Doctoral School of Economics - Scuola di Dottorato in Economia

Working Papers

No. 1 (February 18, 2009)

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

Riccardo Massari

The Doctoral School of Economics

The Doctoral School of Economics is an academic structure that organizes and coordinates the activities of five doctoral programs in economics and related fields in "La Sapienza".

The School's main goals

- the creation of an integrated system of postgraduate teaching;
- the formation of highly skilled researchers in a multidisciplinary perspective;
- to provide a better visibility to postgraduate teaching;
- to facilitate the students' mobility overseas;
- to establish forms of collaboration with national and international institutions.

The five PhD programs of the School

- International Monetary and Financial Markets (Economia dei Mercati Monetari e Finanziari internazionali)
- Economics (Economia Politica)
- Mathematics for Economic-Financial Applications (Matematica per le applicazioni economico-finanziarie)
- Economic Sciences (Scienze Economiche)
- Economic Statistics (Statistica Economica)

Editorial board of the Working Papers

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

Riccardo Massari[♦]

Abstract

During the 1990's and the early 2000's income inequality in Italy shows levels higher than many other OECD countries, not displaying any significant trend, upward or downward. This evidence relies essentially on summary measures of inequality, which may not capture aspects of the whole income probability density, 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. Furthermore, this approach also allows us to decompose the relative density into changes in location and changes in shape, in order to emphasize whether income distribution becomes more polarized or exhibits patterns of convergence toward middle income classes. A similar decomposition enables us to analyze the impact of selected covariates on 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.

[♦] riccardo.massari@gmail.com

1 Introduction and motivation

Researchers and analysts have developed several measures of inequality that satisfy axiomatic properties, in particular the Lorenz criterion, proposed to rank distributions according to their degree of inequality (Atkinson, 1970). However, traditional methods of inequality analysis, that are essentially based on these summary measures, are not suitable to understand the main "factors" that modify the entire distribution. In addition, they fail to distinguish which part of the distribution is more affected by these factors. There are many reasons to explore inequality in different parts of the distribution. Inequality might be more pronounced in the lower tail of the distribution, or in the top deciles, and the same level of inequality would be correlated with 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) have studied 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 bi-modality, 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 sophisticated comparisons of changes in the upper and lower tails of the distributions. As an example, Di Nardo *et al.* (1996) have used the counterfactual distributions in order to account for the effects of institutional changes on earning distributions. More recently, Jenkins and Van Kerm (2005) have developed a similar method to decompose changes in income distribution using subgroup decompositions of 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.

We illustrate the procedure by analyzing how income distribution of Italian households has become more polarized between the late 1980s and the mid 2000s. We also show how this polarization was mainly affected by the horizontal redistribution across socio–demographic groups. The Italian case is very interesting for applying the procedure.

Recent studies have highlighted that during the 1990's and early the 2000's income inequality in Italy has been one of the highest among OECD countries[1]. However, after a sharp rise in the early 1990's, a period of severe economic recession, inequality displays a substantially stable trend, despite many changes experienced by the Italian economy. These results hold true regardless the summary statistics and the source of data utilized(Boeri and Brandolini, 2004).

This evidence is quite surprising, since Italy experienced a prolonged period of recession and a subsequent phase of low growth from 1993 onward. In addition, recent years are

[1]For a comprehensive review, see Brandolini and Smeeding, 2007.

characterized by a serious transformation of the labor market, with an increase of contingent jobs, which has implied a more disperse earnings distribution in the early 1990s (see Brandolini *et al.*, 2001).

In the same period, intra-household redistribution has mitigated the negative effect on inequality of labor and income transfers (D'Alessio and Signorini, 2000). As a matter of fact, the Italian economy has been influenced by relevant social and demographic changes. These changes range from the ageing of the population, to the restructuring of the labor market, with a steady increase of female participation, incrementing the number of income earners in the households. Brandolini *et al.* (2001) showed that being at risk of poverty is more correlated with the number of household members employed other than the head, rather than with low pay.

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 labor 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.

1. The relative distribution approach

The relative distribution is a non-parametric statistical approach introduced by Handcock and Morris (1998, 1999) that compares income (or other) distributions of two populations, the "reference" and the "comparison" population, considering 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. This allows us 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:

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

where r is the proportion of values, and

$$F_0^{-1}(r) = Q_0(r) = inf\left\{ y_0 \,\middle|\, F_0(y) = r \right\} = y_r$$

$$F_0^{-1}(r) = Q_0(r) = inf\left\{ y_0 \,\middle|\, F_0(y) = r \right\} = y_r$$

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:

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

The quantile function $Q_0(r) = y_r$ returns the value of income y in the reference distribution below which a proportion r of 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. The probability of being in the r^{th} quantile of the baseline distribution is higher (lower) for the households that belong to the reference population. In this paper the relative density $g(r)$ is obtained 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 kernel estimates on *P* quantiles y_r of the reference population[2]. The two density functions are estimated at the same points y_r. A non parametric regression is finally applied for smoothing the plug-in estimates $\hat{g}(r)$.

We adopt an adaptive bandwidth in the kernel estimation to take into account data sparseness (Pittau and Zelli, 2004). The relative distribution approach also provides different tools that allows us to isolate which factors have affected the observed changes in income distribution. Differences between the reference and the comparison population could be due to changes in the average (or median) income, but also to differences in shape, that are differences in variation, skewness and other distributional characteristics. It is then possible to distinguish between two effects, a *location* effect, due to a change in the first moment, and a *shape* effect, due to changes in higher order moments of the distribution.

The decomposition of the relative density in *location* and *shape* effect relies on the definition of an *additive* location-adjusted population $Y_{0L} = Y_0 + \rho$. In this analysis, the location-adjusted population is estimated based on the median income. Therefore, Y_{0L} is a counterfactual distribution with the same shape of the reference distribution but 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:

---

[2]We choose a number of point $P = 200$, but the number of quantiles does not significantly affect results.

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

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[3], 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 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. To corroborate the usual impressions of the graphical analysis, the *median relative polarization index* (MRP), introduced by Morris *et al.* (1994), allows us to quantify the degree of polarization.

The index (MRP) is defined as 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 equal to zero indicates no differences in distributional shape.

[3]As observed above, alternative indices as 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).

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[4], 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..

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[5]

Under regularity conditions, $\widehat{MRP}(F, F_0)$ is an asymptotically unbiased and asymptotically normally distributed estimator of $MRP(F, F_0)$.

In addition, the MRP index could be exactly decomposed into two indices, allowing is to estimate separately movements below and above the median, respectively[6]. These two index are defined, respectively, *lower relative polarization index* (LRP) and *upper polarization index* (URP):

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

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

[5]The R_j s 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).

[6]The MRP index could be decomposed into different index in order to observe changes occurred in every percentiles of the distribution.

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

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.

Similarly to 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 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[7]. 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:

[7]Extension to the continuous case is straightforward.

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

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.

2. 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[8] (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[9]. 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

[8]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.

[9]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.

household, regardless of their individual income, their role in the household and other characteristics[10].

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[11]. 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 did not return to the level of 1989.

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 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

[10]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.
[11]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 labor force statistics.

Wolfson index (Wolfson, 1994), the Esteban-Ray (ER) index[12] (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.

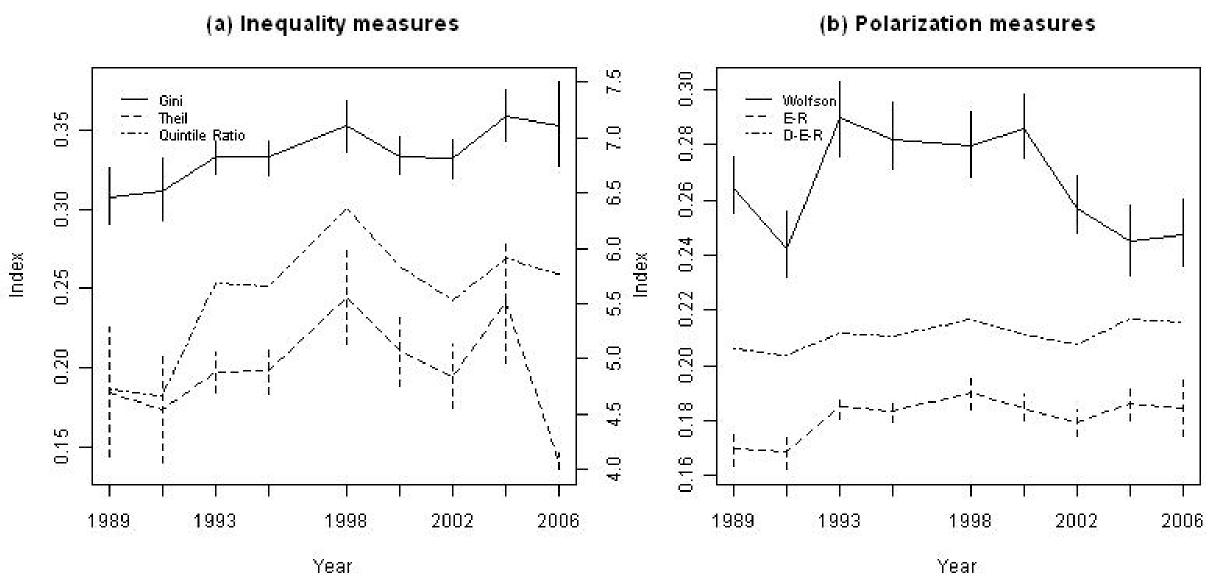
[12]The ER polarization measure implies a regrouping of the population. In analogy with the Wolfson index, we represent the distribution as a bipolar distribution, using the median as the cut-off value that divides the two presumed groups.

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

	1989	1991	1993	1995	1998	2000	2002	2004	2006
Sample size	8274	8188	8089	8135	7147	8001	8011	8012	7768
Mean	22157	22347	21148	20807	22658	22863	23329	25207	26064
Median	18589	19089	17908	17639	18612	19356	19980	20839	21480
					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



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 $\alpha = 0.5$.

3. Main results

2 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 touni-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 density of household income along with the 95% confidence interval[13]. 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

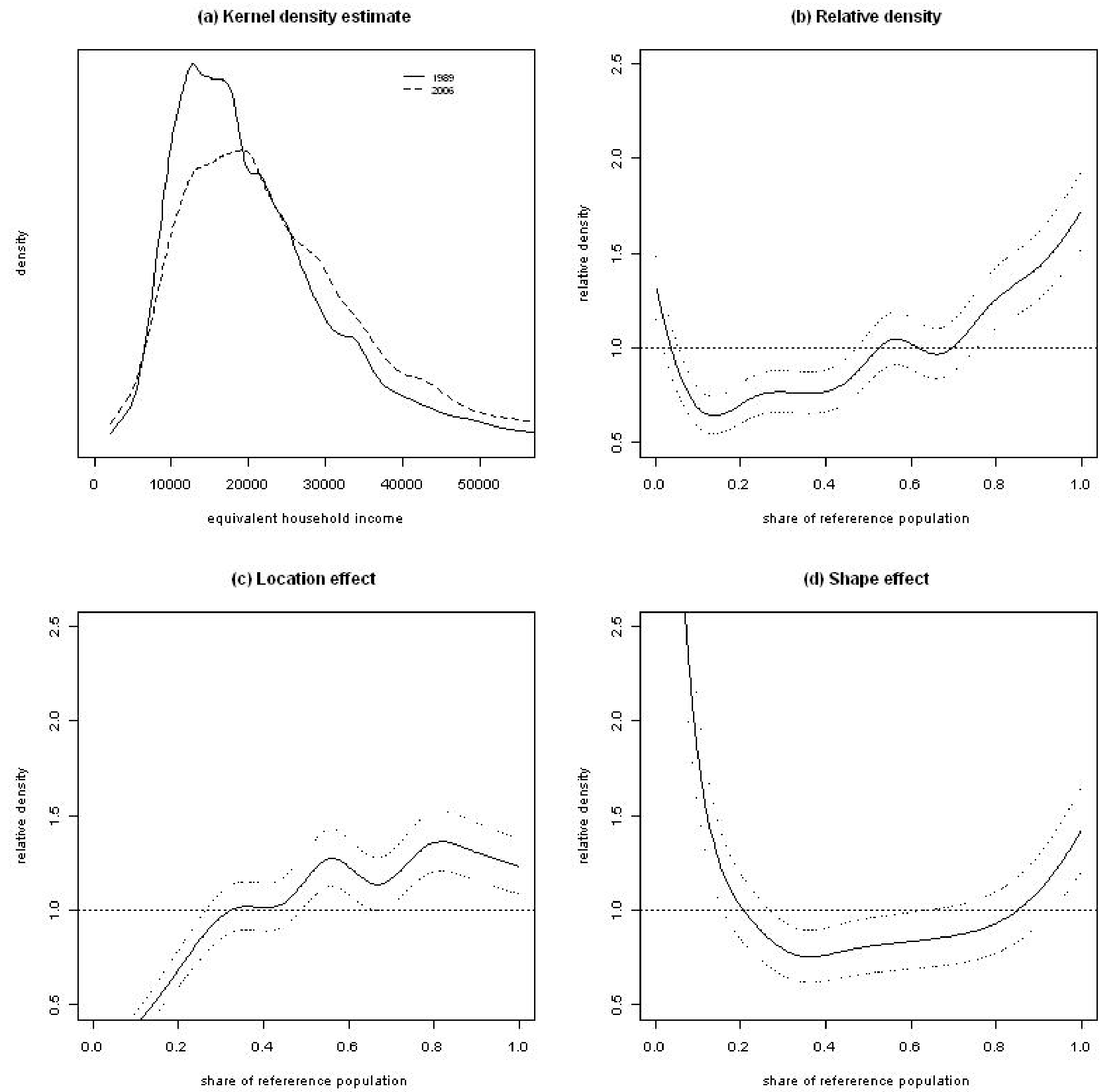
---

[13]Point-wise confidence intervals for the relative density $g(r)$, $0 \le r \le 1$ 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: $0 \le r \le 1$, $F_0(x)$ and $F(x)$ have continuous and differentiable densities, $f_0(x)$. $f(x)$ respectively. In addition, $K(\cdot)$ has to be a twice continuously differentiable kernel function, satisfying $\int_{-1}^{1} K(x)dx = 1; \int_{-1}^{1} xK(x)dx = 0; \int_{-1}^{1} x^2 K(x)dx = \sigma_K^2 > 0$ 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: $g_{n,m}(r) \sim AN\left\{ g(r), \frac{g(r)R(K)}{mh_m} + \frac{g(r)^2 R(K)}{nh_m} \right\}$ where $R(f) = \int f(x)^2 dx$. 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).

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.

Figure 2 Comparison between 1989 and 2006 income distributions

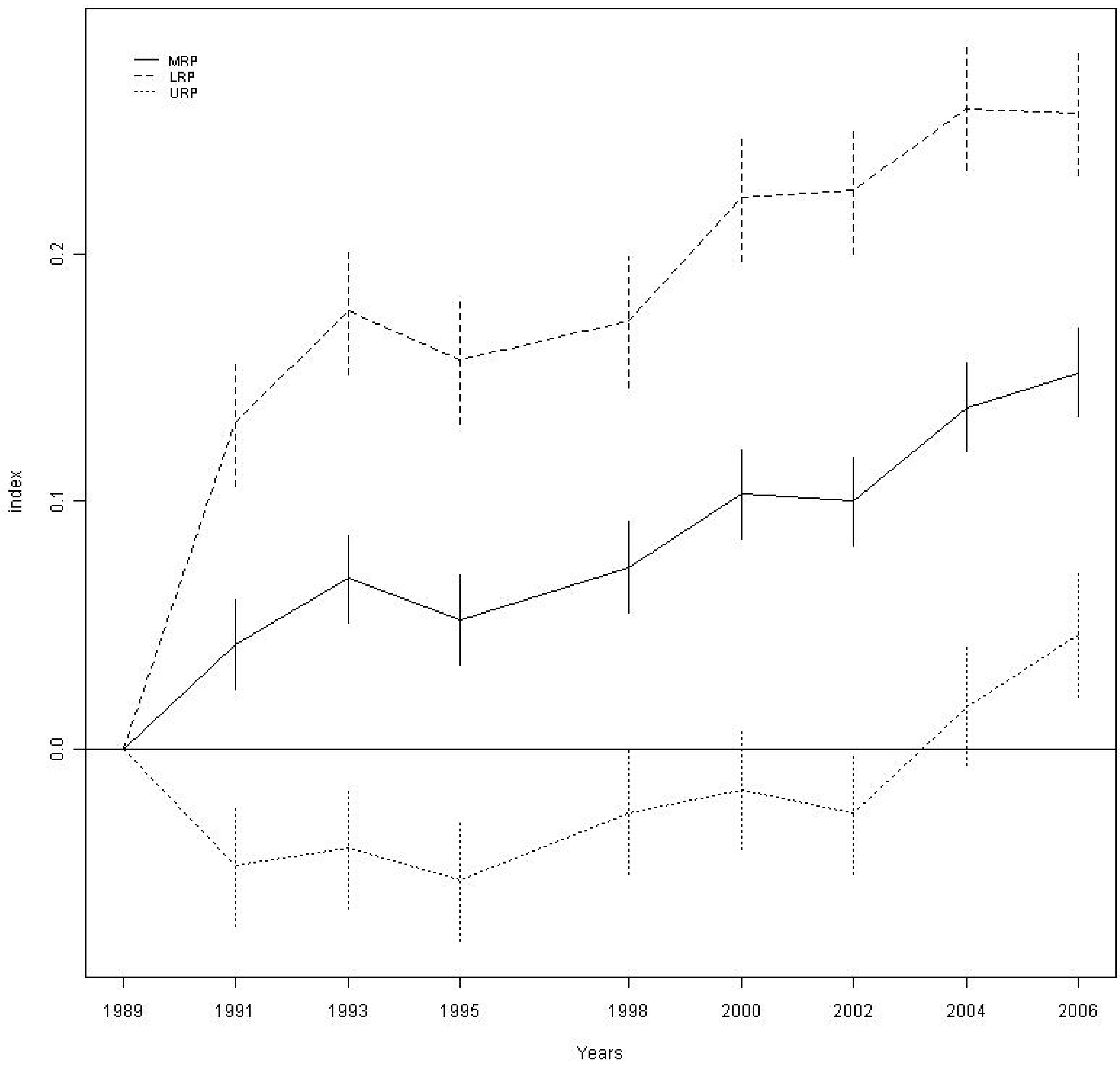


Note: authors' calculation on weighted household income data from SHIW. Income data are size-adjusted and expressed in 2000 prices. Panel (a): the bandwidths for the estimate of 1989 and 2006 kernel density functions are obtained with the Sheather-Jones criterion. Panels (b), (c) and (d): dotted lines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

3 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[14]. Then, the year-to-year comparison, which is 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.

[14]Formally: $RP(F, F_0) = -RP(F_0, F)$ where RP is a generic index of relative polarization.

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 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 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[15]. Boeri and Brandolini (2004) document the importance of classifying the population according to the employment status of household head,

[15]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.

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 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 favor 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 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 labor sectors that are not affected by the international competition, in which is easier to raise profit margins (Baldini, 2008).

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	17456	24236	31516	28700	18202	15945
Median	16175	23037	28016	22405	16020	11036
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.

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.

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

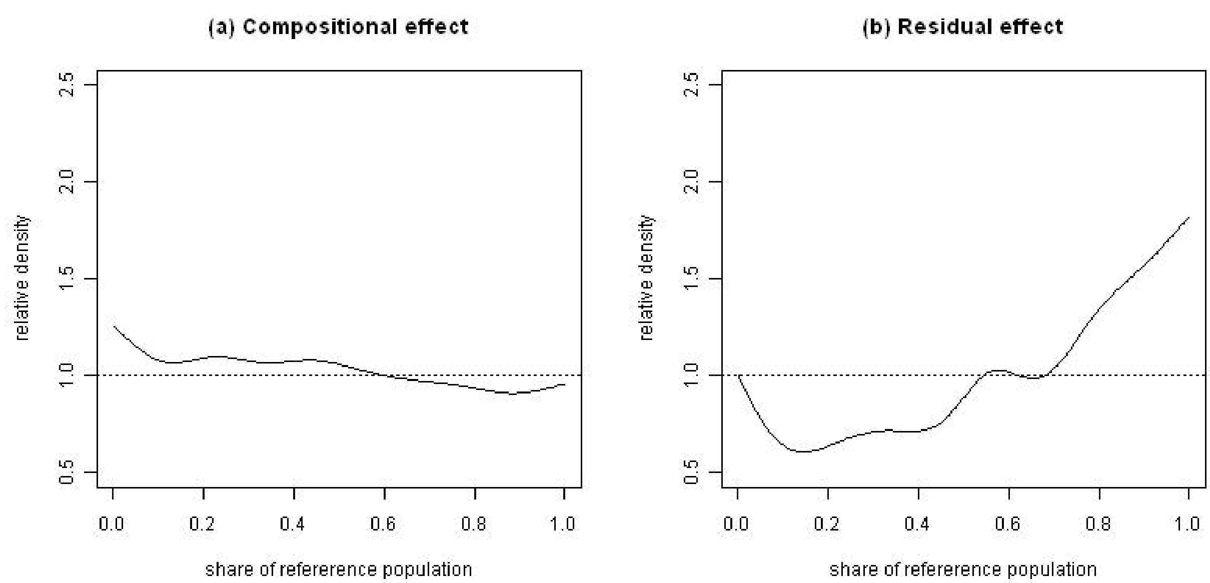
Household head employment status	Quartile			
	I	II	III	IV
Blue Collar	37.02 (-3,97)	30.56 (-1,32)	23.47 (-2.62)	8.95 (-2,67)
White Collar	16.06 (-1,86)	23.66 (-5,26)	33.84 (-3,21)	26.44 (-10,33)
Manager	4.18 (-1,59)	13.38 (-0,84)	21.51 (-4,58)	60.93 (-7,02)
Self-employed	17.69 (-0,89)	16.02 4,09)	21.69 (-1.97)	44.61 (-6,94)
Retired	25.10 (-7,92)	29.13 (-2,53)	24.69 (-2.72)	21.08 (-7,73)
Unemployed (not retired)	71.96 (-17,23)	12.86 (-0,9)	9.13 (-3,77)	6.05 (-12,56)

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

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 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.

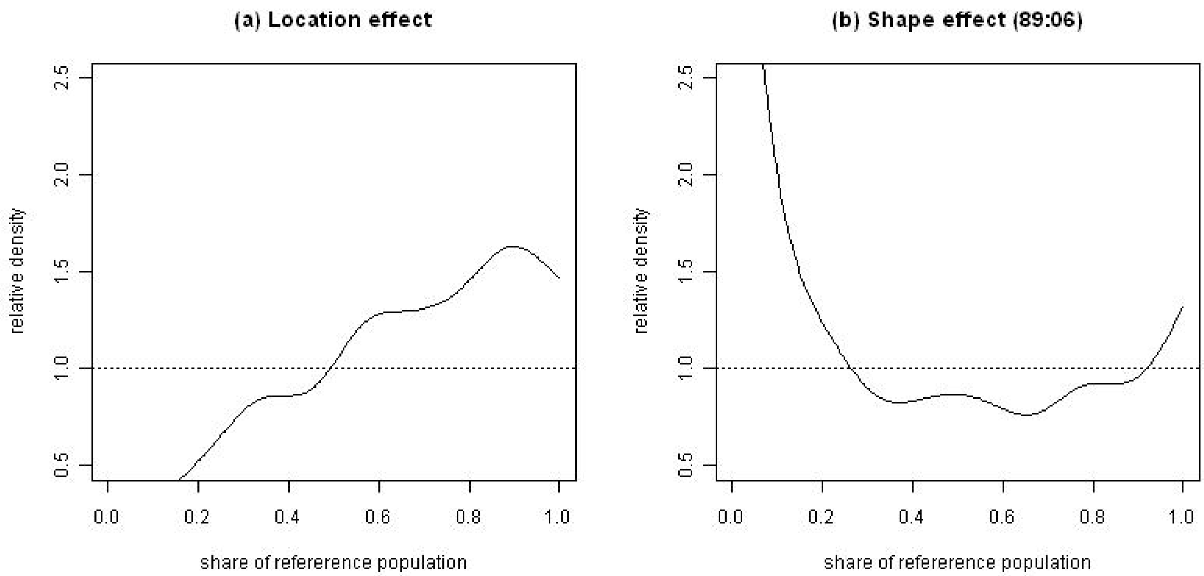
Figure 4 Composition effect



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

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)).

Figure 5 Decomposition of the residual effect



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

4. Conclusions

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

Atkinson A. B., "On the Measurement of Inequality", *Journal of Economic Theory*, 3, pp. 244–63.

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

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

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.

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.

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.

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.

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

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

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.

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.

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

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

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

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

Handcock M.S., Morris M. (1998), “Relative Distribution Methods”, *Sociological Methodology*, 28, pp. 53–97

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

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.

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

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/q57367016865l670/, DOI10.1007/s10888-008-9078-z.

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.

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

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.

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.

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