

# **Income Inequality, Cohesiveness and Commonality in the Euro Area: A Semi-parametric Boundary-free Analysis**

Gordon Anderson  
(University of Toronto, Canada)

Maria Grazia Pittau  
(Sapienza University of Rome, Italy)

Roberto Zelli  
(Sapienza University of Rome, Italy)

Jasmin Thomas  
(University of Toronto, Canada)

Paper prepared for the 35th IARIW General Conference

Copenhagen, Denmark, August 20-25, 2018

Session 2C: Globalization and Inclusive Growth I

Time: Tuesday, August 21, 2018 [Afternoon]

# Income Inequality, Cohesiveness and Commonality in the Euro Area: A Semi-parametric Boundary-free Analysis

Gordon Anderson\*, Maria Grazia Pittau†, Roberto Zelli†  
and Jasmin Thomas\*

## Abstract

The cohesiveness of constituent nations in a confederation such as the Eurozone depends on their equally shared experiences. In terms of household incomes, commonality of distribution across those constituent nations with that of the Eurozone as an entity in itself is of the essence. For the most part income classification has proceeded by employing “hard”, somewhat arbitrary and contentious boundaries. Here, in an analysis of Eurozone household income distributions over the period 2006–2015, mixture distribution techniques are used to determine the number and size of groups or classes endogenously without resort to such hard boundaries. In so doing, some new indices of polarization, segmentation and commonality of distribution are developed in the context of a decomposition of the Gini coefficient and the roles of, and relationships between, these groups in societal income inequality, poverty, polarization and societal segmentation are examined. What emerges for the Eurozone as an entity is a four class increasingly unequal polarizing structure with income growth in all four classes. With regard to individual constituent nation class membership, some advanced, some fell back with most exhibiting significant polarizing behaviour. However, in the face of increasing overall Eurozone inequality, constituent nations were becoming increasingly similar in distribution which can be construed as characteristic of a more cohesive society.

*Keywords:* Income distribution; Inequality; Mixtures; Gini; EU-SILC.

---

\*Department of Economics, University of Toronto.

†Department of Statistical Sciences, Sapienza University of Rome.

# 1 Introduction

As Milanovic (2011) observes, growing inequalities between states in federations such as the Eurozone can be seen as a catalyst for the deterioration of social cohesion and support for the Union’s institutions amongst its citizens. Thus measurements of aspects of wellbeing of the Eurozone as an entity in itself and of its constituent nations are regarded as basic information for evaluating the progress of the Eurozone toward greater social cohesion within and between its various constituencies (Brandolini 2007) and such measurements have become important in core European institutional documents and debates (Filauro 2017).

The sense of cohesion amongst constituent nations hinges on notions of belonging, an absence of alienation from each other or, in the presence of such alienation, that the process going forward is toward a less alienated state (OECD, 2011). This has much to do with concepts of inequality, segmentation and polarization within and between groups and the sense in which they are directionally dynamic processes. When constituent nations are equally unequal with relatively similar income levels there is a commonality of situation which promotes cohesiveness whereas when such inequality and income levels are not equally shared the situation is somewhat more divisive and alienating. Since notions of what constitutes poorness and wellness in income terms may vary across constituencies, these possibilities are perhaps best understood in the context of general income groupings or classifications in the Eurozone. Here, in analyzing the class structure anatomy of an overall euro area distribution and the relationship of its components to its constituent nations, these issue are addressed in terms of household incomes in the euro area and 18 of its constituent nations for the period 2006–2015.

For the longest time in income classification and measurement, income grouping have been identified by employing “hard”, somewhat arbitrary and contentious income cut-offs or boundaries (see for example, Anderson, 2010; Atkinson and Brandolini, 2013; Ravallion 2010, 2012). The main problem being that analysis is overly influenced by boundary choice especially when intertemporal comparisons are being made.<sup>1</sup> Aside from measurement error or data contamination issues (see Deaton, 2010), categorizing poorness and wellness in such an arbitrary fashion can prejudice other aspects of anal-

---

<sup>1</sup>Witness The World Bank, 2017 GNI per capita (\$ US equivalent) thresholds used for classifying nation income status. These were established in 1989–based upon previously established operational criteria—and inflation updated each year or the United Nations \$1 a day or the subsequent changes in the United Nations Development goals \$1 a day poverty measure.

ysis. For example, defining classes by quantiles fixes class sizes over time precluding analysis of poverty reduction strategies and tying class boundaries to some proportion of a location measure ties movement of classes to movements in the overall distribution and assumes away the possibility of independent class variation (incidentally contravening the focus axiom frequently invoked in poverty analysis).

Here, by employing mixture distribution techniques in a general euro area distribution, the number and size of groups or classes is determined by the commonalities in their income patterns and processes without resort to such hard boundaries.<sup>2</sup> This facilitates analysis of individual nation membership of income groupings and the progress of those nations through the overarching Eurozone income class structure. In so doing, some new indices of polarization and segmentation are developed in the context of a decomposition of the Gini coefficient and the roles of, and relationships between, these groups in societal income inequality, poverty, polarization and societal segmentation are examined. In addition since it may be prudent to work with a simple ordinal classification that does not impute cardinal measure to wellbeing, a measure, the Utopia Index, that provides a complete cardinal ordering of wellbeing, though the basis of comparison (class membership) is only an ordinal classification has been developed and implemented. These ideas are applied to an analysis of the Eurozone income distribution over the decade spanning 2006–2015. Implications for the individual constituent nations of the collective are explored. What emerged in the generic Eurozone distribution is a four class increasingly unequal polarizing structure with income growth in all four classes. With regard to individual nation results over the sample period, six nations are seen to be progressing through the class structure with twelve regressing. In terms of class transitions, thirteen nations are seen to exhibit polarizing patterns with four exhibiting converging behavior over the period. With regard to the Utopia index, nation rankings appeared to be fairly stable over time exceptions were Finland and France who made significant advances and Greece which declined. In the following, Section 2 discusses the algebraic relationship between income classes, the Gini coefficient and measures of inequality, polarization and segmentation of subgroups. In Section 3 income data for the European countries are presented and discussed. Section 4 outlines the details of mixture distribution estimation and reports the main empirical results of the empirical analysis. Conclusions are drawn in Section 5.

---

<sup>2</sup>Mixture distributions have also been used to deal with measurement error/data contamination problems see Alvarez-Esteban et al., 2016. On the usefulness of mixture models for distributional analysis see Cowell and Flachaire, 2015.

## 2 Income classes and the Gini coefficient: Inequality, polarization and segmentation of subgroups

### 2.1 Mixture of distribution to identify income classes

In each year, income data are interpreted as a sample from a mixture of  $K$  components in unknown proportions  $w_1, \dots, w_K$ . Each component represents the income distribution of a homogeneous group of households, that is a household belonging to group  $k$  faces income opportunities described by the distribution  $f_k$ . Given some assumptions regarding the nature of the  $f_k$ 's, these components can be specified to belong to some parametric family (normality or log normality are popular specifications that can be theoretically rationalized).

If the components are assumed to belong to the normal family, the mixture density can be written as:<sup>3</sup>

$$f(x; \Psi) = \sum_{k=1}^K w_k f_k(x; \mu_k, \sigma_k^2), \quad (1)$$

where  $f_k(x; \mu_k, \sigma_k) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu_k)^2}{2\sigma_k^2}}$  and  $w_k$ , the mixing weight, represent the proportion of the population in class  $k$ . The vector  $\Psi = (w_1, \dots, w_{K-1}, \xi')'$  contains all the unknown parameters of the mixture model; in this case  $\xi = (\mu_1, \dots, \mu_K, \sigma_1^2, \dots, \sigma_K^2)$ .

Estimation of the triple  $w_k, \mu_k, \sigma_k^2$  for  $k = 1, \dots, K$  yields much information about the structure of the society. For convenience and without loss of generality suppose the types are ordered ( $k > j \Rightarrow \mu_k \geq \mu_j$ ) then  $\mu_1$  corresponds to the average income of the poor type and  $w_1$  represents the proportion of the poor type in society i.e. the relative poverty rate. This does mean that the class membership of a household with income  $x$  cannot be determined with certainty. However, such an analysis has many advantages, classes are determined without resort to arbitrary chosen boundaries, hence they are allowed to vary independently over time in terms of their size, location and scale.

Knowledge of  $f_k(x|\mu_k, \sigma_k^2)$  also facilitates within class inequality measurement and between class polarization and segmentation measurement facilitating the study of such concepts in the context of an overall distribution.

It is also of interest to see how the individual nation states that make up the community have fared in terms of the income classifications. This may be examined

---

<sup>3</sup>These ideas are readily generalized to multidimensional environments (see Anderson, Farcomeni, Pittau and Zelli, 2017).

by generating class membership probabilities for each member state over the period. Once the parameters of the components and the values of the class shares are estimated, posterior or conditional probabilities  $\tau_i$  that household  $i$  with income  $x^*$  is in the  $k$ 'th group can be computed since:

$$\tau_{ik} = Pr(x^* \in k\text{'th class}) = \frac{w_k f_k(x^*)}{\sum_{k=1}^K w_k f_k(x^*)}, \text{ for } k = 1 \dots K. \quad (2)$$

Effectively this provides  $K$  group membership indices for each agent in the population. Note that it is possible for the group distributions to overlap, that is for an household with income  $x^*$  to potentially be a member of more than one group. To the extent that these distributions do not overlap (perfect segmentation in the terminology of Yitzhaki, 1994) knowing the household income will completely determine the household's group and all of the households in a group. To the extent that they do overlap household income will only partially define its group membership in the sense that its probability of being in a particular group is all that can be obtained.

Given the estimated ex-post probabilities of each household  $i$  to belong to a specific class  $k$ ,  $\tau_{ik}$ , the unbiased probability of class membership of each constituent nation  $h$ , based upon the average probability of membership in a particular class in a given nation, can be derived as follows. For country  $h$  (where  $h = 1, \dots, H$ ) with  $n_h$  observations  $x_i$ ,  $i = 1 \dots, n_h$ , the probability that country  $h$  is in class  $k$  is given by:

$$\theta_{hk} = \frac{1}{n_h} \sum_{i=1}^{n_h} \tau_{ik} \text{ for } k = 1 \dots, K. \quad (3)$$

Alternatively, following a traditional "hard" classification, we can assign households to components according to their maximal conditional probability:

$$\mathbb{I}_{i,k} = \begin{cases} 1 & \text{if } \tau_{ik} = \max(\tau_{i1}, \dots, \tau_{iK}) \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where  $\mathbb{I}_{i,k}$  is the indicator variable of the  $i$ -th household to belong to class  $k$ . Then we can calculate the proportion of households resident in country  $h$  simply as:

$$\theta_{hk}^{(a)} = \frac{1}{n_h} \sum_{i=1}^{n_h} \mathbb{I}_{i,k} \text{ for } k = 1 \dots, K. \quad (5)$$

## 2.2 The Gini Coefficient and segmentation of subgroups

An inherent problem with the Gini coefficient, highlighted in Bourguignon (1979), is that it is not generally subgroup decomposable. Following Mookherjee and Shorrocks

(1982) and Anderson and Thomas (2017) the Gini coefficient may be written as the sum of three components as follows:<sup>4</sup>

$$\text{GINI} = \sum_{k=1}^K w_k^2 \frac{\mu_k}{\mu} G_k + \frac{1}{\mu} \sum_{k=2}^K \sum_{j=1}^k w_k w_j |\mu_k - \mu_j| + \text{NSF}, \quad (6)$$

where  $G_k$  is the Gini associated with the  $k$ 'th subgroup and  $\sum_{k=1}^K w_k = 1$  so that  $E_{f(x)}(x) = \mu = \sum_{k=1}^K w_k \mu_k$  and where NSF may be thought of as the “non-segmentation” factor. The NSF may be written as:

$$\text{NSF} = \frac{2}{\mu} \sum_{k=2}^K \sum_{j=1}^{k-1} w_k w_j \int_0^{\infty} f_k(y) \int_y^{\infty} f_j(x)(x-y) dx dy. \quad (7)$$

The Gini is thus a weighted sum of subgroup Ginis plus a weighted sum of subgroup “dominating mean differences” divided by the overall mean (in essence a between group Gini coefficient BGINI) plus a component which is the weighted sum of the extent to which there are individuals in lower group  $j$  who overlap with, i.e. have greater incomes than, individuals in upper group  $k$  weighted by the extent to which they have more. In essence, the Gini is a linear function of within and between group Gini coefficients plus a term measuring the extent to which subgroups are not segmented.

Considering NSF, first note that when subgroups  $k$  and  $j$  are perfectly segmented (so that  $f_k(x) = 0$  for all  $x$  such that  $f_j(x) > 0$  and  $f_j(x) = 0$  for all  $x$  such that  $f_k(x) > 0$ ) the corresponding term in the component vanishes. In the particular case where this is true for all  $j \neq k$ , the Gini is sub-group decomposable (Mookherjee and Shorrocks, 1982).

Noting that in general all three components of GINI are non-negative and that  $0 \leq \text{NSF} \leq \text{GINI}$ , then SI, a segmentation index, may be written as:

$$\text{SI} = 1 - \frac{\text{NSF}}{\text{GINI}}, \quad 0 \leq \text{SI} \leq 1. \quad (8)$$

SI provides a measure of the degree to which constituent groups are segmented or do not overlap. The analysis can be done with respect to particular groups, so the extent to which the poor or the rich are segmented from the rest of society may be readily analyzed.

---

<sup>4</sup>This decomposition is readily extended to the Absolute Gini (Hey and Lambert, 1980; Weymark, 2003) by multiplying these equations by the overall mean from whence it may be seen that the overall Absolute Gini is a weighted sum of subgroup Absolute Ginis, the between group Absolute Gini and the Absolute Non Segmentation factor. Results in Giles (2004) facilitate inference. Derivation of the decomposition in the context of continuous distributions is shown in Appendix A.1.

Considering a generic group  $g \in (1, \dots, K)$ . The specific non-segmentation factor of group  $g$  can be obtained as:

$$\text{NSF}_g = \frac{2}{\mu} \sum_{\substack{k=1 \\ k \neq g}}^K w_k w_g \int_0^\infty f_g(y) \int_y^\infty f_k(x)(x-y) dx dy. \quad (9)$$

This is twice a weighted sum of the (expected) average value of the excess of incomes of people in group  $g$  over those of people in the other groups normalized by average income which is of interest in contemplating the “isolation” of the group.

Clearly  $\text{NSF}_g$  could be inserted in place of NSF in (8) to obtain an index of the segmentation of the specific group.

## 2.3 Comparing constituent distributions: Polarization, Transvariation and Utopia-Dystopia index

### *Polarization*

Conceptually polarization is based upon notions of between group alienation and within group association. Duclos, Esteban and Ray (2004) captured this in an axiomatically developed general polarization index covering many, possibly latent, groups which may be written as:

$$P_\alpha(f) = S \int_0^\infty f(x) \int_0^\infty f(y)^{1+\alpha} |y-x| dy dx \quad (10)$$

Here  $S$  is a standardizing factor and  $\alpha$  is the polarization sensitivity factor confined to  $[0.25, 1]$ . Note that (10) is not unlike a Gini coefficient (if  $\alpha$  is set to 0 and  $S$  is suitably chosen, (10) becomes the continuous distribution version of Gini).  $P_\alpha(f)$  can be interpreted as the scaled expected value of all possible rectangles formed under the distribution with height  $f(y)^{1+\alpha}$  and base  $|y-x|$  where  $f(y)$  reflects the association component (larger  $f(y)$  reflects more association) and  $|y-x|$  reflects the alienation factor.

The index was developed by contemplating “sliding” and “squeezing” translations of basic constituent densities which respectively increased distances between constituent groups or intensified concentration around group means.<sup>5</sup> Slides change the relative locations of groups reflected in BGINI and, under certain circumstances, reduce NSF (i.e. increase the chance of segmentation), squeezes on the other hand simply reduce

---

<sup>5</sup>In the context of mixture distributions these ideas can be explored by considering the component distributions to be the basic densities.



NSF without affecting BGINI. Note that while polarizing slides can be associated with increasing between group inequalities, polarizing squeezes cannot, so a sufficient condition for establishing polarization (convergence) between groups is a combination of increased (decreased) between group inequality, BGINI, and segmentation, SI, suggesting a Gini-based polarization index PG of the form:

$$PG = \left( \frac{\text{BGINI}}{\text{GINI}} \right)^\alpha \text{SI}^{1-\alpha} \text{ for } 0 < \alpha < 1. \quad (11)$$

A group-specific polarization index for each class can also be obtained as:

$$\text{Pol}_g = w_g^\alpha \left( 1 - \frac{\text{NSF}_g}{\text{GINI}} \right)^{1-\alpha} \text{ for } 0 < \alpha < 1. \quad (12)$$

### *Transvariation*

When all subgroups are perfectly segmented the strongest form of stochastic dominance between them prevails, there will be a strict complete first order dominance relationship between all constituent groups consistent with any monotonic non decreasing well-being measure function of income. All such measures would be unambiguous. Perfect segmentation is a sufficient (but not necessary) condition for such an ordering. But normality of constituent distributions precludes such segmentation (all distributions overlap somewhere on the real line).

It is therefore useful to introduce a  $j$ -th index of transvariation (Anderson, Linton and Thomas, 2017) able to capture the degree of overlapping between  $K$  continuous distributions:

$$TR_j = \int_0^\infty \left[ \max(F_1^j(x), F_2^j(x), \dots, F_K^j(x)) - \min(F_1^j(x), F_2^j(x), \dots, F_K^j(x)) \right] dx \quad (13)$$

where  $F^j(x) = \int_0^\infty F^{j-1}(z)dz$ , with  $F^0(x) = f(x)$  and  $j = 0, 1, 2 \dots$ .

Zero order transvariation ( $j=0$ ) is the many distribution version of Gini's classic two distribution transvariation (Gini, 1916; Pittau and Zelli, 2017). Under perfect segmentation  $TR_{(j=0)} = K$  and when all distributions are identical  $TR_{(j=0)} = 0$  so that  $TR_{(j=0)}/K$  provides a good index of the degree of segmentation.

### *Utopia-Dystopia*

When the  $K$  classes  $C_k$ ,  $k = 1, \dots, K$  have only an ordinal ranking (so that  $C_i \preceq C_j$  for  $i < j$  where the operator  $\preceq$  indicates  $C_j$  is at least as good as  $C_i$ ), a countries relative wellbeing measure can still be obtained from the discrete cumulative density

functions  $F_h(C_k)$ ,  $k = 1, \dots, K$ , for countries  $h = 1, \dots, H$  across the  $K$  latent classes since:

$$\begin{aligned} F_{h_m}(C_k) &\leq F_{h_n}(C_k), \text{ for all } k = 1, \dots, K - 1 \text{ with strict inequality somewhere} \\ &\Rightarrow \mathbb{E}_{h_m}(C) \geq \mathbb{E}_{h_n}(C). \end{aligned}$$

That is to say when country  $m$ 's class membership density first order dominates that of country  $n$ ,  $m$  will have a higher expected class membership than country  $n$ .<sup>6</sup> Inference for the comparisons can be conducted using the maximum modulus distributions for multiple simultaneous comparisons (Stoline and Ury, 1979).

A relative Utopia-Dystopia measure for country  $h$ ,  $UI(h)$   $h = 1, \dots, H$  (Anderson, Post and Whang, 2017; Anderson and Leo, 2017) can be developed as:

$$UI(h) = \frac{\sum_{k=1}^K (\max_h F_h(C_k) - F_h(C_k))}{\sum_{k=1}^K (\max_h F_h(C_k) - \min_h F_h(C_k))}, \quad (14)$$

where  $\max_h F_h(C_k)$  ( $\min_h F_h(C_k)$ ) corresponds to the maximum (minimum) value of  $F_h(C_k)$  over all  $h$ . The index can be shown to reside in  $[0,1]$  with 1 representing unequivocal ‘‘Utopia’’ (the best of all nations in that its distribution first order dominates all others) and 0 corresponding to an unequivocal ‘‘Dystopia’’ (such a nations’ distribution is first order dominated by all others) and have many desirable properties of a wellbeing index.

### 3 Data issues

Monitoring income inequality as well as other indicators related to personal income distribution within European countries relies on comparable and internationally harmonized estimates for the member states. The European Union Survey on Income and Living Conditions (EU-SILC) is the harmonized household-level survey that is widely used for these purposes (see e.g. Longford, 2014). The cross-sectional component of EU-SILC is a collection of annual national surveys of socio-economic conditions of individuals and households in the EU countries. All national surveys in EU-SILC have standard questionnaires and procedures for data processing and yield ex-ante harmonized micro-data that allow homogeneous inter-country comparisons using a uniform protocol. The EU-SILC project is carried out under European Union legislation (council regulation No. 1177/2003) and it was formally launched in 2004 for the EU15. In 2006 EU-SILC covered the EU25 Member States as well as Norway and Iceland.

---

<sup>6</sup>Note that only First Order Dominance comparisons can be made here since the ordering is not endowed with cardinal measure.

To analyse the evolution of the Euro area income distribution over time, four temporally equi-spaced waves: 2006, 2009, 2012 and 2015 were chosen and the Euro area defined as those countries that are currently using the euro. Since data for Malta are only available from the 2008 wave, this country is excluded from analysis leaving 18 Euro zone countries.<sup>7</sup>

The income reference period refers to the previous year, consequently analysis with EU-SILC files actually refers to 2005-2014. The income is the total household net disposable income (variable HY020 in the SILC mnemonics), obtained by aggregation of all income sources from all household members net of direct taxes and social contributions. All observations are weighted by cross-sectional weights (variable DB090).

The EU-SILC income definition does not include capital gains, leading to a potential under-estimation of household income, especially top-incomes. Other sources of potential bias in the upper tail of the income distribution derived from EU-SILC are discussed in Törmälehto (2017). However, an increasing number of countries implementing EU-SILC combine interview-based data with register data on incomes. This strategy is expected to mitigate the well-documented low accuracy of sample surveys in estimating top-incomes.

Assuming cohabitation generates economies of scale in consumption and therefore needs do not grow proportionally with members, incomes are age and size-adjusted using the so called modified-OECD equivalence scale. This scale assigns a value of 1 to the household head, of 0.5 to each additional adult member aged 14 and over and of 0.3 to each child aged under 14.

Even if countries share the same currency, the question arises as to whether purchasing power parities should be used to compare incomes from different countries by eliminating the differences in price levels between them. Given significant disparities in the cost of living between countries, for adjusting nominal incomes the PPP index for the household final consumption expenditure is opted for.<sup>8</sup> Households whose income is less than zero were excluded from the sample. Thus the final distribution considered is the real disposable size-adjusted income distribution of a weighted sample of households resident in the euro area.

---

<sup>7</sup>Namely: Austria, Belgium, Cyprus, Estonia, Finland, France, Germany, Greece, Ireland, Italy, Latvia, Lithuania, Luxembourg, Netherlands, Portugal, Slovakia, Slovenia, Spain.

<sup>8</sup>For a discussion on the use of PPPs in the EU income distribution see Brandolini, 2007. On some recent results on the Eu-wide and Eurozone income inequality using EU-SILC data see Filauro, 2017.

## 4 Empirical results

### 4.1 Number of classes and estimation of mixture parameters in the community income distribution

We assume that the overall income distribution in the Eurozone can be described by a mixture of normal distributions. To ensure comparability of inequality measures of the distributions over time, the same origin at zero is taken by excluding negative incomes. Therefore the component densities were taken to be truncated normal, with the number of components to be established. The assumption of normality may be too restrictive, since in principle any functional form can be taken into account. The choice of normality stems from a twofold motivation. Firstly, mixture of normal distributions form a much more general class. In fact, any absolutely continuous distribution can be approximated by a finite mixture of normals with arbitrary precision (Marron and Wand, 1992). Secondly, a mixture model of normals seems to capture better than other functional forms the idea of a polarized economy where relatively homogeneous groups of households are clustered around their expected incomes. The assumption of normality, in fact, results from additive shocks to the expected income of each strata.

The unknown mixture parameters (means, variances and proportions of each component) are estimated by maximum likelihood (ML) via the expectation-maximization (EM) algorithm (Dempster *et al.*, 1977). Starting from a given number of components and an initial parameter  $\Psi^{(0)}$ , the first stage of the algorithm (E-step) is to assign to each data point its current posterior probabilities  $\hat{\tau}_{ik}$  given by (2). The second stage (M-step), comprises estimation of sample means and variances of the normal densities and mixing weights  $\hat{w}_k$  estimated as the means of the probabilities  $\hat{\tau}_{ik}$ . The estimates of the parameters are used to re-attribute a set of improved probabilities of group membership and the sequence of alternate E and M steps continues until a satisfactory degree of convergence occurs to the ML estimates. It is well known that the likelihood function of normal mixtures is unbounded and the global maximizer does not exist (McLachlan and Peel, 2000). Therefore, the maximum likelihood estimator of  $\Psi$  should be the root of the likelihood equation corresponding to the largest of the local maxima located. The solution usually adopted is to apply a range of starting solutions for the iterations. The model was fitted repeatedly using a variety of initial values. Deterministic starting values based on separate models for the outcome based on  $K$  means (Kaufman and Rousseeuw, 1990) were employed and the model fitted.

The model was then fitted 10 additional times based on random jittering of the starting solution, and then another 10 times based on random jittering of the estimates at convergence of the previous runs. The results in the empirical application were fairly stable with respect to the starting solution in the sense that the same maximum for the likelihood or a value very close to it was invariably obtained.

The number of components has been assessed by using the Bayesian's information criterion (BIC). Although regularity conditions do not hold for mixture models, Keribin (2000) showed that BIC is consistent for choosing the number of components in a mixture. In addition, we calculated the Akaike's information criterion (AIC), the consistent Akaike's information criterion (CAIC) and the AIC with a parameter penalty factor of three (AIC3), which is proved to perform well in a mixture context (Andrews and Currim, 2003). Given sample size of between 141,000 and 154,000 observations per year, all the criteria yield similar results and picked a four or five-component mixture as the 'best' parsimonious model for all the years (see Table 1). Specifically, a four-component mixture is selected in the year 2006, while for the remaining years a five-component mixture seems to be preferable according to the criteria. However, the difference of the values taken by all the criteria between four and five component is marginal. Therefore, although the best fitting with five components, we decided to stay with four-components. In fact, adding a fifth component yields a negligible improvement in fit, leaving the first three components unchanged and splitting the fourth component into two classes. Moreover, the fifth component accounts for a very limited proportion of the whole population (from 0.2% to 0.7%) and it largely overlaps the fourth component due to its high variance. The four-component mixture instead is always characterized by distinct means, relatively modest dispersion and non-negligible size. There are no bizarre situations in the model fits such as clusters with very small variance or very flat components with large dispersion and very small probabilities. There is also an absence of components with similar means but different shape due to their disparate variances, etc., i.e. components that can play a role in improving the fit of the whole distribution but may be unacceptable in terms of economic interpretability. The four components can be interpreted as "low"(L), "lower-middle"(LM), "upper-middle"(UM) and "high"(H) income groups.

Table 1: The choice of the number of components according to BIC, CAIC, Akaike and Akaike3

N. of components	loglik	BIC	AIC	CAIC	AIC3
2006					
1	-470056	940136	940114	940138	940138
2	-466189	932437	932383	932442	932393
3	-465139	930373	930286	930381	930302
4	-464693	929516	929397	929527	929419
5	-464695	929556	929404	929570	929432
2009					
1	-510880	1021784	1021762	1021786	1021766
2	-500152	1000363	1000309	1000368	1000319
3	-498302	996699	996612	996707	996628
4	-497977	996084	995965	996095	995987
5	-497864	995894	995742	995908	995770
2012					
1	-528245	1056514	1056492	1056516	1056496
2	-517724	1035507	1035453	1035512	1035463
3	-516046	1032187	1032100	1032195	1032116
4	-515655	1031441	1031321	1031452	1031343
5	-515636	1031438	1031286	1031452	1031314
2015					
1	-562189	1124402	1124380	1124404	1124384
2	-552080	1104220	1104165	1104225	1104175
3	-550175	1100445	1100358	1100453	1100374
4	-549761	1099653	1099533	1099664	1099555
5	-549701	1099569	1099416	1099583	1099444

Figure 1 visually compares the fitted four component mixtures for all the years of the analysis with the corresponding estimated kernel density.<sup>9</sup>

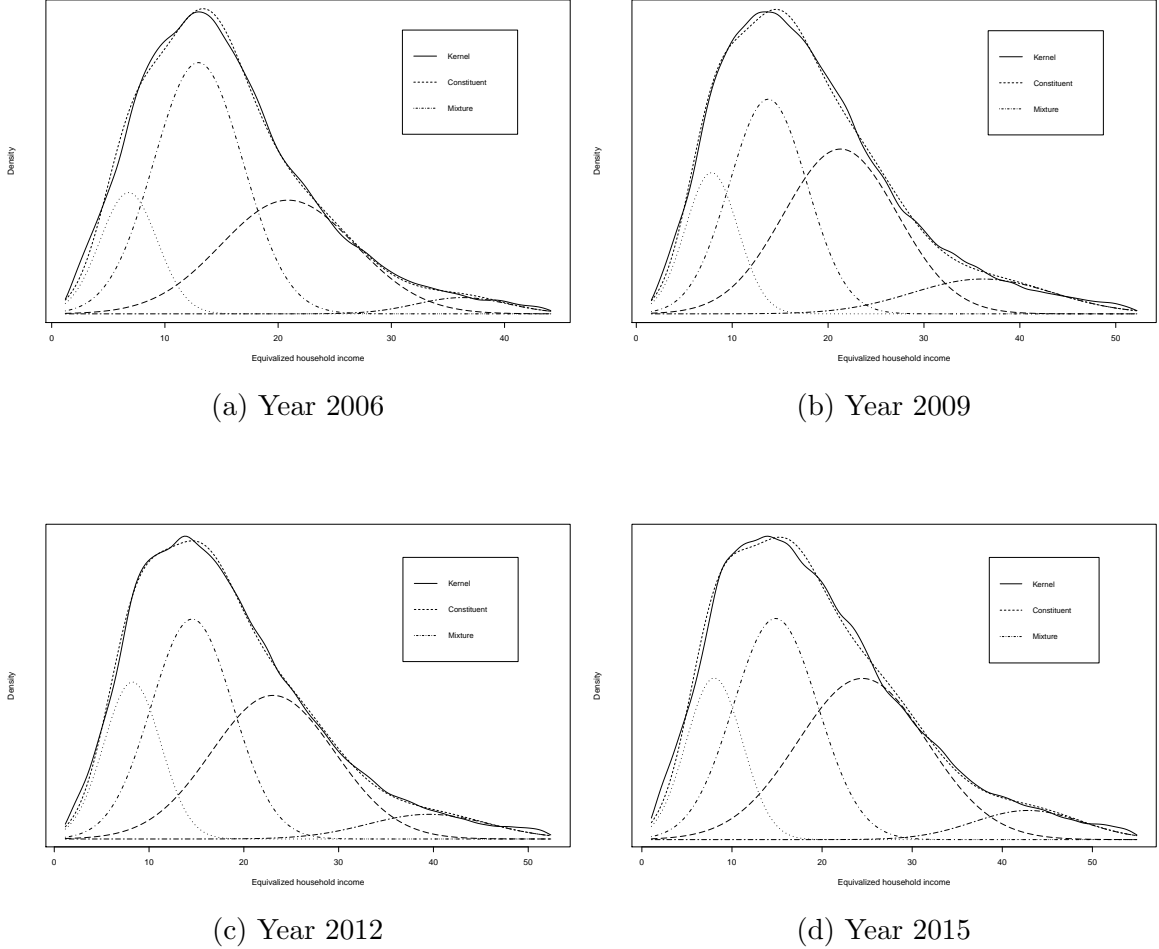


Figure 1: Eurozone income distribution: Four component mixtures with the corresponding estimated kernel density for the years 2006-2015

## 4.2 Inequality, polarization and segmentation in an income class decomposition of the Eurozone income distribution

Table 2 reports the salient statistics compatible with the mixtures shown in Figure 1 for all years: the estimated mean ( $\mu$ ) and standard deviation ( $\sigma$ ) of each truncated normal component along with its corresponding mixing proportion ( $w$ ). As the mixing proportions ( $w$ ) indicate, no single group overwhelms (no group ever accounts for more than half the population). As may be seen mean incomes have generally grown for all groups over the period with growth rates of 1.80% 1.45%, 1.72% and 1.80% respectively

<sup>9</sup>For the purpose of comparison, the variance of each component population was inflated by a factor of  $1 + h^2/\sigma_i^2$  to match that of the kernel density, where  $h$  is the estimated bandwidth of the kernel.

Table 2: Estimated parameters of the components of the mixtures

Year	2006			2009			2012			2015		
Class	$\mu$	$\sigma$	$w$	$\mu$	$\sigma$	$w$	$\mu$	$\sigma$	$w$	$\mu$	$\sigma$	$w$
Low (L)	6.77	2.50	0.15	7.84	2.68	0.16	8.21	2.98	0.19	7.99	3.14	0.18
Lower-Middle (LM)	12.95	3.90	0.49	13.75	3.99	0.35	14.54	4.26	0.38	14.83	4.66	0.36
Upper-Middle (UM)	20.88	5.92	0.33	21.30	5.90	0.39	23.08	6.47	0.37	24.47	7.17	0.40
High (H)	36.27	4.01	0.03	36.11	7.29	0.10	39.52	6.36	0.06	42.85	5.98	0.06

Note:  $\mu$  and  $\sigma$  are expressed in PPA-adjusted Euros.  $w$  are the mixing proportions.

with a slight downturn for the lowest income group in 2015. Income variation has grown steadily over the period for the Low, Lower-Middle and Upper-Middle groups with a Kuznets curve like inverted U shaped profile for the richest group. The Low and High income groups are the smallest in size with the former in the range of 15 to 19% of the population and the latter between 3 and 10% of the population. The size of the poorest group has grown over the period as has the size of the upper middle and high income groups, but the lower middle income group has diminished substantially suggestive of some polarization in the community. To examine this an analogue of the Duclos, Esteban and Ray (2004) polarization measure for mixtures of  $K$  normal distributions<sup>10</sup> can be calculated by the the ERP index, an adaptation of Esteban and Ray's discrete version of (10) computed at distribution modes and their respective abscissa scaled by the respective relative subgroup size where:

$$ERP = \sum_{k=1}^K \sum_{j=1}^K |\mu_k - \mu_j| \left( \frac{w_k}{\sigma_k} + \left( \frac{w_j}{\sigma_j} \right)^{1+\alpha} \right) \frac{1}{\sqrt{2\pi}}. \quad (15)$$

Here  $\alpha$  corresponds to the polarization intensity parameter (chosen by the investigator) should be in the range (0.25, 1]. This corresponds to an aggregation over all possible pairs in many groups of the two group trapezoidal polarization measure proposed in Anderson (2010) and follows the interpretation of the Esteban and Ray (1994) and Duclos Esteban and Ray (2004) polarization measures as the expected value of all possible trapezoids that can be formed from modal differences and their appropriately scaled abscissa. Here it is offered as an alternative to the segmentation based polarization measure proposed in (11) above. Table 3 presents the statistics for a range of polarization intensity parameters and confirms that there appears to be ongoing income polarization throughout the period.

<sup>10</sup>See Pittau et al., 2010, for a discussion on polarization measurements within a normal mixture framework.



Table 3: Polarization coefficients

$\alpha$	2006	2009	2012	2015
0.25	1.070	1.049	0.976	1.133
0.5	0.869	0.871	0.796	0.943
1	0.746	0.768	0.693	0.838

Turning to the Gini decomposition analysis, Table 4 presents various subgroup Gini coefficients and Table 5 records the decomposition results.<sup>11</sup> As measured by the Gini coefficient, income inequality in the euro area had increased by the end of the period overall and in all income subgroups. It dipped in 2009 for all but the high income group (whose inequality peaked at this time) perhaps due to the economic exigencies of the time (recall 2009 refers to the year 2008, the year of the economic crash). Although the four groups appear to be well segmented (there appears to be little group overlap) the degree of segmentation has not changed much over the period. The within group inequality component diminished over the period while between group inequality increased. The overall segmentation-based polarization index coheres with the results in Table 3, but the decomposition results suggest that it is predominantly polarization of the poor and rich that is underlying these effects.

Table 4: Gini coefficients: overall and by subgroups.

	Overall	Low	Lower-Middle	Upper-Middle	High	Non-Lower	Non-High
2006	0.385	0.283	0.239	0.226	0.088	0.330	0.343
2009	0.400	0.267	0.231	0.221	0.158	0.342	0.335
2012	0.404	0.281	0.233	0.224	0.125	0.338	0.349
2015	0.421	0.298	0.249	0.233	0.104	0.352	0.362

Note: Non-Lower stands for all subgroups excluding the lowest.

Non-High stands for all subgroups excluding the highest.

Overall Gini coefficient standard errors (Giles 2004) are always of an order less than 0.002.

A better measure of inequality of distribution between the four classes that captures the lack of commonality in the sub-distributions, is the zero order normalized transvariation ( $TR_0$ ), presented in Section 2.3. It was 0.1403 in 2006 and 0.1398, 0.1401, 0.1454 respectively in subsequent observation years suggesting very little change in the in-

<sup>11</sup>Inferential comparison of Gini coefficients was implemented using Giles (2004) simple regression technique. As Modarres and Gastwirth (2006) and Davidson (2009) both indicate, Giles (2004) overstates the magnitude of the standard error so it can be considered an upper bound. Since it turns out to be very small relative to observed differences in the Gini coefficients rendering differences significant, further more sophisticated computations were deemed to be unwarranted.

Table 5: Gini decomposition, segmentation and polarization of subgroups measured by the within and between components, the non-segmentation factor (NSF), the segmentation index (SI), the Gini-based polarization index (PG), the polarization index of the low income group ( $\text{Pol}_{\text{poor}}$ ) and the polarization index of the high income group ( $\text{Pol}_{\text{rich}}$ ). The latest are obtained fixing  $\alpha = 0.5$ .

Year	Within Gini	Between Gini	NSF	SI	PG	$\text{Pol}_{\text{poor}}$ ( $\alpha = 0.5$ )	$\text{Pol}_{\text{rich}}$
2006	0.084	0.202	0.099	0.743	0.624	0.370	0.172
2009	0.068	0.223	0.109	0.726	0.636	0.375	0.297
2012	0.072	0.223	0.110	0.729	0.634	0.408	0.236
2015	0.078	0.231	0.112	0.734	0.635	0.396	0.232

equality of subgroup distributions over the 2006–2012 period but a substantial increase in subgroup distributional differences in 2015.

### 4.3 The progress of individual constituent nations

Table 6 presents the class membership probabilities for each country in each year of the analysis calculated according to equation (3).<sup>12</sup> Estonia, Lithuania, Latvia and Slovakia all had over 50% membership of the Low income group at the beginning of the period whereas, of that group only Latvia, were in that position (along with Greece, who had well under 50% membership in 2006) at the end of the period. Indeed, the variability of nation poor group membership experience diminished substantially over period suggesting a greater sharing of poor group membership amongst the constituent nations. At the other end of the income spectrum 5 nations (Austria, Germany, Ireland, Luxembourg and Netherlands) enjoyed over 10% of their population in the high income group while Luxembourg’s outstanding 26.7% membership of the high income group at the beginning of the period was somewhat diluted by the end where they were joined by France and Finland as the only members in the over 10% group.

The extent to which income class structures vary across the constituent nations can be assessed by considering the discrete many distribution transvariation (TRMD), analogue of Gini’s Transvariation of class membership distributions (Anderson, Linton and Thomas, 2017) across the 18 nations in Table 6. This statistic, a number between 0 and 1, measures the extent to which a collection of distributions differ. If all distributions are identical it will take on the value 0, if nations are perfectly segmented

<sup>12</sup>Similar results have been obtained adopting the alternative estimate of country membership as in equation (5).

Table 6: Membership probabilities for each country: years 2006–2015. Income groups are labeled Low (L), Lower-Middle (LM), Upper-Middle (UM) and High (H)

Nation	2006				2009				2012				2015			
	L	LM	UM	H	L	LM	UM	H	L	LM	UM	H	L	LM	UM	H
Austria	0.051	0.360	0.483	0.105	0.101	0.355	0.436	0.108	0.086	0.294	0.463	0.157	0.074	0.329	0.518	0.078
Belgium	0.077	0.413	0.433	0.077	0.139	0.395	0.394	0.072	0.118	0.365	0.426	0.092	0.100	0.386	0.463	0.052
Cyprus	0.113	0.400	0.394	0.093	0.157	0.360	0.375	0.108	0.143	0.355	0.375	0.126	0.223	0.410	0.329	0.038
Germany	0.069	0.375	0.452	0.105	0.115	0.357	0.421	0.107	0.099	0.311	0.447	0.143	0.107	0.355	0.467	0.071
Estonia	0.531	0.369	0.095	0.004	0.525	0.348	0.119	0.008	0.487	0.352	0.149	0.013	0.409	0.384	0.194	0.013
Greece	0.243	0.452	0.267	0.038	0.334	0.409	0.228	0.028	0.447	0.384	0.158	0.012	0.528	0.359	0.109	0.004
Spain	0.196	0.440	0.312	0.052	0.248	0.388	0.298	0.066	0.256	0.375	0.304	0.066	0.333	0.411	0.239	0.017
Finland	0.089	0.423	0.413	0.074	0.101	0.344	0.433	0.122	0.077	0.290	0.468	0.165	0.048	0.264	0.548	0.140
France	0.079	0.420	0.419	0.082	0.075	0.320	0.447	0.158	0.071	0.295	0.470	0.164	0.038	0.257	0.572	0.132
Ireland	0.109	0.424	0.365	0.102	0.160	0.392	0.355	0.094	0.179	0.388	0.350	0.083	0.173	0.404	0.377	0.046
Italy	0.126	0.427	0.376	0.071	0.181	0.387	0.351	0.081	0.160	0.351	0.388	0.102	0.171	0.397	0.387	0.044
Lithuania	0.579	0.333	0.085	0.003	0.545	0.326	0.118	0.011	0.533	0.329	0.128	0.009	0.492	0.358	0.141	0.009
Luxembourg	0.032	0.228	0.474	0.267	0.041	0.244	0.457	0.258	0.035	0.214	0.467	0.285	0.042	0.249	0.549	0.160
Latvia	0.604	0.311	0.080	0.005	0.587	0.297	0.105	0.011	0.594	0.291	0.106	0.009	0.537	0.332	0.126	0.006
Netherlands	0.024	0.336	0.523	0.118	0.043	0.320	0.500	0.137	0.049	0.314	0.502	0.135	0.057	0.361	0.522	0.060
Portugal	0.343	0.426	0.191	0.040	0.436	0.360	0.171	0.033	0.413	0.365	0.186	0.035	0.411	0.392	0.183	0.014
Slovenia	0.127	0.517	0.327	0.029	0.179	0.465	0.322	0.033	0.181	0.435	0.345	0.039	0.191	0.480	0.313	0.016
Slovakia	0.619	0.318	0.061	0.002	0.524	0.365	0.106	0.004	0.336	0.442	0.208	0.014	0.376	0.474	0.148	0.002

or different in the extreme (each is in one class and no other and there is at least one nation in each class) the statistic will take on the value 1. Letting  $p_{hk}$  be an element of the  $18 \times 4$  matrix  $P$  whose  $(h,k)$  element is the probability that nation  $h$  is in class  $k$ ,  $h = 1, \dots, 18$ ,  $k = 1, \dots, 4$  and  $P_{\cdot k}$  corresponds to the  $k$ 'th column of  $P$ , then

$$\text{TRMD} = \frac{\sum_{k=1}^4 (\max(P_{\cdot k}) - \min(P_{\cdot k}))}{4}.$$

The TRMD class membership distribution measures for the observation years from 2006 onwards were 0.403, 0.354, 0.365 and 0.338 respectively indicative of some convergence over the period, i.e. nations were becoming more alike in their income class membership distributions.

In order to assess a nation's progress in terms of class membership an ordinal comparison of income class structure can be performed.<sup>13</sup> In essence a first order dominance comparison over discrete ordered states is performed wherein there is no attribution of cardinal measure class differences, all that is asserted is that higher classes are preferred to lower classes.

Let  $F_{h,2006}(C(k)) = \sum_{j=1}^k p_{hj,2006}$  and  $F_{h,2015}(C(k)) = \sum_{j=1}^k p_{hj,2015}$  and consider the difference  $D_h(C(k)) = F_{h,2006}(C(k)) - F_{h,2015}(C(k))$ , for  $k = 1, \dots, K - 1$  and  $h = 1, \dots, H$ . If  $D_h(C(k)) \leq 0$  for all  $k$ , then 2006 dominates 2015 marking an unequivocal downward transition or deterioration in income class. If  $D_h(C(k)) \geq 0$  for all  $k$ , then 2015 dominates 2006 marking an unequivocal upward transition or improvement in

<sup>13</sup>This can be thought of as avoiding data contamination issues that would be present in ordinal comparison.

income classes. The studentized maximum modulus distribution (Stoline and Ury, 1979) indicates an asymptotic 0.01 critical value of 2.934 for the ‘t’ statistic in three way multiple comparisons. Based upon this, Table 7 reports the First Order Dominance comparison of 2006 and 2015. Perhaps most clearly seen in the 2006-2015 comparison diagrams in the Appendix A.2 twelve countries<sup>14</sup> (Austria, Belgium, Cyprus, Germany, Greece, Spain, Ireland, Italy, Luxembourg, Netherland, Portugal and Slovenia) suffered a deterioration in their fortune in that 2006 stochastically dominated 2015 whereas the remaining six countries (Estonia, Finland, France, Lithuania, Latvia and Slovakia) enjoyed advances in their income classification fortunes wherein 2015 dominated 2006. It is of interest to note that with the exception of France and Finland all of these countries joined the union recently in 2004 whereas with the exception of Slovenia the countries that suffered a deterioration in class status all joined the union much earlier in the process and were relatively long time members.

Table 7: First order class comparisons (ordinal comparisons)

Country	$D_h(C(1))$	$D_h(C(2))$	$D_h(C(3))$	t values			
Austria	-0.023	0.008	-0.027	5.14	0.88	5.05	2006 dominates 2015
Belgium	-0.023	0.004	-0.026	4.36	0.43	5.71	2006 dominates 2015
Cyprus	-0.110	-0.120	-0.055	13.22	10.71	9.63	2006 dominates 2015
Germany	-0.038	-0.018	-0.033	10.75	2.90	9.41	2006 dominates 2015
Estonia	0.122	0.107	0.008	12.99	15.86	4.48	2015 dominates 2006
Greece	-0.285	-0.192	-0.034	39.81	28.49	12.98	2006 dominates 2015
Spain	-0.137	-0.108	-0.035	24.37	18.20	14.88	2006 dominates 2015
Finland	0.041	0.200	0.065	11.77	29.93	15.16	2015 dominates 2006
France	0.041	0.204	0.051	12.44	30.38	11.85	2015 dominates 2006
Ireland	-0.064	-0.044	-0.056	9.63	4.63	11.31	2006 dominates 2015
Italy	-0.045	-0.015	-0.026	12.23	2.93	10.91	2006 dominates 2015
Lithuania	0.087	0.062	0.006	8.43	9.28	3.76	2015 dominates 2006
Luxembourg	-0.010	-0.031	-0.106	2.16	2.83	10.65	2006 dominates 2015
Latvia	0.067	0.046	0.000	6.71	7.47	0.00	2015 dominates 2006
Netherlands	-0.033	-0.058	-0.057	11.46	8.06	13.54	2006 dominates 2015
Portugal	-0.068	-0.034	-0.026	7.56	4.39	7.98	2006 dominates 2015
Slovenia	-0.064	-0.027	-0.013	11.76	3.82	5.93	2006 dominates 2015
Slovakia	0.243	0.087	0.000	25.84	14.83	0.00	2015 dominates 2006

A nations propensity for intertemporal polarization or convergence can be assessed using a  $j$  period polarization statistic,  $PS_h(j)$ , which reflects the extent to which the nations incomes are moving from the center to the tails over  $j$  periods. Defining the probability of being in the tails of the distribution (i.e. in low or high income class) in

<sup>14</sup>Note that for Austria and Belgium the second class component is estimated positive, however not significantly so, that is to say one could not reject the hypothesis that the component was negative, thus taken with the significant 1st and 3rd components one could not reject the joint hypothesis that 2006 dominates 2015 for these two countries.

period  $t$  as  $P_{lh,t}$  and the probability of being in the middle of the distribution (i.e. in the lower or upper middle) in period  $t$  as  $P_{mh,t} = (1 - P_{lh,t})$ ,

$$PS_h(j) = 0.5 + (P_{lh,t} - P_{mh,t}) - (P_{lh,t-j} - P_{mh,t-j}) = 0.5 + 2(P_{lh,t} - P_{lh,t-j}).$$

Based on the null hypothesis of no change in polarization, the polarization statistic would equal 0.5 and would be asymptotically  $N(0.5, 0.25((1/n_t) + (1/n_{t-j})))$  where  $n$  is sample size. Values of  $PS(j)$  greater than 0.5 imply polarization, values less than 0.5 imply convergence. These are presented in Table 8. Over the 2006–2015 period Estonia, Lithuania and Slovakia converged significantly (Latvia did not move significantly) whereas all other nations diverged or polarized significantly with Greece making the most striking movement.

Table 8: Nation polarization statistics and standard errors.

	2006-2009		2006-2012		2009-2012		2006-2015		2009-2015		2012-2015	
Country	$PS_h(j)$	s.e.	$PS_h(j)$	s.e.	$PS_h(j)$	s.e.	$PS_h(j)$	s.e.	$PS_h(j)$	s.e.	$PS_h(j)$	s.e.
Austria	0.605	0.009	0.673	0.009	0.568	0.009	0.492	0.009	0.387	0.009	0.319	0.009
Belgium	0.614	0.009	0.611	0.009	0.497	0.009	0.495	0.009	0.381	0.009	0.384	0.009
Cyprus	0.618	0.012	0.627	0.011	0.509	0.012	0.610	0.011	0.492	0.012	0.483	0.011
Germany	0.597	0.006	0.637	0.006	0.540	0.006	0.509	0.006	0.412	0.006	0.372	0.006
Estonia	0.495	0.010	0.428	0.010	0.433	0.010	0.273	0.010	0.278	0.010	0.345	0.010
Greece	0.663	0.009	0.855	0.010	0.692	0.009	1.002	0.008	0.839	0.007	0.647	0.008
Spain	0.632	0.006	0.647	0.006	0.515	0.006	0.704	0.007	0.572	0.006	0.557	0.006
Finland	0.619	0.007	0.657	0.007	0.538	0.007	0.549	0.007	0.430	0.007	0.392	0.007
France	0.644	0.007	0.648	0.007	0.504	0.007	0.519	0.007	0.375	0.007	0.371	0.007
Ireland	0.585	0.010	0.602	0.010	0.517	0.010	0.516	0.010	0.431	0.010	0.414	0.010
Italy	0.630	0.005	0.629	0.005	0.499	0.005	0.537	0.005	0.407	0.005	0.408	0.005
Lithuania	0.448	0.010	0.421	0.010	0.473	0.010	0.338	0.010	0.390	0.010	0.417	0.010
Luxembourg	0.501	0.012	0.542	0.011	0.541	0.010	0.307	0.012	0.306	0.012	0.265	0.011
Latvia	0.478	0.010	0.488	0.010	0.510	0.009	0.367	0.010	0.389	0.009	0.379	0.009
Netherlands	0.577	0.007	0.585	0.007	0.508	0.007	0.451	0.007	0.374	0.007	0.366	0.007
Portugal	0.672	0.011	0.631	0.010	0.459	0.010	0.584	0.009	0.412	0.009	0.453	0.008
Slovenia	0.613	0.007	0.628	0.007	0.515	0.007	0.602	0.007	0.489	0.008	0.474	0.008
Slovakia	0.315	0.010	-0.042	0.010	0.143	0.010	0.014	0.010	0.199	0.010	0.556	0.010

Finally Table 9 presents the results of the Utopia/Dystopia index for the euro area countries across years along with the corresponding ranking of the eighteen countries. As may be seen rankings are fairly stable over time with Luxembourg enjoying Utopian status in two years and Latvia and Slovakia each suffering Dystopian status in one year. Significant movers are Greece, which dropped from 13th to 18th place over the period, and Finland and France who each displayed considerable improvement in ranking over the period.

Table 9: Utopia Index and Rank for each country: years 2006-2015

	2006		2009		2012		2015	
Country	UI	Rank	UI	Rank	UI	Rank	UI	Rank
Austria	0.78	3	0.73	5	0.79	5.00	0.82	4
Belgium	0.69	5	0.62	8	0.65	7.00	0.71	7
Cyprus	0.66	8	0.65	7	0.65	8.00	0.48	10
Germany	0.75	4	0.71	6	0.75	6.00	0.74	6
Estonia	0.08	15	0.06	15	0.11	16.00	0.19	13
Greece	0.43	13	0.30	13	0.14	15.00	0.01	18
Spain	0.50	12	0.47	12	0.44	12.00	0.29	12
Finland	0.67	7	0.75	4	0.81	4.00	0.96	3
France	0.69	6	0.83	3	0.81	2.00	0.98	2
Ireland	0.66	9	0.61	9	0.55	10.00	0.57	9
Italy	0.62	10	0.57	10	0.61	9.00	0.58	8
Lithuania	0.04	16	0.05	16	0.06	17.00	0.07	16
Luxembourg	0.99	1	1.00	1	1.00	1.00	1.00	1
Latvia	0.03	17	0.01	18	0.00	18.00	0.02	17
Netherlands	0.84	2	0.86	2	0.81	3.00	0.80	5
Portugal	0.31	14	0.19	14	0.21	14.00	0.18	14
Slovenia	0.53	11	0.49	11	0.48	11.00	0.46	11
Slovakia	0.00	18	0.04	17	0.25	13.00	0.16	15

## 5 Concluding remarks

By employing mixture distribution techniques to determine the number and size of groups or classes in an income distribution, a four class model of the income distribution of the Eurozone countries over the decade spanning 2006–2015 has been developed without resort to hard class boundaries. Some new indices of polarization and segmentation are developed in the context of a decomposition of the Gini coefficient and the roles of, and relationships between, these groups in societal income inequality, poverty, polarization and societal segmentation are examined. Implications for the individual constituent nations of the collective are explored.

When viewed as an entity in itself what emerged was a four class increasingly unequal polarizing structure with income growth in all four classes for the Eurozone. With regard to individual constituent nation class membership results over the sample period, six nations were seen to be advancing (Estonia, Finland, France, Lithuania, Latvia, Slovenia), twelve falling back (Austria, Belgium, Cyprus, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Slovenia and Spain) all of whom, with the exception of Slovenia were longer time members of the EU. In terms of an ordinally ranked class structure, with the exception of Estonia, Lithuania and Slovakia who converged significantly and Latvia who did not move, all nations exhibited significant polarizing behavior, in the form of significantly divergent behaviour in class transitions over the period. Thus, in the face of increasing overall inequality in the Eurozone indicated in Tables 2, 3 and 4 and the within country increasing inequality/polarization indicated in Table 8, the increasing commonality of class membership distributions across its constituent nations heralded by the across nation Transvariation statistics can be construed as characteristic of a more cohesive society, in essence constituent nations are becoming more alike through their individual increasing variation engendering increasingly overlapping income distributions between nations. Thus while the Eurozone is experiencing increasing income class inequality and polarization its constituent nations are experiencing increasing similarities in their respective income distribution.

While this paper has reached some conclusions on the anatomy of the income distribution in the Euro area, several opportunities for extending the scope of this paper remain. For example, the European Union Survey on Income and Living Conditions provides a rich set of socioeconomic covariates that can be used to model the class

membership of the households, along the lines of Anderson, Farcomeni, Pittau and Zelli (2016). Furthermore, the possibility of using concomitant variables at different levels of aggregation—households nested within EU countries—(Konte, 2016) can contribute to explain variation between and within countries for each income class.



## References

- Atkinson, A.B., and Brandolini, A. (2013). “On the identification of the middle class”, in *Income Inequality* (eds. J. C. Gornick and M. J. Lantti), Stanford University Press, 77-100.
- Alvarez-Esteban, P.C., del Barrio, E., Cuesta-Albertos J.A. and C. Matrán (2016). “A contamination model for the stochastic order”, *Test*, 25, 751–774.
- Anderson G.J. (2010). “Polarization of the Poor: Multivariate Relative Poverty Measurement Sans Frontiers”, *Review of Income and Wealth*, 56(1), 84–101.
- Anderson, G.J., Farcomeni, A., Pittau, M.G., Zelli, R. (2016). “A new approach to measuring and studying the characteristics of class membership: The progress of poverty, inequality and polarization of income classes in urban China”. *Journal of Econometrics*, 191(2), 348–359.
- Anderson, G.J., Farcomeni, A., Pittau, M.G., Zelli, R. (2017). “More equal yet less similar: Human development and the progress of nation wellbeing since 1990. A multidimensional mixture distribution analysis”. *mimeo*.
- Anderson, G.J., Linton, O., Thomas J. (2017). “Similarity, dissimilarity and exceptionality: generalizing Gini’s transvariation to measure ‘differentness’ in many distributions”. *Metron*, 75, 161–180.
- Anderson G.J. and Leo, T.W.(2017). “On Providing a Complete Ordering of Non-Combinable Alternative Prospects”, *Mimeo*, University of Toronto.
- Anderson, G.J., Post T., Whang Y.-J. (2017). “Ranking Incomparable Prospects: The Utopia Index” *Mimeo*, University of Toronto.
- Anderson, G.J., and Thomas, J. (2017). “More unequal yet increasingly similar incomes: Polarization, segmentation and ambiguity in the changing anatomy of constituent Canadian income distributions in the 21st century”. Working Paper 587, Department of Economics, University of Toronto.
- Andrews, R. A. and Currim, I. S. (2003). “A comparison of segment retention criteria for finite mixture logit’s models”, *Journal of Marketing Research*, XL, 235–243.
- Bourguignon F. (1979). Decomposable Income Inequality Measures. *Econometrica*, 47 (4), 901–920.
- Brandolini, A. (2007). “Measurement of income distribution in supranational entities:

- The case of the European Union”, in S.P. Jenkins and J. Micklewright (eds.), *Inequality and Poverty Re-Examined*, 62–83, Oxford, Oxford University Press.
- Cowell, F. and Flachaire, E. (2015). “Statistical methods for distributional analysis”, in *Handbook of Income Distribution*, (eds. A. Atkinson and F. Bourguignon), Elsevier, North Holland, Amsterdam, Vol. 2A, Ch.6, 359–465.
- Davidson, R. (2009). “Reliable inference for the Gini index”, *Journal of Econometrics*, 150(1), 30–40.
- Deaton, A. (2010). “Price Indexes, Inequality, and the Measurement of World Poverty”, *American Economic Review*, 100:1, 5–34.
- Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). “Maximum likelihood from incomplete data via EM algorithm”, *Journal of the Royal Statistical Society*, B, 69, 1–38.
- Duclos, J-Y, Esteban, J., and Ray, D. (2004) “Polarization: Concepts, Measurement, Estimation”, *Econometrica* 72 (6), 1737–1772.
- Filauro, S. (2017). “European incomes, national advantages: EU-wide inequality and its decomposition by country and region”. EERI Research Papers Series No 05/2017.
- Giles, D. (2004). “Calculating a Standard Error for the Gini Coefficient: Some Further Results”. *Oxford Bulletin of Economics and Statistics*, 66 (3), 425–433.
- Gini, C. (1916). “Il concetto di transvariazione e le sue prime applicazioni”, *Giornale degli Economisti e Rivista di Statistica*, in C. Gini (1959): 1–55.
- Hey, J. D. and Lambert, P. J. (1980). “Relative Deprivation and the Gini Coefficient: Comment”, *Quarterly Journal of Economics*, 95, 567–573.
- Kaufman, L. and Rousseeuw P.J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, New York.
- Keribin, S. (2000). “Consistent estimation of the order of mixture model”, *Sankhya*, 62, 49–66.
- Konte, M. (2016). “The effects of remittances on support for democracy in Africa: are remittances a curse or a blessing?”, *Journal of comparative economics*, 44, 1002–1022.
- Longford, N.T. (2014). *Statistical Studies of Income, Poverty and Inequality in Europe: Computing and Graphics in R Using EU-SILC*. Chapman & Hall/CRC Press.

- Marron, J.S. and Wand M.P. (1992). “Exact mean integrated squared error”, *The Annals of Statistics*, 20, 712–736.
- McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. Wiley, New York.
- Milanovic, B. (2011). *The Haves and the Have-Nots: A Brief and Idiosyncratic History of Global Inequality*, New York, Basic Books.
- Modarres, R. and Gastwirth, J.L. (2006). “A cautionary note on estimating the standard error of the Gini index of inequality”, *Oxford Bulletin of Economics and Statistics*, 68, 385–390.
- Mookherjee, D. and Shorrocks, A. (1982). “A decomposition analysis of the trend in UK income inequality”. *Economic Journal*, 92(368), 886—902.
- OECD (2011). *Perspectives on Global Development 2012: Social Cohesion in a Shifting World*, OECD Publishing, Paris.
- Pittau, M.G. and Zelli, R. (2017). “At the roots of Gini’s transvariation: extracts from ‘Il concetto di transvariazione e le sue prime applicazioni’ ”, *Metron*, 75(2), 127–140.
- Pittau, M.G., Zelli, R. and Johnson, P.A. (2010). “Mixture Models, Convergence Clubs and Polarization”, *Review of Income and Wealth*, 56(1), 102–122.
- R Development Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Ravallion, M. (2010). “Mashup indices of development”, *Policy Research Working Paper*, n. 5432, World Bank, Washington DC.
- Ravallion, M., (2012). “Why Don’t We See Poverty Convergence?”, *Economic Review*, 102, 504–523.
- Stoline M.R., and Ury, H.K. (1979). “Tables of the Studentized Maximum Modulus Distribution and an Application to Multiple Comparisons among Means”. *Technometrics*, 21, 87–93.
- Törmälehto, V-M. (2017). “High income and affluence: Evidence from the European Union statistics on income and living conditions (EU-SILC)”. Statistical Working Papers, Eurostat, Luxembourg.
- Yitzhaki, S. (1994). “Economic Distance and overlapping of distributions”, *Journal of Econometrics*, 61 147–159.

Weymark, John (2003). “Generalized Gini Indices of Equality of Opportunity”. *Journal of Economic Inequality*, 1(1), 5–24.

# A Appendix

## A.1 Gini segmentation

To derive the segmentation version of GINI in the context of continuous distributions note that:

$$\begin{aligned}
\text{GINI} &= \frac{1}{\mathbb{E}(x)} \int_0^\infty \int_0^\infty f(y)f(x)|x-y|dx dy = \\
&= \frac{1}{\sum_{k=1}^K w_k \mu_k} \int_0^\infty \int_0^\infty \sum_{k=1}^K w_k f_k(y) \sum_{k=1}^K w_k f_k(x) |x-y| dx dy = \\
&= \frac{1}{\mu} \int_0^\infty \int_0^\infty \left[ \sum_{k=1}^K w_k^2 f_k(y) f_k(x) |x-y| + \sum_{k=2}^K w_k f_k(y) \sum_{j=2}^{k-1} w_j f_j(x) |x-y| \right] dx dy = \\
&= \frac{1}{\mu} \int_0^\infty \int_0^\infty \left[ \sum_{k=1}^K w_k^2 \frac{\mu_k}{\mu} f_k(y) f_k(x) |x-y| \right] dx dy + \\
&+ \frac{1}{\mu} \int_0^\infty \int_0^\infty \left[ \sum_{k=2}^K w_k f_k(y) \sum_{j=2}^{k-1} w_j f_j(x) |x-y| \right] dx dy = \\
&= \sum_{k=1}^K w_k^2 \frac{\mu_k}{\mu} \text{GINI}_k + \frac{2}{\mu} \int_0^\infty \int_0^\infty \left[ \sum_{k=2}^K w_k f_k(y) \sum_{j=2}^{k-1} w_j f_j(x) |x-y| \right] dx dy
\end{aligned}$$

where

$$\text{GINI}_k = \frac{1}{\mu_k} \int_0^\infty \int_0^\infty f_k(y) f_k(x) |x-y| dx dy.$$

From the second component, consider a typical term:

$$\begin{aligned}
&\int_0^\infty \int_0^\infty [w_k f_k(y) w_j f_j(x) |x-y|] dx, dy = \\
&= w_k w_j \int_0^\infty f_k(y) \int_0^\infty f_j(x) |x-y| dx, dy = \\
&= w_k w_j \int_0^\infty f_k(y) \left[ -\int_0^y f_j(x)(x-y) dx + 2 \int_y^\infty f_j(x)(x-y) dx \right] = \\
&= w_k w_j \int_0^\infty f_k(y) \left[ -(\mu_j - y) + 2 \int_y^\infty f_j(x)(x-y) dx \right] = \\
&= w_k w_j \left[ |\mu_j - \mu_k| + 2 \int_0^\infty f_k(y) \int_y^\infty f_j(x)(x-y) dx dy \right].
\end{aligned}$$

Note  $\int_0^\infty f_k(y) \int_y^\infty f_j(x)(x-y) dx dy > 0$ , so that:

$$\begin{aligned} \text{GINI} &= \sum_{k=1}^K w_k^2 \frac{\mu_k}{\mu} G_k + \frac{1}{\mu} \sum_{k=2}^K \sum_{j=1}^k w_k w_j |\mu_k - \mu_j| + \\ &+ \frac{2}{\mu} \sum_{k=2}^K \sum_{j=1}^{k-1} w_k w_j \int_0^\infty f_k(y) \int_y^\infty f_j(x) (x - y) dx dy. \end{aligned}$$

## A.2 National class membership cumulative density

