A dwindling middle class? Italian evidence in the 2000's

Riccardo Massari* Maria Grazia Pittau† Roberto Zelli‡

Abstract

In recent years, increasing attention has been paid to household impoverishment in Italy. Most of works dealing with this issue are based on summary statistics, which do not capture the whole information contained in the income distribution. The paper applies a non-parametric tool, the "relative distribution", to Italian household income data during 2002-2004. The relative density function is a proper density function which compares two distributions observed in different years, in order to describe patterns of differences on the entire income scale. This approach also allows for a decomposition of the relative distribution density function to isolate changes due to differences in location from changes due to differences in shape, thus providing further insights into the analysis of income polarization. During the 2000's there was a significant location effect, but also an increased income polarization, which has particularly affected incomes below the median. Social group analyses, according to the main income source of the household, show significant re-distribution effects within groups.

1 Introduction

Rather than revelling in the *dolce vita*, Italians are battling with the *carovita* (the high cost of living), ... Newspaper headlines warn that the "Middle class has gone to hell" and "Italians don't know how they will make it to the end of the month". The Guardian, Tuesday, December 28, 2004.

In recent years, the so-called "household impoverishment" in Italy, especially with regards to the middle class, has led researchers and policy makers to turn their attention to distributional issues. However, while there is a growing media debate about the vulnerability of the middle-class, the fall of purchasing power of households, their difficulty to make ends meet (the "fourth week syndrome"), on the other hand estimates of inequality indices and poverty measures show a substantial stability in the last decade. Evidence of stability is even more pronounced when the analysis is restricted to the last time period.

Based on consumption data, Trivellato (1998) shows that in 1990-1997 the incidence of poverty has remained substantially stable, while there has been an

*DCNAPS, Sapienza Università di Roma. e-mail address: riccardo.massari@uniroma1.it

†Statistics Department, Columbia University, New York and DCNAPS, Sapienza Università di Roma. e-mail address: grazia.pittau@uniroma1.it

‡DCNAPS, Sapienza Università di Roma. e-mail address: roberto.zelli@uniroma1.it

growth of three percentage point in poverty intensity[1]. Baldini (2005) updates the head-count ratios for the period 1997-2004, using the same data source, confirming a substantial stability, even with some differences between macro-regions. Brandolini (1999) reports the time series of Gini coefficients until 1995, calculated from the Bank of Italy's Survey on Household Income and Wealth (SHIW), showing that income inequality displays a cyclical pattern around a relatively stable trend, during 1980s and early 1990s. Boeri and Brandolini (2004) illustrate that estimates of inequality indices based both on Bank of Italy's SHIW and Istat's Survey of Family Budgets (SFB) remain to large extent unchanged during the period 1993-2002, after a sharp rise in early 1990s. In the same work, the authors note that there is an apparent contradiction between the above results and the consumers' confidence climate[2], because Italians perceive a deterioration of their economic conditions. Their conclusion is that this inconsistency could be due to several factors, like disappointed expectations, and a horizontal re-distribution across social groups. The authors have computed a decomposition of Gini index which allows for detecting separately the within-group and between-group inequality (Lambert and Aronson, 1993). During the period between 1993 and 2002 the within-group inequality remains unchanged, but there was an increase of the distance among social-groups which would have determined a rise of the Gini index.

Most of the analyses carried out are focused on summary measures of location and variation, so that they have fallen short to capture the multi-faceted nature of income distribution, leaving part of the informational content in Household Budget Surveys unexplored.

There are several theoretical motivations why the whole shape of the distribution matters (for a recent survey see Voitchovsky, 2005). When dealing with distributional issues, the behavior at both ends of the distribution is as relevant as the behavior around the mean value. In particular, analysis of polarization which captures a particular form of discrepancy in living standards remains hidden when only trends in average incomes are examined. Therefore, it is worthwhile to explore alternative approaches that permit the whole content of the income distribution to emerge.

The relative distribution approach (introduced and developed by Handcock and Morris, 1999) provides a non-parametric framework for overcoming these deficiencies. Using this methodology, two populations, the "reference" population and the "comparison" one, are directly compared in a way that is based not only on measures on location and variation, but rather takes into account differences throughout the entire income range. Basically, the relative distribution returns the fractions of "comparison" population which fall in each percentile 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. The richness of the relative distribution approach is also in that it provides a number of tools to isolate the effects on the observed outcome. For instance, it is possible to distinguish between shifts in the mean (or median), namely a *location* effect, and changes in variation, skewness and other higher moments,

[1]This means that while the percentage of poor households is approximately the same, they had become relatively poorer during those years.

[2]This analysis is based on data from three sources: the European Commission Business and Consumer Survey (BSC), the Eurobarometer, and the European Community Household Panel (ECHP).

the *shape* effects.

In most studies, the relative distribution analysis has been applied to long-run modifications of distribution, over a time span of ten years or more (Morris *et al.*, 1994; Bernhardt *et al.*, 1995). On the other hand, significant changes could occur in the short run as well, and these are more difficult to be detected, especially with traditional tools of inquiry. These changes could be a signal of underlying structural breaks or the effects of economic and social policies specifically addressed to segments of the population.

In this paper, the relative distribution framework is applied to short-run changes that have recently arisen in the Italian household income distribution. To our knowledge, there are no similar works which apply the relative distribution approach to short-run dynamics of income distribution.

This paper is structured as follows. In the next section the methodology of the relative distribution approach is briefly presented. Section 3 is devoted to the illustration of the main features of the data employed. The main findings of the changes that occurred in the Italian household income distribution from the year 2002 to 2004, applying the relative distribution methodology, are in section 4. Finally, section 5 reports some concluding remarks.

2 Methodology

Let Y_0 be a random continuous variable which represents an outcome attribute (e.g. income) for a population, the *reference* population (e.g. households in 2002). Let F_0 be the cumulative distribution function (CDF) of Y_0 and f_0 its probability density function (PDF), which is assumed to be continuous and derivable. Then consider another population, the *comparison* population (e.g. households in 2004), which generates the random continuous variable Y, and let F and f be respectively its CDF and PDF.

It is possible to show that the relative distribution, $g(r)$, is a proper PDF of a random variable R continuous in $[0, 1]$, where R is obtained from Y by shifting it by the CDF F_0 (Handcock and Morris, 1999). The realizations of R, r, are referred as *relative data*. The relative data can be interpreted as the percentile rank that each observations of the comparison group would have in the reference population[3].

Let $r \in [0, 1]$ such that $F_0(Q_0(r)) = r$, where Q_0 is the quantile function of Y_0: $Q_0(r) = inf\{y_0 \mid F_0(y) = r\} = F_0^{-1}(r) = y_r$. The quantile $Q_0(r)$ can be thought as the value of the outcome y in the reference distribution below which a proportion r of the ordered outcome values fall.

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, i.e.:

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

This density ratio can be interpreted as the ratio of the fraction of households in the comparison population to the fraction of households in the reference population at a given level of the outcome, the quantile $y_r = F_0^{-1}(r)$.

[3]In fact, Ćwik and Mielniczuck (1993) used a similar approach by transforming Y by means of the CDF of Y_0, calling it *grade* transformation

When no changes occur between the two samples, $g(r)$ 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} percentile of the reference population. Put in another way, households of the comparison population have a higher (lower) probability than household in the reference population of having the level of income that corresponds to the r^{th} quantile of the baseline distribution.

Let $Y_{01}, Y_{02}, \ldots, Y_{0n}$ and $Y_1, Y_2, \ldots, Y_n$ two samples of independent observations drawn, respectively, from F_0 and F. A natural estimator of the relative PDF $g(r)$ has been achieved in the following way: as a first step, P percentiles, y_r, of the reference population have been computed. Secondly, the PDFs of Y_0 and Y have been estimated by a kernel density estimator, using a Gaussian kernel and an adaptive bandwidth[4].

To estimate the optimal pilot bandwidth Sheather-Jones plug-in criterion (Sheather and Jones, 1991) has been used. The estimates of $f_0(y)$ and $f(y)$ are:

$$\begin{aligned} \hat{f}_0(y_r) &= \sum_{i=1}^{n} \frac{w_i}{\lambda_i h} K\left(\frac{y_r - Y_{0i}}{\lambda_i h}\right) \\ \hat{f}(y_r) &= \sum_{j=1}^{m} \frac{w_j}{\lambda_j h} K\left(\frac{y_r - Y_j}{\lambda_j h}\right) \end{aligned}$$

where $K(\cdot)$ is the Gaussian kernel, h is the pilot bandwidth, w_i and w_j are the weights attributed to each household in the samples, while λ_i and λ_j are the coefficients used for the computation of the adaptive bandwidth. Note that the points on which the two density function are estimated are the same, namely the percentiles y_r. Then, the quantities $\hat{g}(r) = \hat{f}(y_r)/\hat{f}_0(y_r)$ have been computed. These quantities could be considered as plug-in estimate of $g(r)$. Finally, these fractions have been smoothed applying a local-polynomial model[5].

It should be noted that the estimation procedure of $g(r)$ proposed here is somewhat different from that used by Handcock and Morris (1998) and developed in the `reldist` `R` package (Handcock, 2005), which estimate the relative density by using a local-polynomial smoother to a binned version of the relative data. Our methodology also differs from the procedure proposed by Ćwik and Mielniczuck (1993) in which the relative density is estimated using a kernel density estimator applied to the relative data.

A decomposition of the relative distribution allows one to disentangle a shift effect from a shape effect. With this decomposition it is possible to state how much of overall changes stems from changes due to a simple shift of incomes and changes due to the redistribution of income along the income scale.

First, to obtain the decomposition, a location-adjusted population, Y_{0L}, is defined. The "counter-factual" distribution Y_{0L} is constructed to have the same shape of the reference distribution, and the median[6] of the comparison distri-

[4] An adaptive bandwidth is used to take into account the dispersion of sample data (see, e.g., Pittau and Zelli, 2004).

[5] To estimate a general additive model over the $\hat{g}_r$ the `mgcv` library in `R` has been used (Wood, 2000). The main advantage of the local polynomial smoother is that it is not affected from the boundary bias of the kernel estimator.

[6] In principle, it is possible to consider alternative indices of central tendency, like the mean. However, results obtained using the mean do not differ in a significant way, and are

bution, i.e. $Y_{0L} = Y_0 + \rho$, where $\rho = median(Y) - median(Y_0)$. The CDF of Y_{0L} is defined as $F_{0L}(y) = F_0(y + \rho)$, and its derivative is the PDF f_{0L}.

Being $g(r)$ the relative density of Y to Y_0, it is possible to decompose it in the following way:

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

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, the density ratio for location differences, $g_L(r)$, will be uniform in $[0, 1]$. Conversely, if the two distributions have different location, then $g_L(r)$ displays an increasing (decreasing) trend if the reference median is higher (lower) than the comparison median. Instead, the density ratio for shape differences, $g_S(p)$, represents the relative distribution net of the location effect. The analysis of g_S is important to detect the re-distribution occurred between the reference and the comparison populations. For instance, $g_S(p)$ would take a (reverse) U-shape, if the comparison population is relatively (less) more spread around the median. Then it is possible to observe whether there is an increasing income polarization, or, conversely, a convergence of incomes towards the median income.

The estimation procedure of $g_L(r)$ and $g_S(p)$ is similar to that used to estimate $g(r)$. The only difference is that the density function $f_{0L}(y_r)$ is estimated on the median-adjusted re-scaled data.

With the purpose of inferring about the cause of the observed variation, one or more covariates, that could have induced the shift of the comparison population with respect to the reference one, could be selected. Then, it is possible to analyze, in the framework of the relative distribution approach, the influence of these variables, distinguishing between two effects, namely the *compositional* and the *residual* effects. The compositional effect is due to differences in the structure of the two populations according to the covariates. The *residual* effect reflects the modification of the relationship between the variable under examination, the household income and the covariates.

The estimation of the decomposition for covariates of the relative density can be achieved in a similar way as before. The counter-factual distribution here is the one that have the same shape of the comparison distribution, but the structure of the reference group. This is the hypothetical distribution that one would have observed if the comparison group retains the same profile of the reference group.

It should be noted, however, that in the short run the decomposition of the relative density for covariates is likely redundant, since there will be only

not reported here for the sake of brevity. In addition, since we are dealing with data in the original scale, a multiplicative decomposition can also be applied. However, the multiplicative decomposition has the drawback of affecting the variance and the shape of the distribution. Indeed, the equi-proportionate income changes causes a flattening (or a shrinking) of the shape of the distribution (Jenkins and Van Kerm, 2005). At any rate, results obtained with the multiplicative decomposition do not differ significantly from those presented here (results available from the authors on request).

marginal changes in the sample structure. In any case, provided that the population is exhaustively partitioned in mutually-exclusive sub-groups, according to the covariates, it is possible to analyze differences *within* and *between* groups. This analysis can provide further insights into the differences between the two distributions.

3 Data

Data are drawn from the Bank of Italy SHIW for the years 2002 and 2004 (see Banca d'Italia, 2006, for a detailed description of the survey). The units of observation are households and the variable observed is the annual disposable income of all the household members, net of tax and social security transfers. In particular, the analysis is restricted to the 3604 households in the 2002-04 panel section of SHIW, in order to investigate on the actual re-distribution across the same households.

The observations are re-weighted, with weights $\tilde{w}_i$ proportional to those provided by Bank of Italy in both surveys. Data are adjusted for household size[7] and reported in 2002 prices using the deflator of the final consumption expenditure of households available in national accounts (HED)[8], which indicates a price increase of about 4.8% between 2002 and 2004. Henceforth, we use household income and equivalent household income as synonymous.

To take into account the sharp growth of home value in recent years[9], a similar application is carried out on income net of imputed rents[10]. Indeed, imputed rents could have the effect of artificially inflating the household income and to produce a spurious reduction of inequality[11]. Paiella (2004) observes that, given the very low estimated marginal propensity to consume out of real assets, the rapid increase in real estate prices since the end of the 1990s has affected only marginally household expenditure capability. This is due to the relative scarce liquidity of the real assets, which limits the scope to realize capital gains, and to the strength of the bequest motive.

In Table 1 some summary measures are reported, computed on household income data from SHIW 2002 - 2004. Between 2002 and 2004, both median and mean household income rose in real terms. The income shares of households in bottom percentiles remain substantially unchanged, even if the shares of bottom 5% and 10% slightly increase. Conversely, there is a moderate upsurge of the income shares of the top earners, especially for households in the top decile and in top 5%. Given this results, inequality indices, as well as polarization measures, display a negligible increase. The incidence of poverty measures show a modest fall, that is marginally more pronounced when the the poverty line is computed at 50% of the median income (which is the lowest line between those

[7] 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.

[8] The HED is used for consistency with the definition of income employed in this work, since it is constructed including imputed rents on owner-occupied housing.

[9] The imputed rents have risen, in real terms, by 13% within the observed period.

[10] In this case, data are reported in 2002 prices using the consumer price index (CPI), that does not include the value of imputed rents, that indicates a price rise of about 5%.

[11] According to the 2006 Annual Report of Istat, the Gini coefficient computed in 2003 on Italian household income net of imputed rents is 0.329, but it reduces to 0.312 when imputed rents are included.

considered here). Overall, the income distribution characteristics, as measured with traditional summary measures, do not display significant changes[12]. When income net of imputed rents is considered (see Table 2), substantially few differences arise. The more marked difference is the smaller growth of the mean and the median of the distribution.

Table 1: Summary measures of Italian household income distribution (total household income): 2002-2004

	2002	2004
Mean (2002 euros)	24,621	25,578
Median (2002 euros)	21,277	21,592
Income shares (per cent)		
Bottom 5%	1.06	1.14
Bottom 10%	2.85	2.96
Bottom 20%	7.60	7.58
Top 20%	39.72	39.88
Top 10%	24.91	25.21
Top 5%	15.45	16.10
Inequality measures		
Gini Index	0.32	0.31
Theil Index	0.20	0.29
Quintile ratio	5.26	5.23
Polarization measures		
Wolfson Index	0.26	0.27
Esteban-Ray (α=1.3)	0.17	0.18
Head-count poverty ratio (per cent)		
line at 50% of median	11.12	10.34
line at 60% of median	16.96	16.60
line at 50% of mean	15.54	15.07
line at 60% of mean	23.94	23.55

Note: authors' calculation on weighted household income data from 2002-04 panel subset of SHIW. Income data are household size-adjusted and are expressed in 2002 prices.

4 Main results of the relative distribution analysis

Reported in Figure 1, are the kernel estimates of the PDF's of 2002 and 2004 income distribution, using household income (panel (a)) and income net of imputed rents (panel (c)). Panel (a) shows a bi-modal distribution in 2002, with

[12]Bootstrapped confidence intervals at 95% of the inequality indices almost overlap. For instance, the upper bounds of the Gini coefficients are 0.332 and 0.338, respectively in 2002 and 2004; the lower bounds are 0.297 and 0.304

Table 2: Summary measures of Italian household income distribution (income net of imputed rents): 2002-2004

	2002	2004
Mean (2002 euros)	19,859	20,404
Median (2002 euros)	17,060	17,439
Income shares (per cent)		
Bottom 5%	0.90	1.05
Bottom 10%	2.69	2.93
Bottom 20%	7.61	7.57
Top 20%	39.66	40.27
Top 10%	24.89	25.68
Top 5%	15.35	16.60
Inequality measures		
Gini Index	0.32	0.33
Theil Index	0.19	0.21
Quintile ratio	5.21	5.32
Polarization measures		
Wolfson Index	0.26	0.27
Esteban-Ray (α=1.3)	0.17	0.18
Head-count poverty ratios (per cent)		
line at 50% of median	10.70	9.87
line at 60% of median	17.10	16.93
line at 50% of mean	16.43	15.65
line at 60% of mean	24.94	25.10

Note: authors' calculation on weighted household income data from 2002-04 panel subset of SHIW. Income data are household size-adjusted and are expressed in 2002 prices.

a peak located at about 22000 euros, and a second one located at lower income level. The 2004 distribution appears to be tri-modal, with a first mode at an income level lower than the 2002 mode, the second one located at the same level of the mode of 2002, and a third one placed at a higher income level. Most important, it is noticeable a sharp decline of the PDF, at the middle-income range. In the case of household income net of imputed rents, the shape of the 2002 distribution is similar to that computed for total income, while the insurgence of three distinct modes is evident in 2004. As shown in Pittau and Zelli (2006), the insurgence of these modes, and the gap between them could be interpreted as an upsurge in polarization, especially in a context of stable inequality.

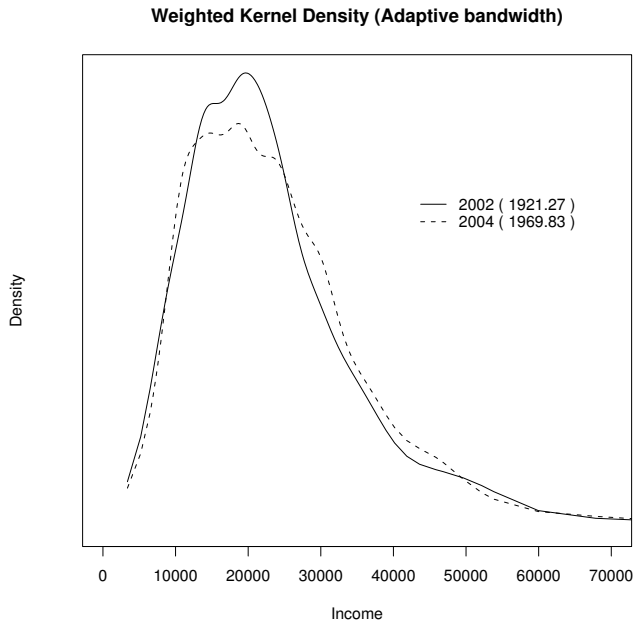
Further insights can be gained employing the relative density function to directly compare the two PDFs. In fact, it is possible to state whether the lower and the upper tails of the distribution are growing at the same rate, and if the difference in the growing rate is due to changes in location and/or in shape. Panel (b) displays the relative distribution, between 2004 and 2002, of the total household income. It confirms a clear shrink of the middle mass of the distribution towards both the tails of the distribution. For instance, in correspondence of the median, the value of the relative distribution is about 0.8, then approximately in 2004 there are 20% of households less than in 2002 that earn an amount of income equal to the 2002 median. In fact, the share of households with total income in 2004 between the 20^{th} and the 64^{th} percentile (of the 2002 distribution) decreased, while the mass at the lower tail (between the 9^{th} and the 19^{th} percentile) and the mass at the upper tail (between the 65^{th} and the 95^{th} percentile) increased. At the 80^{th} percentile, the relative distribution is roughly equal to 1.2, hence households are more likely to receive this income level in 2004 than in 2002. At the extreme ends, instead, the relative density is below 1, indicating than in 2004 there are less households at the extreme values of the income scale than in 2002. This figure highlights an upsurge of the income polarization which has taken place during this period. The share of households at both ends of the distribution rose from 2002 to 2004, while the share of households in the middle deciles fell. In addition, there is a shift of the poorer households (those in the bottom decile) towards the second decile of the income distribution. Taking the second decile[13] as a relative poverty line, in panel (b) the rise in the second decile is nearly compensated by the decrease in the first decile.

Reported in panel (d) there is the relative density computed on income net of imputed rents. A decline of the middle-income mass of the distribution is still evident, but, the growth at the bottom percentiles is more pronounced than the one observed at the other end of the income distribution. The peak of the relative distribution is at around the 10^{th} percentile. Households in 2004 are 20% more likely to fall at that level of 2002 income net of imputed rents, than households in 2002. When total income is taken into account, instead, the increase from 2002 to 2004 of probability to fall at the 10^{th} percentile is no more than 10%.

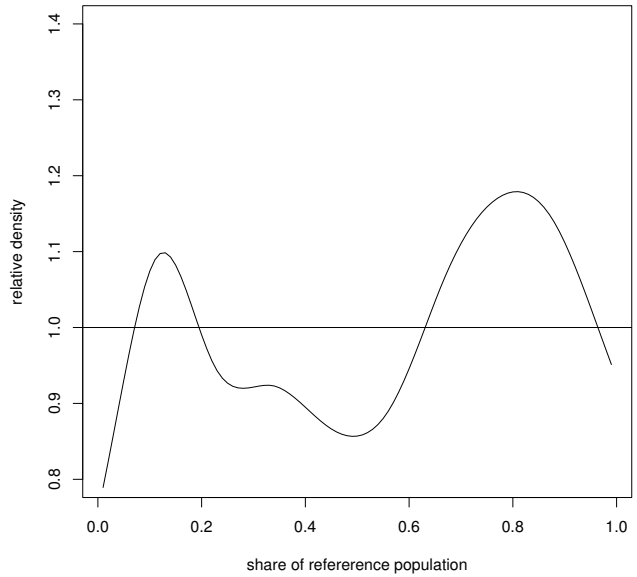
As observed before, it is possible to decompose the overall effect into two sources: the location shift and the shape effects. These two effects are reported in Figure 2. The relative density reported in panel (a) represents the effect only due to the median shift. In other words, the pattern that the relative density

[13] Deciles are computed on the 2002 income distribution.

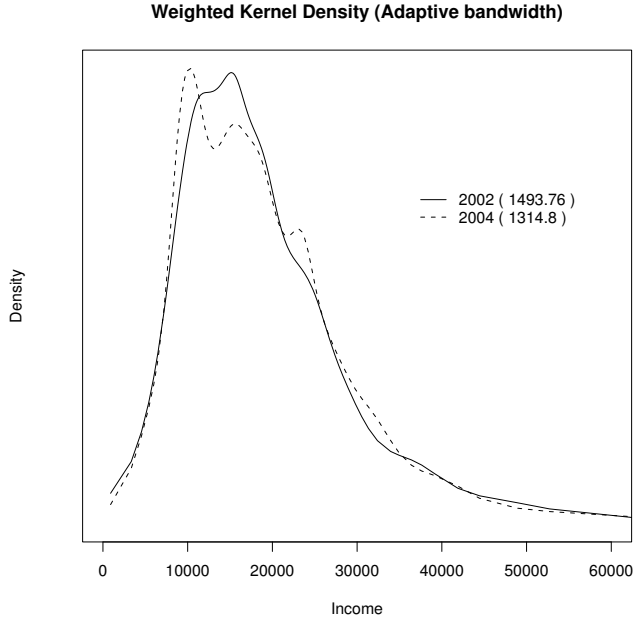
Figure 1: **Comparison between 2002 and 2004 income distributions**



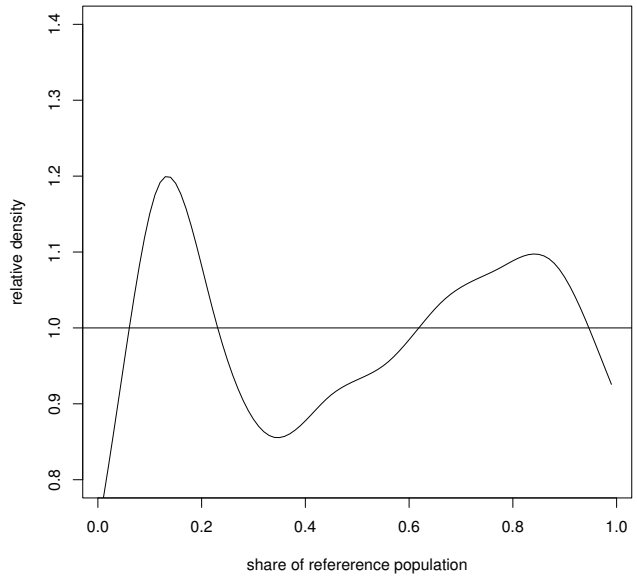
(a) *Total household income*



(b) *Total household income*



(c) *Household income net of imputed rents*



(d) *Household income net of imputed rents*

Note: authors' calculation on weighted household income data from 2002-04 panel subset of SHIW. Income data are household size-adjusted and are expressed in 2002 prices. Panel (a) and (c): The bandwidths for the estimate of 2002 and 2004 kernel density functions are obtained with the Sheather-Jones criterion. Their values are reported in brackets.

would have displayed if there had been no change in distributional shape but only a shift of location of the density. The location effect reduces the share of households in bottom percentiles and augments those in higher percentiles. For total income, the effects of the median shift are large.

For instance, at the 10th percentile the relative density is about 0.9, indicating that if only the location shift would have affected the change of income distributions, 10% less households in 2004 would have gained the same income at that percentile rank than in 2004. Nonetheless, it should be noted that only the lower tail of the relative distribution in panel (b) of Figure 1 is rather well reproduced by the location shift. The bottom decile of the observed relative distribution fall below 0.8 as the lower decile of the relative distribution which represents the median shift, while the upper decile of the former is about 0.95, while for the latter relative distribution is about 1.1. The same can be said when income is net of imputed rents, even if the effect is much less relevant (panel (c)). Panel (b) and (d) show the effect of the redistribution across households, net of location shift. Given what has been just observed, not surprisingly when only the shape effect operates there is an upsurge of relative density at bottom percentiles and a relatively small decrease at the top income, with respect to the observed relative distributions. Overall, both figures display, once again, a diminishing weight of the median deciles income, and an increase of the share of households at both end of the distribution, i.e. an increase of polarization of household income.

4.1 A closer look on polarization

Even if a graphical analysis of the relative distributions shows whether polarization occurs, one would also be interested to quantify the observed patterns. Handcock and Morris (1998) propose a scale-invariant measure of polarization that isolates shape changes from level changes in the relative distribution, the median polarization index. This polarization index is computed on the relative distribution, and can be decomposed to examine and to contrast the growth in the upper and lower tails. The lower and upper polarization indices represent the contributions to the overall polarization made by observations above and below the median of the relative distribution, respectively[14]. To isolate differences in distributional shape it is necessary to remove differences in location. Hence the polarization index is computed on the location matched relative data $R_L = F_0(Y - \rho)$, where ρ is the difference between the two medians.

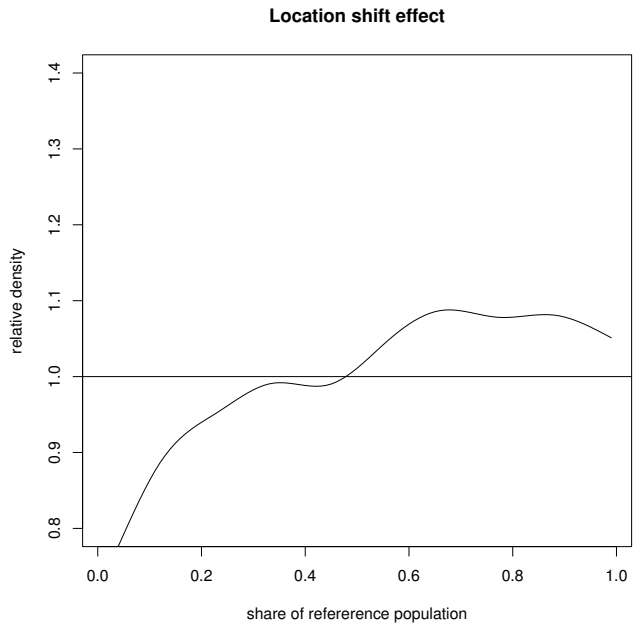
The median relative polarization index of Y respect to Y_0 is defined as:

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

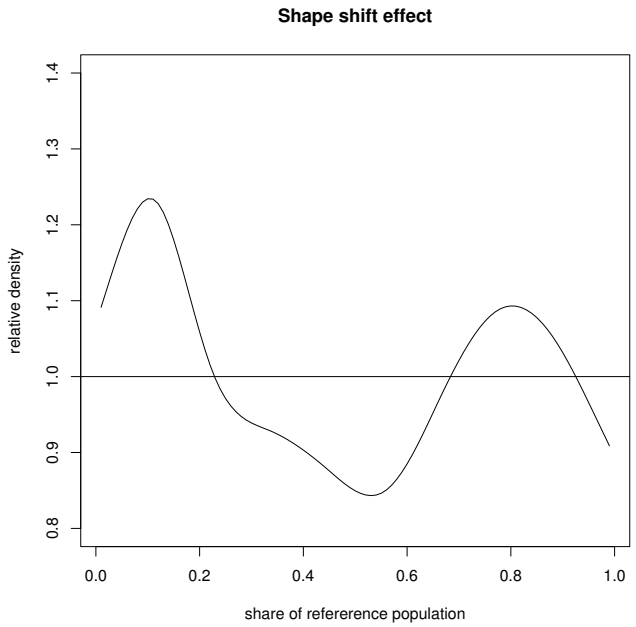
The median relative polarization index is the mean absolute deviation from the median of the location-matched relative distribution g_S, and it is re-scaled in order to have an index that varies between -1 and 1. Positive values represent an upsurge of the income polarization, negative values indicate a convergence of incomes towards the median, and zero represents no differences in distributional shape.

[14]In contrast to other well-known polarization measures, the relative polarization index directly measures the change of polarization between two distributions. Moreover, it allows one to investigate on selected sub-sections of the distribution.

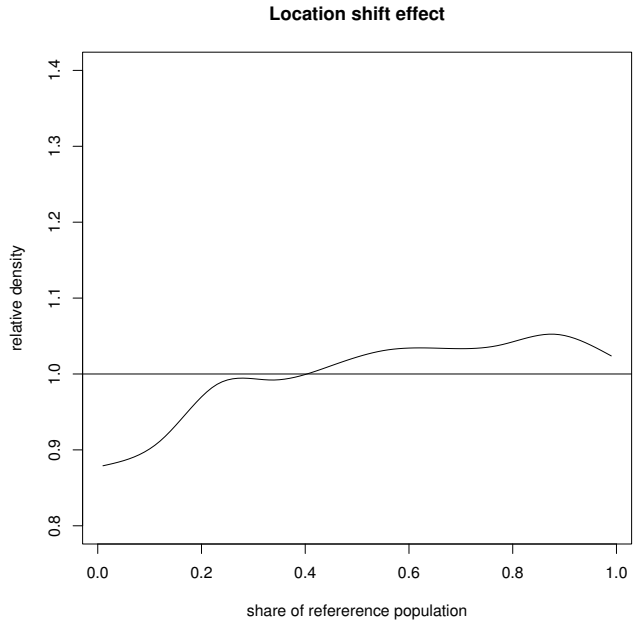
Figure 2: **Location shift and shape effects**



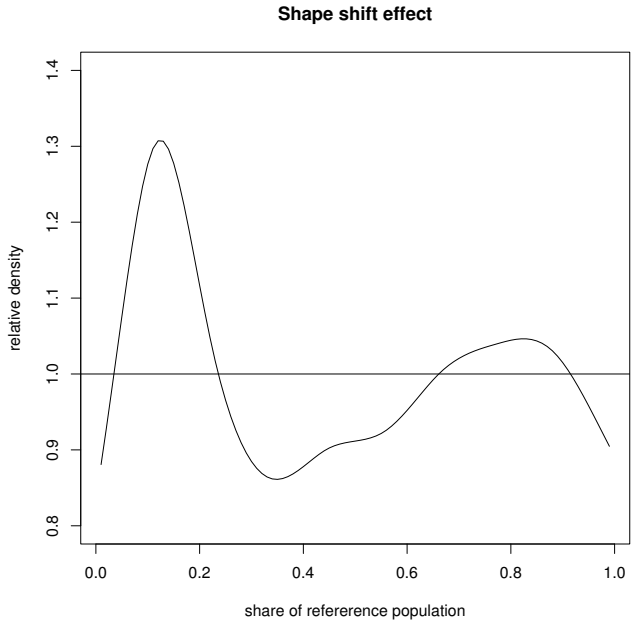
(a) *Total household income*



(b) *Total household income*



(c) *Household income net of imputed rents*



(d) *Household income net of imputed rents*

Note: authors' calculation on weighted household income data from 2002-04 panel subset of SHIW. Income data are household size-adjusted and are expressed in 2002 prices. 2002 data are adjusted to have the same mean of 2004 data, using an additive decomposition. Panel (c) and (d): household income is net of imputed rents

The estimator of the median relative polarization index can be expressed as:

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

where $\hat{R}_j = F_{n0}(Y_j - \rho)$ are the estimate of the location-matched relative data[15]. Under regularity condition, $\widehat{MRP}(F, F_0)$ is an asymptotically unbiased estimator for $MRP(F, F_0)$, and it is asymptotically normally distributed. Hence it is possible to test whether the polarization index is zero, against a complementary hypothesis.

The lower and upper polarization indices are defined respectively by:

$$\begin{aligned} LRP(F, F_0) &= 8 \int_0^{1/2} \left| r - \frac{1}{2} \right| g_S(r) dr - 1 \\ URP(F, F_0) &= 8 \int_{1/2}^{1} \left| r - \frac{1}{2} \right| g_S(r) dr - 1. \end{aligned}$$

These indices have properties similar to the median relative polarization index. They vary between -1 and 1, with similar interpretations. Then it is possible to see whether an income downgrading is prevalent with respect to an income upgrading. The natural estimator of these two indices are, respectively:

$$\begin{aligned} \widehat{LRP}(F, F_0) &= \frac{8}{m} \sum_{j=1}^{m} \left| \hat{R}_j - \frac{1}{2} \right| \mathcal{I}\left(\hat{R}_j \leq \frac{1}{2} \right) - 1 \\ \widehat{URP}(F, F_0) &= \frac{8}{m} \sum_{j=1}^{m} \left| \hat{R}_j - \frac{1}{2} \right| \mathcal{I}\left(\hat{R}_j > \frac{1}{2} \right) - 1 \end{aligned}$$

where $\mathcal{I}(x)$ is the indicator function, which is equal to 1 when the condition x is verified, and 0 elsewhere[16].

Table 3 reports the three polarization indices (second column) for the location-matched relative distribution, computed for both the definition of household income considered[17]. All indices are positive, hence denoting an increasing of polarization from 2002 to 2004. However, the upper relative polarization index is not significant, both for total household income and for income net of imputed rents. Therefore, during the observed period there was a significant downgrading of household incomes, which dominates the upgrading.

[15] In this work, contrary to the procedure followed by Handcock (2005), which is based on the weights of the sample unites, these quantities are estimated with a kernel-type estimate of F_0, namely:

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

where $\mathbf{M}$ is the cumulative distribution function of the kernel M, and h_0 is the bandwidth used to estimate F_0 (Molanes-López and Cao, forthcoming).

[16] Under the same conditions needed to derive the statistical properties of $\widehat{MRP}(F, F_0)$, $\widehat{LRP}(F, F_0)$ and $\widehat{URP}(F, F_0)$ are asymptotically unbiased estimators of $LRP(F, F_0)$ and $URP(F, F_0)$, respectively, and are jointly asymptotically distributed.

[17] To compute the location-adjusted distribution we have used the additive decomposition, as explained before. When the multiplicative decomposition is used, the relative polarization is less evident, but the sign of the indices does not change.

Table 3: Polarization indices

[a] Total household income

Index	Value	95% Confidence interval	p-value
Median relative polarization index	0.0370	[0.0105; 0.0635]	0.00
Lower relative polarization index	0.0616	[0.0248; 0.0983]	0.00
Upper relative polarization index	0.0124	[-0.0258; 0.0505]	0.26

[b] Household income net of imputed rents

Index	Value	95% Confidence interval	p-value
Median relative polarization index	0.0240	[-0.0025; 0.0506]	0.04
Lower relative polarization index	0.0465	[0.0095; 0.0835]	0.01
Upper relative polarization index	0.0016	[-0.0365; 0.0396]	0.47

Note: authors' calculation on weighted household income data from 2002-04 panel subset of SHIW. Income data are household size-adjusted and are expressed in 2002 prices. The p-value refers to the null hypothesis that there is no change in polarization.

4.2 Decomposition for covariates

In order to investigate the determinants of changes occurred, it is possible to take into account the impact of covariates on household income. Since there are multiple income sources and multiple individuals per household, it could be of some interest to analyze the effect of each income source on the household income. When dealing with households as units of analysis, it is not convenient, and also rather difficult, to analyze separately the different income typologies, while much information could be drawn by studying household characteristics. Thus, the sample has been broken into three sub-groups according to the employment status of the household head: employee, self-employed, retired from work or unemployed. Households whose head is a contingent worker are included in the employee sub-group.

This classification is used to study the horizontal re-distribution that it may have occurred between households in different socio-economic groups, as suggested by Boeri and Brandolini (2004)[18]. The composition of the three sub-groups, in both samples, is reported in Table 4.

Table 4: Composition of sample by household head employment status

	Head employee	Head self employed	Head retired or unemployed	All households
2002	45.9	13.8	40.3	100
2004	45.5	13.3	41.2	100

Note: authors' calculation on weighted household income data from 2002-04 panel subset of SHIW.

The subdivision of households into these three sub-groups is consistent with the main income source of households (see Table 5). For instance, households

[18]There are many other possible decompositions. However, previous results quoted in Boeri and Brandolini (2004) underline 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 which the head is an employee earn a great deal of their resources from wages. Therefore, changes in the wages distribution are likely to affect this sub-sample more than the other two. Shifts observed for these households could be taken as indicative of changes in the wage distribution. The same could be said for respectively for the remaining sub-samples, respectively self-employment incomes and transfers.

Table 5: Household income source per household head employment status (in percentage of total household income)

	[a] 2002		
	Employee	Self employed	Retired or unemployed
Wages	70.8	10.8	8.1
Transfers	8.1	6.5	61.7
Self employment	2.4	62.6	2.9
Assets revenue*	1.9	3.6	4.9
Imputed rents	16.9	16.6	22.4

	[b] 2004		
	Employee	Self employed	Retired or unemployed
Wages	71.5	10.9	7.4
Transfers	8.2	6.7	64.1
Self employment	1.8	63.4	1.6
Assets revenue*	1.0	2.7	3.7
Imputed rents	17.4	16.4	23.2

Note: authors' calculation on weighted household income data from 2002-04 panel subset of SHIW. Income data are expressed in 2002 prices.
* Net of imputed rents.

In Table 6 some summary statistics are reported for households classified according to the employment status of the head. At a first glance, it can be noted that, as expected, incomes of households whose head is self employed are far higher, on average, than incomes of all households. This holds true when median income is considered. As to the whole sample, income shares of households in bottom 5% and 10% arise a little, but considering all households in the bottom quintile, the income share reduces. Income share of the top earners augment of a little amount. Inequality and polarization indices remains practically unchanged. With regards to households with head employee, income shares decrease both for low-income and high-income classes. Considering households with head self employed, poorest households receive a smaller fraction of income in 2004 with respect to 2002. On the contrary, richer households, especially those in the top decile and in the top 5%, earn a higher fraction of the total income. Consequently, inequality indices considerably rise. It is worthwhile noting that the income distribution of these households is the most unequal and polarized amongst the other. Conversely, households with head retired or unemployed in the bottom tail of income distribution increase their income shares, while those in the top tail receive a lesser part of the total income in 2004. Naturally, inequality and polarization indices show a small reduction.

Table 6: Summary measures for household incomes, by employment status of household head (total household income)

[a] 2002				
	All households	Employee	Self employment	Retired or unemployed
Mean (2002 euros)	24,621	24,104	30,472	23,213
Median (2002 euros)	21,277	21,560	23,576	20,061
Income shares				
Bottom 5 %	1.06	1.30	0.91	0.90
Bottom 10 %	2.85	3.23	2.33	2.74
Bottom 20%	7.60	8.31	5.68	7.84
Top 20%	39.72	37.25	46.01	39.00
Top 10%	24.91	22.49	30.88	23.92
Top 5%	15.45	13.44	20.78	14.32
Inequality measures				
Gini Index	0.31	0.28	0.38	0.31
Theil Index	0.19	0.14	0.33	0.17
Quintile ratio	5.23	4.48	8.10	4.97
Polarization measures				
Wolfson Index	0.26	0.24	0.35	0.25
Esteban-Ray (α=1.3)	0.17	0.16	0.21	0.17

[b] 2004				
	All households	Employee	Self employment	Retired or unemployed
Mean (2002 euros)	25,578	24,378	36,737	23,286
Median (2002 euros)	22,168	22,983	28,032	20,229
Income shares				
Bottom 5%	1.10	1.19	0.73	1.22
Bottom 10%	2.94	3.06	2.04	3.34
Bottom 20%	16.10	12.21	24.53	14.30
Top 20%	39.90	36.14	46.71	38.06
Top 10%	25.24	20.98	33.82	23.37
Top 5%	16.10	12.21	24.53	14.30
Inequality measures				
Gini Index	0.32	0.28	0.43	0.29
Theil Index	0.20	0.13	0.38	0.15
Quintile ratio	5.30	4.64	8.85	4.50
Polarization measures				
Wolfson Index	0.27	0.25	0.33	0.27
Esteban-Ray (α=1.3)	0.18	0.16	0.22	0.16

Note: authors' calculation on weighted household income data from 2002-04 panel subset of SHIW. Income data are household size-adjusted and are expressed in 2002 prices.

In table 7 are reported the same summary measures as in the previous table, computed on household income net of imputed rents. Considering all the households in the sample, differences in inequality indices are negligible. The same can be said with regards to households with head employee. Thus, irrespective of the income definition used, figures in Table 6 and 7 show that inequality remains unchanged in the whole sample during the biennial 2002-2004, and the same occurs for households with head employee. But looking at income shares at the bottom and at the top of income distribution, these results are due to different behavior of incomes: in the first case there is an admittedly small rise of the fraction of income attributed to households at both tail of the distribution, while in the latter case a reduction of the share of smaller and higher occurs. In relation to households with head self employed, once again during 2002-2004 is observed a decrease of income shares of poorer households, and an upsurge of fraction of total income achieved by richer households, with the difference that in this case the augment is more relevant. Both in 2002 and in 2004, income distribution of these households is more unequal when income net of imputed rents is employed. Finally, what has been observed for households with head retired or unemployed in Table 6 can be said in this case.

Displayed in Figure 3 are the kernel estimates of income distribution density for both income definitions, and the density distribution functions of the three household typologies[19]. Irrespective of the income definition, the overall PDF seems to be affected mainly by the PDF's of households with head employee and those with head retired or unemployed. For households with head employee (dashed lines), the PDF shifts from a bi-modal distribution in 2002 to an approximate tri-modal distribution in 2004. The three modes of 2004 distribution are more evident when income net of imputed rents is taken into consideration. In addition, the modes are more spread in 2004 than in 2002, that could be interpreted as an upsurge of inequality, which is not captured by summary measures (see Table 6). Finally, as to income net of imputed rents, the three modes observed in 2004 are approximately located at the same level of the modes of the overall PDF[20]. With regards to households with head retired or unemployed (dashed and dotted lines), in 2004 there is a rise of households with low income. Comparing households with head employee with those with head retired or unemployed, the latter distribution appears to be more left-shifted, in both year. Broadly, modes of overall density functions seem to correspond to the peaks of the distributions of households with head employee and with head retired or unemployed.

In a fashion similar to that observed for the decomposition of location shift and shape shift effects, it is possible to distinguish between a compositional effect, due to a change in the population shares, and a residual effect, which is due to the change in the relationship between the covariate (employment status of household head) and the response variable (household income). Clearly, the first effect is negligible, since the same households are observed over a time span of only two years[21]. Then, the behavior of the relative PDF, observed before, is due to the fact that the conditional distribution of household income

[19]The PDF's of the household sub-groups have been re-weighted with the share of households in each sub-group.

[20]To state whether these three modes are significant goes beyond the scope of this paper.

[21]Results are not reported here for the sake of brevity, but they are available from the authors.

Table 7: Summary measures for household incomes, by employment status of household head (household income net of imputed rents)

[a] 2002

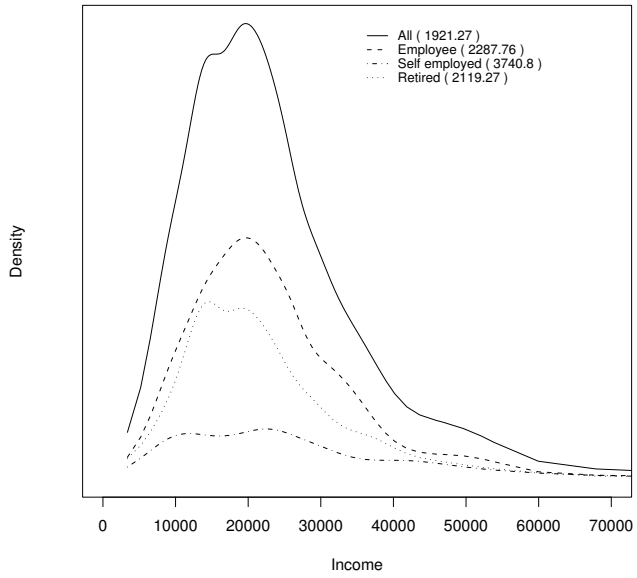
	All households	Employee	Self employment	Retired or unemployed
Mean (2002 euros)	19,859	19,995	25,207	17.878
Median (2002 euros)	17,060	18,029	19,462	15,578
Income shares				
Bottom 5%	0.90	1.32	0.63	0.54
Bottom 10%	2.69	3.32	1.77	2.45
Bottom 20%	7.61	8.45	5.25	7.67
Top 20%	39.66	36.40	47.27	38.88
Top 10%	24.89	21.93	31.66	23.72
Top 5%	15.35	13.14	21.67	14.25
Inequality measures				
Gini Index	0.32	0.27	0.40	0.31
Theil Index	0.19	0.14	0.36	0.16
Quintile ratio	5.21	4.31	9.01	5.07
Polarization measures				
Wolfson Index	0.26	0.24	0.36	0.24
Esteban-Ray (α=1.3)	0.17	0.16	0.22	0.17

[b] 2004

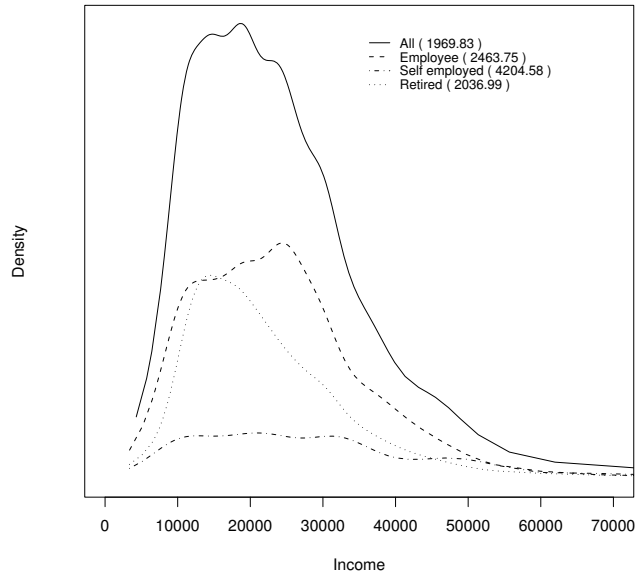
	All households	Employee	Self employment	Retired or unemployed
Mean(2002 euros)	20,404	20,106	30,545	17,444
Median (2002 euros)	17,439	18,936	21,262	15,362
Income shares				
Bottom 5%	0.98	1.14	0.43	1.13
Bottom 10%	2.84	3.13	1.52	3.33
Bottom 20%	7.50	8.01	4.35	9.04
Top 20%	40.28	35.53	50.92	37.39
Top 10%	25.69	20.94	36.02	22.68
Top 5%	16.64	11.84	26.02	13.51
Inequality measures				
Gini Index	0.33	0.28	0.46	0.28
Theil Index	0.21	0.13	0.44	0.14
Quintile ratio	5.37	4.43	11.70	4.13
Polarization measures				
Wolfson Index	0.27	0.25	0.39	0.26
Esteban-Ray (α=1.3)	0.18	0.16	0.23	0.16

Note: authors' calculation on weighted household income data from 2002-04 panel subset of SHIW. Income data are household size-adjusted and are expressed in 2002 prices.

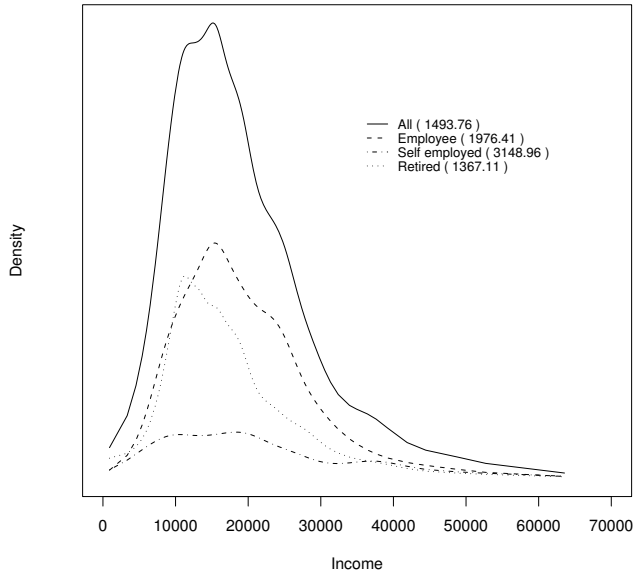
Figure 3: **Comparison between overall PDF and sub-groups PDF's**



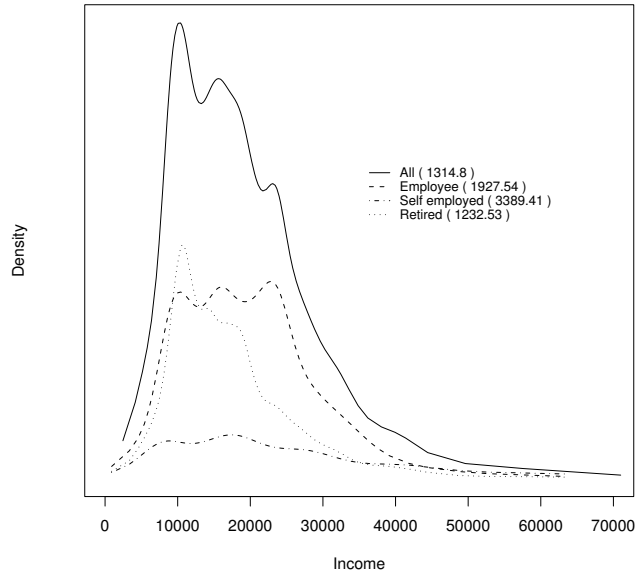
(a) *Total household income*, 2002



(b) *Total household income*, 2004



(c) *Household income net of imputed rents*, 2002



(d) *Household income net of imputed rents*, 2004

Note: authors' calculation on weighted household income data from panel subset of 2002 and 2004 SHIW. Income data are household size-adjusted and are expressed in 2002 prices.

by household head's employment status changed over these two years.

Further insight is provided by an analysis of income polarization, separately for each sub-sample (see Table 8. For households with head employee, irrespective of income definition used, the median polarization index displays an increase of the polarization, due in particular to incomes below the median. For higher incomes, the index shows a decrease of the polarization, which is not significant.

Considering households with head self employed, the growth of the median polarization is significant. This is mainly due to a shift from the median income towards the bottom of the income distribution, when the total income is considered. However, the analysis of income net of imputed rents shows a significant rise of the upper polarization index, while the income downgrading is negligible.

Finally, as to households with head retired or unemployed, the median relative polarization index is negative, denoting a shift of incomes towards the median income, but only in the case of income net of imputed rents this result is statistically significant.

5 Concluding remarks

To provide additional guidance for the interpretation of the modification of income distribution that have occurred in recent years, in this paper we have proposed an alternative approach which take into consideration shifts that affect not only the median, or the variance, but every quantiles of the distribution.

The "relative distribution" approach reveals that there has been a significant location effect, which is more evident for total household incomes rather than for incomes net of imputed rents. At the same time, there has been an increase of income polarization, mainly due to a downgrading of lower incomes. In other words, there is a higher fraction of households which fall below the median income, in particular in the lower classes, in 2004 with respect to 2002. In this sense, if it is not appropriate to talk about a general "household impoverishment", this evidence indicates a clear dwindling of the middle-income class. Similar comments apply when income is measured net of imputed rents.

The decomposition for covariates shows that the compositional effect in covariates is negligible, and that differences observed between 2002 and 2004 are due to the change in the relationship between covariates and household income.

A within group analysis show a substantial stability of income inequality, with the notable exception of the rise of inequality for household with self employed head, and also a little increase of concentration for those whose head is retired or unemployed.

The analysis of relative polarization for each household typology points out a significant increase of the polarization for both households with head employee and with head self employed, when total household income is considered. For both these household typologies the increase of polarization is due to a downgrading of incomes. At the same time, there is a movement of incomes of households with head retired or unemployed towards the median, which nonetheless is not statistically significant.

When incomes net of imputed rents are analysed, the reduction of polarization observed for the latter household typology became significant. Moreover, the increase of the polarization observed for the former household sub-groups

Table 8: Polarization indices per household head employment status

Household head employee

[a] Total household income

Index	Value	95% Confidence interval	p-value
Median relative polarization index	0.0565	[0.0150; 0.0980]	0.00
Lower relative polarization index	0.1363	[0.0795; 0.1930]	0.00
Upper relative polarization index	-0.0232	[-0.0830 0.0365]	0.22

[b] Household income net of imputed rents

Index	Value	95% Confidence interval	p-value
Median relative polarization index	0.0475	[0.0064; 0.0887]	0.01
Lower relative polarization index	0.1187	[0.0618; 0.1755]	0.00
Upper relative polarization index	-0.0236	[-0.0824; 0.0352]	0.22

Household head self employed

[a] Total household income

Index	Value	95% Confidence interval	p-value
Median relative polarization index	0.1103	[0.0353; 0.1854]	0.00
Lower relative polarization index	0.1909	[0.0907; 0.2910]	0.00
Upper relative polarization index	0.0298	[-0.0808; 0.1405]	0.30

[b] Household income net of imputed rents

Index	Value	95% Confidence interval	p-value
Median relative polarization index	0.0745	[-0.0006; 0.1496]	0.03
Lower relative polarization index	0.0352	[-0.0721; 0.1424]	0.26
Upper relative polarization index	0.1138	[0.0089; 0.2187]	0.02

Household head retired or unemployed

[a] Total household income

Index	Value	95% Confidence interval	p-value
Median relative polarization index	-0.0245	[-0.0633; 0.0143]	0.11
Lower relative polarization index	-0.0238	[-0.0782; 0.0305]	0.20
Upper relative polarization index	-0.0251	[-0.0805; 0.0303]	0.19

[b] Household income net of imputed rents

Index	Value	95% Confidence interval	p-value
Median relative polarization index	-0.0395	[-0.0783; -0.0007]	0.02
Lower relative polarization index	-0.0397	[-0.0939; 0.0145]	0.08
Upper relative polarization index	-0.0394	[-0.0948; 0.0161]	0.08

Note: authors' calculation on weighted household income data from 2002-04 panel subset of SHIW. Income data are household size-adjusted and are expressed in 2002 prices. The p-value refers to the null hypothesis that there is no differences in income distribution shape.

remains significant. Nevertheless, while for households with head employee the driving force behind the increase polarization is still a downgrading, for households with head self employed the main reason is an upgrading of incomes, that is the lower tail of the distribution has been stable while the upper tail of the distribution has grown away from the median.

6 Acknowledgment

This draft benefited from helpful comments of Paul Johnson, Andrea Brandolini, participants of CEIS seminar and DSE-ISFOL workshop. Obviously, responsibility for remaining errors is ours.

References

[1] Baldini M. (2005) "Le molte conferme sulla povertà in Italia", www.lavoce.info, 13 October.

[2] Banca d'Italia (2006), Italian Household Budgets in 2004, Supplements to the Statistical Bulletin, Banca d'Italia, Vol. XVI, No. 17.

[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*, 63, No. 3/4, pp. 449-487.

[4] Bernhardt A., Morris M., Handcock M.S. (1995), "Women's Gains or Men's Losses? A Closer Look at the Shrinking Gender Gap in Earnings", *The American Journal of Sociology*, Vol.101, No. 2, pp. 302-328.

[5] 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, No. 2, pp. 183–239.

[6] Cowell F.A. (1977), *Measuring Inequality*, Oxford, Philip Allan.

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

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

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

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

[11] Istat (2006), *Rapporto Annuale. La situazione del Paese nel 2005*, Rome.

[12] Jenkins S.P., Van Kerm P. (2005), "Accounting for Income Distribution Trends: A Density Function Decomposition Approach", *Journal of Economic Inequality*, 3, pp. 43-61.

[13] Lambert P.J., Aronson J.R. (1993). "Inequality Decomposition Analysis and the Gini Coefficient Revisited", *Economic Journal*, vol. 103, pp. 1221-1227.

[14] Molanes-López E.M., Cao R. (forthcoming), "Plug-in bandwidth selectors for the kernel relative density estimator", *Annals of the Institute of Statistical Mathematics*.

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

[16] Paiella M. (2004), "Does Wealth Affect Consumption? Evidence for Italy", Banca d'Italia, *Temi di discussione* No. 510, Roma.

[17] 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.

[18] Pittau M.G., Zelli R. (2006), "Empirical Evidence of Income Dynamics Across EU Regions", *Journal of Applied Econometrics*, 21, pp. 605-628.

[19] 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.

[20] Trivellato U. (1998), "Il monitoraggio della povertà e della sua dinamica: questioni di misura ed evidenze empiriche", *Statistica*, vol. 58, No. 4, pp. 549-574.

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

[22] Wood S.N. (2000), "Modelling and smoothing parameter estimation with multiple quadratic penalties", *Journal of the Royal Statistical Society (Series B)*, 62(2), pp. 413-428.