Estimation of the Income Distribution and Detection of Subpopulations: an Explanatory Model

by

Emmanuel Flachaire

Université Paris I Panthéon-Sorbonne

and

Olivier Nuñez

Universidad Carlos III de Madrid

April 2006

Abstract

Empirical evidence, obtained from nonparametric estimation of the income distribution, exhibits strong heterogeneity in most populations of interest. It is common, therefore, to suspect that the population is composed of several homogeneous subpopulations. Such an assumption leads us to consider mixed income distributions whose components feature the distributions of the incomes of a particular homogeneous subpopulation. We develop a model with mixing probabilities that are allowed to vary with exogenous individual variables that characterize each subpopulation. This model simultaneously provides a flexible estimation of the income distribution, a breakdown into several subpopulations and an explanation of income heterogeneity.

JEL classification : C13, D31

Key-words : income distribution, mixture models.

1 Introduction

In inequality analysis, parametric and non-parametric estimation often suggests heavy-tails or bi-modality in the income distribution (Marron and Schmitz 1992, Schluter and Trede 2002, Davidson and Flachaire 2004). This suggests heterogeneity in the underlying population. To model this heterogeneity it is natural is to assume that the population can be broken down into several homogeneous subpopulations. This is the starting point of our paper. Empirical studies on income distribution indicate that the Lognormal distribution fits homogeneous subpopulations quite well (Aitchison and Brown 1957, Weiss 1972). And the theory of mixture models indicates that, under regularity conditions, any probability density can be consistently estimated by a mixture of normal densities (see Ghosal and van der Vaart 2001 for recent results about rates of convergence). Thus, from the relationship between the Normal and Lognormal distributions, we see that any probability density with a positive support (as for instance income distribution) can be consistently estimated by a mixture of Lognormal densities. We expect, then, to be able to estimate closely the true income distribution with a finite mixture of Lognormal distributions and so to identify the subpopulations.

In this paper, we analyse conditional income distributions using lognormal mixtures. Our contribution is to propose a conditional model by specifying the mixing probabilities as a particular set of functions of individual characteristics. This allows us to characterize the distinct homogeneous subpopulations: we assume that an individual's belonging to a specific subpopulation can be explained by his individual characteristics. For instance, households with no working adult are more likely to be nearer the bottom of the income distribution than those with all-working adults. The probability of belonging to a given subpopulation, then, may vary among individuals as explained by individual characteristics.

The method is applied to disposable household income, as obtained from a survey studying changes in inequality and polarization in the United Kingdom in the 1980s and 1990s. This empirical study demonstrates the usefulness of our method and, although the results are all confirmed by previous studies, they do not lead to conclusions as rich as those achieved here. We find that our method produces results that are readily given to economic interpretation.

The paper is organized as follows: we present our explanatory mixture model in Section 2 and illustrate its use in Section 3.

2 The Explanatory Mixture Model

We assume that the population can be broken down into K homogeneous subpopulations with a proportion p_k of the population, each being a logarithmic-transformation of the Normal distribution with mean μ_k and standard deviation σ_k. Thus, the density function of the income distribution in the population is defined as,

$$f(y) = \sum_{k=1}^{K} p_k \, \Lambda \left(y \, ; \mu_k, \sigma_k \right) \tag{1}$$

where $\Lambda(.;\mu,\sigma)$ is the Lognormal distribution with parameters μ and σ. Note that, as with the number of modes used to detect heterogeneity, the number of components in the mixture is invariant under a continuous and monotonic transformation of income Y. So, if Y is a mixture of K Lognormal densities, then $\log(Y)$ is a mixture of K Normal densities.

A conditional model can be constructed by letting the mixing probabilities vary with exogenous individual characteristics. Given a vector of individual characteristics X, we consider that the income of an individual with these characteristics is distributed according to the mixture

$$f\left(y\,|X\right) = \sum_{k=1}^{K} p_k(X)\,\Lambda\left(y\,;\mu_k,\sigma_k\right), \tag{2}$$

where $p_k(X)$ is the probability of belonging to the homogeneous subpopulation k. We can typically assume that these mixing probabilities depend on a linear index of X. Note that this model is more flexible than the classical analysis of variance, since the probability of belonging to one subpopulation is not necessarily 1 or 0. Moreover, the range of values of the household characteristics which characterize the subpopulation are not pre-fixed but determined by the sample.

For a fixed number of components K, we can estimate $f(y)$ by maximum likelihood (Titterington et al. 1985 and Lindsay 1995), and $f(y\,|X)$ with a specific algorithm, the details of which are given below. In practice, the number of components K is unknown and can be chosen as that which minimizes some criterion. There is a large number of criteria and the literature on this subject is still in progress (McLachlan and Peel 2000). The optimal criterion for our model requires more study, which we leave to future work. For the moment, we select the K that minimizes the BIC criterion (Schwarz 1978), which is known to give consistent estimation of K in mixture models (Keribin 2000).

An alternative conditional model could be constructed by assuming the individual characteristics influence the magnitude of the group-specific earnings μ_k. Then, the individual characteristics could be used to model the mean of the subpopulations rather than the probabilities of belonging to a subpopulation. This conditional model could be written

$$f\left(y\left|X\right.\right)=\sum_{k=1}^{K}p_k\,\Lambda\left(y\,;\mu_k\left(X\right),\sigma_k\right), \tag{3}$$

where the conditional mean is typically assumed to depend linearly on X, i.e., $\mu_k\left(X\right)=X\beta_k$. Conditioning means is relevant when one wishes to analyse the intra-group variability, whereas conditioning probabilities applies when focusing on inter-group variability. In inequality measurement, the major concern is more often to detect the individual characteristics which discriminate between "rich and poor" individuals, rather than to explain the differences between the "rich". Since, in model (3), a different vector of parameters β_k is required for each subpopulation, model (2) provides a potentially more effective framework to analyse inequalities.

2.1 Model

Our principal interest is to explain the distribution of individuals across groups by means of individual characteristics, as in a regression analysis.

Define the variables $U_i = X_i^c\beta + \varepsilon_i, \ (i = 1, 2, \ldots, n)$, where X_i^c is a centered vector of individual characteristics, β is an l-vector of parameters and the ε_i are i.i.d. random variables with a common continuous distribution - we assume $N(0, 1)$ without loss of generality. Now, for $k = 1, 2, , \ldots, K$, let

$$Z_{ik} = \begin{cases} 1 & \text{if} \quad U_i \in \left[\gamma_{k-1}, \ \gamma_k\right[\\ 0 & \text{if} \quad U_i \notin \left[\gamma_{k-1}, \ \gamma_k\right[\end{cases},$$

where $-\infty = \gamma_0 < \gamma_1 < \ldots < \gamma_{K-1} < \gamma_K = +\infty$.

The unobserved vector $Z_i = (Z_{i1}, Z_{i2}, \ldots, Z_{iK})$ of dummy variables has the value 1 at the coordinate of the group the individual i belongs to. Moreover, it is assumed that, given the vectors Z_i, the observed logarithmic transformations of income Y_i are independent and distributed according to the density

$$f(y_i|Z_i) = \sum_{k=1}^{K} Z_{ik} \ \varphi(y_i; \mu_k, \sigma_k), \tag{4}$$

where $\varphi(.; \mu, \sigma)$ is the Normal density function with mean μ and standard deviation σ. To avoid the classical "label switching" problem (Redner and Walker 1984), the following identifiability constraint is imposed: $\mu_1 < \mu_2 < \ldots < \mu_K$.

Note that the Z_i's are independent and distributed according to the multinomial distributions $M\left(1;p_{i1},p_{i2},\ldots,p_{iK}\right)$, where

$$p_{ik}\equiv\mathbf{E}\left(Z_{ik}\right)=\Phi\left(\gamma_{k}-X_{i}^{c}\beta\right)-\Phi\left(\gamma_{k-1}-X_{i}^{c}\beta\right). \tag{5}$$

Consequently, for each individual, the probability of belonging to the k-th group is the probability that a standard normal variable belongs to a certain interval with bounds depending on the values of that individual's characteristics.

From the previous model, it follows that, marginally, the Y_i are independent and distributed according to the mixture densities

$$f\left(y_{i}|X_{i}\right)=\sum_{k=1}^{K}p_{ik}\ \varphi\left(y_{i};\mu_{k},\sigma_{k}\right). \tag{6}$$

Letting $\mu=\left(\mu_{1},\ldots,\mu_{K}\right)$, $\sigma=\left(\sigma_{1},\ldots,\sigma_{K}\right)$, $\gamma=\left(\gamma_{1},\ldots,\gamma_{K-1}\right)$ and $\theta=(\mu,\sigma,\gamma,\beta)'$, the log-likelihood function of the parameters is equal to

$$\ell_{n}(\theta,y)=\sum_{i=1}^{n}\log f\left(y_{i}|X_{i}\right). \tag{7}$$

The maximum likelihood estimator (MLE) can be found by equating the first derivatives of $\ell_n(\theta,y)$ with respect to the different parameters to zero. There is no explicit solution to this system of equations and an iterative algorithm may be used.

2.2 Estimation

The log-likelihood function (7) is not necessarily globally concave with respect to the unknown parameters θ, and so Newton's methods can diverge. Another approach is often used to estimate mixture models: for a fixed K, an easy scheme for estimating θ is the Expectation-Maximisation (EM) algorithm (Dempster *et al.*, 1977), the "missing data" being Z_i's. However, this algorithm often exhibits slow linear convergence. We use it therefore only initially, to take advantage of its good global convergence properties, and then switch to a direct Maximum Likelihood (ML) estimation method (Redner and Walker 1984 and McLachlan and Peel 2000) exploiting the rapid local convergence of Newton-type methods.

The full log-likelihood for our model is

$$\ell_n(\theta, Z, y) = \sum_{i=1}^{n} \sum_{k=1}^{K} Z_{ik} \left(\log \varphi(y_i; \mu_k, \sigma_k) + \log p_{ik} \right).$$

Since $\ell_n(\theta, Z, y)$ is linear in Z, the expectation step in the EM algorithm is carried out by substituting for the missing data Z_{ik} their respective conditional expectations

$$\widehat{p}_{ik} \equiv \mathrm{E}\left(Z_{ik}|\theta, y_i\right) = \frac{p_{ik}\ \varphi(y_i; \mu_k, \sigma_k)}{\sum_{j=1}^{K} p_{ij}\ \varphi\left(y_i; \mu_j, \sigma_j\right)}.$$

Then, in the maximisation step, an estimate of θ is obtained by maximizing the predicted log-likelihood $\ell_n(\theta, \widehat{p}, y)$ with respect to its first argument. The equations $\partial \ell_n(\theta, \widehat{p}, y) / \partial \mu =$

0 and $\partial \ell_n (\theta, \widehat{p}, y) / \partial \sigma = 0$ give the explicit estimates

$$\widehat{\mu}_k = \frac{1}{N_k} \sum_{i=1}^{n} \widehat{p}_{ik} y_i, \qquad \text{and} \qquad \widehat{\sigma}_k = \sqrt{\frac{1}{N_k} \sum_{i=1}^{n} \widehat{p}_{ik} \left(y_i - \widehat{\mu}_k\right)^2},$$

where $N_k = \sum_{i=1}^{n} \widehat{p}_{ik}$, is the current estimate of the number of observations in the k^{th} cluster, $k = 1, 2, \ldots, K$.

Current estimates of β and γ are computed using iteration on a Newton algorithm based on the first derivatives:

$$\frac{\partial \ell_n (\theta, \widehat{p}, y)}{\partial \beta_j} = - \sum_{i=1}^{n} X_{ij}^c \sum_{k=1}^{K} \frac{\widehat{p}_{ik}}{p_{ik}} \left[\varphi \left(\gamma_k; X_i^c \beta, 1\right) - \varphi \left(\gamma_{k-1}; X_i^c \beta, 1\right) \right]$$

for $j = 1 \ldots l$, and

$$\frac{\partial \ell_n (\theta, \widehat{p}, y)}{\partial \gamma_k} = \sum_{i=1}^{n} \varphi \left(\gamma_k; X_i^c \beta, 1\right) \left[\frac{\widehat{p}_{ik}}{p_{ik}} - \frac{\widehat{p}_{i(k+1)}}{p_{i(k+1)}} \right]$$

for $k = 1, 2, \ldots, K - 1$.

These two steps are iterated until some convergence criterion is met.

The Newton-Raphson method is then used to refine the estimates obtained from the EM, and the standard errors of the parameter estimates are approximated by the square root of the diagonal elements of the inverse of the observed information matrix.

Note that, in Normal mixture model with unequal variance components, the likelihood is usually unbounded. Nevertheless, assuming that the variances are not too disparate (Hathaway 1985), the maximum likelihood is well defined. Typically, the unboundedness problem arises when the estimation procedure assigns a certain component to an outlier. Following Policello (1981), we solve this problem by requiring that there be at least two observations from each subpopulation present in the sample ($N_k \geq 2,\ k = 1, 2, \ldots, K$).

2.3 Simulations

In mixture models, the presence of significant multimodality in finite samples has several important consequences (Lindsay 1995). The first is that the solution from the algorithm employed can depend on the initial values chosen. Starting values can be chosen in different ways. Finch, Mendell, and Thode (1989) suggest using multiple random starts. Furman and Lindsay (1994) investigate using moment estimators. However, there is no best solution. In our experiments, we estimate initial values of the mean μ and of the standard deviation σ with robust statistics: from a sorted subsample, we compute the median and the interquartile range in K subgroups with the same number of observations. This choice works well in many simulation experiments.

The second consequence is that the results of a simulation study can depend on the stopping rules and search strategies employed, so it can be difficult to compare simulation studies. In mixture models, convergence problems can be encountered when the proportion of observations in a subgroup is too small, when the initial parameter values are too far from the true values, or when K, the number of components chosen, is too large. We reduce the number of components when the current estimate of the number of observations in a subpopulation is smaller than 2 ($N_k < 2$).

In our simulations, we consider the explanatory mixture model defined in (6) and (5) with the following values,

$$\mu_k = 2\,k \qquad \sigma_k = 0.5 + (k/100)(-1)^k \qquad \gamma_k = -3 + 6\,k/K \quad \text{and} \quad \beta_j = (-1)^j \tag{8}$$

for $j = 1, \ldots, l$. These values are chosen to have distinct Lognormal distributions with quite similar, but different, variances and proportions of individuals in each distribution. We define the $n \times l$ matrix of regressors X by drawing observations from the Normal distribution $N(0, 1)$. In our experiments, the number of observations ($n = 2000$) and the number of regressors ($l = 5$) are fixed, and the number of components is varied according to $K = 2, 4, 6, 8$. For each value of K, we conduct 1000 replications.

In a first set of experiments, K is assumed to be known. The mean and standard deviation of the 1000 realizations obtained for each parameter are presented in Table 1, with the true values given in the second column. Note that the true values of γ_k are not given because they are not the same for different values of K. From this table, we can see that the unknown parameters are very well estimated with the explanatory mixture model: means are very close to the true values and standard deviations are small.

In practice, the number of components K is unknown and has to be selected. The selection criterion used here is the BIC (Schwarz 1978):

$$\mathbf{BIC} = -2\ell_n(\widehat{\theta}, y) + (3K - 1 + l)\log n$$

In our experiments, the rates of correct selection by the BIC, for $K = 2, 4, 6, 8$, are respectively 100%, 99%, 97% and 65%. These results suggest that the BIC performs well when K is not too large. When K is large, we need to examine the robustness of the method. Table 2 presents simulation results with K unknown and selected with the BIC (they are given for the parameter vector β only, because this parameter does not depend on K). These results show that the estimation method performs quite well. However, compared to the results obtained with K known (Table 1), we can see small biases, with a similar magnitude, and greater standard deviations, for large values of K.

While additional experiments could be done, it is not our goal here to conduct a complete simulation study. We see from our experiments that explanatory mixture model estimation works quite well when the observed population is defined as a mixture of sufficiently distinct subpopulations.

2.4 Interpretation

From our explanatory mixture model, we can make a few observations about its use in practice.

- Let us consider model (6), with individual probabilities p_{ik} defined in (5). Under the null hypothesis $H_0 : \beta_j = 0$, the individual characteristic X_{ij} is not significant in p_{ik}. A t-test can easily be computed: we divide the parameter estimate by its standard error, as is done in standard linear regression. If we reject the null hypothesis $\beta_j = 0$, it means that individual probabilities are not the same and therefore that the characteristic X_{ij} is statistically significant in explaining inter-group variability.

• We can interpret the parameter β_j, $j = 1, \ldots, l$, as explaining the individual's position in the income distribution based on his characteristics X_{ij},

If $\widehat{\beta}_j > 0$ (resp. $\widehat{\beta}_j < 0$), then the individual's position moves toward the upper part of the income distribution (resp. bottom) as X_{ij} increases.

To describe this result formally, we consider the expected individual income (in logarithm scale) $P_i = \sum_{k=1}^{K} \widehat{p}_{ik}\ \widehat{\mu}_k$, where $\widehat{\mu}_1 < \widehat{\mu}_2 < \ldots < \widehat{\mu}_K$. Then, the partial derivatives of P_i with respect to the X_{ij} measure the influences on P_i of a change in the value of X_{ij},

$$\frac{\partial P_i}{\partial X_{ij}} = -\widehat{\beta}_j \left[\sum_{k'=1}^{K} \left(\varphi\left(\widehat{\gamma}_k; X_i\widehat{\beta}, 1\right) - \varphi\left(\widehat{\gamma}_{k-1}; X_i\widehat{\beta}, 1\right)\right) \widehat{\mu}_k \right] \tag{9}$$

$$= \widehat{\beta}_j \left[\sum_{k'=1}^{K-1} \varphi\left(\widehat{\gamma}_k; X_i\widehat{\beta}, 1\right) \left(\widehat{\mu}_{k+1} - \widehat{\mu}_k\right) \right] \tag{10}$$

The righthand term, in brackets, is always positive, so we see that, if $\widehat{\beta}_j$ is positive, P_i increases if X_{ij} increases, ceteris paribus. In addition, we can see that the first term $\widehat{\beta}_j$ does not depend on the component k, and the last term, in brackets, is specific to the component k. Thus, we can view $\widehat{\beta}_j$ as the overall influence of the characteristic j on the position of the individual i in the income distribution.

• To provide a plot of the whole income distribution, we can use an estimate of the marginal distribution,

$$\hat{f}(y) = \sum_{k=1}^{K} \bar{p}_k \, \Lambda\left(y\,; \widehat{\mu}_k, \widehat{\sigma}_k\right) \quad \text{with} \quad \bar{p}_k = \frac{1}{n} \sum_{i=1}^{n} \widehat{p}_{ik} \tag{11}$$

where $\bar{p}_k$ is the average proportion of individuals in subpopulation k, calculated as the mean of the estimated individual probabilities of belonging to this subpopulation.

3 Application

Clearly, the method developed above is useful only if it works well with real data. To investigate its application, we use known data and compare its results with those obtained in the literature with different techniques. The data are from the Family Expenditure Survey (FES), a continuous survey of samples of the UK population living in households. The data are made available by the data archive at the University of Essex: Department of Employment, Statistics Division. We take disposable household income (i.e., post-tax and transfer income) before housing costs, divide household income by an adult-equivalence scale defined by McClements, and exclude the self-employed, as recommended by the methodological review produced by the Department of Social Security (1996). To restrict the study to relative effects, the data for each year are normalized by the arithmetic mean of the year. For each person in the households we know the sex, age and labour force status (employee, unemployed, inactive, student). For a description of the data and equivalent scale, see the annual report produced by the Department of Social Security (1998).

Based on these data, Jenkins (2000) and the annual report produced by the Department of Social Security (1998) show that, while increasing during the 1980s, inequality appears to have fallen slightly during the 1990s. Table 3 shows the Theil, Mean Logarithmic Deviation, and Gini indexes, with their standard errors in parentheses, for the years 1979, 1988, 1992 and 1996. All these inequality measures increase considerably from 1979 to 1988 and decrease from 1992 to 1996.

Here, we analyse this evolution of inequality using our method, a mixture estimation with explanatory variables. An individual is an adult if aged 19 or over, or if aged 16 to 18 but not a student; otherwise (s)he is a child. We consider the following characteristics:

X_{i1} - *Pensioner* : the head of the family is a person of state pension age or above.

X_{i2} - *Lone parent family* : a single non-pensioner adult with children.

X_{i3} - *All-working* : non-pensioner household with all adults working.

X_{i4} - *Non-working* : non-pensioner household with all adults not working.

X_{i5} - *Number of children.*

Note that X_{i1}, X_{i3} and X_{i4} are mutually exclusive variables (a pensioner household cannot be a non-working or all-working household). We use the explanatory mixture estimation with the dummy variables X_{i1}, X_{i2}, X_{i3}, X_{i4} and X_{i5} as the set of regressors.

3.1 The Shape of the Income Distribution

Figures 1, 2, 3 and 4 plot the marginal distribution of our estimation by mixture with explanatory variables (`mixture`) and the several Lognormal distributions that constitute the mixture, `pLog`$k = \bar{p}_k \, \Lambda(\widehat{\mu}_k, \widehat{\sigma}_k)$, for $k = 1, \ldots, K$, for the years 1979, 1988, 1992 and 1996. See equation (11) and estimation results in Table 4. Restricting our attention to the global curve, we see in all figures a multimodal distribution, which is slightly modified over the years. However, from the estimation of the income distribution alone, no clear conclusion can been drawn to explain the inequality evolution. Our method allows us to break down the income distribution into several distinct Lognormal distributions, so we can analyse the relative evolution of these distinct distributions over the years.

Initially, we see from the figures that a mixture of K Lognormal distributions does not necessarily mean that the observed population is composed of K homogeneous subpopulations. This may arise from the choice of the BIC criterion. As discussed in Section 2, the selection of the number of components is a difficult task and a rigorous study of this issue should be investigated. An optimal choice of K should pair the number of components with the number of homogeneous subpopulations. However, even with a suboptimal K, we obtain interesting results using our approach.

Let us compare the income distributions in 1979 and 1988 (Figures 1 and 2). First, we see five distinct homogeneous subpopulations in 1979 and six in 1988 - a new small distribution appears at the bottom. And we see that the lowest distributions move leftwards ($\widehat{\mu}_3 = 0.6184$ in 1979 and $\widehat{\mu}_4 = 0.5550$ in 1988, see Table 4). Secondly, we see that the upper single Lognormal distribution has significantly increased: more people are in the upper distribution, $\bar{p}_5 = 0.2106$ in 1979 becomes $\bar{p}_6 = 0.3240$ in 1988, meaning that the "richest" subpopulation comprises 21.06% of the population in 1979 and 32.40% in 1988. Finally, we see two disparate changes: the number of people at the top of the distribution increases and the gap between upper and lower distributions widens. This suggests increasing inequality in the 1980s.

Let us compare the income distributions in 1988 and 1992 (Figures 2 and 3). We detect six homogeneous subpopulations in 1988 and seven in 1992. The lowest distribution has significantly increased ($\bar{p}_1 = 0.0280$ in 1988 and $\bar{p}_1 = 0.0419$ in 1992) and the upper distribution has significantly decreased ($\bar{p}_6 = 0.3240$ in 1988 and $\bar{p}_7 = 0.2104$ in 1992). This suggests that there are fewer very "rich" people and more very "poor" people, and so explains increasing inequality with fewer changes than in the 1980s.

Comparing the income distributions in 1992 and 1996 (Figures 3 and 4), note that the lowest distribution - and so, the bottom of the global curve - moves significantly to the right: the condition of life for the "poorest" people gets better. In addition, from the shape of the global curve, we see a narrowing of the gap between the two major modes. This suggests decreasing inequality.

We can see, from these figures, K varying over time. For instance, in 1979 we select $K = 5$ and in 1988 $K = 6$, the analysis suggesting increasing inequality with the forming of a small subpopulation of very poor people. Here, we select K with the BIC criterion in order to obtain a better fit of the income distribution. Note that, if panel data were available, it could be more appropriate to focus the analysis on individual trajectories and thus to fix K over time using a mixture autoregressive model (Wong and Li 2000).

3.2 The structure of the income distribution

The parameter estimates of the explanatory variables X_{i1}, X_{i2}, X_{i3}, X_{i4} and X_{i5}, based on mixture estimation, for the years 1979, 1988, 1992 and 1996 are given in Table 5, with standard errors in parenthesis. These results allow us to analyse the position of households in the income distribution. In 1979, the largest negative values are successively associated with *pensioners* ($X_{i1} : \widehat{\beta}_1 = -1.770$) and *non-working* ($X_{i4} : \widehat{\beta}_4 = -1.160$), the largest positive value is associated with *all-working* ($X_{i3} : \widehat{\beta}_3 = 0.611$). Thus, households with no working adult and pensioners are strongly over-represented in the bottom of the distribution, while households with all working adults are over-represented in the top of the distribution. If we restrict our attention to the most significant variables, from Table 5, major changes over years can be reduced to:

1. The income position of *pensioners* improves: parameter estimates $\widehat{\beta}_1$ decrease over time, from -1.770 in 1979 to -0.999 in 1996.

2. The gap between the income position of *all-working* and *non-working* households increases in the 1980s and decreases slightly in the 1990s: $\widehat{\beta}_3 - \widehat{\beta}_4$ is, respectively, equal to 1.771, 2.221, 1.957, 1.865.

3. The income position of *non-working* households becomes less than that of *pensioners*: respectively, -1.160 vs. -1.770 in 1979 and -1.107 vs. -0.999 in 1996.

These results must be interpreted conditionally on the value of the other parameters and explanatory variables staying the same, since their meaning comes from the partial derivatives (9). They show that, in the 1980s, the polarization between *all-working* and *non-working* households increased and then decreased slowly in the 1990s. By contrast, the position of pensioners improved steadily over the years.

From these studies on the shape and structure of the income distribution over the years, we can now explain the increasing inequality in the 1980s as due to an increasing polarization between working and non-working households and increasing numbers in the upper part of the distribution. We can explain the slight decrease in inequality during the 1990s as due to a small decrease in this polarization: the number of people in the upper part of the distribution decreased and the income position of non-working households improved slightly. The income position of pensioners, however, has improved. All of these results are supported in one or another of the previous studies using different methods, see Cowell et al. (1996), Jenkins (1995, 1996, 2000) and the descriptive statistical studies by the Department of Social Security (1998).

4 Conclusion

In this paper, we propose a new method for analysing the income distribution, based on mixture models. It allows us to estimate the density of the income distribution, to detect homogeneous subpopulations, and to analyse the position of individuals with specific characteristics. The method is illustrated using income data in the United Kingdom in the 1980s and 1990s. We are able to analyse not only the shape and structure of the income distribution, but also to see at the same time how inequality and polarization have changed over years. Our empirical results demonstrate that this method can be successfully used in practice.

References

Aitchison, J. and J. A. C. Brown (1957). *The Lognormal Distribution.* Cambridge University Press, London.

Cowell, F. A., S. P. Jenkins, and J. Litchfield (1996). "The changing shape of the U.K. income distributions: kernel density estimates". In *New Inequalities: The Changing Distribution of Income and Wealth in the United Kingdom*, pp. 49–75. J. R. Hills (ed), Cambridge University Press.

Davidson, R. and E. Flachaire (2004). "Asymptotic and bootstrap inference for inequality and poverty measures". working paper Greqam 2004-20.

Department of Social Security (1996). Households Below Average Income: Methodological Review Report of a Working Group. Corporate Document Services, London.

Department of Social Security (1998). Households Below Average Income 1979-1996/7. Corporate Document Services, London.

Finch, S. J., N. R. Mendell, and H. C. Thode (1989). "Probabilistic measures of adequacy of a numerical search for a global maximum". *Journal of the American Statistical Association 84*, 1020–1023.

Furman, D. and B. G. Lindsay (1994). "Measuring the relative effectiveness of moment estimators as starting values in maximizing mixture likelihoods". *Comput. Statist. Data Anal. 17*, 493–507.

Ghosal, S. and A. W. van der Vaart (2001). "Entropies and rates of convergence for maximum likelihood and bayes estimation for mixtures of normal densities". *Annals of Statistics 29*(5), 1233–1263.

Hathaway, R. J. (1985). "A constrained formulation of maximum-likelihood estimation for normal mixture distributions". *Annals of Statistics 13*, 795–800.

Jenkins, S. P. (1995). “Did the middle class shrink during the 1980s? U.K. evidence from kernel density estimates”. *Economics Letters 49*, 407–413.

Jenkins, S. P. (1996). “Recent trends in the U.K. income distribution: what happened and why”. *Oxford Review of Economic Policy 12*, 29–46.

Jenkins, S. P. (2000). “Trends in the UK income distribution”. In *The personal Distribution of Income in an International Perspective*. Springer-Verlag, Berlin.

Keribin, C. (2000). “Consistent estimation of the order of mixture models”. *Sankhyä Series A 62*, 49–66.

Lindsay, B. (1995). *Mixture Models: Theory, Geometry and applications*. Regional Conference Series in Probability and Statistics.

Marron, J. S. and H. P. Schmitz (1992). “Simultaneous density estimation of several income distributions”. *Econometric Theory 8*, 476–488.

McLachlan, G. J. and D. Peel (2000). *Finite Mixture Models*. New York: Wiley series in Probability and Mathematical Statistics: Applied Probability and Statistics Section.

Policello, G. (1981). "Conditional maximum likelihood estimation in gaussian mixtures". In C. T. et *al.* (Ed.), *Statistical Distributions in Scientific Work*, Volume 5, pp. 111–125.

Redner, R. and H. F. Walker (1984). “Mixture densities, maximum likelihood and the EM algorithm”. *SIAM Review 26*, 195–239.

Schluter, C. and M. Trede (2002). “Tails of Lorenz curves”. *Journal of Econometrics 109*, 151–166.

Schwarz, G. (1978). “Estimating the dimension of a model”. *Annals of Statistics 6*, 461–464.

Titterington, D. M., U. E. Makov, and A. F. M. Smith (1985). *Statistical Analysis of Finite Mixture Distributions.* J. Wiley, New-York.

Weiss, Y. (1972). “The risk element in occupational and educational choices”. *Journal of Political Economy 80*, 1203–1213.

Wong, C. and W. Li (2000). “On a mixture autoregressive model”. *Journal of the Royal Statistical Society: Series B 62*, 91–115.

	true	$K=2$		$K=4$		$K=6$		$K=8$	
$\hat{\mu}_1$	2	1.986	(0.017)	1.988	(0.021)	2.011	(0.028)	2.023	(0.029)
$\hat{\mu}_2$	4	4.003	(0.021)	3.982	(0.030)	3.978	(0.040)	4.012	(0.048)
$\hat{\mu}_3$	6			6.012	(0.027)	6.010	(0.035)	6.020	(0.043)
$\hat{\mu}_4$	8			7.966	(0.025)	7.998	(0.040)	7.963	(0.046)
$\hat{\mu}_5$	10					9.951	(0.035)	10.020	(0.038)
$\hat{\mu}_6$	12					12.002	(0.030)	11.946	(0.048)
$\hat{\mu}_7$	14							13.977	(0.042)
$\hat{\mu}_8$	16							15.990	(0.034)
$\hat{\sigma}_1$	0.49	0.499	(0.013)	0.483	(0.017)	0.477	(0.020)	0.475	(0.021)
$\hat{\sigma}_2$	0.52	0.522	(0.012)	0.532	(0.027)	0.495	(0.039)	0.533	(0.048)
$\hat{\sigma}_3$	0.47			0.438	(0.023)	0.489	(0.031)	0.460	(0.038)
$\hat{\sigma}_4$	0.54			0.532	(0.027)	0.511	(0.039)	0.537	(0.049)
$\hat{\sigma}_5$	0.45					0.463	(0.031)	0.480	(0.036)
$\hat{\sigma}_6$	0.56					0.551	(0.023)	0.515	(0.051)
$\hat{\sigma}_7$	0.43							0.399	(0.038)
$\hat{\sigma}_8$	0.58							0.592	(0.028)
$\hat{\gamma}_1$		0.012	(0.048)	-1.537	(0.057)	-2.014	(0.064)	-2.237	(0.067)
$\hat{\gamma}_2$				0.016	(0.044)	-1.012	(0.051)	-1.506	(0.052)
$\hat{\gamma}_3$				1.389	(0.058)	0.039	(0.046)	-0.729	(0.046)
$\hat{\gamma}_4$						0.994	(0.050)	0.024	(0.043)
$\hat{\gamma}_5$						1.958	(0.065)	0.763	(0.047)
$\hat{\gamma}_6$								1.498	(0.056)
$\hat{\gamma}_7$								2.296	(0.066)
$\hat{\beta}_1$	-1	-0.987	(0.054)	-0.969	(0.038)	-0.986	(0.035)	-1.027	(0.031)
$\hat{\beta}_2$	1	0.982	(0.053)	0.963	(0.040)	0.985	(0.035)	1.027	(0.033)
$\hat{\beta}_3$	-1	-1.028	(0.060)	-0.988	(0.038)	-0.982	(0.035)	-1.005	(0.033)
$\hat{\beta}_4$	1	1.038	(0.059)	0.981	(0.038)	0.976	(0.034)	1.047	(0.034)
$\hat{\beta}_5$	-1	-1.046	(0.066)	-0.992	(0.039)	-1.009	(0.034)	-1.019	(0.033)

Table 1: Simulation results with K known

	true	$K=2$		$K=4$		$K=6$		$K=8$	
$\hat{\beta}_1$	-1	-0.987	(0.054)	-0.984	(0.038)	-1.104	(0.056)	-1.169	(0.171)
$\hat{\beta}_2$	1	0.982	(0.053)	0.986	(0.040)	1.133	(0.055)	1.169	(0.169)
$\hat{\beta}_3$	-1	-1.028	(0.060)	-1.047	(0.038)	-1.131	(0.053)	-1.168	(0.167)
$\hat{\beta}_4$	1	1.038	(0.059)	0.975	(0.038)	1.122	(0.053)	1.170	(0.171)
$\hat{\beta}_5$	-1	-1.046	(0.066)	-1.006	(0.039)	-1.149	(0.055)	-1.169	(0.168)

Table 2: Simulation results with K unknown

	Theil		MLD		Gini	
1979	0.1066	(0.0023)	0.1056	(0.0020)	0.2563	(0.0023)
1988	0.1619	(0.0053)	0.1542	(0.0036)	0.3074	(0.0034)
1992	0.1794	(0.0065)	0.1743	(0.0046)	0.3214	(0.0037)
1996	0.1507	(0.0046)	0.1457	(0.0036)	0.2976	(0.0033)

Table 3: Inequality measures over years

	1979	1988	1992	1996
$\hat{\mu}_1$	0.4096 (0.0041)	0.3080 (0.0218)	0.2828 (0.0168)	0.3369 (0.0100)
$\hat{\mu}_2$	0.4967 (0.0065)	0.3657 (0.0056)	0.3304 (0.0086)	0.3962 (0.0098)
$\hat{\mu}_3$	0.6184 (0.0070)	0.4458 (0.0068)	0.4102 (0.0090)	0.4869 (0.0075)
$\hat{\mu}_4$	0.7910 (0.0116)	0.5550 (0.0118)	0.5010 (0.0134)	0.5928 (0.0103)
$\hat{\mu}_5$	0.9053 (0.0129)	0.6949 (0.0132)	0.6307 (0.0129)	0.7228 (0.0156)
$\hat{\mu}_6$	-	0.8918 (0.0127)	0.8014 (0.0182)	0.8973 (0.0255)
$\hat{\mu}_7$	-	1.3216 (0.1167)	0.9550 (0.0208)	0.9846 (0.0253)
$\hat{\mu}_8$	-	-	1.4536 (0.1879)	-
$\hat{\sigma}_1$	0.0507 (0.0024)	0.1117 (0.0107)	0.1094 (0.0076)	0.0649 (0.0061)
$\hat{\sigma}_2$	0.0426 (0.0034)	0.0418 (0.0034)	0.0325 (0.0053)	0.0455 (0.0041)
$\hat{\sigma}_3$	0.0668 (0.0044)	0.0407 (0.0038)	0.0372 (0.0036)	0.0421 (0.0046)
$\hat{\sigma}_4$	0.1109 (0.0069)	0.0552 (0.0064)	0.0473 (0.0050)	0.0501 (0.0063)
$\hat{\sigma}_5$	0.2349 (0.0077)	0.0889 (0.0067)	0.0718 (0.0058)	0.0834 (0.0087)
$\hat{\sigma}_6$	-	0.2086 (0.0075)	0.1258 (0.0104)	0.1491 (0.0206)
$\hat{\sigma}_7$	-	0.4358 (0.0443)	0.2419 (0.0113)	0.3398 (0.0280)
$\hat{\sigma}_8$	-	-	0.6068 (0.0781)	-
$\hat{\gamma}_1$	-1.2964 (0.0831)	-2.6619 (0.1500)	-2.3222 (0.1027)	-1.9912 (0.1821)
$\hat{\gamma}_2$	-0.6855 (0.0573)	-1.3767 (0.1060)	-1.5818 (0.1309)	-1.1308 (0.0971)
$\hat{\gamma}_3$	0.1538 (0.0728)	-0.6687 (0.0640)	-0.8137 (0.0932)	-0.4395 (0.0740)
$\hat{\gamma}_4$	1.1937 (0.1098)	-0.1540 (0.0835)	-0.3227 (0.0752)	0.0629 (0.0751)
$\hat{\gamma}_5$	-	0.6188 (0.0772)	0.2897 (0.0794)	0.7316 (0.1116)
$\hat{\gamma}_6$	-	2.8623 (0.1930)	1.0760 (0.1276)	1.6041 (0.2285)
$\hat{\gamma}_7$	-	-	3.0681 (0.1747)	-
$\bar{p}_1$	0.1893	0.0280	0.0419	0.0687
$\bar{p}_2$	0.1328	0.1421	0.0792	0.1309
$\bar{p}_3$	0.2131	0.1559	0.1554	0.1724
$\bar{p}_4$	0.2543	0.1329	0.1310	0.1450
$\bar{p}_5$	0.2106	0.2002	0.1740	0.1850
$\bar{p}_6$	-	0.3240	0.1995	0.1799
$\bar{p}_7$	-	0.0170	0.2104	0.1181
$\bar{p}_8$	-	-	0.0086	-

Table 4: Estimation by explanatory mixture: numerical results.

	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_5$
1979	-1.770 (0.059)	-0.672 (0.106)	0.611 (0.050)	-1.160 (0.086)	-0.439 (0.020)
1988	-1.329 (0.058)	-0.694 (0.106)	0.781 (0.053)	-1.440 (0.068)	-0.352 (0.022)
1992	-1.109 (0.053)	-0.546 (0.083)	0.717 (0.050)	-1.240 (0.060)	-0.345 (0.019)
1996	-0.999 (0.055)	-0.616 (0.078)	0.758 (0.053)	-1.107 (0.062)	-0.384 (0.020)

Table 5: Parameter estimates $\widehat{\beta}_j$ of individual characteristics X_j

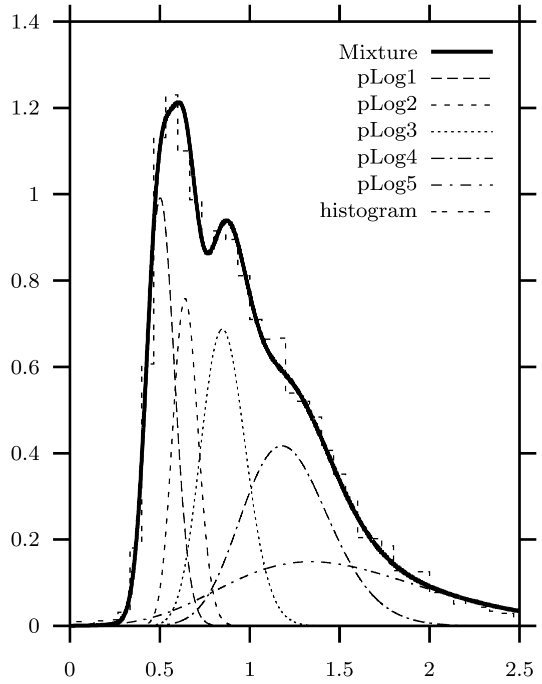

Figure 1: Income distribution in 1979

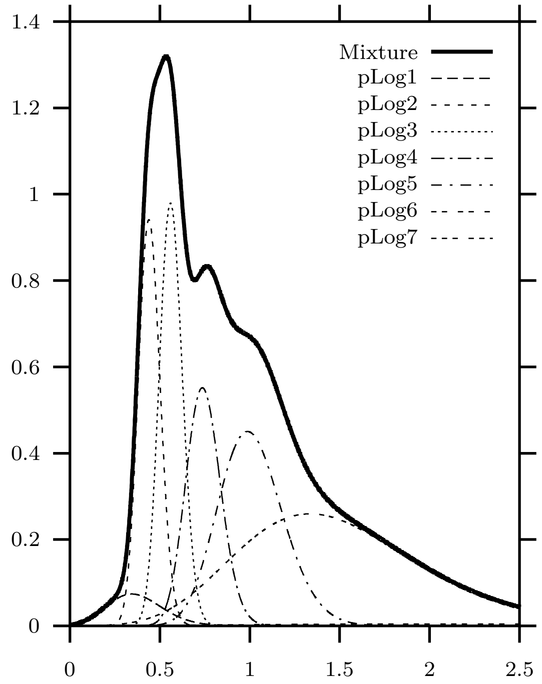

Figure 2: Income distribution in 1988

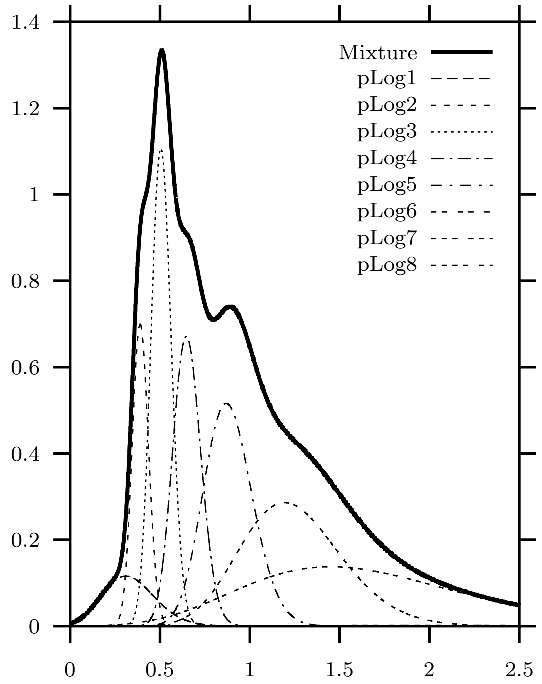

Figure 3: Income distribution in 1992

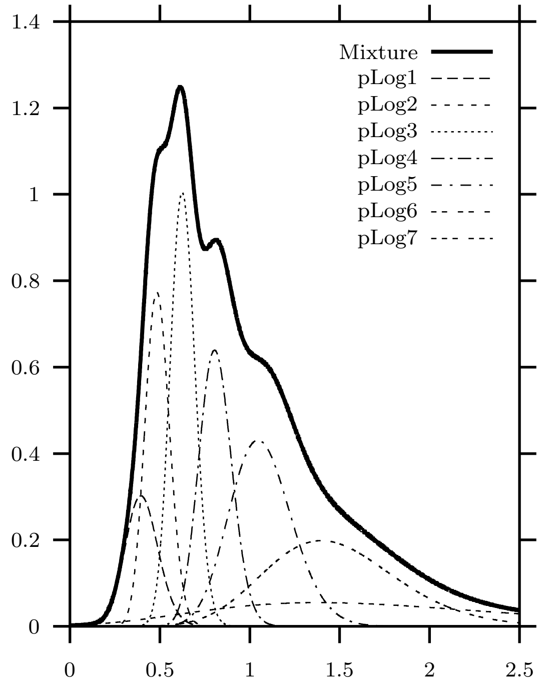

Figure 4: Income distribution in 1996