Is income becoming more polarized in Italy? A closer look with a distributional approach

Riccardo Massari*

January 7, 2009

PRELIMINARY VERSION
Please do not quote without permission from the author

Abstract

During 1990's and early 2000's income inequality in Italy has been higher than that observed in many other OECD countries, but has not displayed any significant trend, upward or downward. For the most part, this evidence relies on summary measures of inequality. By their very nature, summary indices of inequality may not capture aspects such as multi-modalities and polarization. This paper applies a non-parametric tool, the "relative distribution", to describe patterns of changes on the entire Italian household income distribution over the period 1989–2006. This approach also allows for a decomposition of the relative density to isolate changes due to differences in location from changes due to differences in shape. This is done in order to distinguish whether the changes in distribution have taken the form of a polarization or of a convergence of incomes toward the middle income classes. A similar decomposition enables one to analyze the impact of selected covariates on the income distribution. During the period Italy experienced a significant increase of household income polarization, which has particularly affected incomes below the median. In addition, this relative polarization is mainly correlated to changes in the returns to household-head occupational status.

Key words: Income distribution, Relative Distribution, Polarization.

JEL classification: D31; C14.

---

*Sapienza University of Rome, e-mail address: riccardo.massari@uniroma1.it. I wish to thank Maria Grazia Pittau and Roberto Zelli for their suggestions and their support. This paper has also benefited from helpful comments of the participants of the 1st Alumni's Workshop held by the Doctoral School of Economics, Sapienza University of Roma, and of the Conference on Income Distribution and the Family, held by University of Kiel. Responsibility for remaining errors is mine.

1 Introduction and motivation

Many recent studies have highlighted that during 1990's and early 2000's income inequality in Italy has been one of the highest among OECD countries (for a comprehensive and updated review about this topic, see Brandolini and Smeeding, 2007). However, after a sharp rise in early 1990's, in correspondence of a severe recession, inequality displays a substantially stable trend, despite the many changes occurred in the economy. These results hold true regardless of the summary statistics and of the source of data employed (Boeri and Brandolini, 2004).

This evidence appears to be quite surprising, since Italy underwent a prolonged period of recession and a subsequent phase of low growth from 1993 onward. During early 2000's another, less severe, recession has affected Italian economy. In addition, recent years have seen a serious transformation of the labour force, with an increase of contingent jobs, which has implied a widening of earnings distribution in the early 1990s (see Brandolini *et al.*, 2001).

All the same, intra-household redistribution has moderated the negative effect on inequality of labour and income transfers (D'Alessio and Signorini, 2000). As a matter of fact, Italian economy has been influenced by relevant social and demographic changes, during recent years, that ranges from the ageing of the population, to the restructuring of the labour force, with a steady increase of female participation and in the tendency of adult working children to leave the family later, that augmented the number of income earners in an average households. Brandolini *et al.* (2001) show that probability of being at risk of poverty was more closely correlated with the amount of household members employed other than the head, rather than with low pay.

These results are mainly based on summary measures of inequality, like Gini index, or on the comparison of specific quantiles of the distribution, like quintile ratio, that focus on particular characteristics of the income distribution, leaving much of the information contained in survey income data unued. Income trends could be characterized by complex changes, such as multi-modalities or polarization, and also comparisons of changes in the upper and lower tails of the distributions. If the purpose of the analysis is to comprehend which factors modified the entire distribution and on which part of the distribution these factors had an effect, then the classical methods based on summary measures are not the most appropriate.

There are many reasons to explore inequality in different parts of the distribution (Voitchovsky, 2005). Had the inequality more pronounced in the lower tail of the distribution, or in the top deciles, the same level of inequality degree would involve different economic outcomes (Morris *et al.*, 1994).

Even if summary statistics do not indicate any significant changes between two or more situations, modifications in the horizontal allocation of income across groups could have took place in recent decades. Pittau and Zelli (2004) study the changing shape of income distributions during recent years in Italy with a kernel-based analysis, revealing underlying movements along the income scale and the insurgence of a bimodality, after the 1993 recession. The emergence of the modes, and the gap between them, is interpreted as an increase in polarization, in a context of stationary inequality.

Distributional differences range from mean-shifts and changes in variance to more subtle comparisons of changes in the upper and lower tails of the distributions. As an example of this kind of analysis, Di Nardo *et al.* (1996) have used the general approach of forming counterfactual distributions in order to account for the effects of institutional changes on the earning distribution. More recently, Jenkins and Van Kerm (2005) have developed a similar method to decompose changes in the income distribution using subgroup decompositions of the income density function. With this method overall changes are linked to changes in subgroup shares and changes in subgroup densities. As for Italy, D'Ambrosio (2001) studies the polarization of Italian household income during 1987–1995, within each geographic areas (North, Center, South) given the relevant economic differences existing between these areas. In particular, the author studies the determinants of the observed polarization, by means of a non parametric decomposition of income densities, able to measure the distance between and within given groups. Most of the observed variation in the period under examination can be attributed to the within-group income schedule.

In this paper we propose an application based on a statistical approach to full distributional comparison between two distribution based on the definition of the "relative distribution", introduced by Handcock and Morris (1998, 1999). This approach takes into account all of the distributional differences that could arise in the comparison of two distributions. The probability density function of the relative distribution, the relative density, is a re-scaled density ratio of the two distributions. It has simple intuitive meaning, and preserves all of the information necessary to compare two distributions. The relative distribution approach also provides the potential for decomposition that allows one to examine complex hypotheses regarding the origins of distributional changes within and between groups. By contrast with the method proposed by Di Nardo *et al.* (1996) and Jenkins and Van Kerm (2005), this method first illustrates overall differences between the two distributions and then connects these changes to differences in location or in shape, or differences in subgroup shares or subgroup densities, in order to isolate the marginal effects of covariates on the relative density of incomes. Based on this method, Massari *et al.* (2008) show that during the 2000's in Italy there was an increase of the income polarization, which has particularly affected incomes below the median.

Then, the purpose of the present paper is to investigate whether the early 1990's recession and the subsequent prolonged period of stagnation that have affected the Italian economy, combined with the many changes occurred at vary levels, especially in the labour market, have produced significant movements across the income scale. In addition, it is possible to see whether these movements have taken the form of a convergence of the top and bottom percentiles toward the middle income classes or of a shrinking of the latter. In the last case, there could have been a polarization of household incomes, if there is an increase in both tails, or a downgrading (upgrading) if there is an increase in the lowest (highest) percentiles.

The paper is organized as follows: in Section 2 the relative density approach is illustrated. Section 3 presents a brief description of data. Section 4 reports the main results of the application. Section 5 concludes the paper with some remarks.

2 The relative distribution approach

The relative distribution is a non-parametric statistical approach introduced by Handcock and Morris (1998, 1999) that compares the income (or other) distributions of two populations, the "reference" and the "comparison" population, in a way that considers differences throughout the entire income range. Basically, the relative distribution returns the fractions of the "comparison" population which fall in each quantile of the "reference" population. Thus, it is easy to locate and to identify the shifts that have occurred along the income distribution between the two populations.

More formally, let Y_0 be a continuous random variable which represents income for the *reference* population. Let F_0 be the cumulative distribution function (CDF) of Y_0 and f_0 its probability density function (PDF). The *comparison* population generates the continuous random variable Y with F and f its CDF and PDF, respectively. The relative distribution, or *grade transformation*, of Y to Y_0 is defined as the random variable (Ćwik and Mielniczuck, 1989):

$$R = F_0(Y)$$

i.e, R is obtained from Y by mapping values of Y itself to the percentile ranking of Y_0. The realization of R, r, are often referred as the *relative data*, and they represent the rank of the comparison value in terms of the reference group's CDF. The CDF of R is then defined as:

$$G(r) = F(F_0^{-1}(r)) \qquad 0 \leq r \leq 1,$$

where r is the proportion of values, and

$$F_0^{-1}(r) = Q_0(r) = inf\{y_0 \mid F_0(y) = r\} = y_r$$

is the quantile function of F_0. The relative density $g(r)$ is defined as the ratio of the density of the comparison population to the density of the reference population evaluated at the r^{th} quantile of the reference distribution:

$$g(r) = \frac{f(F_0^{-1}(r))}{f_0(F_0^{-1}(r))} = \frac{f(y_r)}{f_0(y_r)} \qquad 0 \leq r \leq 1 \ , \ y_r \geq 0. \tag{1}$$

The quantile function $Q_0(r) = y_r$ returns the value of the income y in the reference distribution below which a proportion r of the ordered income values fall. Thus, $g(r)$ can be interpreted as the ratio of the fraction of households in the comparison population to the fraction of households in the reference population evaluated at the quantile y_r. The relative density $g(r)$ is the PDF of the random variable R. The *rescaling* imposed by the quantile function ensures that the density ratio is a proper PDF.

When no changes occur between the two distributions, $g(r)$ is uniform in $[0, 1]$. A value of $g(r)$ higher (lower) than 1 means that the share of households in the comparison population is higher (lower) than the corresponding share in the reference population, at the r^{th} quantile of the reference population. Broadly speaking, households of the comparison population have a higher (lower) probability than households in the reference population of having the level of income that corresponds to the r^{th} quantile of the baseline distribution.

In this paper the relative density $g(r)$ is calculated as the ratio (Massari *et al.*, 2008):

$$\hat{g}(r) = \hat{f}(y_r)/\hat{f}_0(y_r) \tag{2}$$

where $\hat{f}$ and $\hat{f}_0$ are obtained as kernel estimates on P quantiles y_r of the reference population[1]. Note that the points y_r on which the two density functions are estimated are the same. A local-polynomial model is finally applied for smoothing the plug-in estimates $\hat{g}(r)$. The main advantage of the local polynomial smoother is that it is not affected from the boundary bias of the kernel estimator.

This estimation procedure allows us to include sample weights and to use an adaptive bandwidth in the kernel estimator. The weights are attached to each household according to the sample design. The adaptive bandwidth is advised for income data because of sample sparseness (Pittau and Zelli, 2004).

The relative distribution approach also provides a number of tools to isolate different factors which have affected the observed changes. Differences between the reference and the comparison population could be attributed to a change in the average (or median) income, but they could also be due to differences in shape, that include differences in variation, skewness and other distributional characteristics. It is possible to distinguish between these two effects, namely a *location* effect, due to a change in the first moment, and a *shape* effect, due to a change in higher order moments of the distribution.

The decomposition of the relative density in *location* and *shape* effect goes through the definition of an *additive* location-adjusted population $Y_{0L} = Y_0 + \rho$. In this work, the location-adjusted population is constructed with reference to the median income. Therefore, Y_{0L} is a counterfactual distribution with same shape as the reference distribution but with the median of the comparison distribution. The value ρ is the difference between the medians of the two distributions Y and Y_0. The CDF of Y_{0L} is defined as $F_{0L}(y) = F_0(y + \rho)$, and its derivative is the PDF f_{0L}.

Hence, the decomposition can be written as:

$$\underbrace{\frac{f(y_r)}{f_0(y_r)}}_{g(r)} = \underbrace{\frac{f_{0L}(y_r)}{f_0(y_r)}}_{g_L(r)} \times \underbrace{\frac{f(y_r)}{f_{0L}(y_r)}}_{g_S(p)}, \tag{3}$$

where p is the percentile rank in the location-adjusted population Y_{0L} which corresponds to y_r. If the comparison and the reference distributions have the same median[2], the density ratio for location differences, $g_L(r)$, will be uniform in $[0, 1]$. Conversely, if the two distributions have different median, then $g_L(r)$ is increasing (decreasing) in r if the comparison median is higher (lower) than the reference median. The density ratio for shape differences, $g_S(p)$, represents the relative density net of the location effect. The analysis of g_S detects re-distribution that has occurred between the reference and the comparison populations. For instance, $g_S(p)$ would take a (inverse) U-shape, if the comparison population is relatively (less) more spread around the median than the location-adjusted population. It is thus possible to determine whether there is an increasing income polarization, a downgrading - defined

[1] In this work, $P = 200$. However, the choice of the number of quantiles does not significantly affect results.

[2] As observed above, alternative indices like the mean can be considered. The corresponding results do not differ in a significant way, and are not reported here. A *multiplicative* median location shift can also be applied. However, the multiplicative shift has the drawback of affecting the variance and the shape of the distribution. Indeed, the equi-proportionate income changes cause a flattening (or a shrinking) of the shape of the distribution (Jenkins and Van Kerm, 2005).

as the movement of households into the lower tail of the income distribution - an upgrading, or a convergence of incomes towards the median.

The graphical analysis of the relative density provides a detailed description of polarization patterns. A link between what one observes in the graphical analysis and the quantification of the degree of polarization is yielded by the *median relative polarization index*, as introduced by Morris *et al.* (1994).

The median relative polarization (MRP) index is the mean absolute deviation from the median of the location-matched relative density g_S, re-scaled in order to vary between -1 and 1. Positive values represent an increase in income polarization, while negative values imply a convergence of incomes towards the median. A value of zero indicates no differences in distributional shape.

This index is effective for inference analysis, and provides detailed information about distributional changes. The median relative polarization index keeps track of changes in the shape of the distribution and measures the direction and the magnitude of these changes, assessing whether they take the form of a movement of incomes towards the median of the distribution[3], or a shift towards one or both tails, i.e. an increasing in polarization. The MRP index is also effective in detecting situations different from bi-polarization. In addition, summary measures based on the relative density are robust to the presence of outliers (Handcock and Morris, 1998).

Formally, the median relative polarization index of Y with respect to Y_0 is defined as:

$$MRP(F, F_0) = 4 \int_0^1 \left| r - \frac{1}{2} \right| g_S(r) dr - 1 \tag{4}$$

and can be estimated as:

$$\widehat{MRP}(F, F_0) = \frac{4}{m} \sum_{j=1}^{m} \left| \hat{R}_j - \frac{1}{2} \right| - 1 \tag{5}$$

where $\hat{R}_j = F_{0n}(Y_j - \rho)$ are the estimates of the location-matched relative data and m is the size of the comparison population[4].

Under regularity conditions, $\widehat{MRP}(F, F_0)$ is an asymptotically unbiased and asymptotically normally distributed estimator of $MRP(F, F_0)$. Therefore, it is possible to conduct inference on the polarization index.

In addition, the relative polarization index could be exactly decomposed into two indices, in order to observe separately movements below and above the median[5]. These two index are defined, respectively, *lower relative polarization index* (LRP)

[3] More generally speaking, towards a point of accumulation of the distribution, say the mean, or the mode.

[4] The R_js are here estimated with a kernel-type estimator of F_0 as in Molanes-López and Cao, (2007):

$$F_{0n} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{M} \left(\frac{y - Y_{0i}}{h_0} \right)$$

where $\mathbf{M}$ is the cumulative distribution function of the kernel M, h_0 is the bandwidth and n the size of the reference population (Massari *et al.*, 2008.

[5] In principle, the median relative polarization index could be decomposed in order to observe changes occurred in every percentiles of the distribution, but for a first inspection, this decomposition into two indices would suffice.

and *upper polarization index* (URP):

$$LRP(F, F_0) = 8\int_0^{1/2} \left|r - \frac{1}{2}\right| g_S(r)dr - 1 \quad (6)$$

$$URP(F, F_0) = 8\int_{1/2}^{1} \left|r - \frac{1}{2}\right| g_S(r)dr - 1. \quad (7)$$

with $MRP(F, F_0) = 0.5 * [LRP(F, F_0) + URP(F, F_0)]$. Their statistical properties are similar to the median relative polarization index. They vary between -1 and 1 and can be estimated in a similar way.

In a similar fashion of what observed for location and shape decomposition, it is possible to adjust the relative density for changes in the distribution of other covariates, thus allowing one to separate the impacts of changes in population composition from changes in the covariate-outcome relationship. This decomposition according to covariates draws on the the definition of the counterfactual density of income in the reference year, if household characteristics had been adjusted at the level of the comparison year.

Assume that the covariate X is discrete[6]. Let $\{\pi_k^t\}_{k=1}^K$, where K is the number of categories of the covariate, be the probability mass function of X at time t, i.e. the population composition according to the covariate. The conditional density of Y_t given that $X_t = k$ is:

$$f_{Y_t|X_t}(y|k) \qquad k = 1, \ldots, K$$

and the marginal density of Y_t can be written as:

$$f_t(y) = \sum_{k=1}^{K} \pi_k^t f_{Y_t|X_t}(y|k).$$

Then, the counterfactual density that has the composition of the comparison year, but retains the conditional densities of the reference population is:

$$f_{0C}(y) = \sum_{k=1}^{K} \pi_k f_{Y_0|X_0}(y|k).$$

The relative density is then decomposed into a component that represents the effect of changes in the marginal distribution of the covariate (the composition effect), and a component that represents the residual changes, i.e. the composition-adjusted relative density:

$$\underbrace{\frac{f(y_r)}{f_0(y_r)}}_{g(r)} = \underbrace{\frac{f_{0C}(y_r)}{f_0(y_r)}}_{g_C(r)} \times \underbrace{\frac{f(y_r)}{f_{0C}(y_r)}}_{g_R(p)}. \quad (8)$$

The composition effect detects if changes are due to the different composition of the population, under the assumption that the conditional distribution of income remain unchanged. The residual component reveal changes of income distribution due to the fact that returns to the selected covariates changed over time, when the population composition is held constant.

[6] Extension to the continuous case is straightforward.

3 Data and summary statistics

Data source is the Historical Archive of Bank of Italy Survey of Household Income and Wealth (SHIW-HA) and covers the period 1989-2006[7] (see, Banca d'Italia, 2008, and Brandolini, 1999, for further details).

The variable observed is the annual disposable income of all household members. The definition of household income used in this work is the same used in similar works and includes wages and salaries, income from self-employment, pensions, public assistance, private transfers, income from real properties, imputed rental income from owner-occupied dwellings, and yields on financial assets net of interest paid on mortgages. All figures are net of tax and social security transfers. The unit of analysis is the household, which is both the *economic unit of aggregation* and the *welfare unit*. The household is defined as a group of individuals living together who, independently of their kinship, share their income wholly or in part. Following this definition, when two or more inter-related legal families live together, only one sharing unit is recorded.

Data are household size adjusted, in order to compare households with different composition[8]. The use of this equivalence scale entails the assumptions that intra-household allocation is egalitarian, since the same share of income is impute to all members of the household, regardless of their individual income, their role in the household and other characteristics[9].

All data are deflated to 2000 prices using the national accounts household expenditure deflator. The choice of the deflator is consistent with the definition of income used, inclusive of imputed rents.

Table 1 provides summary measures for annual household incomes from 1989 to 2006[10]. In 1993-95 there has been a considerable fall in average and median households income, and only in 1998 mean and median incomes have returned to level comparable with the pre-recession period. As for households income shares, the most notable feature is that income shares of the poorest percentiles of the population in 2006 are lower than those observed in 1989, on the contrary of what observed for the richest percentiles. The recovery of average and median income occurred in 1998 has been for the most part determined by the growth of higher incomes, since the income shares of the bottom quantiles have reached their minimum in that year. Since 1998, there has been a slow improvement in the shares of the poorest households, that however

[7] In 1986 the sample design went through a profound revision, so that the size was more than doubled compared to previous surveys. In 1987 there was an over-sampling of richer households that is likely to affect overall results in that year. Therefore, some cautions have to be taken with temporal comparison before those years. To ensure the comparability of data, we use data observed from 1989 onwards. During this period surveys has been affected only from minor changes of sample design and size.

[8] The equivalence scale is the Italian official scale, which assigns a unitary weight to a 2-member household, and then weights of 0.599, 1.335, 1.632, 1.905, 2.150 and 2.401 to households of one, three, four, five, six and seven or more members, respectively.

[9] In order to evaluate the impact of the choice of equivalence scale, two similar applications have been carried out, employing the OECD modified equivalence scale and the LIS equivalence scale, respectively. Results are substantially similar, so that the analysis is not sensitive to the equivalence scale used.

[10] All data are weighted with adjusted weights, provided by the SHIW-HA (Variable "PESOFL2."). These weights are computed post-stratifying the samples in order to obtain the marginal distributions of components by sex, age group, type of job, geographical area and demographic size of the municipality of residence, as registered in population and labour force statistics.

did not return to the level of 1989.

Table 1: *Summary measures of Italian household disposable income: 1989–2006*

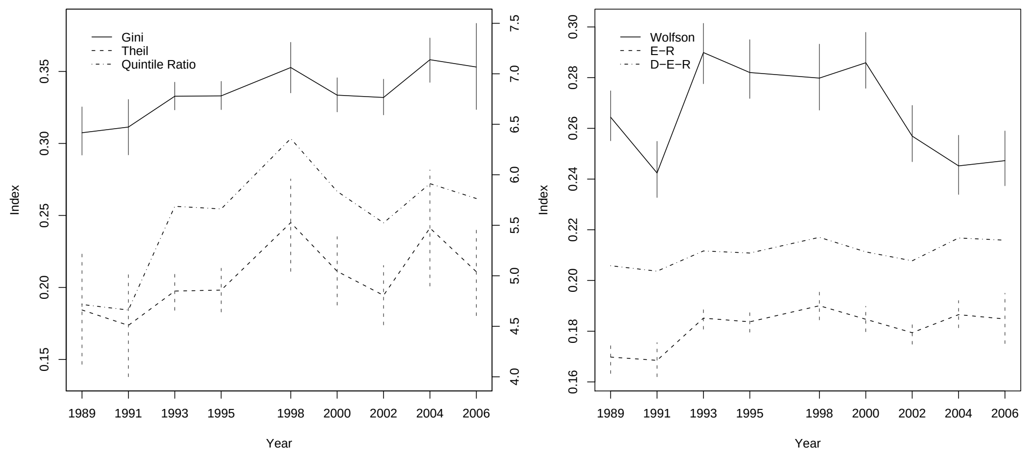
	1989	**1991**	**1993**	**1995**	**1998**	**2000**	**2002**	**2004**	**2006**
Sample size	8,274	8,188	8,089	8,135	7147	8,001	8,011	8,012	7,768
Mean	22,157	22,347	21,148	20,807	22,658	22,863	23,329	25,207	26,064
Median	18,589	19,089	17,908	17,639	18,612	19,356	19,980	20,839	21,480
Income shares									
Bottom 5%	1.42	1.36	0.88	0.94	0.76	0.86	0.92	1.04	1.02
Bottom 10%	3.44	3.39	2.61	2.65	2.37	2.54	2.64	2.75	2.75
Bottom 20%	8.38	8.33	7.18	7.18	6.70	7.02	7.31	7.12	7.23
Top 20%	39.53	38.81	40.86	40.68	42.58	40.96	40.37	42.10	41.69
Top 10%	24.87	23.99	25.70	25.60	27.93	26.07	25.43	27.51	27.14
Top 5%	15.73	15.02	15.98	16.11	18.41	16.56	15.82	18.07	18.00

Note: author's calculation on weighted household income data from SHIW.
Income data are size-adjusted and expressed in 2000 prices.

Figure 1 displays the temporal profiles of income inequality and polarization, alongside with the bootstrapped confidence intervals of some of the indices, represented by vertical bars. We have examined three inequality indices: the Gini coefficient, the Theil Index and the quintile ratio, defined as the ratio of the income share of the top population fifth to that of the bottom fifth. The three inequality indices have nearly the same temporal profile, even if quintile ratio and the Theil index are less stable (Panel (a)). As shown by the vertical bars, between 1989 and 2006 none of the pairwise comparisons of the Gini coefficients and of the the Theil indices is statistically significant. All of the three indices show an increase of inequality from 1989 to 1998. The change between 1998 and 1989 is statistically significant only for the Gini coefficient. From 1998 on, both Gini and Theil indices display little differences. As for polarization, we used the Wolfson index (Wolfson, 1994), the Esteban-Ray (ER) index[11] (Esteban and Ray, 1994), with the *polarization sensitivity parameter* α set at 1.3, and the DER index proposed by Duclos, Esteban and Ray (Duclos *et al.*, 2004), with $\alpha = 0.5$. The latter index, on the contrary of the first two, is defined on the continuous space and relies on the kernel estimated income distribution. The Wolfson polarization measure indicates a significant decline between 2000 and 2006 not detected by the other two indices (Figure 1(b)). As for the ER index, differences between 1991 and any one of the subsequent years is significant at the 5 per cent level. Therefore our data suggests that there has been a spread of income polarization in 1993, but subsequently there are not marked differences. The DER index, displays a temporal profile similar to that of ER index, even if it is slightly less variable.

[11]The ER polarization measure implies a regrouping of the population. In analogy with the Wolfson index, we represent the distribution as a bipolar distribution, using the median as the cut-off value that divides the two presumed groups.

Figure 1: *Income inequality and polarization*



(a) *Inequality measures* (b) *Polarization measures*

Note: authors' calculation on weighted household income data from SHIW. Income data are size-adjusted and expressed in 2000 prices. Vertical bars represent 95% bootstrapped confidence intervals.
Panel (a): quintile ratio values are reported on the right-hand scale.
Panel (b): Esteban-Ray index is computed with α set at 1.3. As for Duclos - Esteban - Ray polarization index, figures are divided by 2 (see Duclos *et al.*, 2004, for details) and are computed for $\alpha = 0.5$.

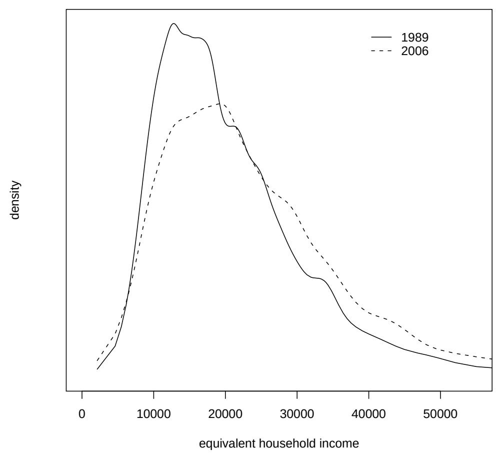
4 Main results

4.1 Changes in income distributions

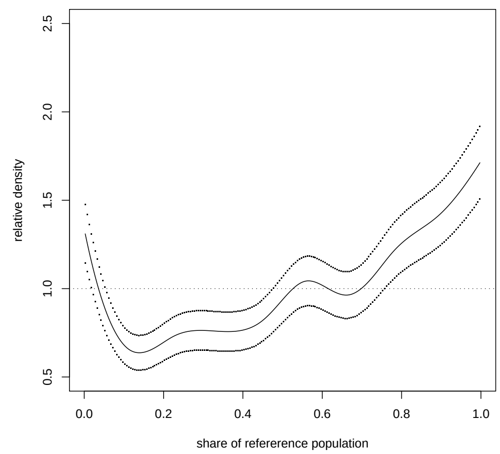
Figure 2(a) reports the kernel density estimates of 1989 and 2006 income distributions. The two densities have been estimated with a weighted kernel function. Both bandwidths have been selected with the Sheather-Jones criterion (Sheather and Jones, 1991). To take into account the variability across densities, a pooled bandwidth was also computed as the weighted mean of the two original bandwidths (Marron and Schmitz, 1992). The shape of the income distribution changed from being near-bimodal to uni-modal. There was also a shift from the 1989 peak down to the right, and a large decrease of decline in the mass at the middle-income classes, combined with a slight increase of concentration at the very lowest incomes.

Further insights can be gained by observing the relative density function that directly compares the two densities and indicates whether the upper and the lower tails of the distribution are growing at the same rate. Panel (b) displays the relative den-

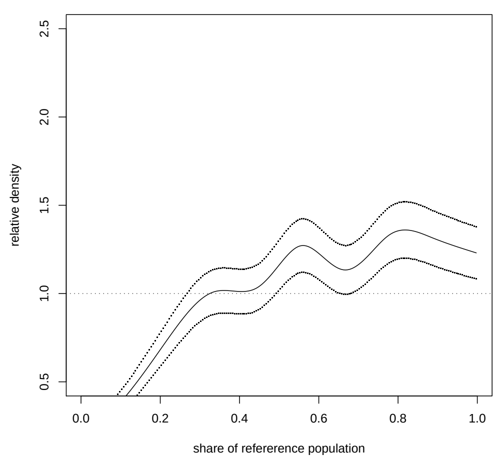
Figure 2: ***Comparison between 1989 and 2006 income distributions***



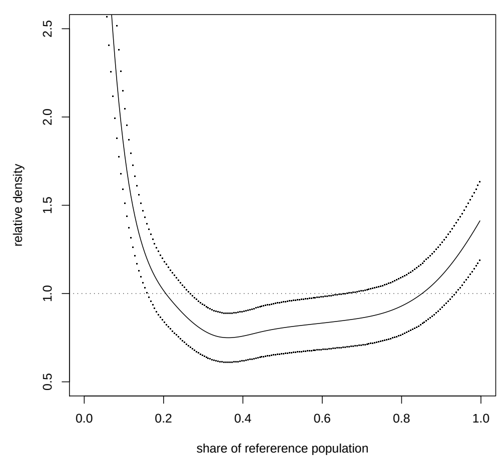
(a) *Kernel density estimate*



(b) *Relative density*



(c) *Location shift effect*



(d) *Shape effect*

Note: authors' calculation on weighted household income data from SHIW. Income data are size-adjusted and expressed in 2000 prices. Panel (a): the bandwidths for the estimate of 1989 and 2006 kernel density functions are obtained with the Sheather-Jones criterion. Panels (b), (c) and (d): dotted lines represent 95% confidence intervals.

sity of household income along with the 95% confidence interval[12]. Households in the low and middle income classes moved toward high and, to a less extent, lowest deciles. Indeed, if we choose any percentile approximately between the 5^{th} and the 55^{th} in the 1989 distribution, the percentage of households in 2006 that earn an amount of income corresponding to the chosen percentile is less than the corresponding percentage of households in 2000.

The relative impact of location and shape shifts to the whole changes in the period examined can be seen in Panel (c) and (d), respectively. The median upshift between 1989 and 2006 impacts the whole range of the distribution with varying intensity, more positively affecting households in the bottom deciles, as shown by the location effect (Panel (c)). Panel (d) display the shape effect, that represent the shape and the magnitude of the income redistribution across households. Having isolated changes of the shape, a rise of relative density, with respect to the observed relative density, at bottom percentiles and a relatively small decrease at the top incomes is detected. The median-adjusted relative density of incomes, indicates a marked change for the incomes below the median, with a decline of the mass between approximately the 20^{th} and the 85^{st} percentile and a prominent increase of the fraction of households below the 20^{th} percentile, indicating a clear downgrading of the distribution. This shows that while the greater part of the households experienced a growth in their real income, a noticeable fraction of them fails to catch up with the rest of population.

4.2 A closer look on polarization

Reported in Figure 3 are the relative polarization indices described in Section 2. These indices traces changes in the shape of the distribution, and they measure the amount and the tendency of these changes. In 1989 the three indices are set equal to 0. The choice of another baseline year would have not affected overall results. The value of the indices would have changed, but the trend would have remained the same. Had the base-year been 2000, for instance, then the zero-point on the y-axis would have shift from 1989 to 2000[13]. Then, the year-to-year comparison, which is

[12]Point-wise confidence intervals for the relative density $g(r)$, $0 \leq r \leq 1$ are based on the asymptotic normal (AN) approximation (Handcock and Morris, 1999, p. 144). The normal asymptotic properties of the estimator $g_{n,m}(r)$ of $g(r)$ are derived under regular assumptions: $0 \leq r \leq 1$, $F_0(x)$ and $F(x)$ have continuous and differentiable densities, $f_0(x)$ $f(x)$ respectively. In addition, $K(.)$ has to be a twice continuously differentiable kernel function, satisfying

$$\int_{-1}^{1} K(x)dx = 1; \int_{-1}^{1} xK(x)dx = 0; \int_{-1}^{1} x^2K(x)dx = \sigma_K^2 > 0$$

and vanishing outside the bounded interval $[-1; 1]$. Choosing a bandwidth h_m, such that, as $m, n \to \infty$, $h_m \to 0$ with $mh_m^3 \to \infty$, $mh_m^5 \to 0$, $m/n \to k^2 < \infty$, then:

$$g_{n,m}(r) \sim AN\left\{g(r), \frac{g(r)R(K)}{mh_m} + \frac{g(r)^2R(K)}{nh_m}\right\}$$

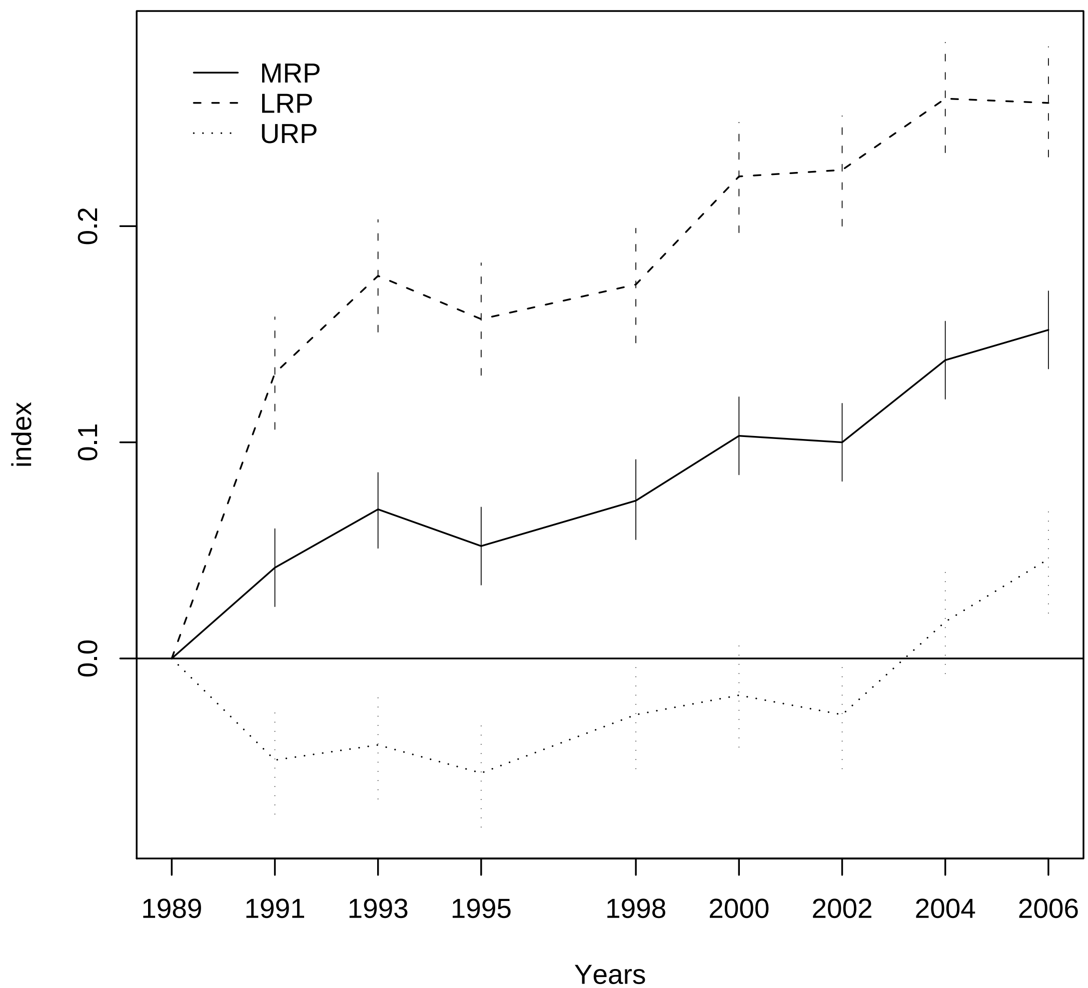
where $R(f) = \int f(x)^2dx$. The second term in the asymptotic variance for $g_{n,m}(r)$ is due to the fact that F_0 is unknown and it must be estimated. In this paper, we use the biweight kernel density function that satisfies the above properties and we estimate h_m using the Sheather-Jones criterion (Sheather and Jones, 1991).

[13]Formally:

$$RP(F, F_0) = -RP(F_0, F)$$

the major concern here, is not affected by the choice of the base year.

Figure 3: *Relative polarization indices 1989-2006*



Note: author's calculation on weighted household income data from 1989-2006 SHIW-HA. Data are household size-adjusted and are expressed in 2000 prices. Vertical bars represent 95% confidence intervals

The rising trend of MRP indicates a strong increase of polarization from the very beginning of the period examined. During 1995 to 2002 polarization of income rises slowly, following the downshift of the index in 1995 with respect to 1993. In 2004 there is a significant change in the shape of the distribution, indicating a further increase of income polarization. This process continues in 2006, even if the change is not statistically significant.

This total outcome results from two contrasting effects that can be detected by the lower and upper indices, LRP and URP. For the whole period the lower index is larger than the upper index, suggesting that the increase in the polarization is generated for

where RP is a generic index of relative polarization.

the most part by the downgrading in household incomes. Polarization in the lower tail displays a trend similar to that observed for the median index. For the most part of the period the upper index shows no real trends, after a sharp decrease in 1991 with respect to 1989. Only in 2002 the upper tail begins to rise more steeply. The net result is that while households at the top of the distribution held on their positions and began to experience an upgrading of their incomes only in recent years, households at the poorest deciles lost ground.

4.3 Linking the changes in employment status of household head to changes in household income

Survey data provide additional covariates which could have significant influence on the variable of interest. In the relative distribution framework this impact could take two forms. The first is a composition effect due to the change of the composition of the population according to these covariates. The second measures the change in the conditional distribution of the response variable, i.e. the change in the relationship between the variable of interest and the covariates. The first effect quantifies the impact of the change of the composition of the population on the income distribution. The second effect detects how the overall income distribution would have changed if the composition of population had been stable.

In this work, we concentrate on the impact of the occupational status of household head on the observed increase of the income polarization[14]. Boeri and Brandolini (2004) document the importance of classifying the population according to the employment status of household head, to explain the cyclical evolution of income inequality in Italy in the last decades. In the SHIW household head is identified as the highest income recipient in the household.

The analysis of homogenous population groups reveals important changes, in the period under investigation, in the distribution of income among social groups as defined by the occupational status of the household's head. The population composition according to the employment status has changed between 1989 and 2006 (see Table 2) especially for unemployed, not retired, and, to a less extent, for managers (both public and private) and self-employed. Between 1989 and 2006 both mean and median income increases ranges between 1.6 and 2.3% per year for the the households of self-employed, managers and retired persons and by only 0.4% for the households of blue and white collars (including school teachers). At the same time, the income share of households headed by a manager or a self-employed diminished less than their share in the population, on the contrary of what happened to households whose head is a blue- or a white-collar. Indeed, the population share of households headed by white collars increases a little, while their income share decline. For households with a retired head the increase in income share was higher than that in the population share, whereas the opposite occurred for households headed by an unemployed, not retired.

Table 2, shows that income distribution shifted in favour of the households of managers self-employed and retired, and to against the households of blue- or white-collar. In turn, these findings are possibly related to several factors, such as the

[14]Other factors have been investigated, such as, sex, age, geographic area of residence and level of education of household head, but their effects are less clear.

Table 2: *Distribution of income by social group*

	Blue collar	White collar	Manager	Self-employed	Retired	Unemployed (not retired)
			1989			
Population share (%)	21.70	18.13	6.93	20.22	32.03	0.99
Income share (%)	17.10	19.83	9.86	26.19	26.31	0.71
Mean	17,456	24,236	31,516	28,700	18,202	15,945
Median	16,175	23,037	28,016	22,405	16,020	11,036
			2006			
Population share (%)	20.21	18.15	4.86	16.55	37.28	2.94
Income share (%)	14.27	18.04	7.44	25.29	33.60	1.35
Mean	18398	25900	39872	39842	23493	11989
Median	17467	24437	36630	28774	20566	8010
			Per-year percentage change			
Population share (%)	-0.40	0.01	-1.76	-1.07	0.96	11.59
Income share (%)	-0.97	-0.53	-1.44	-0.20	1.63	5.30
Mean	0.32	0.40	1.56	2.28	1.71	-1.46
Median	0.47	0.36	1.81	1.67	1.67	-1.61

Note: authors' calculation on weighted household income data from SHIW.
Income data are size-adjusted and expressed in 2000 prices.

wage moderation that took place in the early 1990's, as a consequence of agreements reached by government and social parts, the trade unions and the Confederation of Italian Industry, on the mechanism of wage determination, or to the presence of labour sectors that are not affected by the international competition, in which is easier to raise profit margins (Baldini, 2008).

Table 3 reports the distribution of social groups by income classes in 2006. Households with head unemployed and households whose head is a blue collar are more likely to fall in the first quartile. Households headed by a white collar are concentrated at middle income classes, while those whose head is a managerial worker or is self-employed are relatively more likely to fall in the top quartile. Households with head retired are more equally distributed among income classes. As for composition within groups, there is a marked decrease of the percentage of households headed by a white collar in the top quartile, that is counterbalanced by a noticeable increase of households whose head is a manager or self-employed. There is also a shift from the bottom quartiles to the quartiles above the median for households whose head is retired, while the opposite happened for households headed by an unemployed.

As discussed in Section 2, the relative distribution approach allows one to use the covariate adjustment technique to determine whether differences in the occupational profile between the two populations explain some of the changes in disposable income distribution. Figure 4 represents the adjustment of the relative density for occupational status of household head composition changes. Panel (a) represents the population composition effects, while Panel (b) represents the occupational-adjusted relative density of disposable, that is, the expected relative density of income had the occupational profiles of the two populations been identical.

Figure 4(a) is rather close to a uniform distribution. The implication is that the difference in occupational composition between the two populations had little effect on the observed relative density of income. There was a slight increase in low income deciles associated with the composition change, but the observed increase of income

Table 3: *Distribution of social groups by income quartiles: Year 2006 (%)*

	Quartile			
	I	II	III	IV
Household head employment status				
Blue Collar	37.02	30.56	23.47	8.95
	(3.97)	(1.32)	(-2.62)	(-2.67)
White Collar	16.06	23.66	33.84	26.44
	(1.86)	(5.26)	(3.21)	(-10.33)
Manager	4.18	13.38	21.51	60.93
	(-1.59)	(-0.84)	(-4.58)	(7.02)
Self-employed	17.69	16.02	21.69	44.61
	(-0.89)	(-4.09)	(-1.97)	(6.94)
Retired	25.10	29.13	24.69	21.08
	(-7.92)	(-2.53)	(2.72)	(7.73)
Unemployed	71.96	12.86	9.13	6.05
(not retired)	(17.23)	(-0.9)	(-3.77)	(-12.56)

Note: authors' calculation on weighted household income data from SHIW.
Absolute variation with respect to 1989 in parentheses.

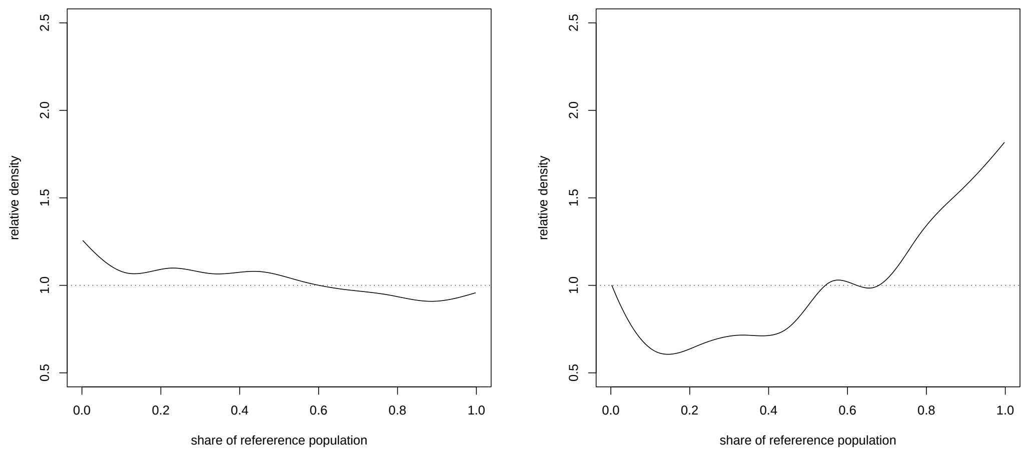
polarization is not being driven by changes in the occupational profile. Figure 4(b) represents the occupation-adjusted relative income distribution. Given the absence of major composition effects, the adjusted distribution is not much different than the original relative density. The shrink of the lower-middle incomes is still evident, embracing a range between the 1^{st} and the 55^{th} percentile. The changing returns to household-head occupational status dominates the changing composition of the population. Given the evidence shown in Tables 2 and 3, the growth in upper tail of relative density is mainly due to the increase of the relative income gap households headed by a manager or a self-employed, and households whose head is a blue- or white-collar or is unemployed.

Both composition and residual components can be further decomposed into location and shape effects. However, since composition effect is negligible, we apply the decomposition only to the residual effect. The location shift in the residual component captures the impact of the changing returns to the occupational status of household head. Figure 5(a) indicates that the upgrading observed in Figure 4(b) is mainly location driven. Once accounting for the impact of the changing returns to the covariate across deciles, the residual effect represents changes in the distribution of these returns. Hence, those differences in the distribution of returns to household head occupation are correlated to the increase of income polarization in 2006, with respect to 1989 (Panel (b)).

5 Conclusions

We have used the relative density method to analyse changes in the Italian household income distribution between 1989 and 2006. In contrast to methods that rely on

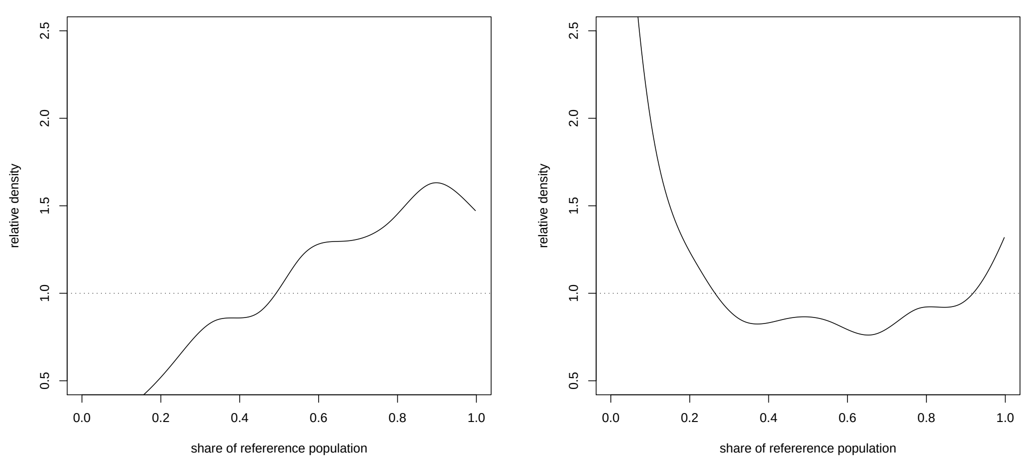
Figure 4: *Composition effect*



(a) *Composition effect* (b) *Residual effect*

Note: author's calculation on weighted household income data from SHIW. Income data at 2000 prices are household size-adjusted.

Figure 5: *Decomposition of the residual effect*



(a) *Location effect* (b) *Shape effect*

Note: author's calculation on weighted household income data from SHIW. Income data at 2000 prices are household size-adjusted.

summary statistics, this non-parametric method summarizes multiple features of the income distribution. This paper documents relevant changes in the income distribution, despite substantial stability in income inequality and traditional polarization measures. The analysis of the size-adjusted household incomes indicates a relevant location effect, an overall upshift of the distribution, that partly masks a tendency to polarization in household incomes. In fact, having controlled for the median increase, a more clear rise in polarization is detected, mainly due to a downgrading of lower incomes.

A temporal comparison of the relative polarization indices indicates that the downgrading of lower incomes has took place all over the period in exams. This effect overcompensated the convergence of higher incomes toward the median.

The change in the relationship between the response variable (the household income) and the distribution of households according to the occupational status of household head have produced an horizontal redistribution across households. The growth in both tails of the relative density is mainly due to the increase of the relative income gap between wealthier households and the lowest income groups, rather than to changes of the composition of the population according to household head employment.

Several extensions of this work are possible. First, the different components of household income can be analyzed separately. Second, the decomposition of relative density according to the covariates could be improved, allowing one to detect the contribution of each categories to the observed changes. Finally, these findings could be linked to specific changes in economic institutions.

References

[1] Baldini M. (2008), "Quali redditi sono rimasti al palo", www.lavoce.info, 13 January.

[2] Banca d'Italia (2008), *Household Income and Wealth in 2006*, Supplements to the Statistical Bulletin, Banca d'Italia, Vol. XVIII, No. 7.

[3] Boeri T., Brandolini A. (2004), "The Age of Discontent: Italian Households at the Beginning of the Decade", *Giornale degli Economisti e Annali di Economia*, Vol. 63, No. 3/4, pp. 449–487.

[4] Brandolini A. (1999), "The distribution of personal income in post-war Italy: source description, data quality and the time pattern of income inequality", *Giornale degli Economisti e Annali di Economia*, 58, pp. 183–239.

[5] Brandolini A., Cipollone P., Sestito P. (2001), "Earnings dispersion, low pay and household poverty in Italy, 1977–1998", in D. Cohen, T. Piketty, G. Saint-Paul (eds.), *The Economics of Rising Inequalities*, pp. 225–264, Oxford University Press, Oxford.

[6] Brandolini A., Smeeding T.M. (2007), "Inequality Patterns in Western-Type Democracies: Cross-Country Differences and Time Changes", CHILD Working Papers wp08-07, CHILD - Centre for Household, Income, Labour and Demographic economics - ITALY.

[7] Ćwik J., Mielniczuck J. (1989), "Estimating Density Ratio with Application to Discriminant Analysis" *Communications in Statistics*, 18, pp. 3057–69.

[8] Ćwik J., Mielniczuck J. (1993), "Data-dependent bandwidth choice for a grade density kernel estimate", *Statistics & Probability Letters*, 16, pp.397–405.

[9] D'Alessio G., Signorini L. F. (2000), "Disuguaglianza dei redditi individuali e ruolo della famiglia in Italia", Banca d'Italia, *Temi di discussione* No. 390, Roma.

[10] D'Ambrosio C. (2001), "Household characteristics and the distribution of income in Italy: An application of social distance measures", *Review of Income and Wealth*, vol. 47(1), pages 43–64.

[11] DiNardo J., Fortin N., Lemieux T. (1996) "Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach" *Econometrica*, 64, pp. 1001–1044.

[12] Duclos J.I., Esteban J., Ray D. (2004), "Polarization: Concepts, Measurement, Estimation", *Econometrica*, Vol. 72, No. 6, pp. 1737–1772.

[13] Esteban J., Ray D. (1994), "On the Measurement of Polarization", *Econometrica*, Vol. 62, No. 4, 819–851.

[14] Handcock M.S. (2005), reldist: Relative Distribution Methods, R package version 1.5-3.
(http://www.stat.washington.edu/ handcock/RelDist)

[15] Handcock M.S., Morris M. (1998), "Relative Distribution Methods", *Sociological Methodology*, 28, pp. 53–97

[16] Handcock M.S., Morris M. (1999), *Relative Distribution Methods in the Social Sciences*, Springer-Verlag, New York.

[17] Jenkins S.P., Van Kerm P. (2005), "Accounting for income distribution trends: A density function decomposition approach", *Journal of Economic Inequality*, vol. 3, No. 1, pp. 43–61.

[18] Marron J.S., Schmitz H.P. (1992), "Simultaneous Density Estimation of Several Income Distributions", *Econometric Theory*, 8, pp. 476–488.

[19] Massari R., Pittau M.G., Zelli R. (2008), "A dwindling middle class? Italian evidence in the 2000's", *Journal of Economic Inequality*, published on line: `http://www.springerlink.com/content/q573670168651670/`, DOI 10.1007/s10888-008-9078-z.

[20] Molanes-López E.M., Cao R. (2008), "Plug-in bandwidth selectors for the kernel relative density estimator", *Annals of the Institute of Statistical Mathematics*, Vol. 60, No. 2, pp. 273–300.

[21] Morris M., Bernhardt A.D., Handcock M.S. (1994), "Economic Inequality: New Methods for New Trends", *American Sociological Review*, 59, pp. 205–219.

[22] Pittau M.G., Zelli R. (2004), "Testing for Changing Shapes of Income Distribution: Italian Evidence in the 1990s from Kernel Density Estimates", *Empirical Economics*, 29, pp. 415–330.

[23] Sheather S.J., Jones M.C. (1991), "A Reliable data-based bandwidth selection method for kernel density estimation", *Journal of the Royal Statistical Society (Series B)*, 53, pp. 683–690.

[24] Voitchovsky S. (2005), "Does the profile of income inequality matter for economic growth?", *Journal of Economic Growth*, 10, pp. 273–296.

[25] Wolfson M.C. (1994), "When Inequalities Diverge", *American Economic Review Papers and Proceedings*, vol. 84, pp. 353–358.