0,045
2011	0,193	0,140	0,103	0,438	0,382	0,056	0,136	0,084	0,046
2012	0,196	0,142	0,103	0,458	0,403	0,055	0,141	0,087	0,048
Four groups									
<i>ER</i>							<i>EGR</i> ($\beta=1$)		
Year	$\alpha=1$	$\alpha=1,3$	$\alpha=1,6$	$G(f)$	$G(\rho^*)$	ε	$\alpha=1$	$\alpha=1,3$	$\alpha=1,6$
2004	0,151	0,101	0,069	0,118	0,085	0,033	0,118	0,068	0,036
2005	0,147	0,100	0,068	0,114	0,081	0,033	0,114	0,067	0,035
2006	0,145	0,100	0,069	0,112	0,079	0,033	0,112	0,067	0,036
2007	0,143	0,098	0,068	0,110	0,077	0,033	0,110	0,065	0,035
2008	0,143	0,097	0,066	0,110	0,078	0,032	0,110	0,065	0,034
2009	0,148	0,098	0,065	0,118	0,089	0,030	0,118	0,068	0,036
2010	0,151	0,100	0,066	0,123	0,095	0,028	0,123	0,071	0,038
2011	0,153	0,101	0,067	0,126	0,100	0,027	0,126	0,075	0,040
2012	0,157	0,104	0,069	0,131	0,105	0,026	0,131	0,078	0,044

Table 4: Group population shares, mean salaries by group and inter-group cut-off salaries.
Mixed distribution of normalized net monthly salaries, 2004-2012.

Two groups											
Year	Group population shares		Mean salaries by group		Inter-group cut-off salaries						
	π_1	π_2	y_1	y_2	z_1						
2004	0,160	0,840	0,000	1,000	0,000						
2005	0,135	0,865	0,000	1,000	0,000						
2006	0,112	0,888	0,000	1,000	0,000						
2007	0,106	0,894	0,000	1,000	0,000						
2008	0,121	0,879	0,000	1,000	0,000						
2009	0,195	0,805	0,000	1,000	0,000						
2010	0,219	0,781	0,000	1,000	0,000						
2011	0,244	0,756	0,000	1,000	0,000						
2012	0,273	0,727	0,000	1,000	0,000						
Three groups											
Year	Group population shares			Mean salaries by group			Inter-group cut-off salaries				
	π_1	π_2	π_3	y_1	y_2	y_3	z_1	z_2			
2004	0,160	0,520	0,320	0,000	0,682	1,517	0,000	1,000			
2005	0,135	0,521	0,344	0,000	0,680	1,493	0,000	1,000			
2006	0,112	0,533	0,355	0,000	0,683	1,473	0,000	1,000			
2007	0,106	0,550	0,345	0,000	0,692	1,490	0,000	1,000			
2008	0,121	0,522	0,357	0,000	0,686	1,460	0,000	1,000			
2009	0,195	0,483	0,322	0,000	0,689	1,470	0,000	1,000			
2010	0,219	0,467	0,315	0,000	0,686	1,471	0,000	1,000			
2011	0,244	0,458	0,298	0,000	0,700	1,473	0,000	1,000			
2012	0,273	0,444	0,283	0,000	0,706	1,482	0,000	1,000			
Four groups											
Year	Group population shares				Mean salaries by group				Inter-group cut-off salaries		
	π_1	π_2	π_3	π_4	y_1	y_2	y_3	y_4	z_1	z_2	z_3
2004	0,160	0,321	0,343	0,177	0,000	0,564	0,985	1,819	0,000	0,782	1,259
2005	0,135	0,309	0,358	0,198	0,000	0,558	0,965	1,756	0,000	0,776	1,224
2006	0,112	0,350	0,347	0,190	0,000	0,586	0,991	1,778	0,000	0,788	1,264
2007	0,106	0,334	0,357	0,203	0,000	0,584	0,971	1,737	0,000	0,784	1,230
2008	0,121	0,324	0,340	0,215	0,000	0,578	0,968	1,688	0,000	0,778	1,225
2009	0,195	0,282	0,324	0,199	0,000	0,568	0,960	1,678	0,000	0,778	1,208
2010	0,219	0,277	0,309	0,195	0,000	0,566	0,963	1,676	0,000	0,775	1,215
2011	0,244	0,276	0,302	0,178	0,000	0,579	0,977	1,690	0,000	0,787	1,227
2012	0,273	0,265	0,291	0,171	0,000	0,578	0,984	1,678	0,000	0,791	1,232

Table 5: Bipolarization annual variation rate.

Complete Distribution			
<i>EGR ($\beta=1$)</i>			
Year	$\alpha=1$	$\alpha=1,3$	$\alpha=1,6$
2005	-0,07	-0,07	-0,08
2006	-0,06	-0,06	-0,06
2007	-0,04	-0,04	-0,05
2008	0,05	0,06	0,08
2009	0,21	0,23	0,27
2010	0,10	0,11	0,15
2011	0,07	0,09	0,11
2012	0,11	0,14	0,18

Mixed Distribution			
<i>EGR ($\beta=1$)</i>			
Year	$\alpha=1$	$\alpha=1,3$	$\alpha=1,6$
2005	-0,191	-0,137	-0,104
2006	-0,162	-0,125	-0,101
2007	-0,018	-0,011	-0,007
2008	0,122	0,102	0,089
2009	0,575	0,455	0,375
2010	0,375	0,224	0,153
2011	0,729	0,361	0,233
2012	2,522	0,517	0,270

Table 6: Normalized bipolarization indexes

Complete Distribution			
$NormEGR_{2G}(\beta=1)$			
Year	$\alpha=1$	$\alpha=1,3$	$\alpha=1,6$
2004	0.806	0.836	0.848
2005	0.779	0.813	0.831
2006	0.755	0.794	0.819
2007	0.742	0.783	0.811
2008	0.761	0.801	0.825
2009	0.839	0.866	0.875
2010	0.880	0.906	0.911
2011	0.915	0.939	0.941
2012	0.974	0.997	0.995

Mixed Distribution			
$NormEGR_{2G}(\beta=1)$			
Year	$\alpha=1$	$\alpha=1,3$	$\alpha=1,6$
2004	0.118	0.126	0.143
2005	0.064	0.072	0.092
2006	0.009	0.017	0.036
2007	0.002	0.011	0.032
2008	0.051	0.063	0.086
2009	0.251	0.270	0.294
2010	0.307	0.325	0.347
2011	0.375	0.394	0.415
2012	0.438	0.458	0.476

Table 7:

$Norm(EGR_{MD})/Norm(EGR_{CD})$			
Year	$\alpha=1$	$\alpha=1,3$	$\alpha=1,6$
2004	0.146	0.151	0.169
2005	0.082	0.089	0.110
2006	0.012	0.021	0.044
2007	0.003	0.014	0.039
2008	0.067	0.078	0.104
2009	0.300	0.311	0.336
2010	0.349	0.359	0.381
2011	0.409	0.420	0.441
2012	0.450	0.459	0.478

Figure 1: Normalized net monthly salaries' kernel density functions, 2004-2012.

$P_0=0.2712$

Appendix

Figure 2: Measures of *EGR* polarization.
Complete distribution of normalized net monthly salaries, 2004-2012.

Two groups

EGR ($\beta=1$)

Three groups

EGR ($\beta=1$)

Four groups

EGR ($\beta=1$)

Figure 3: Measures of *EGR* polarization.
Mixed distribution of normalized net monthly salaries, 2004-2012.