4 THE PARETO LAW AND THE DISTRIBUTION OF LABOUR INCOME IN ITALY*

This chapter provides an empirical examination of the distribution of labour earnings in Italy. Using data drawn from the 2005 and 2006 waves of the PLUS (See chapter 9) to analyzed the shape of the observed distributions to be right-skewed and display a long right tail, which is adequately described by a Pareto-type model. This chapter also address the question of earnings dispersion by applying a nested decomposition procedure of the Theil inequality measure, which combines into a unified framework the standard decompositions by population subgroups and income sources. The empirical evidence obtained points to the key role played by the self-employees in shaping labour income inequality, especially at the upper extreme of the earnings distribution, and the emergence of non-standard forms of employment as an important feature of the contemporary workplace. The structure of the chapter is: after a brief introduction, data and methodological decisions are discussed; then it presents the empirical results obtained in the above analysis; finally, some concluding remarks and policy implications are drawn.

4.1 Introduction[1]

Theories of labour earnings distributions have always had strong empirical motivations. The earliest empirical studies of earnings distribution (e.g. Lydall, 1968, and Harrison, 1981) discovered remarkable regularities that are found in all observed distributions in large populations. In particular, earnings distributions (and income distributions more generally) tend to be skewed to the right and display long right tails. Furthermore, mean earnings generally exceed median ones, and the top percentiles of their distribution account for a significant share of the total.

---

* Fabio Clementi and Michele Giammatteo.

1. We acknowledge the Italian Institute for the Development of Vocational Training for Workers (ISFOL) for providing us with data from the 2005 and 2006 waves of the Participation Labour Unemployment Survey (PLUS). Microdata use authorization code ISFOL PLUS 2006/428. The PLUS data are available at no cost by sending a request e-mail to plus@isfol.it. Usual disclaimer applies.

The assembly and empirical analysis of data on personal incomes were pioneered by Pareto (1897), who apparently was responsible for the first attempt at defining a general "law" that tried to explain the regularities of observed distributions. Let $\bar{F}(x) = \Pr(X \geq x)$ be the complementary cumulative distribution of X denoting the percentage of individuals with incomes greater than or equal to x. Then, the (strong) Pareto law asserts that

(1) $$\bar{F}(x) = \begin{cases} (x/x_{\min})^{-\alpha} & \text{when} \quad x_{\min} \leq x < \infty \\ 1 & \text{when} \quad x < x_{\min} \end{cases}, \quad x_{\min}, \alpha > 0$$

$x_{\min}$ being the minimum possible value of X.

Later, extensive development of microeconomic data sources led to a more comprehensive treatment of the subject. Indeed, available empirical work leaves little doubt that the Pareto law, as it stands, does not account satisfactorily for a wide range of incomes. Subsequently, the use of other density functions to model the income distribution, such as the lognormal (Aitchison and Brown, 1954, 1957) or gamma (Salem and Mount, 1974), has been advocated. However, rapidly accruing evidence showed that the lognormal and gamma distributions fit the data relatively well in the middle range of income, but tend to exaggerate the skewness and perform poorly in the upper end (McDonald and Ransom, 1979; McDonald, 1984). Furthermore, if one's attention is restricted to the upper tail of the distributions, the evidence does not contradict the Pareto law, provided that the chosen $x_{\min}$ is large enough. This suggests that observed distributions obey a weak version of the Pareto law (Mandelbrot, 1960), e.g.

(2) $$\lim_{x \to \infty} \frac{\bar{F}(x)}{(x/x_{\min})^{-\alpha}} = 1,$$

and some well-known density functions that have been proposed and implemented in the literature asymptotically approach (rather than coincide with) the Pareto distribution. Among these, the Singh-Maddala (1976) and Dagum (1977) distributions have shown them to be a good compromise between parsimony and goodness-of-fit in many instances.

This chapter tackles the issue of the shape of the labour earnings distribution in Italy. Using data drawn from the 2005 and 2006 waves of the Participation Labour Unemployment Survey (PLUS), a sample survey on the Italian labour market supply carried out by ISFOL[2], we find that the Italian labour income distribution in any one year is highly

2. The Italian Institute for the Development of Vocational Training for Workers (ISFOL) is a research institute connected to the Italian Ministry of Labour and Social Affairs and member of the Italian National Statistical

skewed to the right, with an upper tail very well described by the Pareto distribution[3]. In particular, the pattern noted in the analysis of the top of the distribution reveals that inequality amongst the rich decreased between 2005 and 2006, with the Pareto coefficient rising from 1.96 to 2.22. This finding adds support to the story told by summary inequality indices that document a slight decline in overall inequality from one year to the other, and reveals the effect of the shape of earnings distribution on the inequality of labour market outcomes. Therefore, in order to understand the detail behind this change, we also consider how much of the dispersion in earnings concentrated in different parts of the distribution might be accounted for by alternative sources of labour income. To this end, a nested decomposition of the Theil inequality measure by population subgroups and income sources is performed as proposed by Giammatteo (2007), and the results seem to corroborate recent findings pointing to a deterioration in the Italian labour market situation due to the widening gap between the incomes of employees and self-employees and the increased job precariousness[4]. Indeed, in both the years examined we observe a significant positive contribution to the between-group component of inequality arising from self-employment income, which results to be more highly concentrated (and thus responsible for the inequality level) in the upper end of the distribution, as opposed to the inequality-decreasing effect in terms of between-group differences exerted by income from standard employment, which instead appears more concentrated in the bulk. Earnings from atypical employment, in turn, have seen their share of both the total population and income increases from one year to the other, thus arising as an important feature of the contemporary workplace. The rest of this work is organized as follows: Section 4.2 describes the data used in the study and outlines the approach to estimating the upper tail of the observed distributions and the implied amount of inequality. Section 4.3 presents the results of the analysis. Section 4.4 summarizes the main results and draws some implications for policy.

System (SISTAN). The PLUS survey is included in the Italian National Statistical Programme (NSP), the SISTAN tool for planning statistical activity of public interest. For a collection of various research results on the Italian labour market conducted by ISFOL using this dataset, see Mandrone and Radicchia (2006).

3. There exists a number of earlier theories that are directly relevant to the functional description of the upper tail of earnings distribution through the Pareto model. One such set of theories is that dealing with executive remuneration in a hierarchical structure (Simon, 1957; Lydall, 1959). For a comprehensive review of the main theoretical approaches addressing the stylized facts of the distribution of earnings, see e.g. Neal and Rosen (2000).

4. As far as the impact of the increase in precarious employment on the Italian labour market distributional outcomes is concerned, Boeri and Brandolini (2004) claim that «here the reasoning is tentative». Indeed, despite the trend increase in non-standard forms of employment and its influence on rising earnings inequality have been acknowledged from a number of points of view, there has until now been only limited empirical evidence on the subject (see e.g. Sciulli, 2006, and Lucidi and Raitano, 2009). In this respect, Rani (2008), when looking at how changing employment patterns might explain the rise in inequality observed in the majority of countries over the past two decades, recognizes that «it is difficult to empirically establish the relationship between widening inequality and wage differentials between standard and non-standard work due to lack of data». The special emphasis given in the PLUS survey to the investigation of atypical contracts could play a major role in filling this gap, at least for the case of Italy.

4.2 Data and Methodology

The analysis of the distribution of labour earnings in Italy is based on data obtained from the PLUS. The database is described in Chapter 9 Complementary to other key national statistical sources[5], the core objective of the PLUS is that of providing reliable estimates of rare and only marginally explored labour market issues, such as the distribution of contract types (employee/self-employed status and their articulated subclassifications), job search activity, young and women employment participation, old-age activity and retirement choice, pattern of education and other training, intergenerational dynamics, etc. Therefore, despite its limited time span, this dataset may be useful to pin down the role that alternative sources of labour earnings play as determinants of income distribution and inequality among workers, particularly for the special emphasis given to the investigation of atypical contracts.

The income concept used is the monthly "gross income" normalized on annual basis earned by workers classified according to the following categories: *standard* full-time workers with open-ended contracts, *self-employed* and *atypical* workers - this latter category including workers with fixed-term contracts and other non-standard jobs. This variable is in current year euros (€), and we use the consumer price index for the whole nation (NIC) based on the year 1995 in order to obtain distributions of "real" income[6]. Furthermore, because of the complex sampling design of the PLUS survey, data make use throughout the analysis of appropriate sampling weights to produce representative estimates and correct standard errors and statistical tests[7]. A set of basic statistics calculated from these data is given in Table 4.1.

5. In Italy, information on labour market characteristics can be obtained from various sources. Two prominent examples are the Labour Force Survey (LFS, http://www.istat.it/en/archive/36394), conducted quarterly by the National Institute for Statistics (ISTAT), and the Work Histories Italian Panel (WHIP, http://www.laboratoriorevelli.it/whip/whip_datahouse.php?lingua=eng&pagina=home), built from a sample of microdata from the administrative archives of the National Institute of Social Security (INPS). However, while the former considers the household as sampling unit, the latter includes microdata on private sector employees only.

6. The series of the NIC index is publicly available on the website of the Italian National Statistical Institute (ISTAT) at the address http://www.istat.it/prezzi/precon/dati/ indici_nazionali_1_nic.xls.

7. The expansion weights coming with the PLUS survey are calibrated using GREG estimation (Deville and Särndal, 1992), which guarantees reduction of sample selection bias, small estimation variance, and large consistency with the standard labour market indicators derivable from the Italian Labour Force Survey (RCFL) conducted by ISTAT.

Table 4.1 - Sample statistics, 2005 and 2006

Wave	Statistics	Gross inc.	Standard	Self-employed	Atypical
2005	Obs	15,868	10,777	2,506	2,585
	Pop. ('000)	21,570	13,975	4,816	2,779
	Min	472	1,727	472	944
	p25	11,802	12,619	9,441	7,553
	Med.	14,612	14,612	17,309	11,955
	p75	18,597	18,464	31,471	14,612
	Max	236,035	158,585	236,035	141,621
	Mean	17,967	16,118	26,626	12,258
	St. dev.	16,786	7,981	30,595	8,071
	Skewness	5.97	5.38	3.29	4.00
	Kurtosis	55.87	65.35	17.58	37.17
2006	Obs	16,475	10,185	2,380	3,910
	Pop. ('000)	22,619	14,254	4,282	4,083
	Min	231	715	231	715
	p25	11,094	13,087	9,245	6,504
	Med.	14,458	15,144	16,949	11,094
	p75	18,574	18,574	30,817	13,867
	Max	288,906	184,413	288,906	96,559
	Mean	17,182	16,295	25,719	11,328
	St. dev.	15,195	8,150	28,899	7,630
	Skewness	6.87	6.09	3.91	4.26
	Kurtosis	82.81	83.83	26.45	39.47

Source: authors'own calculations using the PLUS 2005 and 2006 data

As we have said, it is normally the case that income data, if they follow the Pareto law (1) at all, do so only for values of x above some lower bound x_{min}. Therefore, before calculating the estimate of the shape parameter α, we need first to discard all observa-

tions below this point so that we are left with only those for which the Pareto model is a valid one. Perhaps the most common way of estimating parameters for the classical Pareto distribution is to choose visually the point $x_{\min}$ beyond which the empirical complementary cumulative distribution of the data becomes roughly straight on a doubly logarithmic plot and extract the magnitude of α by least-squares linear regression[8].
Unfortunately, this method and other variations on the same theme, by imposing an arbitrary threshold above which the Pareto relationship is valid, generate significant systematic errors under relatively common conditions and the results they give can not be trusted (Aigner and Goldberger, 1970; Weron, 2001; Goldstein et. al., 2004; Sornette, 2004; Brizio and Montoya, 2005; Clauset et. al., 2007). Therefore, more accurate and robust approaches are desirable. Among these, the maximum likelihood estimator α of introduced by Hill (1975) - which is known to be asymptotically normal (e.g. Hall, 1982) and consistent (Mason, 1982) - does not assume a parametric form for the entire distribution function, but focuses only on the tail behaviour.
That is, if $x_n \geq x_{n-1} \geq \ldots \geq x_{n-m+1} \geq \ldots \geq x_2 \geq x_1$ are the sample elements put in descending order, then the Hill estimator for α based on the $m = n - k + 1$ largest sample values is

(3) $$\alpha_{\mathrm{H}} = \left[\frac{1}{m} \sum_{i=1}^{m-1} \left(\ln x_{n-i+1} - \ln x_{n-m+1} \right) \right]^{-1},$$

where n is the sample size and k is the rank of the order statistic x_{n-m+1}.
The Hill estimator is based on the assumption that m is known. In practice, m is unknown and needs to be estimated. The most common way to choose the value of m is to plot the estimates $\hat{\alpha}_H$ against m, yielding the so-called Hill plot, and look for a region where the plot levels off to identify the optimal number of observations in the upper tail to be used in the estimation of α (e.g. Beirlant et. al., 2004). However, the Hill plot typically is far from being constant, which makes it difficult to use the estimator in practice without further guideline on how to choose the value m. Moreover, the finite-sample properties of the Hill estimator depend crucially on the choice of m, or, equivalently, the estimate of the lower bound on Pareto behaviour $\hat{x}_{\min}$: indeed, if we choose too low a value for $x_{\min}$ we will get a biased estimate of the shape parameter, since we will be attempting to fit a Pareto model to non-Pareto data; on the other hand, if we choose too high a value for $x_{\min}$ we are effectively throwing away legitimate data points $x_i < \hat{x}_{\min}$, which increases both the statistical error on the shape parameter and the bias from finite size effects. Thus, if we wish our estimate of to be accurate, we also need an accurate method for estimating $x_{\min}$.

8. For a considerable in-depth discussion on this and other inference procedures for the classical Pareto distribution see e.g. Arnold (1983), Johnson et. al. (1994), Quandt (1966) and Kleiber and Kotz (2003).

Here we adopt a numerical technique for selecting $x_{\min}$ proposed by Clauset et al. (2007b) that is based on minimizing the "distance" between the Pareto model and the empirical data. The fundamental idea behind this method is simple: we choose the estimate of $x_{\min}$ that makes the probability distributions of the measured data and the best-fit Pareto model as similar as possible above $\hat{x}_{\min}$. Specifically, for each $x_{\min}$ we first obtain by using the estimate $\hat{\alpha}_H$ of the shape parameter over the data $x \geq x_{\min}$, and then compute the Kolmogorov-Smirnov (K-S) goodness-of-fit statistic

(4) $$D = \max_{x \geq x_{\min}} \left| \hat{F}(x) - F(x; x_{\min}, \hat{\alpha}_{\mathrm{H}}) \right|$$

between the empirical cumulative distribution of the data points being fit, $\hat{F}(x)$, and the theoretical Pareto cumulative distribution function with parameters $x_{\min}$ and $\hat{\alpha}_H$, e.g. $F(x; x_{\min}, \hat{\alpha}_H)$. Our optimal estimate of the lower bound, $\hat{x}^*_{\min}$, is then the value of $x_{\min}$ where D attains its minimum, from which we infer the optimal sample fraction, m^*, and the optimal estimate of the shape parameter, $\hat{\alpha}^*_H$. Once the parameters have been estimated, by exploiting the asymptotic distribution theory of the Hill estimator we calculate the standard error of the shape parameter as $\frac{\hat{\alpha}^*_H}{\sqrt{m^*}}$ (e.g. Lux, 1996), whereas the uncertainty in the estimate $x_{\min}$ for is derived by making use of a nonparametric bootstrap method (Efron and Tibshirani, 1993). That is, given our n measurements, we generate a synthetic dataset by drawing a new sequence of points, x_i, $i = 1, \ldots, n$, uniformly at random from the original data. Using the method described above, we then estimate $x_{\min}$ for this surrogate dataset. By taking the standard deviation of all the estimates over a large number of repetitions of this process, we can quantify our uncertainty in the original estimated parameter.

Finally, we also perform a K-S goodness-of-fit test of the Pareto distribution for the observations above $\hat{x}^*_{\min}$ by generating a *p*-value that quantifies the plausibility of the hypothesized model[9]. In detail, our procedure is as follows. First, we fit our empirical data to the Pareto model using the method described above and calculate the K-S statistic for this fit. Next, we generate a large number of synthetic datasets having m^* observations randomly drawn from a Pareto distribution with shape parameter α and lower bound $x_{\min}$ equal to those of the distribution that best fits the observed data. We

9. One of the features of the K-S statistic is that its distribution is known for datasets truly drawn from any given distribution. This allows one to write down an explicit expression in the limit of large for the *p*-value. Unfortunately, this expression is only correct so long as the underlying distribution is fixed (see e.g. Stephens, 1986). If, as in our case, the underlying distribution is itself determined by fitting to the data and hence varies from one dataset to the next, we can not use this approach, which is why the Monte Carlo procedure described here is instead recommended.

fit each synthetic dataset individually to the Pareto distribution and calculate the K-S statistic for each one relative to its own model[10]. Then we simply count what fraction of the time the resulting statistic is larger than the value for the empirical data. This fraction is the *p*-value for the fit, and can be interpreted in the standard way: if it is larger than the chosen significance level, then the difference between the empirical data and the model can be attributed to statistical fluctuations alone; if it is smaller, the model is not a plausible fit to the data.

With regard to the inequality analysis, the methodology we follow here is based on a *nested* procedure of decomposition of the Theil (1967) index (Giammatteo, 2007) that combines into a simultaneous approach the standard decompositions by population subgroups (which separates total inequality in within- and between-group components) and income sources (which divides overall inequality into proportional factor contributions).

Despite the Gini-based multidecomposition of inequality proposed by Mussard (2004), the choice of the Theil index as the reference measure of inequality is motivated by two main reasons: *i*) it allows perfect (subgroups) decomposability[11] and *ii*) satisfies the fundamental property of uniform addition for source-based decomposition[12]. A third, not trivial, advantage is given by its simple and very "smart" structure. More precisely, it is derivable as a linear function of three basic elements: (pseudo-)Theil sub indices of inequality (for groups and income sources), population shares and income shares. In other words, it allows to separate "size" and "spread" determinants of inequality both at the subgroup and income source level through the explicit reference to aggregates with political and economic relevance.

As shown in Giammatteo (2007), we can enclose into a unified framework the standard subpopulation and income source decompositions by deriving the following (weighted) bidimensional formulation of the Theil index.

10. Note crucially that for each synthetic dataset we compute the K-S statistic relative to the best-fit Pareto model for that dataset, not relative to the original distribution from which the dataset was drawn. In this way we ensure that we are performing for each synthetic dataset the same calculation that we performed for the real dataset, a crucial requirement if we wish to get an unbiased estimate of the *p*-value.

11. See e.g. Cowell (1980a,b) and Shorrocks (1984).

12. Following Morduch and Sicular (2002), a rule of factor decomposition satisfies the property of uniform addition if it registers strictly negative contributions to overall inequality for any income component equally distributed and positive. In this regard, Podder (1993) claims that «it is reasonable to think the addition of a constant to all incomes leading to a reduction in inequality if we accept relative measures». See also Shorrocks (1982, 1983) and Paul (2004) on this issue.

$$T(Y)=\sum_{m=1}^{M}\left[\sum_{k=1}^{K}P_k\frac{\mu_{k(w)}^m}{\mu_{(w)}}\ln\frac{\mu_{k(w)}}{\mu_{(w)}}\right]$$
$$+\sum_{m=1}^{M}\left\{\sum_{k=1}^{K}P_k\frac{\mu_{k(w)}}{\mu_{(w)}}\left[\sum_{i=1}^{nk}p_i\frac{y_{ik}^m}{\mu_{k(w)}}\ln\frac{y_{ik}}{\mu_{k(w)}}\right]\right\}$$
$$=\sum_{m=1}^{M}Tb_w(m)+\sum_{m=1}^{M}Tw_w(m)$$
$$=Tb_w+Tw_w, \quad (5)$$

where p_i represents the individual weight[13], P_k is the sum of the sample weights p_i $(i=1,\ldots,n_k)$ for group k, while, $\mu_{(w)}$, $\mu_{k(w)}$ and $\mu_{k(w)}^m$ are, respectively, the weighted means for the total, k^{th} subgroup and m^{th}source of the k^{th} subgroup distributions[14]. Expression (5) implicitly defines the *pseudo*-Theil of the Y^m istribution, $T_w(m)=Tb_w(m)+Tw_w(m)$, e.g. the absolute contribution to total inequality of the component m. It is important to observe that $T_w(m)$ does not measure the m source inequality[15], as incomes in total and partial distributions have different ranks and the weights are those corresponding to the total distribution. Note also that while the global index $T(Y)$ is always positive, the generic absolute contribution $T_w(m)$ can assume both positive and negative values. Hereafter, we shall use the expression of *inequality increasing* (*decreasing*) *sources* for the income components showing positive (negative) values of $T_w(m)$. Similarly, we can define $Tb_w(m)$ as the generic m source contribution to between-group inequality ("between-group *pseudo*-Theil") and $Tw_w(m)$ as the generic m source contribution to within-group inequality ("within-group *pseudo*-Theil").
The bidimensional decomposition (5) provides a wider set of possible inequality determinants than those that would be obtained by applying separated decompositions. In particular, we are able to distinguish among positive and negative sub effects on within- and between-group inequality components independently on the sign of the overall source contributions. More precisely:
standard subgroup decomposition provides aggregated within and between components of total inequality declining any information on additional source-based determinants;

13. The weights are proportional to the actual population of the strata from which the sample observations are drawn from. In the PLUS survey, strata are defined by region, type of city (metropolitan/not metropolitan), age (5 classes), sex and employment status (employed, unemployed, student, retired, other inactive/housewife). A detailed description of the sampling design and strategy of the survey is contained in Giammatteo (2009).
14. Notice that when the *not weighted* formulation is adopted we simply have $p_i=\frac{1}{n_k}$ and $P_k=\frac{n_k}{n}$.
15. The m^{th} source inequality is, instead, given by $T_m=\frac{1}{n\mu_m}\sum_{i=1}^{n}y_i^m\ln\left(\frac{y_i^m}{\mu_m}\right)$.

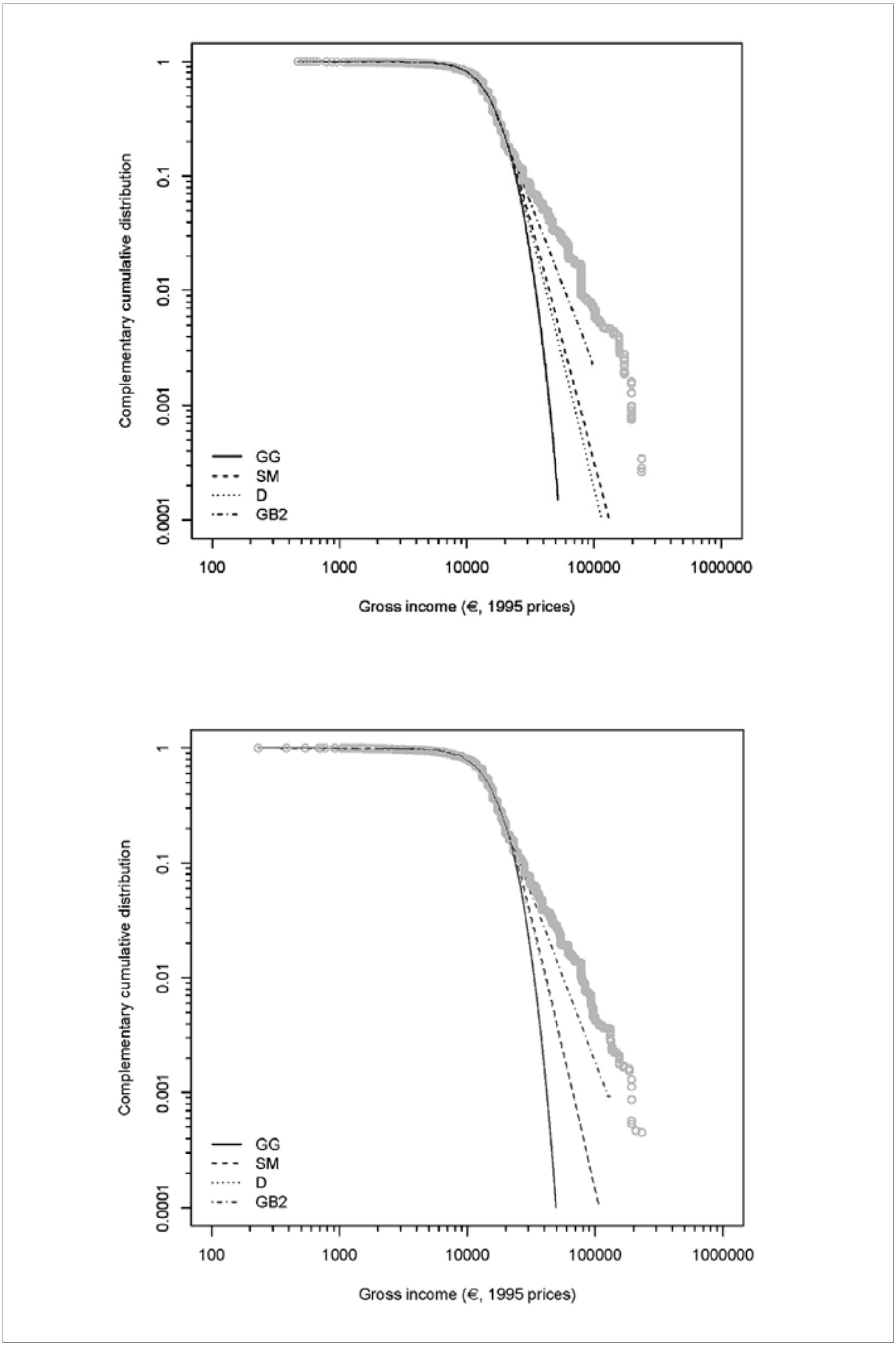
simple income source decompositions fail to distinguish in which way total income subcomponents affect total inequality through (equalising or not equalising) effects within subpopulations or between them.
The nested approach enforces both the subpopulation and income source decompositions, also representing a useful instrument for the analysis of the inequality consequences of specific government policies (transfers or tax programs, labour market reforms, etc.)[16].

4.3 Empirical Results

Using the data and methods described earlier, in this section we fit the classical Pareto model (1) to the upper tail of the Italian labour income distribution and analyze the extent to which the level of inequality in the tail and the whole of the distribution is affected by the earnings accruing to different workers.
The summary statistics in Table 4.1 suggest that the Pareto distributional assumption may be appropriate in our case. Indeed, there are two noticeable features. First is the positive skewness, which suggests that labour income distribution in any one year is highly skewed to the right. This can also be inferred by looking at the difference between median and mean income, the former being consistently lower than the latter in each year. Secondly, the level of kurtosis is well above the normal threshold both in 2005 and 2006, hinting to the presence of a thick upper tail. Figure 4.1 reveals the extent of what suggested by Table 4.1. In the panels, the horizontal axis represents the $\log_{10}$ of the annual gross income reported in the PLUS survey and the vertical axis is the $\log_{10}$ of the corresponding complementary cumulative probability. The signature feature of distributions that follow the Pareto law in the upper tail - e.g. the approximate linearity above some lower bound of the tail of their complementary cumulative distributions charted on a double logarithmic scale - is clearly evident by simple qualitative appraisal of the data.

16. Simpler but less precise approaches are given by: i) analyses of the relation between inequality and public policies through the use of dispersion graphs between inequality indices and country expenditures for social security (see e.g. Beblo and Knaus, 2001); ii) pre- and post-transfer inequality computations in order to assign factor contributions as relative difference between the two values (see e.g. Keane and Prasad, 2002, and Forster et al., 2003). As emphasized by Lerman (1999), the latter approach «may lead to misleading results».

Figure 4.1 - Observed and predicted probabilities, 2005 and 2006



Interestingly enough, this evidence is not accounted for by "super" models that obey the asymptotic Pareto law - or one of the other versions proposed by Kakwani (1980) and Esteban (1986) and discussed e.g. in Krämer and Ziebach (2004). Indeed, the transition between the bulk of the distribution and the right tail is not smooth for the data shown in Figure 4.1, so such models would not be useful in our case. For example, McDonald (1984) introduced the generalized beta II distribution (GB2), a four-parameter distribution with density given by

$$\mathrm{GB2}(x;a,b,p,q)=\frac{ax^{ap-1}}{b^{ap}B(p,q)\left[1+(x/b)^{a}\right]^{p+q}},\quad x>0,\quad a,b,p,q>0$$

where $B(\cdot,\cdot)$ is the beta function, which is not only very successful in fitting the data (Bordley et al., 1996), but also includes many of the models proposed to describe the size distribution of incomes.

Of these, the three-parameter Singh-Maddala (SM) and Dagum Type I (D) distributions, corresponding to the special cases

$$\mathrm{SM}(x;a,b,q)=\mathrm{GB2}(x;a,b,1,q)$$

and

$$\mathrm{D}(x;a,b,p)=\mathrm{GB2}(x;a,b,p,1),$$

and the generalized gamma (GG) distribution, which is obtained by

$$\mathrm{GG}(x;a,\beta,p)=\lim_{q\to\infty}\mathrm{GB2}\left(x;a,b=q^{1/a}\beta,p,q\right),$$

have shown them to be a good compromise between parsimony and goodness-of-fit in many instances[17]. However, when fitted to both years of the data, these models fail to accommodate the empirical distribution at the upper extreme, as can be seen by visual inspection of Figure 4.1, where the observed probabilities are overlaid with those predicted by the fitted distributions (see Table 4.2)[18]. This suggests that a separate treatment of the extreme tail of empirical distributions is in order, and we therefore apply the estimation method described in Section 2 to make a strong case for the Pareto

17. The nested relationship of these distributions can be seen in greater detail in McDonald and Xu (1995).

18. The parameters have been estimated by using a nonlinear least-squares method as supplied by the R function nls (R Development Core Team, 2009). The results reported in Table 4.2 include the sum of squared errors (SSE), the sum of absolute errors (SAE), the negative log-likelihood value ($-\ln L$), and the values of Akaike (1973) and Schwarz (1978) information criteria (AIC and BIC). All of these measures agree that the GB2 is the best model in both the years. According to a likelihood ratio test, its additional parameter provides a statistically significant (at the 0.1% level) improvement over its nested three-parameter distributions.

hypothesis. The results from the fitting of a Pareto distribution to each of the years of data using the maximum-likelihood approach combined with the goodness-of-fit test based on the K-S statistic are shown in Figures 4.2 and 4.3.

Using this method, we obtain for the 2005 wave that the best possible fit for the largest values in the upper tail is attained for $\hat{x}^*_{\min} = 19,925 \pm 4,264$, where the minimum value of the K-S statistic is 0.06; the estimate of the shape parameter is $\hat{\alpha}^*_{\mathrm{H}} = 1.96 \pm 0.03$ based on $m^* = 3,291$ the largest observations (about 21% of the sample). For the 2006 wave, the minimum value of the K-S statistic is found to be 0.07, yielding $\hat{x}^*_{\min} = 19,946 \pm 5,117$; the estimated value of α is $\hat{\alpha}^*_{\mathrm{H}} = 2.22 \pm 0.04$ is, based on the $m^* = 3,345$ largest observations (about 20% of the sample).

The use of these estimates produces the fits shown by the solid lines in panels (c), where the complementary cumulative distributions of the empirical data are plotted on doubly logarithmic axes. A look at the Hill plots displayed in panels (b) suggests that these fits are a good match to the data, since beyond the chosen $x_{\min}$ the estimate $\hat{\alpha}^*_H$ of the shape parameter appears roughly stable. Nonetheless, as a more objective indication of how plausible the Pareto model is as a fit to the tail data, we give in panels (a) the *p*-values for the K-S statistic, which as can be seen are large enough that the data can be firmly considered to follow the Pareto distribution in the upper tail. The linear behaviour emerging from the Pareto Q-Q plots of the sample quantiles above $\hat{x}^*_{\min}$ shown in panels (d) strongly supports the quantitative results obtained by hypothesis testing.

Having provided strong evidence for the presence of a Pareto tail in the Italian labour income distribution, we now turn to examining its role in shaping overall inequality. The situation is summarized in Tables 4.3 and 4.4, which are divided into four main parts: summary distribution statistics (namely, population and income share and relative mean), standard measures (Gini and Theil indices) of the amount of inequality in the income distribution, and standard as well as nested decomposition of the Theil index by population subgroups and type of income receivers.

Table 4.2 - Estimated distribution functions, 2005 and 2006

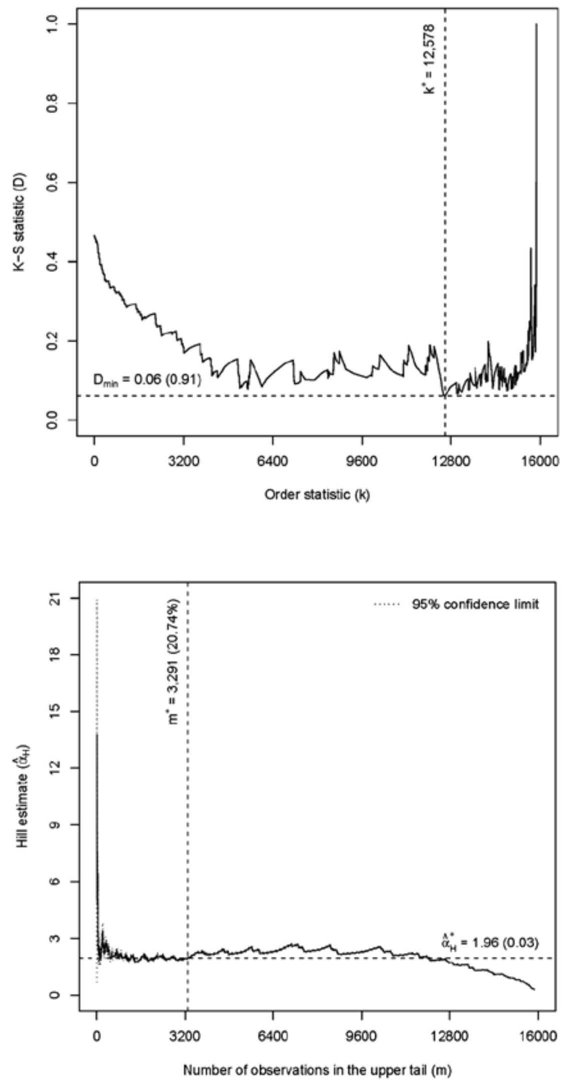
Wave	Model	Parameters[a]				SSE	SAE	– ln L[b]	AIC	BIC
		a	*b* [β]	p	q					
2005	GG	0.71	[443]	12.19	-	19,345	13,141	235,770,998	471,542,003	471,542,048
		(0.03)	(111)	(0.94)						
	SM	3.95	15096	-	1.08	14,497	11,585	227,248,892	454,497,789	454,497,834
		(0.02)	(68)		(0.01)					
	D	4.56	16386	0.72	-	13,612	11,126	226,905,664	453,811,334	453,811,378
		(0.02)	(50)	(0.01)						
	GB2	21.9	15010	0.12	0.13	7,335	8,181	225,048,367	450,096,743	450,096,802
		(0.18)	(15)	(1e-03)	(1e-03)					
2006	GG	1.12	[3,841]	4.76	-	17,170	11,598	245,429,328	490,858,661	490,858,706
		(0.02)	(268)	(0.20)						
	SM	3.61	16191	-	1.35	13,211	10,236	236,567,335	473,134,676	473,134,721
		(0.01)	(82)		(0.02)					
	D	4.70	16836.00	0.63	-	11,677	9,495	236,119,720	472,239,446	472,239,490
		(0.02)	(41)	(0.01)						
	GB2	17.3	15145	0.14	0.17	6,543	6,756	234,784,043	469,568,095	469,568,155
		(3e-06)	(1e-03)	(1e-04)	(2e-04)					

Notes: [a] Numbers in round brackets: estimated standard errors

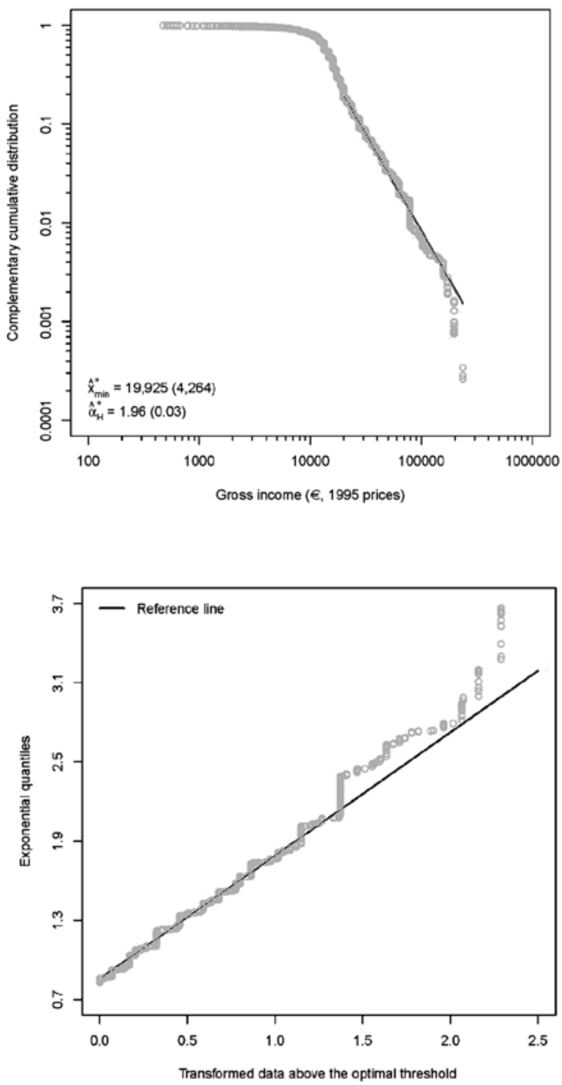
[b] Asterisks placed next to the negative log-likelihood values of the GG, SM and D distributions signals that, according to the likelihood ratio test for comparing nested models, the GB2 with its further parameter provides a statistically significant (at the 0.1% level) better fit over them.

Source: authors'own calculations using the PLUS 2005 and 2006 data

Figure 4.2 - Estimation of the Pareto upper tail, PLUS 2005

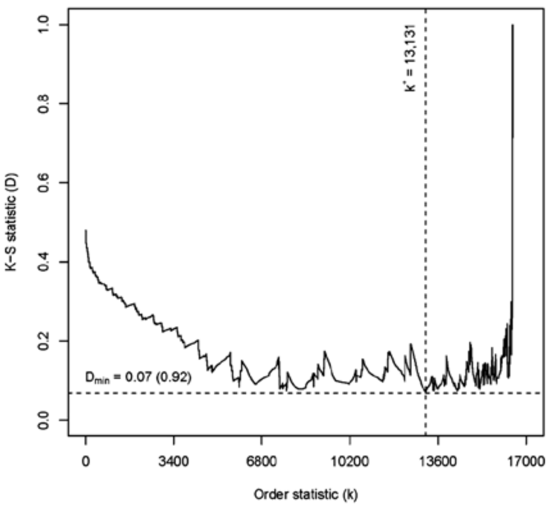


Notes: (a) Sequence of K-S values versus the rank of ordered sample values. The minimum value, $D_{\min} = 0.06$, is attained for $k^* = 12{,}578$, corresponding to the optimal estimate of the scale parameter $\hat{x}^*_{\min} = 19{,}925$. The number in parentheses denotes the bootstrap *p*-value for the best-fit Pareto model based on 5000 resamples. (b) Hill estimate as a function of the n. of obs. in the upper tail. The dotted lines represent the 95% confidence limits around the point estimates given by $\hat{\alpha}_H \pm f_{95\%}\left(\hat{\alpha}_H / \sqrt{m}\right)$, where $f_{95\%}$ is the 95% point of the normal distribution.

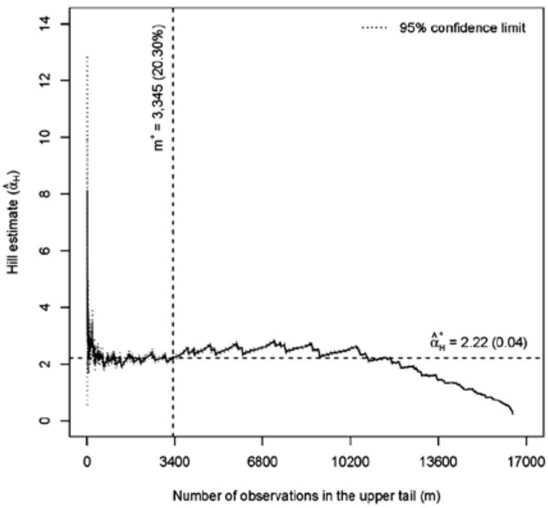


Notes: (c) Empirical (points) and Pareto (line) complementary cumulative distribution functions for the upper tail in log-log scale using the estimated optimal values for $x_{\min}$ and α. The standard error on $\hat{\alpha}^*_H$ (in parentheses) is derived from the asymptotic distribution theory of the Hill estimator and is given by, whereas the uncertainty in the estimate for $x_{\min}$ is derived by making use of a nonparametric bootstrap method based on 100 repetitions. (d) Q-Q plot for the observations above the estimated optimal threshold value using the transformation.

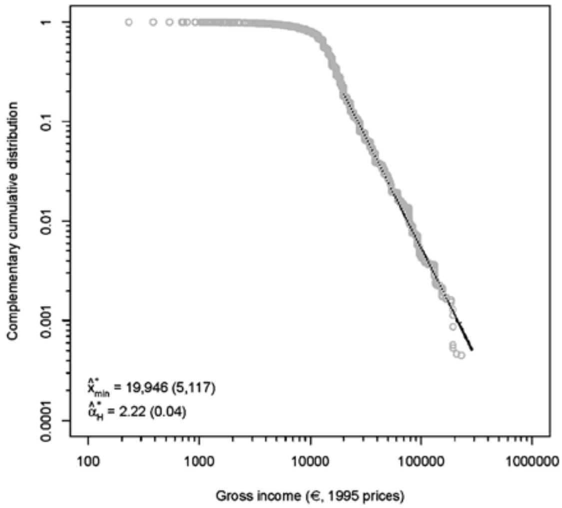
Figure 4.3 - Estimation of the Pareto upper tail, PLUS 2006



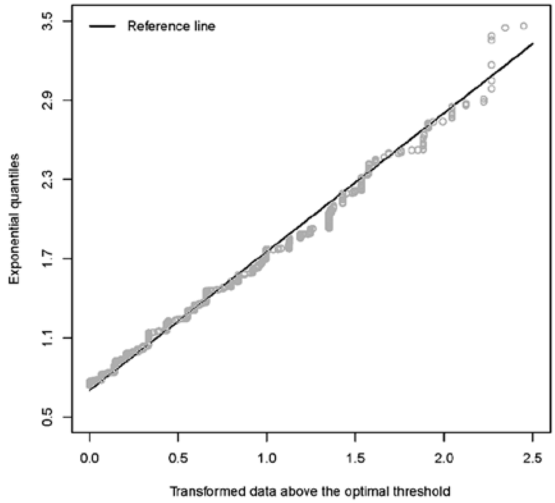
(a) K-S statistics for the best-fit Pareto model



(b) Hill plot



(c) Best-fit Pareto model for the upper



(d) Q-Q plot for the upper tail

Table 4.3 - Distribution of labour income in Italy, 2005: summary statistics, inequality measures, and standards as well as nested decomposition of the Theil index by population subgroups and type of income receivers. Population subgroups - namely, group I and II - include individuals with income $< \hat{x}^*_{\min}$ and $\geq \hat{x}^*_{\min}$ respectively[a]

	Group I	Group II	Gross inc.	Standard	Self-empl.	Atypical	Gross inc.
Summary distribution statistics[b]							
Pop. Share	0.804	0.196	**1.000**	0.648	0.223	0.129	**1.000**
	(0.005)	(0.005)	**-**	(0.006)	(0.006)	(0.004)	**-**
Inc. share	0.579	0.421	**1.000**	0.581	0.331	0.088	**1.000**
	(0.010)	(0.010)	**-**	(0.010)	(0.011)	(0.004)	**-**
Rel. mean	0.721	2.145	**1.000**	0.897	1.482	0.682	**1.000**
	(0.009)	(0.034)	**-**	(0.012)	(0.036)	(0.016)	**-**
Inequality measures[b]							
Theil	0.068	0.187	**0.249**	0.090	0.450	0.165	**0.249**
	(0.002)	(0.017)	**(0.014)**	(0.005)	(0.025)	(0.012)	**(0.014)**
Gini	0.189	0.303	**0.337**	0.210	0.498	0.295	**0.337**
	(0.003)	(0.012)	**(0.007)**	(0.004)	(0.012)	(0.010)	**(0.007)**

	Group I	Group II	Gross inc.	Standard	Self-empl.	Atypical	Gross inc.
Standard decomposition[b]							
				Absolute values			
Within	0.039	0.079	**0.118**	0.052	0.149	0.014	**0.216**
	(0.001)	(0.008)	**NA**	(0.003)	(0.011)	(0.001)	**NA**
Between	-0.190	0.321	**0.131**	-0.063	0.130	-0.034	**0.033**
	NA	NA	**(0.001)**	NA	NA	NA	**(4e-04)**
			Percent values				
Within	15.8	31.5	**47.3**	20.9	59.8	5.8	**86.6**
Between	-76.2	128.9	**52.7**	-25.3	52.2	-13.7	**13.4**
Nested Theil decomposition							
						Absolute values	
Inequality				**- 0.011**	**0.279**	**- 0.019**	**0.249**
Within				**0.018**	**0.109**	**-0.009**	**0.118**
Group I				0.050	- 0.003	- 0.008	**0.039**
Group II				- 0.032	0.112	- 0.002	**0.079**
Between				- 0.029	0.170	- 0.010	**0.131**

	Group I	Group II	Gross inc.	Standard	Self-empl.	Atypical	Gross inc.
Group I				- 0.142	- 0.025	- 0.023	**-0.190**
Group II				0.113	0.195	0.013	**0.321**
					Percent values		
Inequality				**-4.4**	**112.0**	**-7.6**	**100.0**
Within				**15.5**	**92.4**	**-7.9**	**100.0**
Group I				42.2	-2.3	-6.6	**33.3**
Group II				-26.7	94.8	-1.3	**66.7**
Between				**-22.2**	**129.6**	**-7.5**	**100.0**
Group I				-108.1	-18.9	-17.6	**-144.6**
Group II				86.0	148.5	10.1	**244.6**

Notes
[a] Figures might not add up because of rounding.
[b] Numbers in round brackets: estimated standard errors (set to NA when not available).
Source: authors'own calculations using the PLUS 2005 and 2006 data

Table 4.4 - Distribution of labour income in Italy, 2006: summary statistics, inequality measures, and standards as well as nested decomposition of the Theil index by population subgroups and type of income receivers. Population subgroups - namely, group I and II - include individuals with income $< \hat{x}^*_{min}$ and $\geq \hat{x}^*_{min}$ respectively[a]

	Group I	**Group II**	**Gross inc.**	**Standard**	**Self-empl.**	**Atypical**	**Gross inc.**
Summary distribution statistics[b]							
Pop. Share	0.807	0.193	**1.000**	0.630	0.189	0.181	**1.000**
	(0.005)	(0.005)	**-**	(0.007)	(0.006)	(0.005)	**-**
Inc. share	0.598	0.402	**1.000**	0.598	0.283	0.119	**1.000**
	(0.010)	(0.010)	**-**	(0.010)	(0.011)	(0.004)	**-**
Rel. mean	0.742	2.078	**1.000**	0.948	1.497	0.659	**1.000**
	(0.009)	(0.034)	**-**	(0.012)	(0.043)	(0.014)	**-**
Inequality measures[b]							
Theil	0.068	0.171	**0.225**	0.090	0.414	0.173	**0.225**
	(0.002)	(0.021)	**(0.015)**	(0.005)	(0.033)	(0.015)	**(0.015)**
Gini	0.193	0.280	**0.323**	0.211	0.477	0.305	**0.323**
	(0.003)	(0.013)	**(0.007)**	(0.004)	(0.014)	(0.010)	**(0.007)**

	Group I	**Group II**	**Gross inc.**	**Standard**	**Self-empl.**	**Atypical**	**Gross inc.**
Standard decomposition[b]							
	Absolute values						
Within	0.041	0.069	**0.110**	0.054	0.117	0.021	**0.192**
	(0.001)	(0.009)	**NA**	(0.003)	(0.012)	(0.002)	**NA**
Between	-0.179	0.294	**0.115**	-0.032	0.114	-0.050	**0.033**
	NA	NA	**(0.001)**	NA	NA	NA	**(4e-06)**
			Percent values				
Within	18.2	30.7	**48.9**	24.0	52.0	9.3	**85.3**
Between	-79.6	130.7	**51.1**	-14.2	50.7	-22.2	**14.7**
Nested Theil decomposition							
				Absolute values			
Inequality				**0.022**	**0.232**	**- 0.029**	**0.225**
Within				**0.025**	**0.097**	**- 0.013**	**0.110**
Group I				0.055	- 0.001	- 0.013	**0.041**
Group II				- 0.029	0.098	- 0.000	**0.069**
Between				**- 0.003**	**0.134**	**- 0.016**	**0.115**

	Group I	Group II	Gross inc.	Standard	Self-empl.	Atypical	Gross inc.
Group I				- 0.128	- 0.021	- 0.030	**-0.179**
Group II				0.125	0.155	0.014	**0.294**
				Percent values			
Inequality				**9.8**	**103.1**	**-12.9**	**100.0**
Within				**23.0**	**89.0**	**-12.0**	**100.0**
Group I				49.9	-0.9	-11.6	**37.4**
Group II				-26.9	89.9	-0.4	**62.6**
Between				**-2.8**	**116.5**	**-13.8**	**100.0**
Group I				-111.1	-18.5	-25.9	**-155.5**
Group II				108.3	135.0	12.2	**255.5**

Notes

[a] Figures might not add up because of rounding.

[b] Numbers in round brackets: estimated standard errors (set to NA when not available).

Source: authors'own calculations using the PLUS 2005 and 2006 data

As revealed by the first part of Table 4.3, in 2005 standard employees represented almost 65% of total population and received less than 60% of total income, while the self-employees made up about 22.3% of the whole population and their income share was nearly 1/3 of the total income; conversely, atypical workers accounted for 13% of the overall population and for 8.8% of total labour income. In 2006 (see the first part of Table 4.4) this pattern did not vary significantly, except for a not trivial increase (decrease) in the population and income share of atypical workers (self-employees), while a slight fall in the population share and a concurrent increase in the share of total income is observed for the group of standard employees. Looking at the relative means, the mean income from standard employment relative to that of the whole population was about 89.7% in 2005 against 94.8% in 2006; the corresponding yearly percentages were approximately 148.2% and 149.7% for self-employees, and 68.2% and 65.9% for atypical workers. Furthermore, by considering the subgroups made up of individuals with income $< \hat{x}^*_{\min}$ (group I) and $\geq \hat{x}^*_{\min}$ (group II), we observe that: i) in 2005 the population share of group I was just over 80% and accounted for about 58% of the overall income amount, while in 2006 the corresponding shares were about 80.7% and 59.8%, respectively; ii) this evidence is reversed for group II, whose fractions of the overall population and income in 2005 were approximately 19.6% and 42.1%, while in 2006 they amounted to about 19.3% and 40.2%; iii) the relative mean income of the two groups was respectively around 72.1% and 214.5% in 2005 against 74.2% and 207.8% in 2006. Finally, for what concerns inequality, we computed an estimate of the Theil index of 0.249 in 2005 and 0.225 in 2006; the estimated Gini was, respectively, 0.337 and 0.323. As can be depicted from the second part of the tables, in both the years the two measures showed sharp inequality heterogeneity at the population subgroup and income earner type levels. Nevertheless, it is worthwhile to underline that either the ranks and changes of the inequality measured by the two indices exhibited strong consistency across the years, thus suggesting robustness of our findings.
The penultimate part of Tables 4.3 and 4.4 presents the results of the standard Theil decomposition by subgroups (I and II) of the population and type of income receivers. In the first case we observe that inequality in both the years was impulsed with almost the same intensity by the within- and the between-group components, which were estimated to be respectively 47.3% and 52.7% of overall inequality in 2005 and showed only very slight changes in 2006 (the corresponding estimates are 48.9% and 51.1%). We also find a negative component of between-group inequality ascribable to group I (just over -76% in 2005 and below -80% in 2006) and a positive one to group II (around 130% on average in the two years), which also played a significant role in shaping within-group inequality (31.7% in 2005 and 30.7% in 2006). Looking at the decomposition results in terms of type of income receivers, the difference inside the subpopulations accounted for almost 87% of overall inequality in 2005 and 85% in 2006, while the mean divergence between groups increased from 13.4% of the total

inequality in 2005 to 14.7% in 2006. Furthermore, in both the years the between-group component was negative for the incomes accruing to standard and atypical workers (decreasing for the former group and increasing for the latter from 2005 to 2006) and positive for self-employment income (around 51% on average in the two years), whereas the within-group component was mainly shaped by the distribution of labour income among standard and self-employed workers.

Finally, the results arising from applying the nested Theil decomposition procedure are displayed in the last part of Tables 4.3 and 4.4. As stated before (see Section 2), this method allows to discern the way in which every single component of labour income impacts on the within- and between-group inequalities (as estimated by partitioning of the population into groups I and II). We emphasize in this regard the following aspects.

i. The contribution to total inequality attributable to income from standard work was -4.4% in 2005, implying an inequality-reducing effect. This result arises as a combination of two opposite effects: a positive one on the within-group inequality (15.5%) and a negative one on the between-group component (-22.2%). Furthermore, income from standard work contributed positively to within differences when we look to the group I (42.2%) but negatively considering the group II (-26.7%). A similar result is also obtained in the between-group context, where we observe an inequality-decreasing (increasing) effect of -108.1% (86.0%) attributable to group I (II) afferents.
ii. Self-employment income represented the most important inequality-increasing source of income: its impact on total inequality amounted to 112% in 2005, with the two components of within- and between-group inequality contributing respectively 92.4% and 129.6%, both markedly influenced by group II differences.
iii. Income stemming from atypical work made marked negative contributions of -7.6% in 2005 and -12.9% in 2006.

Moving to compare the nested Theil decomposition results over the two years, a somewhat interesting empirical evidence seems to emerge. Firstly, in 2006 compared to the previous year income from standard work shifted from a negative to a positive contribution (9.8%) to overall inequality, in consequence of an increased weight of the within-group differences (23.0%) and a weaker inequality-decreasing effect of the between-group component (-2.8%). Secondly, with respect to the previous year income from self-employment, while preserving the sign, exerted effects of different magnitude on total inequality (103.1%, 89% and 116.5%), whereas income from atypical work reinforced its inequality-decreasing effect. The nested decomposition highlights that great part of this equalising overall effect seems to be justified by the low average incomes and their simultaneous high concentration among group I afferents. We also note that the negative between-group contribution to total inequality from atypical work changes from -7.5% to -13.8%, providing evidence of an increasing mean income gap between this group of workers and the regular ones. Despite this, it is interesting to observe positive (and increasing) contributions to between-group inequality coming from atypical incomes concentrated in the upper tails of the two yearly distributions (10% in 2005 and 12% in 2006).

4.4 Concluding Remarks and Policy Implications

Our investigation of the shape of the Italian labour income distribution over the years 2005 and 2006 gives added support to the "conventional wisdom" that earnings are highly skewed to the right with an upper tail that is suitably modeled by a Pareto distribution. The fitted Pareto functions of the upper tail reveal that inequality amongst the rich slightly decreased from one year to the other, and this appears to reflect in the pattern traced by standard inequality measures that show a reduction of a few percent in overall labour income inequality between 2005 and 2006.

In order to uncover the effect of the underlying earnings distribution shape on the inequality of labour market outcomes, we have performed a nested decomposition of the Theil inequality measure that emphasizes the twofold role played by sources of labour income and their distribution among the two groups of earners identified by splitting the samples in correspondence of the estimated lower quantiles above which the Pareto distribution was fitted. The results indicate a negative contribution from standard employment income to overall inequality for the year 2005, which can be divided into an inequality-increasing effect exerted mainly within the group of earners in the bulk of the distribution and an inequality-decreasing effect in terms of between-group differences. The contribution from this source of labour income has switched from negative to positive in 2006, mainly because of the rise in relative mean - which has resulted in a weaker negative effect on the between-group component. As regards earnings from self-employment, they represent the most important inequality-increasing source of labour income in both the years of investigation, as a result of highest relative mean income and dispersion, especially at the upper extreme of the distributions. Finally, earnings arising from non-standard forms of employment have experimented from one year to the other an increase in both the population and income shares, as well as worsening relative mean and increasing source-specific inequality.

The policy implications of the current findings emerge quite naturally. The empirically documented Pareto distribution at high income levels implies rather high inequality. This possibly suggests more redistribution, considering that during the mid-2000s Italy achieved a modest redistribution from top to bottom of the income ladder (OECD, 2008). However, any policy disregarding the diverse impact on inequality that standard employment, self-employment and atypical incomes have in different parts of the earnings distribution would likely have only limited success in reducing overall inequality through tax and transfer systems. Policy-makers should therefore avoid a uniform approach for all types of income and adopt a set of well-targeted policies properly taking into account these heterogeneities in terms of inequality.