Session Number: 8B
Session Title: Issues in Income Distribution
Paper Number: 14
Session Organizer: Edward Wolff

Paper Prepared for the 28th General Conference of
The International Association for Research in Income and Wealth
Cork, Ireland, August 22 – 28, 2004

TRENDS IN INCOME DISTRIBUTION IN ITALY: A NON PARAMETRIC AND A SEMI-PARAMETRIC ANALYSIS

Maria Grazia Pittau and Roberto Zelli

For additional information please contact:
M. Grazia Pittau
Dept Contabilità Nazionale e APS Università di Roma "La Sapienza",
BOX 83 ROMA 62
P.le Aldo Moro, 5, 00185 ROMA, ITALY
grazia.pittau@uniroma1.it

This paper is posted on the following websites: http://www.iariw.org
http://www.econ.nyu.edu/iariw
http://www.cso.ie/iariw/iariwhome.html

TRENDS IN INCOME DISTRIBUTION IN ITALY: A NON PARAMETRIC AND A SEMI-PARAMETRIC ANALYSIS

Maria Grazia Pittau and Roberto Zelli [1]

Abstract

Using kernel density estimation and mixture models, household size-adjusted income distributions in Italy are cross-sectionally examined over the period 1987-2000. Nonparametric tests assess the shape time invariance and the presence of modes in the distributions. Evidence shows that income tend to cluster around more than one point, giving good reasons to model the shapes by a finite mixture density with an appropriate choice of components which represent homogeneous subpopulations. Effects of socio-demographic factors on the probability of households to belong to one of the component of the mixture are identified by a compositional data analysis.

Keywords: Income distribution, Kernel estimation, Mixture models, CODA analysis
JEL classification: D31; C14.

[1]Dipartimento di Contabilita' Nazionale e Analisi dei Processi Sociali, University of Rome "La Sapienza", Italy, DCNAPS, Box 83 Roma 62, P.le Aldo Moro 5 00185 Roma, Italy, e-mail: grazia.pittau@uniroma1.it, Fax: +39-06-4453246.

1 Introduction

In Italy, the eighties and nineties were decades marked by significant social end economic changes. During the eighties Italy experienced a long period of economic growth but also the public expenditure increased enormously in relation to GDP. This proved unsustainable in the years that followed. Since the early nineties, there was a difficult process of restoration of public finances and it had restrictive effects on households' disposable income. The corrective measures implemented in the nineties relied primarily on public revenues. Fiscal receipts rose by 5 percentage points in relation to GDP since the beginning of the decade, expanding the gap between gross and net disposable income. In the labour market, the 1993 agreement between government, unions and industrial employers had a curbing effect on wage and salary growth. The reforms in the pension system from 1992 onwards also contributed to reduce transfers to households. Disposable income was also influenced by the dynamics of returns on capital, especially of net interests, which had a remarkable contraction since 1993. This reduction was the result of the consolidation of disinflation process and of the attainment of government targets for public accounts. The reduction in interest rates encouraged households' investments in more hazardous and remunerative activities, such as investment trusts and share certificates. All these factors had an impact not only on the level and growth of the disposable income, but also on the distribution of income across households.

While there is a considerable body of literature on the evolution of income growth and inequality in Italy, relatively little work has been carried out in terms of examining changes in the whole shape of the distribution. The shape of the income distribution provides a picture from which three important distributional features can be observed simultaneously (Cowell, Jenkins and Litchfield, 1996): income levels and changes in the location of the distribution as a whole; income inequality and changes in the spread of the distribution; clumping and polarisation and changes in patterns of clustering at different modes.

Density functions might be estimated by fitting a parametric specification, such as the commonly used log-normal, Singh-Maddala or Dagum, but there is always a trade-off between degree of fitness and easy interpretation and utilization of the parameters. This is also due to the fact that empirical distributions of income are generally multimodal, while the mathematical functions fit well for unimodal shapes. Therefore, a nonparametric approach based on kernel density estimation has been recently preferred in the empirical analyses.

The aim of the paper is to track the evolution of the distribution of income in Italy over a critical period, the nineties, characterized by the worst recession since 1975 and by economic and institutional reforms, by looking at the evolution of the whole shape of the distribution using kernel density estimation instead of focusing on one or more summary measures of it. The analysis based on kernel density, as a matter of fact, relies heavily on the visual impression of the income distribution. To give statistical support to the stylised facts emerging from the representation of the densities, we performed some testing procedures which are helpful to reveal certain aspects of the shapes that otherwise may be unmarked.

First, important hypotheses of "convergence" and "vanishing of middle class" can be further examined by analysing shifts in the shape of the income distribution over time. To establish if the observed changes in the overall kernel-estimated densities are significant, a non parametric changing shape test has been used to discriminate between significant and not significant shifts in the distributions.

Second, when the visual impression seems to corroborate the presence of more than one mode in the distribution, further investigation should be devoted in identifying the sub-populations cluster around the modes (Hildebrand 1998). Generally, to assess the degree of population heterogeneity in the income distribution, two main approaches have been proposed in the literature. One approach is to detect the number of modes in the estimated income density, as in Bianchi (1997) and in Pittau and Zelli (2001), since each mode may help to ascertain separate underlying income distributions. The other approach is to approximate the empirical distribution by a finite mixture of (unimodal) distributions, as in Bakker and Creedy (1999) and in Flachaire and Nuñez (2002), so that relatively simple distributions can be combined intuitively in order to match particular features of the empirical distributions. Since, as shown by Izenman and Sommer (1988), there is no guarantee of a biunivocal correspondence between the number of modes and the number of unimodal components in the mixture, in this paper both the approaches have been followed.

The remainder of the paper is organised as follows. In the next section, a brief description of the Bank of Italy survey on Household Income and Wealth survey is presented, along with a discussion of the economic unit selected and the data utilised in the analysis. Section 3.1 briefly explains the kernel-based methodology. The kernel density estimates are obtained by using both an adaptive bandwidth to handle data sparseness and a weighting variable to account for sample design. The differences in the overall unknown densities are tested by a non parametric test (Li, 1996), while the number of modes in the income distribution is investigated following the proposal of Silverman (1981) based on bootstrap calculation of critical values. In Section 3.2 the representation, interpretation and estimation of mixture models are illustrated. Section 4 reports the main empirical results, while Section 5 summarises and concludes.

2 Data

The data were obtained from the Bank of Italy survey on Household Income and Wealth (SHIW) over the period 1987-2000. The SHIW is conducted every two years[2] on about 8000 households and is largely used in empirical analyses of income and wealth in Italy and in general on household spending and saving behaviour. Before 1987, the survey was every year. A significant revision of the sample design and of the definition of income was conducted in 1986. Therefore temporal comparison before that year should be made with caution.

The survey unit is the household, defined as a group of individuals linked by ties of blood, marriage or affection, sharing the same dwelling and pooling all or part of their incomes. On the basis of this definition, individuals who live together for economic reasons are not regarded as members of the same household, while only one sharing unit is recorded when two or more inter-related legal families live together (Brandolini, 1999, p.25).

Our focus of interest is the total household income from all sources, that is the total of the personal income of all the members of the family including five main components: income from employment; income from self-employment; cash and imputed rents; net interest and dividends; transfers. Income is recorded after the payment of taxes and social security contributions. No information on taxation is collected.

The sampling scheme did not substantially change over the period under analysis. The samples are drawn in two stages (municipalities and households), with the stratification of the primary sampling units (municipalities) by region and size. All municipalities with a population of more than 40,000 are included, while the other are randomly drawn. Households were then selected randomly. As described in Banca d'Italia (2000), each household is assigned a weight inversely proportional to its probability of inclusion in the sample, which is constant at the municipality level. The weights are subsequently modified to align the structure of the sample with that of the population in respect of several known characteristics.

Since 1989 a panel section has been introduced. The proportion of the sample of household already interviewed in the previous survey rose from 15 per cent in 1989 to 48 per cent in the year 2000.

Since the comparison is made across years, household incomes are reported in 1987 prices by correcting for inflation using the Italian consumer price index (CPI). As pointed out by Burkhauser et *al.*, (1997), income distribution analysis is sensitive to the selected price index. Although the CPI has been criticised in recent years for overstating inflation,

[2]The 1997 survey was shifted to 1998.

especially in USA (Boskin et *al.*, 1997), no widely accepted alternative exists. Moreover, the Italian CPI is less affected by outlet substitution and substitution bias effects (CGIS, 2000).

The analysis has been carried out on household income, adjusted for different household types using an equivalence scale. In fact, if we look at the income distribution of an indicator of the personal well-being an adjustment for taking into account the heterogeneity of household types is more appropriate to evaluate households' income as the demographic structure of the population changes. The equivalence scale is the Italian official scale which employs the Engel method. Therefore, the Italian official scale follows as the ratio of total expenditure of the demographically varying households with the same budget share of food. In particular, households differ in their composition from one to another only with respect to the number of components.

Standard summary statistics of real equivalent income distributions using SHIW weighted data are reported in Table 1. In the calculous of these measures, households with income below zero were excluded.

Table 1: Summary measures of real size-adjusted income distribution in Italy during the period 1987-2000

	1987	1989	1991	1993	1995	1998	2000
Theil	26.33	17.92	16.66	19.43	19.33	24.04	23.09
Gini	33.52	31.02	30.22	33.36	32.94	35.95	35.72
Ratio 90/10	4.03	3.88	3.82	4.55	4.36	5.41	5.24
Mean (000 lira)	22005	24982	23704	22727	22989	25057	25397
Median (000 lira)	17759	21270	20471	19348	19588	21443	22323
Sample size	8027	8261	8185	8066	8122	7112	7958

Source: own calculation on SHIW microdata.

Initially, the mean real equivalent household income grew quickly, from 22 million lira in 1987 to 25.0 million in 1989, corresponding to a 13.5 per cent increase. Since the 1991 a declining trend took place. The average income fell to a low 22.7 in the recession of 1993 and rose only modestly in 1995 and more consistently in 1998 and 2000. In fact, in 1998 the mean equivalent real income reached the level of 25.1 million lira, reflecting an increase of 9 per cent over the period 1995-98, and in the year 2000 the level of almost 25.4 million, with a further increase of 1.4 percentage points.

Fluctuations of the mean income around a flattened trend were accompanied by a slight downward trend in income inequality in the mid-eighties, followed by a rise in the early nineties. In fact, the inequality of size-adjusted income, measured by the Gini coefficient[3] (in percentage terms), increased over the whole period, raising from 33.5 in 1987 to 35.7 in the year 2000. Summing up the year to year changes, the inequality dropped slightly in the sub-period 1987-91 (from 33.5 to 30.2), increased in 1993 reaching about the same level of 1987, fell in 1995, it rose again more consistently in 1998 (from 32.9 in 1995 to 35.9 in 1998) and then it stabilised in the year 2000.

These results reflect the above mentioned economic condition in Italy during the last two decades. However, the traditional measures of income inequality summarising the entire distribution with one value produce an incomplete view of the underlying distribution of interest. Moving beyond these summary measures to look at the whole income distribution can reveal some more interesting matters of detail.

3 Analytical framework

3.1 Density estimation, comparison of densities and mode testing

The income distribution of a population of economic units is traditionally dealt with statistical models which tend to provide an accurate description of the phenomenon. The identification of a particular mathematical form for the model depends on the properties it is supposed to fill. A list of eleven properties to guide the choice of the model to be adopted (as models foundation, goodness of fit of the whole income range, economic significance of parameters, principle of parsimony) has been illustrated in Dagum (1980).

The most widely used models until the seventies, the Pareto and the lognormal, satisfy very few of these properties. The Pareto model, according to its goodness of fit, functional simplicity and the economical interpretation of its parameters, continued to be considered as the model of the very high-income groups, while the lognormal fits the whole range of income distribution but it is quite poor in describing the tails of the actual distribution. During the seventies both the models were outperformed, in terms of a range of goodness-of-fit properties, by parsimonious alternatives, including the Gamma (Salem and Mount, 1974), the Singh and Maddala (1976) and the Dagum (1977) model. Attempts to take into account various observed empirical regularities as well as increased computational capacity have lead to the adoption of more general statistical models (McDonald and Mantrala, 1995). In particular, McDonald (1984) has shown that the Gamma, the Singh and Maddala and the Dagum models are special cases of a more

[3] Analogous comments can be formulated using the Theil entropy measure or the ratio between the 90th and the 10th quantiles.

general generalized Beta of the second kind (GB2). Recently, Parker (1999) gave also an explicit theoretical micro-foundation of the GB2 (model foundation is the first property in the Dagum list) based on a neoclassical model of optimising firm behaviour.

In general, these models are flexible enough to deal with unimodal and zeromodal distribution and with data coming from some heavy tailed distribution with asymmetry. However, the underlying distribution is unknown in practise, and the assumed model may not be suitable for fitting the data or it may not be generally true. In particular, empirical evidence (see for instance Park and Marron, 1990) shows that the income distribution is frequently bimodal, or even multimodal.

Therefore, a nonparametric approach based on kernel density estimation has been preferred in the recent body of literature (see, among others, Jenkins, 1995; Daly *et al.*, 1997; Burkhauser *et al.*, 1999; Biewen, 2000; D'Ambrosio, 2001). The kernel density estimator can be considered as a viewing window that slides over the data and the estimate of the density depends on the number of observations that fall into the window. In detail, given a sample $X_{1,}X_{2,}...,X_n$ of independent and identically distributed observations, the *fixed bandwidth* kernel estimator for the density function $f(x)$ at point x is:

$$\widehat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{1}$$

where $h > 0$ is the bandwidth and $K(\cdot)$ is the kernel function[4]. The bandwidth h governs the degree of smoothness of the density estimate and the choice of the optimal bandwidth is an important empirical issue not easy to solve (Silverman 1986). Ideally, h should vary with the amount of information available, according to the data sparseness in the distribution. This feature can be achieved by using an adaptive bandwidth. The term adaptive refers to local smoothing of the estimated density to obtain an improved global estimate (Sain and Scott, 1996). An adaptive bandwidth smoothes out the "wiggles" in areas where the density of x is sparse (generally, being wide at the edges of the distribution), without leading to oversmooth the dense parts of the distribution (generally, being narrow in the middle).

The assessment of the adaptive bandwidth requires a two-step procedure. In the first step a fixed bandwidth following a statistical "optimal" rule on which a pilot estimate of the density is obtained. The second step starts with the calculation of the bandwidth weighting factors λ_i, defined as:

[4]For large samples, any kernel function will be close to an optimal one, thus rendering the choice of kernel a minor issue (Silverman, 1986). In this paper the kernel function used is the Gaussian one.

$$\lambda_i = \sqrt{\frac{\exp\left(\frac{1}{n}\sum_{j=1}^{n}\log \widetilde{f}\left(x_j\right)\right)}{\widetilde{f}\left(x_i\right)}},$$

where $\widetilde{f}\left(x_j\right)$ is the pilot estimate of the density. Then, the adaptive kernel density estimate for the point x takes the form:

$$\widehat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h\lambda_i}K\left(\frac{x - X_i}{h\lambda_i}\right). \tag{2}$$

The estimator (2) is the kernel estimator used in the analysis. In order to account for the sampling design, a weighting variable, which is a weight inversely proportional to the household probability of inclusion in the sample, has been included in the kernel estimation.

The optimal pilot bandwidth has been chosen according to the Sheather and Jones (SJ) plug-in method (1991). This method is statistically optimal in the sense of minimizing the mean squared error of the estimator. Jones, Marron and Sheather (1996) reported the results of a comprehensive simulation study showing the superior performance, theoretical and computational, of the SJ bandwidth selector for smooth densities.

The use of kernel densities to trace out the evolution of the income distribution provides a powerful visualization framework from which location, (multi)modality and spread can be observed simultaneously. Since the analysis based on kernel density relies to a great extent on the visual impression, in addition to this graphical approach statistical tests are also helpful to reveal certain features of the shapes that otherwise may be undetected.

As pointed out by Hildebrand (1998), the number of modes can be suggestive of the number of separate underlying income distributions, each of which may refer to certain economically important population subgroups. To statistically assess the presence of more than one mode[5] in the distributions, Silverman (1981) proposed a test based on the "critical smoothed bandwidth" $h = h_m$ of the kernel density estimator, that is the smallest possible value of h producing a density with m modes.

The intuition of the test is that if the underlying density has $m + 1$ modes, a "large" value of h_m is expected because a considerable amount of smoothing is required to obtain a m-mode density from a $(m + 1)$-mode density. This suggests that h_m can be used as a statistic to test the null hypothesis that $\widehat{f}(x)$ has m modes versus the alternative that

[5] A mode here is defined as a point at which the gradient of the density changes from positive to negative

$\widehat{f}(x)$ has more than m modes. A large value of h_m indicates more than m modes, thus rejecting the null. How "large" h_m should be, is assessed by the bootstrap test[6].

The test proceeds as follows. From the density, which has been rescaled[7] so as to have the same variance as the original sample, B samples with replacement are drawn. For each bootstrap sample $x^*(b) = (x_1^*, x_2^*, ...x_n^*)$ the critical bandwidth $h_m^*(b)$ consistent with m-modality is computed. An estimate of the achieved significance level of the test is obtained as

$$\widehat{ASL}_m = \# \left\{ h_m^*(b) > h_m \right\} / B, \qquad b = 1, 2, ...B. \tag{3}$$

The null hypothesis of m modes in the density is rejected whenever $\widehat{ASL}_m$ is smaller than the standard levels of significance.

To take into account the adaptive bandwidth, this procedure has been modified as the following:

- the critical bandwidth for each m has been selected as the minimum initial value that produces m modes in the density, estimated by adapting the critical bandwidth.
- for each bootstrap sample, the density has been estimated by adapting in the usual way the previously calculated initial bandwidth. On this estimated density, the number of modes has been finally counted.

In order to verify whether two density functions are the same, Li (1996) proposed a simple test statistic for testing the closeness between two unknown density functions f and g from which two equally sized samples, $\{x_i\}$ and $\{y_i\}$ are drawn. The test statistic is based on the integrated squared difference between the kernel estimated density functions of a population observed in two different periods $\widehat{f}(x)$ and $\widehat{g}(x)$:

$$\widetilde{I}_n = \int \left(\widehat{f}(x) - \widehat{g}(x) \right)^2 dx.$$

Since $\widetilde{I}_n$ can be decomposed into a U-statistics plus other terms, it is possible to apply the central limit theorem for degenerate U-statistics with variable kernels. In fact,

[6] The critical bandwidth is uniquely defined if the Gaussian kernel is used because in that case the number of modes of the estimated density is non-increasing in h (see Silverman, 1981).

[7] This small adjustment rescales in order to have the same variance as the original sample since the kernel estimation in (1) artificially increases the variance of the estimate (Efron and Tibshirani, 1993). This is, rather than sampling with replacement from the data, a sampling from a smooth estimate of the population. For this reason it is called smoothed bootstrap.

under the null hypothesis that $f(x)$ is equal to $g(x)$, assuming $h \longrightarrow 0$ and $nh \longrightarrow \infty$ as $n \longrightarrow \infty$, Li (1996) has shown that:

$$L = nh^{1/2} \frac{\left(\widetilde{I}_n - c(n)\right)}{\widehat{\sigma}} \longrightarrow N\left(0, 1\right), \tag{4}$$

where:

$$c(n) = 2\frac{K(0)}{nh},$$

$$\widehat{\sigma}^2 = \frac{1}{n^2 h} \sum_{i=1}^{n} \sum_{j=1}^{n} \left[K\left(\frac{x_i - x_j}{h}\right) + K\left(\frac{y_i - y_j}{h}\right) + K\left(\frac{x_i - y_j}{h}\right)\right] \left[\int K^2(u) du\right].$$

It is interesting to observe that the test is also valid when x_i and y_i are not equally sized and are dependent as in the case of a panel of households over two periods, which is the case for a subsample of the SHIW survey.

4 Mixture modelling

With a mixture model-based approach, the observed data set can be viewed as originating from a mixture of g component distributions in unknown proportions $\pi_1, ..., \pi_g$, where the mixing proportions π_i are nonnegative and sum to one. Mixture models can be viewed as a semi-parametric alternative to the non-parametric densities, especially when the non parametric density exhibits more than one mode, and provide greater flexibility and precision in modelling the underlying distributions of sample data.

Formally, the probability density function of the random vector X_j under a g-component mixture model is defined as:

$$f\left(x_j, \mathbf{\Psi}\right) = \sum_{i=1}^{g} \pi_i f_i \left(x_j, \theta_i\right), \tag{5}$$

where the vector $\mathbf{\Psi} = \left(\pi_1, ..., \pi_{g-1}, \boldsymbol{\xi}'\right)'$ contains all the unknown parameters of the mixture model; π_i, $i = 1, ..., g$ represent the mixing proportions and the vector $\boldsymbol{\xi}$ contains all the parameters $(\theta_1, ..., \theta_g)$ known *a priori* to be distinct; $f_i\left(x_j, \theta_i\right)$ denotes the values of the univariate density specified by the parameter vector θ_i; in case of normality of the components, $\theta_i = (\mu_i, \sigma_i^2)$ denotes the mean and the variance of each normal component i .

The mixing proportions $\pi_1, ..., \pi_g$ give the prior probability that an economic unit belongs to the ith component of the mixture, representing an *endogenous* parameter which determines the relative importance of each component in the mixture. One of the main advantage in using mixture models is that, once a model is generated, conditional probabilities that one household with a certain income comes from a component of the mixture can be computed for each household. Formally, the posterior or conditional probability τ_{ij} is:

$$\tau_{ij} = \tau_i\left(x_j, \mathbf{\Psi}\right) = \frac{\pi_i f_i\left(x_j, \theta_i\right)}{\sum_{i=1}^{g} \pi_i f_i\left(x_j, \theta_i\right)}, \tag{6}$$

which represents the probability that the jth household with income x_j comes from the ith component of the mixture.

As pointed out by McLachlan and Peel (2000), fitting an unknown density by a mixture of $g = n$ components in equal proportions $1/n$, where n is the size of the observed sample, is equivalent to a non-parametric kernel estimation. Therefore, mixture models can be viewed as a compromise between parametric models, represented by a single parametric family ($g = 1$), and a non-parametric model, represented (in the case of $g = n$) by the kernel density estimator.

Assuming that the distribution within each component is normal, the finite mixture is a reasonable approximation when the groups are well-separated. The assumption of normality may too restrictive, since in principle any functional form can be considered. Bakker and Creedy (1999), for example, combine a gamma (or lognormal) with an exponential distribution to approximate the male New Zealand income distribution for wage and salary earners. The first functional form is chosen to model the middle component of the mixture, while the second one should capture a lower mode while also assisting in fattening the tail of upper incomes. However the family of normal mixture densities of g components can approximate a wide variety of density shapes (Marron and Wand, 1992) and for this reason it has been initially preferred in our analysis. Moreover, while in other analyses (see also Paap and van Dijk, 1998) the number of components is chosen *a priori* and therefore the focus is the selection of the functional form, in our analysis, since we have no information *a priori* regarding how many components include in the mixture model, the selection of the smallest number g components compatible with the data becomes more crucial.

Even though different methods can been used to estimate the parameters, the iterative fitting by maximum likelihood (ML) via the expectation-maximization (EM) algorithm (Dempster *et al.*, 1977) seems to be superior to the other procedures in finding a local maximum of the likelihood function, as reported in McLachlan and Peel (2000).

Starting form an initial parameter $\mathbf{\Psi}^{(0)}$, an iteration of the EM algorithm consists in computing the current conditional probabilities $\tau_i\left(x_j, \mathbf{\Psi}\right)$, $j = 1, 2, ..., n$; and $i = 1, 2, ..., g$

that x_j arises from the i-th component of the mixture for the current value of $\mathbf{\Psi}$, according to the equation (6) (E-step). Then, the ML estimates are computed using the conditional probabilities $\tau_i(x_j, \mathbf{\Psi})$ as conditional mixing weights (M-step). The sequence of alternate E and M steps continues until convergence occurs to the ML estimates[8].

As previously pointed out, special attention should be paid to the choice of the appropriate number of the mixture components or subpopulations from which the sample arises. The choice of the number g of component densities compatible with the data is a difficult problem that has not been completely solved (Aitkin and Wilsonin, 1980; Richardson and Green, 1997). An obvious way to test the null hypothesis $H_0 : g = g_0$ versus $H_1 : g = g$ for some $g > g_0$, could be to use the likelihood ratio test statistic (LRTS). Unfortunately, as pointed out by Quinn, McLachlan and Hjort, (1987), with mixture models, the regularity conditions do not hold and the asymptotic distribution of the LRTS is not necessarily chi-squared. Consequently, McLachlan (1987) proposed a bootstrap procedure to estimate its distribution under the null hypothesis. For very large samples, however, the computational effort may not allow one to apply this procedure. Therefore, classical criteria based on the integrated likelihood, as the Bayesian Information Criterion (BIC) of Schwarz (1978), seem a reasonable shortcut. The BIC provides, under regularity conditions, a reliable approximation of the integrated likelihood. In the BIC, a term is added to the likelihood penalizing the complexity of the model, so that it may be maximized for more parsimonious parameterization. Although the regularity conditions for BIC do not hold for assessing the number of components in a mixture model (Aitkin and Rubin, 1985), there is an increasing practical support for its use in this context (Fraley and Raftery, 1998; Biernacki, Celeux and Govaert, 2000).

It should be noted that the components reflect distinct groups in the population, which cannot necessarily correspond to the number of modes detected in the distribution. For instance, even if the number of modes can be suggestive of the number of separate underlying income distributions, a mixture distribution can be also unimodal when the component are not sufficiently far apart. On the other hand, bimodality does not necessary imply that the data have been sampled from a two-component mixture distribution.

---

[8]The EM algorithm is relatively easy to implement. However, it is well known that the likelihood function for mixture models usually has local maxima (see among others, McLachlan and Peel 2000) and the likelihood function is unbounded and the location of the root of the likelihood equation is not straightforward. The associated practical problem regarding the uniqueness of the maximum likelihood estimation is the selection of suitable starting values for the vector $\mathbf{\Psi}$ of the unknown parameters and for the posterior probabilities for the EM algorithm. In this paper, to prevent convergence to local maxima in the likelihood surface, randomly selected starts, large sample hierarchical (Kaufman and Rousseeuw,1990) clustering-based starts have been selected for initialization of the EM estimation procedure.

5 Empirical results

5.1 Assessing changing shapes

Figures 1.1-1.7 report Gaussian kernel estimates of household adjusted real income frequency functions of the total population for the years 1987-2000[9]. A great shift towards the right occurred between 1987 and 1989, accompanied by a reduction of the number of modes. In particular, the two modes in the middle of the 1987 income distribution, corresponding to 11 million lira and 15 million lira, tend to converge to one peak in 1989 at around 14.8 million lira. On the other hand, the two peaks in the right tail of the 1987 distribution, at around 21 and 26 million lira, seem to link to the corresponding peaks in 1989. Apparently, the shape of the distribution did not alter significantly in 1991, even if an erosion of the second peak at around 21 million is marked. In 1993 the effects of recession are evident since there is a shift in the density towards lower income levels and the incomes related to the peaks are lower than those detected in the peaks of 1991. The structure of the distribution in 1995 tend to unimodality at around 13 million. The close-unimodal shape of 1995 slightly changed in 1998: at least two modes are evident, along with a polarisation of the distribution. In the year 2000, there is a further shift of the mass towards the right and the bimodality is much less evident.

The visual impression suggests that the income frequency densities modified over time. To determine whether the observed changes in the income distribution are statistically significant a non parametric test has been carried out. Applying the Li test on pairwise comparisons of the real income distribution across time, the hypothesis of changes in the structure of income distributions is strongly supported. However, these changes can be due to shifts of the position of the whole distribution (e.g. shifts rightward mean that all the population is better off) and/or changes in the tails and in the mass of the distribution. For this reason, in order to emphasise the observed differences only in the shapes, incomes are normalised by the population median. The densities of relative income are now comparable on a common scale of multiples of median income. The results of the pairwise comparison of the 1987-2000 Italian distributions, obtained by considering the sampling weights and different initial bandwidths of the distributions, are reported in Table 2.

[9]Following the common procedure, in the kernel estimation, the largest one per cent of observations was removed to avoid the effects of the outliers.

Table 2: The Li test statistics for differences in relative size-adjusted household income distributions for Italy across years 1987 to 2000.

	1987	1989	1991	1993	1995	1998	2000
1989	3.17*	–					
1991	4.78*	0.05	–				
1993	3.38*	5.84*	5.14*	–			
1995	3.43*	4.30*	2.85*	-0.48	–		
1998	1.84**	3.38*	3.42*	1.69**	0.02	–	
2000	3.03*	2.06*	1.51	0.85	-0.38	0.07	–

Note: $*$ means H_0 is rejected at the 5%, $**$ means H_0 is rejected at the 10%.

According to the Li test, in twelve out of twenty cases the null hypothesis of time invariance is rejected (at 5% level) for the household equivalent income. In particular, the most peculiar shape is the 1987 one, which is significantly different (at least at 10 per cent level) from the all the succeeding distributions.

Looking at the changes of the shapes of household income distributions between subsequent surveys, two significant alterations in the distribution are detected. The first one at the end of the eighties, between 1987 and 1989, in a phase characterised by significative growth and reduction in inequality. The second one during the phase of recession between 1991 and 1993, suggesting that households responded differently to the economic and social events which affected Italy in that period. For household size-adjusted relative income, the lower-middle income mass, located in the [0.5, 1.0] range and the upper-middle mass, located in the range [1.0, 1.2], shrank in 1993 with respect to the previous years. Instead the lower and upper end of the distribution do not reveal important changes. After 1993, in terms of absolute level, the whole distribution shifted rightward, but without significant alteration of the shape. This implies that the intensification of polarisation and inequality occurred in the watershed year 1993 has not been reabsorbed in the following years of slow recovery.

Figure 1.1-1.7: Kernel density estimates of households size-adjusted income distribution at constant price 1987.

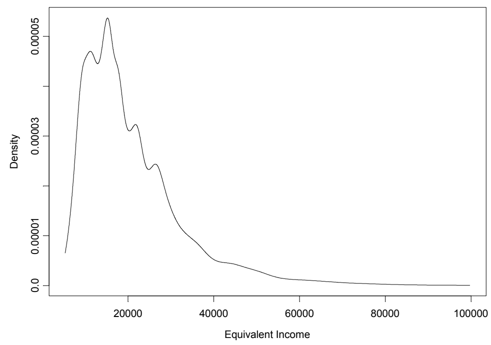


Figure 1.1: Year 1997

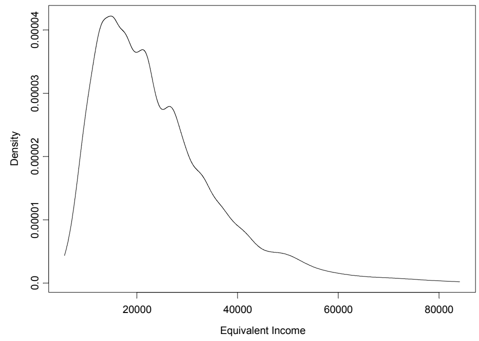


Figure 1.2: Year 1989

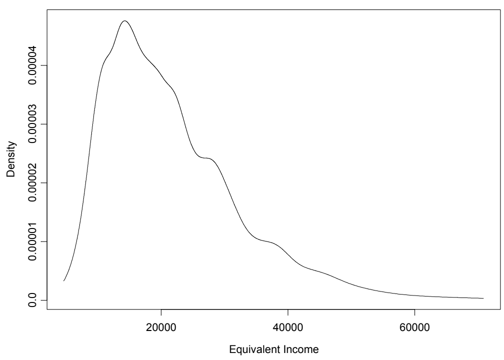


Figure 1.3: Year 1991

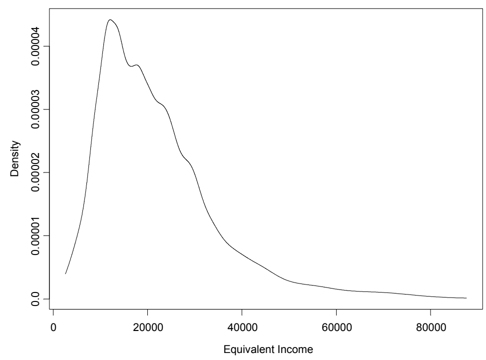


Figure 1.4: Year 1993

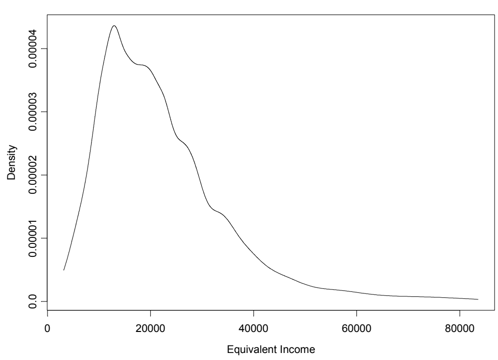


Figure 1.5: Year 1995

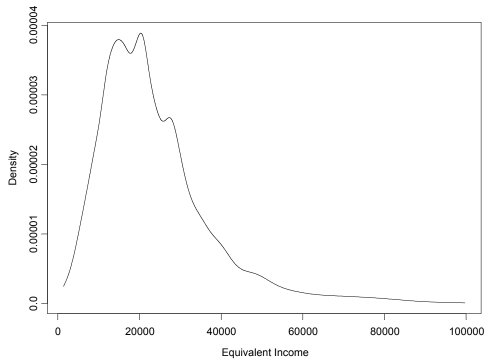


Figure 1.6: Year 1998

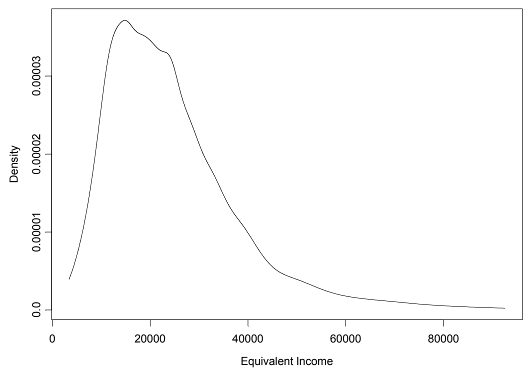


Figure 1.7: Year 2000

5.2 Testing for population heterogeneity

The presence of modes in the distribution are tested by means of the bootstrap test described in Section 3.1. In Table 3 the initial critical bandwidth in the two-step procedure of the density estimation and the p-values of the bootstrap test are reported for each year of the survey obtained with $B = 500$ replications. The unimodality hypothesis is clearly rejected in in 1987, 1989, 1993 and 1998; in 1991, 1995 and 2000 this hypothesis cannot be rejected at 5% significance level. On the other hand, the presence of more than one mode has not been rejected in all the distributions under test, apart from the equivalent income distribution in 1995, which sounds unimodal (see Fig. 1.5). As pointed out by Efron and Tibshirani (1993), this test is conservative, that is the test is less likely to falsely reject the null hypothesis. Specifically, the statistical test for unimodality is based on the smallest possible value for and this makes the probability in (3) as large as possible. This may explain why the test does not reject different hypotheses.

However, even if the low power of the test does not allow us to precisely assess the number of modes in the distribution, the fact that we can not reject the hypothesis of multimodality for almost all the distributions may suggest that fitting the income distribution by a unimodal parametric model may be misleading and further investigation should be devoted in identifying the sub-populations clusterising around the modes.

Table 3: Bootstrap test for m-modality in the size-adjusted real income distributions: critical bandwidths and estimated ASL

Year	H_0:m modes	h_m	ASL_m	Year	H_0:m modes	h_m	ASL_m
1987	$m = 1$	1843	0.014	1995	$m = 1$	1297	0.432
	$m = 2$	1410	0.274		$m = 2$	1105	0.064
	$m = 3$	1406	0.058		$m = 3$	1099	0.020
	$m = 4$	913	0.002		$m = 4$	976	0.038
1989	$m = 1$	1448	0.032	1998	$m = 1$	2203	0.048
	$m = 2$	1440	0.364		$m = 2$	1568	0.128
	$m = 3$	971	0.005		$m = 3$	1236	0.312
	$m = 4$	923	0.000		$m = 4$	1194	0.024
1991	$m = 1$	1331	0.180	2000	$m = 1$	1658	0.358
	$m = 2$	1155	0.464		$m = 2$	1349	0.082
	$m = 3$	1090	0.000		$m = 3$	1278	0.038
	$m = 4$	1062	0.000		$m = 4$	1048	0.022
1993	$m = 1$	1491	0.034				
	$m = 2$	1095	0.214				
	$m = 3$	1008	0.026				
	$m = 4$	920	0.008				

An alternative way to describe the population heterogeneity is to investigate on the number of components of a mixture model of the income distributions. Having assumed that a mixture of normal densities can fit the data well, one of the most important questions is to assess the number g of components. Table 4 reports the values of the maximized loglikelihood and the values of the Bayesian Information Criterion. In order to keep the model as parsimonious as possible, we adopted the heuristic rule of accepting the mixture with less components even when the increase of the BIC value did not exceed 0.1 in percentage terms. According to the BIC values, Table 5 provides the number of components chosen, and the corresponding mean (μ), standard error (σ) and mixing proportion (π). No "structural breaks" seem to occur in the period under analysis, in the sense that the 1993 recession did not systematically changed the number of components of the mixture.

One thing that should be noted first is that the number of components chosen is usually greater than the number of modes selected with the Silverman bootstrap test. A mixture with three or four components is found the most appropriate. This may suggest that the component densities in the mixture are not separated enough to give rise to the same number of modes. These results suggest that it is quite difficult to detect well-defined groups of 'poor' and 'rich' households from the data. Moreover, the location of the modes in the kernel estimated densities does not systematically correspond to those of the component densities of the mixture and this makes the interpretation of the multimodality test more complex.

Table 4: The choice of the number of components: BIC values

years	$g = 1$	$g = 2$	$g = 3$	$g = 4$	$g = 5$
1987	172960.4	169721.2	168979.3	168778.7	168778.4
1989*	176054.0	173870.9	173326.2	173178.4	173158.5
1991	172960.4	169721.2	170163.6	170105.4	170212.2
1993	172688.2	170491.8	170087.4	170084.4	170082.1
1995	173124.1	171498.9	171125.8	171068.3	173153.4
1998	154167.2	152045.6	151810.3	151857.9	151854.2
2000	171305.4	169372.8	169034.4	168953.2	168962.4

*The BIC value for g=6 is 173149.7

Figures 2.1 - 2.3 report the estimated kernel densities and the fitted mixtures for the years 1989, 1991, 1993. From Table 5 we can also see that the estimated mixture models do not have a dominant modal component, but the single components are generally associated with high mixing proportions. Usually, three components are sufficient to model the distribution until the upper-middle range, with the exception of the 1989 distribution which required a component to model the poorest households in the lower tail of the distribution. A two-component mixture is compatible only with the 1991 data. When detected, a small cluster of rich households seems to be generated by a fourth or fifth component of the mixture, as evident for instance in the years 1989, 1995 and 2000 (for this last year see Figure 3). Note that the kernel density estimates were not able to capture this feature of the distributions.

Table 5: Estimated parameters of the normal components

Year		Components				
		g_1	g_2	g_3	g_4	g_5
1987	μ	12794.4	24115.5	47458.3		
	σ	3760.0	7683.9	18243.2		
	π	0.38	0.48	0.14		
1989	μ	8940.6	13184.7	21010.3	32522.5	52868.0
	σ	1598.8	2746.9	5166.3	8661.6	14173.7
	π	0.07	0.21	0.39	0.26	0.07
1991	μ	17521.2	39474.6			
	σ	6939.0	17547.7			
	π	0.74	0.26			
1993	μ	12464.7	23512.4	44533.8		
	σ	4378.1	7862.3	16114.6		
	π	0.35	0.50	0.15		
1995	μ	10873.5	18927.7	29747.7	51542.9	
	σ	3702.8	5460.9	8436.4	14683.6	
	π	0.25	0.39	0.28	0.08	
1998	μ	14764.8	26555.8	51095.9		
	σ	5775.0	9137.9	18814.5		
	π	0.42	0.47	0.11		
2000	μ	11357.7	20815.5	32676.4	57725.1	
	σ	3794.6	6009.7	9314.4	15696.3	
	π	0.21	0.41	0.31	0.07	

Figure 2.1-2.3: Kernel density estimates and fitted mixture model

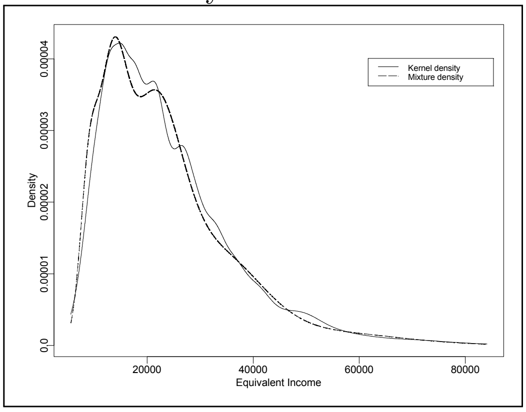


Fig. 2.1 kernel and mixture of 5 components, year 1989

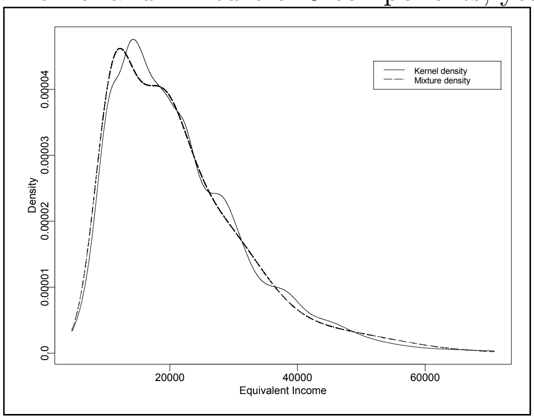


Fig. 2.2 kernel and mixture of 2 components, year 1991

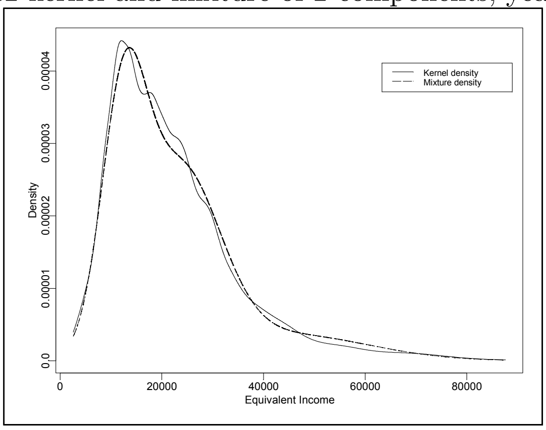
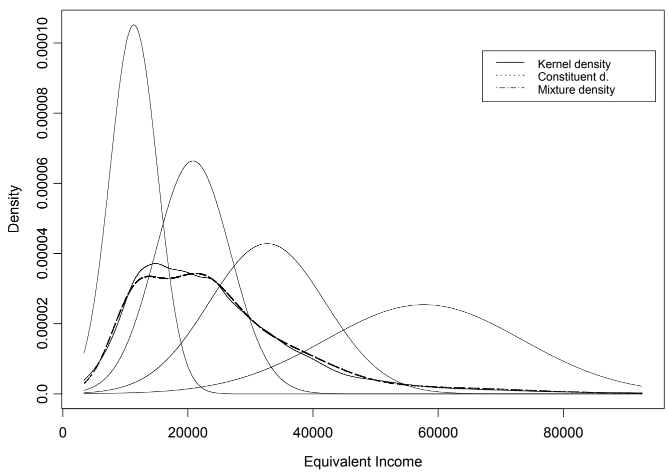


Fig. 2.3 kernel and mixture of 3 components, year 1993

The *ex post* estimated probabilities are helpful for characterising each component and for analysing the mobility between groups. The detection of the subpopulation may follow a twofold strategy. The first one is to allocate each household in one of the mixture components according to a discretional rule, as in Paap and Van Dijk (1998), who defined "poor" the countries belonging to the first component of the estimated mixture with a conditional probability greater than 50 per cent. The second one is to explicitly model the conditional probabilities, by regressing for example these estimated probabilities on explicative social, economic and demographic variables. In order to find which variables are statistically significant in distinguishing each component from the others.

Figure 3: Kernel density estimates and fitted mixture model with its four normal components, year 2000



Adopting the latter approach on the 2000 income data, if the criterion of 70 per cent is employed, only the 61 per cent of household can be clearly located, confirming that the groups are not well separated. If the threshold drops to 50 per cent, we are able to locate the 97.4 per cent of the sample in one of the four estimated components. Once the households have been allocated, the main features of the underlying economic subgroups can be now enhanced.

Table 6 shows the main economic and demographic characteristics of the households belonging to each component of the mixture according to the 50 per cent criterion of allocation[10]. On average, the real adjusted income of all the households is 25,4 million lira. Around this average, there is a big dispersion from 9.8 million lira for the poorest group (40 per cent of the mean income) up to 77.9 million lira for the richest group, three times higher than the contemporary average income. The second group, the lower-middle income group, is characterised by a real adjusted income 16.3 per cent lower than the

[10]The estimated values in each group and in the full sample have been obtained by the Horwitz-Thompson unbiased estimator. The estimates for the full sample could be slightly different from those released by the Bank of Italy (2002) since the latter are obtained by an optimal estimator that takes into account simultaneously the panel and non panel share of the sample.

average household income, while the third one, the upper-middle group, by a real income 44 per cent greater than the national average. As expected, there is a negative relation between the size adjusted income and the number of household components, while there is a positive relation between the adjusted income and the number of earners. A notable exception is the richest group with an average number of earners per household less than the value of the upper-middle group. On the contrary, the age of the head of the household seems not able to discriminate between the income groups. Table 7 reports the distribution of households by gender, education, work status, town size, geographical area and branch of activity of the chief economic supporter. As evident from the table, the level of education and the residential area seem to better identify the economic groups of households. Regarding the work status, while there is a considerable difference across income groups in the structure of employment status, there is no evidence of a concentration of elderly (retired) household's heads on one particular income group. These tables can only provide a descriptive sketch of the social demographic features of the fitted components but they are not able to clearly identify the separate effect of each factor on the *probability* of households to belong to one of the component of the mixture. At this regard, estimating a system of equations that relate to the ex-post probabilities τ_{ij} may help to identify which factors significantly contribute in explaining the degree of association of each household to the components. In such modelling situation, the dependent variables (i.e. the probabilities τ_{ij}) are naturally restricted to lie between zero and one and their sum with respect to j is equal to one. Therefore it is expected that also the modelled probabilities should satisfy these restrictions. These restrictions do not allow one to use a traditional modelling framework based on the multinormal distribution of the stochastic component[11]. Following Fry et al. (1996), to ensure that the stochastic component of the model will satisfy the restriction that the modelled probabilities should be constrained to the unit simplex, the compositional data analysis (hereafter CODA) will be used. The CODA technique relies on the use of "log-ratio" in the statistical analysis (Aitchison, 1986). In fact, since the restriction to the unit simplex lead the normal distribution inappropriate, a one-to-one transformation is required to map the data on probabilities to transformed data suitable for analysis based on multivariate normal techniques (e.g. multivariate regression). Such transformation is the additive log-ratio defined as:

$$y_{ij} = \log\left(\frac{\tau_{ij}}{\tau_{ik}}\right), \ j = 1, ..., k-1.$$

[11] This problem is analogous to the estimation of a system of demand share equations (see Fry et al., 1996).

The inverse transformation is the additive logistic transform and the probabilities can be obtained as:

$$\tau_{ij} = \frac{\exp(y_{ij})}{1 + \exp(y_{i1}) + ... + \exp(y_{ik-1})}, \; j = 1, ..., k-1$$

$$\tau_{ik} = 1 - \tau_{i1} - \tau_{i2} - ... - \tau_{ik-1}$$

It should be noted that the unit sum constraint reduces the dimension of the probability space. However, all the statistical procedures involving the log-ratio covariance matrix are invariant to the probability order and to the choice of the probability used as denominator in the log-ratios (see Aitchison, 1986, 93-98).

Table 6: HH groups by economic and demographic characteristic, year 2000 (average)

Household characteristics*	g_1	g_2	g_3	g_4	*All*
Real adj. income(000lira)	9871.9	21258.3	36567.3	77976.5	25396.5
Real wealth (000lira)	102510	215950	465100	1450490	319170
Household size	3.08	2.68	2.51	2.38	2.72
Number of earners	1.32	1.76	1.97	1.81	1.71
Age	53.2	53.9	51.1	53.8	53.2

*Household characteristics refer to the head of the household defined as the person earning the highest income.

Table 7a: HH groups by social and demographic characteristics, year 2000 (percentage)

Household characteristics	g_1	g_2	g_3	g_4	*All*
Gender					
Male	73.0	72.4	68.9	78.3	71.8
Female	27.0	27.6	31.1	21.7	28.2
Education					
up to elementary school	48.7	36.3	16.0	9.8	33.3
middle school	39.4	37.8	30.6	14.5	35.0
high school	10.7	21.4	36.8	36.9	23.2
university degree	1.2	4.5	16.6	38.8	8.5
Work status					
Employee	39.6	45.6	49.9	33.7	44.3
blue-collar worker	29.0	23.8	11.8	1.9	20.9
office worker or school teacher	8.9	18.5	27.8	11.6	17.9
cadre or manager	1.7	3.3	10.3	20.2	5.5
Self-employed	13.9	11.0	14.2	28.3	14.4
sole proprietor, member of arts or professions	3.2	4.5	7.2	17.8	6.0
other self-employed	10.7	6.5	7.0	10.5	8.4
Not employed	46.7	43.4	32.9	33.0	41.3
retired	37.5	41.8	32.2	31.0	38.1
other	9.2	1.6	0.7	2.0	3.2
Town size					
up to 20.000 inhabitants	49.5	50.7	42.4	27.5	47.5
from 20.000 to 40.000	15.5	12.8	11.0	14.2	13.2
from 40.000 to 500.000	24.8	23.4	30.3	37.4	26.0
more than 500.000	10.2	13.0	16.2	20.9	13.3

Table 7b: HH groups by social and demographic characteristic, year 2000 (percentage)

Household characteristics*	g_1	g_2	g_3	g_4	*All*
Geographical area					
North-West	12.9	29.9	36.7	39.7	27.9
North-East	7.8	19.3	27.0	31.6	19.0
Centre	12.8	23.8	20.2	16.3	19.7
South	43.0	19.4	11.0	7.6	22.4
Islands	23.5	7.6	5.1	4.8	11.0
Branch of Activity					
agriculture	7.0	2.8	1.2	1.7	3.3
industry	19.9	23.4	21.7	19.3	21.8
public administration	10.4	14.4	20.1	14.1	14.7
other sector	16.1	16.0	24.2	31.8	18.9
not employed	46.7	43.4	32.8	33.1	41.3
Total	100	100	100	100	100

The households' characteristic reported in tables 6 and 7 are considered as the independent variables used in the CODA model. The qualitative explanatory variables have been transformed into dummy variables. For each qualitative variable the number of dummy variables is always one less than the number of the categories to avoid perfect multicollinearity[12]. Therefore, a reference category has been chosen for each variable. To overcome the asymmetry in the treatment of probabilities, the estimating equations have been specified in terms of centred log-ratios by applying the following transformation (McLaren et al. 1995):

$$y_{ij} = \log\left(\frac{\tau_{ij}}{\widetilde{\tau}_i}\right), \; j = 1, ..., k$$

where $\widetilde{\tau}_i$ is the geometric mean of the k probabilities associated to $i-th$ household. After the log-ratio transformation the model can be treated as a multiple linear regression model. Table 8 reports the regression coefficients of the independent variables and their standard errors for each income group. R-squared values indicate that the chosen variables can statistically explain from 40 per cent to 90.6 per cent of the variation in the values of the probabilities τ_{ij}. For models with such large number of observations, the explanatory power is quite satisfactory. The effects on τ_{ij} of the number of earners, of

[12]The initial assumption of the model is that the dummies allows for different intercepts in the the regression line but the slope of the regression line is always the same, i.e. there is no interaction between dummy and non-dummy variables.

the net wealth and of the households size are always significant, with the expected signs. Moreover, even if the age of the head of the household sounded not important in the preliminary analysis is indeed always significant, also for the rich group. Head's education appears to be, *ceteris paribus*, of considerable importance in explaining the probability of belonging to each group. These results are as expected and are consistent with the findings of innumerable studies. Controlling for other social- economic and demographic characteristics, what is remarkable is that there is a significant positive difference in the marginal effect of the head of the household being female on the probability of belonging on the poor income group. Instead, a significant negative difference is evident in the case of the upper-middle and the rich groups, while gender is not significant for the lower-middle income group. Household location is also influent in terms both of town size and geographical residence. It is worth noticing that households living in Centre-North, but not in Islands, are less likely to belong to the poor group compared with households living in the South of Italy. In contrast, living in the Centre-North increases the probability of belonging to the upper-middle group, while the only significant finding for the richest group is a higher tendency for Northern, but not Centre residents, to be a member to that group, with respect to Southern residents. For the poorest group, the effect of being not employed (retired or other) of the household's head is to augment the likelihood of being poor compared to households with the head working as blue collar. Instead, there are no differences in the probabilities between blue collars and office workers. It is interesting to note that being office worker increases the likelihood of belonging to the lower-middle income group but not to the upper-middle one. In the richest group, only the retired households have less probability of membership, while major chances are given to cadre, manager and self employed. Moreover, in this group there are no statistically significant difference in sector of activity, apart from public sector. These preliminary results suggest looking at the income groups without imposing a predetermined breakdowns by population subgroups defined by social structures. Furthermore, modelling the probabilities estimated by mixture models could provide a framework for evaluating which households, characterised by a specific socio-demographic pattern, will be less likely to maintain over time their position in the income groups.

Table 8: CODA estimation results for each income group.

Mixture components	g_1		g_2		g_3		g_4	
Household characteristics*	Value	Std.Error	Value	Std.Error	Value	Std.Error	Value	Std.Error
Real wealth (000lira)	-0.02*	0.00*	0.00*	0.00	0.01*	0.00	0.01*	0.00
Household size	3.69*	0.14*	-0.25*	0.01	-1.44*	0.06	-1.99*	0.08
Number of earners	-3.61*	0.22*	0.54*	0.02	1.48*	0.09	1.58*	0.12
Age	-0.10*	0.01*	0.03*	0.00	0.05*	0.00	0.02*	0.01
Gender								
Male								
Female	1.41*	0.36	-0.03	0.03	-0.54*	0.15	-0.84*	0.19
Education								
up to elementary school	3.34*	0.44	-0.43*	0.03	-1.37*	0.18	-1.54*	0.23
middle school								
high school	-4.66*	0.45	0.58*	0.03	5.47*	0.27	2.14*	0.24
university degree	-13.38*	0.66	1.23*	0.05	1.93*	0.18	6.69*	0.35
Work status								
Employee								
blue-collar worker								
office worker or school teacher	-0.10	0.61	0.22*	0.05	0.06	0.25	-0.18	0.32
cadre or manager	-6.31*	0.82	0.66*	0.06	2.56*	0.33	3.09*	0.44
Self-employed								
sole proprietor, member of arts or professions	-2.06*	0.77	0.37*	0.06	0.86*	0.31	0.83*	0.41
other self-employed	-1.74*	0.65	0.24*	0.05	0.72	0.26	0.77*	0.34
Not employed								
retired	2.00*	0.64	-0.32*	0.05	-0.86*	0.26	-0.82*	0.34
other	2.12*	1.03	-0.23*	0.08	-0.78	0.42	-1.11*	0.55
Town size								
up to 20.000 inhabitants	2.62*	0.36	-0.01	0.03	-0.99*	0.15	-1.62*	0.19
from 20.000 to 40.000	1.03*	0.50	0.08*	0.04	-0.36	0.20	-0.76*	0.27
from 40.000 to 500.000								
more than 500.000	-0.03	0.52	0.14*	0.04	0.07	0.21	-0.19	0.27
Geographical area								
North-west	-3.91*	0.42	0.81*	0.03	1.68*	0.17	1.42*	0.22
North-east	-3.61*	0.46	0.87*	0.04	1.58*	0.19	1.16*	0.24
Centre	-1.44*	0.46	0.58*	0.04	0.68*	0.19	0.18	0.25
South								
Islands	-0.74	0.55	0.23*	0.04	0.36	0.22	0.14	0.29
Branch of Activity								
agricolture	0.52	0.90	0.01	0.07	-0.15	0.36	-0.38	0.48
industry								
public administration	1.97*	0.60	0.00	0.05	-0.75*	0.24	-1.22*	0.32
other sector	-1.18*	0.50	0.25*	0.04	0.54*	0.20	0.39	0.26
Multiple R-Squared	0.5042		0.9058		0.5419		0.4004	
F statistic	329.4		3115.0		383.1		216.3	

* indicates statistical significance at the 5% level.

6 Concluding remarks

Using nonparametric density estimation and mixture models, the most relevant changes in Italy over the period 1987-2000, have been empirically examined from the perspective of the whole income distribution. The kernel density estimates have been obtained by using both an adaptive bandwidth to handle data sparseness and a weighting variable to account for sample design. The analysis based on kernel density, however, relies heavily on the visual impression of the income distribution. To obtain a better insight into the estimated cross-sectional distribution, statistical tests have been conducted to assess the shape time invariance and the presence of modes in the distributions. The changing shape test allowed us to discriminate between significant and not significant shifts in the relative distributions. The results indicate that the Italian real income distribution significantly shifted rightward during the eighties, positively affecting the well being of a large fraction of the households. Moreover, the mass of the distribution tends to be less dispersed over time. Instead, the 1993 recession altered shape and location of the size-adjusted income density, increasing spread and polarisation of the distribution. In the following period of slow recovery, the whole distribution tended to maintain the same shape, eventually increasing the income dispersion. This may suggest that there was a general improving of the well-being of the population in absolute terms but the relative positions of the households did not significantly changed.

The multimodality test was able to assess that size-adjusted household income tends in general to cluster around more than one point along the income scale. This evidence gives good reason for modelling the income distribution by a finite mixture density, which provides a semiparametric framework to model unknown distributional shapes through an appropriate choice of components. This allows one to identify homogeneous subpopulations directly from the data, without imposing any *a priori* assumption, as exemplified for the case of the year 2000. Our results indicate that the number of components chosen is usually greater than the number of modes selected with the multimodality test. In fact, a mixture of three or four normal components seems to fit the data well for almost all the Italian income distributions in the period under analysis, apart from the 1989 distribution, where an extra component was necessary to model the poorest group of household, and the 1991 distribution, where a two-component mixture was able to capture the features of the empirical distribution. The results obtained seem encouraging enough to further investigation on the identification of the mixture components and the subsequent classification of households and give support to theoretical models capable to explain multimodal income distributions.

References

[1] Aitchison, J. (1986), *The Statistical Analysis of Compositional Data*, Chapman and Hall, London.

[2] Aitkin, M. and Rubin D.B., (1985), Estimation and hypothesis testing in finite mixture models, *Journal of Royal Statistical Society*, B, 47, 67-75.

[3] Aitkin, M. and Wilson G.T., (1980), Mixture models, outliers, and the EM algorithm, *Technometrics*, 22, 325-331.

[4] Banca d'Italia (2000), Italian Household Budgets in 1998, *Supplements to the Statistical Bulletin*, X, 22.

[5] Bakker A. and Creedy J., (1999), Macroeconomic variables and income inequality in New Zealand: an exploration using conditional mixture distributions, *New Zealand Economic Papers*, 33(2), 59-80.

[6] Bianchi, M., (1997), Testing for convergence: evidence from non-parametric multimodality tests, *Journal of Applied Econometrics*, 12, 393-409.

[7] Biernacki, C., Celeux G. and Govaert G. (2000), Assessing a mixture model for clustering with the integrated completed likelihood, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, 719-725.

[8] Biewen, M. (2000), Income Inequality in Germany During the 1980s and 1990s, *Review of Income and Wealth*, 46, 1-19.

[9] Boskin, M.J., Dulberger E.R., Gordon R.J., Griliches Z. and Jorgenson D. (1997), The CPI Commission: Findings and Recommendations, *American Economic Review* (Papers and Proceedings), 87, 78-83.

[10] Brandolini, A. (1999), The Distribution of Personal Income in Post-War Italy: Source Description, Data Quality, and the Time Pattern of Income Inequality, *Temi di discussione* n. 350, Banca d'Italia, Roma.

[11] Burkhauser, R. V., Crews A.D. and Daly M.C. (1997), Recounting Winner and Losers in the 1980s: a Critique on Income Distribution Measurement Methodology, *Economic Letters*, 54, 35-40.

[12] Burkhauser, R. V., Cutts A.C., Daly M.C. and Jenkins S. P. (1999), Testing the Significance of Income Distribution Changes Over the 1980s Business Cycle: A Cross-National Comparison, *Journal of Applied Econometrics*, 14, 253-272.

[13] CGIS, Commisione per la Garanzia dell'Informazione Statistica (2000), Validità e qualità degli indici dei prezzi al consumo, *Rapporto di Ricerca* n. 99.01.

[14] Cowell, F.A., Jenkins S.P. and Litchfield J.A. (1996), The Changing Shape of the UK Income Distribution: Kernel Density Estimates, in New Equalities, The Changing distribution of Income and Wealth in the United Kingdom, Edited by John Hills, Cambridge University Press, Cambridge.

[15] Dagum, C. (1977), A New Model for Income Distribution, *Economie Appliquée*, 30, 413-437.

[16] Dagum, C. (1980), The Generation and Distribution of Income, the Lorenz Curve and the Gini Ratio, *Economie Appliquée*, 33, 327-367.

[17] Daly M.C., Crews A.D. and Burkhauser R.V. (1997), A New Look at the Distributional Effects of Economic Growth During the 1980s: A Comparative Study of the United States and Germany, *Economic Review*, 2, 18-31.

[18] D'Ambrosio, C. (2001), Household Characteristics and the Distribution of Income in Italy: an Application of Social Distance Measures, *Review of Income and Wealth*, 47, 1, 43-64.

[19] 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.

[20] Efron B. and Tibshirani R.J. (1993), *An Introduction to the Bootstrap*, Monographs on Statistics and Applied Probability, 57, Chapman and Hall, New York.

[21] Flachaire, E. and Nunez O. (2002), Estimation of income distribution and detection of subpopulations: an explicative model, presented at the International Workshop on Income Distribution and Welfare, Università Bocconi, Milano, June.

[22] Fraley, C. and Raftery A. (1998), How many clusters? Which clustering method? Answer via model-based cluster analysis, *The Computer Journal*, 41(8), 578-588.

[23] Fry, J.M., Fry T.R.L. and Mclaren K.R., (1996), The stochastic specification ofdemand share equations: restricting budget shares to the unit simplex, *Journal of Econometrics*, 73, 377-385.

[24] HILDENBRAND, W. (1998), How Relevant are Specifications of Behavioural Relations on the Micro-Level for Modelling the Time Path of Population Aggregates?, *European Economic Review*, 42, 1-21.

[25] IZENMAN, A.J. AND SOMMER C.J. (1988), Philatelic Mixture and Multimodal Densities, *Journal of the American Statistical Association*, 83, 941-953.

[26] KAUFMAN, L. AND ROUSSEEUW, P.J. (1990), *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, New York.

[27] JENKINS, S.P. (1995), Did the Middle Class Shrink during the 1980s? UK Evidence from Kernel Density Estimates, *Economics Letters*, 49, 407-413.

[28] JONES, M.C., MARRON, J.S. AND SHEATHER, S.J. (1996), A Brief Survey of Bandwidth Selection for Density Estimation, *Journal of the American Statistical Association*, 91, 401-407.

[29] LI, Q. (1996), Nonparametric Testing of Closeness Between Two Unknown Distribution Functions, *Econometric Reviews*, 15, 261-274.

[30] MARRON, J.S. AND WAND, M.P. (1992), Exact mean integrated squared error, *The Annals of Statistics,* 20, 712-736.

[31] MCDONALD, J.B. (1984), Some Generalized Functions for the Size Distribution of Income, *Econometrica*, 52, 647-663.

[32] MCDONALD, J.B. AND MANTRALA, A. (1995), The Distribution of Personal Income: Revisited, *Journal of Applied Econometrics*, 10, 201-204.

[33] MCLACHLAN, G.J. (1987), On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture, *Journal of the Royal Statistical Society,* Series C, 36, 318-324.

[34] MCLACHLAN, G.J. AND PEEL, D. (2000), *Finite Mixture Models*, New York, Wiley.

[35] MCLAREN, K.R , FRY, J.M. AND FRY, T.R.L. (1995), A Simple Nested Test of the Almost Ideal Demand System, *Empirical Economics*, 20, 149-61

[36] PAAP, R. AND VAN DIJK, H.K. (1998), Distribution and mobility of wealth of nations, *European Economic Review*, 42, 1269-1293.

[37] Park, B.U. and Marron, J.S. (1990), Comparison of Data-Driven Bandwidth Selectors, *Journal of the American Statistical Association*, 85, 67-72.

[38] Parker, S.C. (1999), The Generalized Beta as a Model for Distribution of Earnings, *Economics Letters*, 62, 197-200.

[39] Pittau, M.G. and Zelli, R. (2001), Income distribution in Italy: a nonparametric analysis, *Statistical Methods and Applications*, 1-3, 175-190.

[40] Richardson, S. and Green P.J. (1997), On Bayesian analysis of mixtures with Unknown number of components, *Journal of Royal Statistical Society*, **B**, 731-792.

[41] Sain, S. R. and Scott, D. W. (1996), On Locally Density Estimation, *Journal of the American Statistical Association*, 91, 1525-1534.

[42] Salem, A.B.Z. and Mount, T.D. (1974), A Convenient Descriptive Model for Income Distribution, *Econometrica*, 42, 1115-1127.

[43] Schwarz, G. (1978), Estimating the dimension of a model, *Annals of Statistics*, 6, 461-474.

[44] Sheather, S.J., and Jones, M.C. (1991), A Reliable Data-based Bandwidth Selection Method for Kernel Density Estimation, *Journal of Royal Statistical Society*, B, 53, 683-690.

[45] Silverman, B.W. (1981), Using Kernel Density Estimates to Investigate Multimodality, *Journal of Royal Statistical Society*, B, 43, 97-99.

[46] Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

[47] Singh, S.K. and Maddala, G.S. (1976), A Function for the Size Distribution of Incomes, *Econometrica*, 44, 963-973.