Università degli Studi
di Macerata

Dipartimento di Studi
sullo Sviluppo Economico

Working paper n. 42

June/2012

The labour market and the distribution of income: an empirical analysis for Italy

Fabio Clementi, Michele Giammatteo
University of Macerata, Bank of Italy

ISSN: 1971-890X

The labour market and the distribution of income: an empirical analysis for Italy*

Fabio Clementi[1], Michele Giammatteo[2]
University of Macerata, Bank of Italy

Abstract

This paper provides an empirical examination of the distribution of labour earnings in Italy. Using four waves of data from the Participation Labour Unemployment Survey, a database of information on the Italian labour market supply, we find the shape of the observed distributions to be positively skewed with a "fat" and long tail on the right. We 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.

JEL classification: D33, D63
Keywords: labour income, size distribution, inequality

Corresponding author: Fabio Clementi (fabio.clementi@unimc.it)

Department Informations:
Piazza Oberdan 3, 62100 Macerata – Italy; phone: +39 0733 258 3960; fax: +39 0733 258 3970; e-mail: csampaoli@unimc.it

*We acknowledge the Italian Institute for the Development of Vocational Training for Workers (ISFOL) for providing us with data from 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.

[1]University of Macerata, Department of Studies on Economic Development, Piazza G. Oberdan 3, 62100 Macerata, Italy.

[2]Bank of Italy, Via Otricoli 41, 00181 Rome, Italy. The opinions expressed in this paper are those of the author only and in no way involve the responsibility of the Bank of Italy. The usual disclaimer applies.

1 Introduction

In the social world, size distributions tend to be asymmetric and leptokurtic. The distribution of earnings is a leading practical example of this attitude. Indeed, earlier studies on the subject (e.g. Lydall, 1968 and Harrison, 1981) discovered a number of empirical regularities that are found in all observed distributions of large populations. Specifically, earnings distributions are positively skewed and have a "fat" right tail, or, otherwise stated, the mean usually exceeds the median and the top percentiles of earners account for a conspicuous share of the total. These stylized facts are relevant for a variety of reasons, mostly because they affect inequality positively (Maasoumi and Theil, 1979).

Economic theory has proposed various explanations of the observed regularities.[1] A vast literature considers the returns to human capital investment and the related issue of skill-biased technological change as decisive factors in shaping the distribution of labour earnings (see e.g. Machin and Van Reenen, 1998, Acemoglu, 2002, Moore and Ranjan, 2005 and Lemieux, 2006). Others studies place more emphasis on institutional features, such as the role of unions in the wage-setting process and the degree of centralization and coordination in wage bargaining (Blau and Kahn, 1996; Gottschalk and Joyce, 1998; Kahn, 2000; Barth and Lucifora, 2006). Economic openness has also been viewed as a factor affecting the observed patterns of earnings inequality (Wood, 1994; Leamer, 1995; Acemoglu, 2002; Anderson, 2005). Finally, changes in the composition of workforce (as a result of the shift in employment from manufacturing to services, population aging, increased participation of women, and so on) as well as labour market conditions (country-specific supply and demand factors, market rigidities, etc.) can clearly contribute to define the earnings distribution structure (Bertola and Ichino, 1995; Gregg and Manning, 1997; Moore and Ranjan, 2005).

In Italy, the research literature has principally focused on two possible determinants of earnings distribution: the composition of employment and the institutional framework.

As regards the first issue, some studies have paid explicit attention to the increasing role of self-employment as a specific feature of the Italian workforce. For instance, Torrini (2005, 2006) and Rani (2008) found that in Italy self-employment accounts for a substantial share of total labour force with respect to other developed countries and emphasized its importance in accounting for inequality.[2]

For what concerns institutional factors, the past two decades have witnessed some very important reform experiences which have been influencing the recent trend of Italy's labour market.[3] Key among them were the 1993 reform of the bargaining system and the measures aimed at improving labour market flexibility.

The bargaining pattern set out in 1993 consists of a first level nation-wide sectoral bargaining and a second level of bargaining decentralised at regional or firm level. While national contracts are expected to set the minimum wage and preserve the purchasing power of workers, decentralised bargaining should be devoted to distribute productivity and/or profitability gains. Second-level bargaining, however, is not compulsory and it can

[1] In the following we pay explicit attention to the distribution of labour earnings, interchangeably referred to as "wages" or "incomes".

[2] Most recently, this finding has also been maintained by the OECD (2011, p. 3): «Changes in self-employment income were important drivers of increased earnings inequality: their share in total earned income has increased by 10% since the mid-1980s, and self-employment income seems more predominant among high earners, to the contrary of many other OECD countries».

[3] For a discussion of the main reform measures implemented in Italy since the early 1990s and how they have impacted on the labour market outcomes we refer the reader to Schindler (2009).

not define wages lower than the national minimum. This has limited to some extent the use of decentralised bargaining and resulted in a wage distribution more compressed than it was expected (see e.g. Casadio, 2003, Checchi and Pagani, 2005 and Dell'Aringa and Pagani, 2007).[4]

With regards to the pursuit of labour market efficiency, legislative measures specifically directed at fostering flexibility have been undertaken via the increase of the so-called atypical or non-standard forms of employment, while leaving the regulation of existing employment relations largely unchanged. Despite substantial improvements in employment rates, this has led to a strong segmentation of the Italian labour market, where highly protected and well-paid open-end jobs coexist along with risky and low-paid temporary occupations. The shift towards non-standard forms of work could exacerbate existing earnings inequality, as wage differentials between standard and non-standard forms of employment widen.[5]

Drawing from these recent labour market developments, in the present work we provide new empirical evidence on the distribution of earnings in Italy. The analysis is conducted with the data of the Participation Labour Unemployment Survey (PLUS), a sample survey on the Italian labour market supply carried out by ISFOL[6] for the years 2005, 2006, 2008 and 2010. 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.

Relying on this source of data, we find that the empirical distribution in any one year is highly skewed to the right, so that the proportion of workers earning more than the modal wage is larger than the proportion of those earning less. The "fat" and long tail on the right of the observed distributions also points to the existence of a relatively

[4]Traditional wage-setting institutions like collective bargaining affect workers predominantly at the bottom or middle of the wage distribution. By contrast, wage-setting mechanisms of high executives (the "working rich") concern workers at the very top of the distribution. The importance of executive compensations to explain the rise in top income shares during the last quarter of the twentieth century has been a standard result in all the studies analysing income concentration within the top groups in Anglo-Saxon countries. A tentative explanation explored by Piketty and Saez (2003, 2006, but see also Lemieux *et al.*, 2007 and Lemieux, 2008) is that the growth in performance-related schemes that affect the compensation of high executives and the change in social norms regarding inequality and the acceptability of very high wages have removed some implicit barriers to the rise of incomes for the very highest earners. However, the surge experienced by top incomes in continental Europe and other advanced countries like Japan has been small relative to existing estimates for English-speaking countries, and even the results for Italy are fairly modest. Nevertheless, the Italian experience shows a persistent increasing pattern in top income shares since the mid-1980s, mainly driven by top wages and self-employment income (Alvaredo and Pisano, 2010).

[5]See Rani (2008) for an attempt to assess the extent to which changes in employment patterns are associated with the rise in income inequality observed over the past two decades in the majority of countries. With regard to wage differentials between temporary and permanent workers in the Italian case, a quantitative appraisal of the phenomenon can be found in Picchio (2006), Cutuli (2008) and Lucidi and Raitano (2009).

[6]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 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).

small number of very well-paid individuals, a fact that is confirmed by a mean wage exceeding the median. Since this pattern is likely to impact on the inequality of labour market outcomes, we undertake a more-in-depth investigation by considering how much of the dispersion in earnings concentrated in different parts of the distribution ("rich" and "non-rich") 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. 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-employed and the increased job precariousness. Indeed, in all the years examined we observe significant positive contributions to within- and between-group components of total inequality arising from self-employment revenues, which results to be more highly concentrated (and thus responsible for the inequality level) in the upper end of the distribution. This trend combines with the effects exerted by income from standard employment, which instead appears more concentrated in the bulk. Earnings from non-standard forms of work, in turn, have seen their shares of both the total population and income increasing over time, thus arising as an important feature of the contemporary workplace.

The rest of the paper is organized as follows: Section 2 describes the data used in the study and outlines the approach to estimating observed distributions and the implied amount of inequality. Section 3 presents the results of the analysis. Section 4 summarizes the paper.

2 Data and methods

2.1 The data

The PLUS survey consists of four waves of data conducted in 2005, 2006, 2008 and 2010 on around 38,000 individuals—of which 16,000 workers of both public and private sectors—belonging to the Italian population aged 18–64.[7] Complementary to other key national statistical sources,[8] the core objective of PLUS is that of providing reliable estimates of rare and only marginally explored labour market issues. In particular, it is devoted to the study of 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. Some of the key prerogatives of the PLUS survey that seem worthwhile highlighting here are as follows:

i) it is planned with the chief purpose of providing accurate estimates of very small-scale phenomena, in that it allows to produce consistent evaluations of population aggregates of about 100,000 individuals with a coefficient of variation lower than

[7] The age range was 15–64 for the 2005 wave. Furthermore, starting with the 2006 wave, a panel section consisting of a large number of observations (about 65%) was additionally provided.

[8] 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.

10% (for example, the contract type composition of Italian total employment is annually estimated at a degree of desegregation that allows reliable analyses of fixed-term/atypical job distribution);[9]

ii) consistent labour income variables are derived through the implementation of appropriate techniques in the questionnaire design (e.g. with differentiation of the interview submission process by type of worker), consolidation of respondents loyalty (for panel units), and thorough data processing (multiple data check and imputation);

iii) only survey respondents are included (absence of proxy interviews), reducing in this way the extent of measurement errors and partial non-responses.

The variable chosen for the analysis 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.[10] 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.[11] Furthermore, because of the complex sampling design of the PLUS survey, data make use troughout the analysis of appropriate sampling weights to produce representative estimates and correct standard errors and statistical tests.[12]

2.2 Modelling the Italian labour income distribution

As we have said, one of the main objectives of this paper is to determine how much of the dispersion in earnings concentrated in different parts of the distribution may be accounted for by alternative sources of labour income. For this purpose we shall distinguish in the following between two groups of high- and low-income earners, or "rich" and "non-rich", the divide being represented by the minimum possible income in the classical Pareto (1895, 1896, 1897a,b) distribution

$$F(x) = 1 - \left(\frac{x}{x_{\min}}\right)^{-\alpha}, \quad x_{\min} \le x < \infty, \quad x_{\min}, \alpha > 0, \tag{1}$$

which is usually considered as a good approximation of the distribution of incomes among the rich and the moderately rich.[13] In order to avoid imposing an arbitrary threshold above which the Pareto relationship (1) is valid, the lowest income $x_{\min}$ is estimated from the data by adopting a numerical technique proposed by Clauset *et al.* (2007, 2009) that

[9]See e.g. Corsetti and Mandrone (2010) and Mandrone and Marocco (2012) for applications related to this issue.

[10]Non-standard jobs include *economically dependent jobs*, i.e. "legally autonomous" but "semi-dependent" jobs. The workers performing these jobs depend on a single employer for their income, are subject to compulsory daily presence, use employer's equipment and perform the same tasks as some of their fellows. They are contractually treated as "professional" workers, but any specific skills, professional knowledge or competence is actually needed.

[11]The series of the NIC index is publicly available on the ISTAT's website at the address `http://www.istat.it/it/files/2011/02/indici_nazionali_nic_tuttilivaggr.xls`.

[12]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 ISTAT's LFS survey.

[13]An extensive historical survey of the use of the Pareto distribution in the context of income and wealth distributions can be found e.g. in Arnold (1983).

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 the estimate of the shape parameter α over the data $x \geq x_{\min}$ by using the conditional maximum likelihood estimator introduced by Hill (1975)

$$\hat{\alpha}_{\mathrm{H}} = \left[\frac{1}{m} \sum_{i=1}^{m-1} \left(\ln x_{n-i+1} - \ln x_{n-m+1} \right) \right]^{-1}, \tag{2}$$

where $m = n - k + 1$ is the number of extreme sample values above the threshold, n is the sample size and k is the rank of the order statistic x_{n-m+1}, and then compute the Kolmogorov-Smirnov (K-S) goodness-of-fit statistic

$$D = \max_{x \geq x_{\min}} \left| \hat{F}(x) - F(x; x_{\min}, \hat{\alpha}_{\mathrm{H}}) \right| \tag{3}$$

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 $\hat{\alpha}_{\mathrm{H}}$, i.e. $F(x; x_{\min}, \hat{\alpha}_{\mathrm{H}})$. 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, $\hat{\alpha}^*_{\mathrm{H}}$.

Once the parameters have been estimated, by exploiting the asymptotic distribution theory of the Hill estimator (2) we calculate the standard error of the shape parameter as $\frac{\hat{\alpha}^*_{\mathrm{H}}}{\sqrt{m^*}}$ (e.g. Lux, 1996), whereas the uncertainty in the estimate for $x_{\min}$ 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,[14] 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.[15] 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 (3) 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 fit each synthetic dataset individually to the Pareto distribution and calculate the K-S statistic for each one relative to its own model.[16] Then we simply count what fraction of the time the resulting

[14] In practice, we perform 100 such bootstrap samplings.

[15] 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 n 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.

[16] 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.

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.

2.3 The inequality decomposition analysis

With regard to the inequality analysis, the methodology we shall follow is based on a *nested* procedure of decomposition of the Theil (1967) index 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[17] and *ii*) satisfies the fundamental property of uniform addition for source-based decomposition.[18] 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 subindices 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 the Appendix, 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

$$\begin{aligned} T(Y) &= \sum_{m=1}^{M}\left[\sum_{k=1}^{K} P_k \frac{\mu_{k(w)}^{m}}{\mu_{(w)}} \ln \frac{\mu_{k(w)}}{\mu_{(w)}}\right] + \sum_{m=1}^{M}\left\{\sum_{k=1}^{K} P_k \frac{\mu_{k(w)}}{\mu_{(w)}}\left[\sum_{i=1}^{nk} p_i \frac{y_{ik}^{m}}{\mu_{k(w)}} \ln \frac{y_{ik}}{\mu_{k(w)}}\right]\right\} \\ &= \sum_{m=1}^{M} Tb_w(m) + \sum_{m=1}^{M} Tw_w(m) \\ &= Tb_w + Tw_w, \end{aligned} \tag{4}$$

where p_i represents the individual weight,[19] 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.[20] Expression (4) implicitly defines the *pseudo*-Theil of the Y^m distribution, $T_w(m) = Tb_w(m) + Tw_w(m)$,

[17]See e.g. Cowell (1980a,b) and Shorrocks (1984).

[18]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.

[19]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).

[20]Notice that when the *unweighted* formulation is adopted we simply have $p_i = \frac{1}{n_k}$ and $P_k = \frac{n_k}{n}$.

i.e. 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,[21] 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) source* for the income component showing a positive (negative) value 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 m source contribution to within-group inequality ("within-group *pseudo*-Theil").

The bidimensional decomposition (4) 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 subeffects on within- and between-group inequality components independently on the sign of the overall source contributions. More precisely:

i) standard subgroup decomposition provides aggregated within and between components of total inequality declining any information on additional source-based determinants;

ii) 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.).[22]

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 distributions and analyze the extent to which the level of inequality within and between the two groups that we consider, respectively, as "the rich" and "the non-rich" is affected by earnings accruing from different sources.

3.1 The size distribution of labour earnings in Italy

The summary statistics in Table 1 suggest that the Pareto distributional assumption may be appropriate in our case.

[Table 1 about here.]

[21]The m^{th} source inequality is, instead, given by $T_m = \frac{1}{n}\sum_{i=1}^{n} \frac{y_i^m}{\mu_m} \ln \frac{y_i^m}{\mu_m}$.

[22]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 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 Keane and Prasad, 2002, and Förster *et al.*, 2003). As emphasized by Lerman (1999), the latter approach «may lead to misleading results».

Indeed, there are two noticeable features. First, the labour income distribution in any one year displays statistically significant evidence for skewness. 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. Second, the level of kurtosis is significantly above the normal threshold in any one of the years concerned, hinting to the presence of a thick upper tail.

The Pareto diagrams shown in panel (c) of Figures 1 through 4 reveal the extent of what suggested by the table above. These diagrams are plots of the annual gross income x, charted on a logarithmic scale, against the complementary cumulative distribution of individuals with annual gross income greater than or equal to x (also on a log scale). The distinctive feature of distributions that follow the Pareto model in the upper tail—i.e. the approximate linearity above some lower bound of their complementary cumulative distributions plotted on a double logarithmic scale—is clearly evident when examining these graphs, and we can therefore apply the estimation method discussed in Section 2 to make a stronger case for the Pareto hypothesis.

The results of fitting the Pareto distribution to each of the years of data are summarized in Table 2.

[Table 2 about here.]

As can be seen, the model fit varied slightly across years but was generally excellent. This is demonstrated first by the precision of the parameter estimates. All t-ratios were indeed significant at the 0.1% level and relatively large—for example, the smallest t-ratio for any estimate of $x_{\min}$ was slightly less than 4 and was typically at least 7 times larger for α. Excellent goodness of fit is also demonstrated by the complementary cumulative distribution plots shown in panel (c) of Figures 1 to 4, where the Pareto model (solid line) exhibits a remarkable agreement with the data in the upper tail of the distributions, even when the latter gets quite noisy (as, for example, in 2008).

[Figure 1 about here.]

[Figure 2 about here.]

[Figure 3 about here.]

[Figure 4 about here.]

Furthermore, a look at the Hill plots displayed in panel (b) of the same figures confirms that this model is a good match to the data, since beyond the cut-off income values used the estimates of the shape parameter appear roughly stable.[23]

As a more objective indication of the suitability of the Pareto model, Table 2 reports for each wave the K-S statistic that yields the best fit to the tail data (dashed line in panel (a) of Figures 1–4) and the Monte Carlo p-value for the goodness-of-fit test. Notice how all p-values are very close to unity, meaning that in all cases our data can be firmly considered to follow the Pareto distribution in the upper tail. This is confirmed by visual inspection of the Pareto Q-Q plots of the sample quantiles above $\hat{x}^*_{\min}$, shown in panel (d)

[23] The so-called "Hill plot" is a visual diagnostic tool charting the Hill estimate of the shape parameter $\hat{\alpha}_{\mathrm{H}}$ for each $x_{\min}$. The idea is to visually identify a region where the plot levels off, representing a stable estimate of α, and then choose $x_{\min}$ as the beginning of that region (see e.g. Beirlant *et al.*, 2004).

of the figures.[24] As it can be seen, every plot lies extremely close to the reference line, and much closer than is typically observed in plots of this type.

It must also be noticed that the size of the group here considered as "the rich" shrank dramatically in 2010. Indeed, based on the results reported in Table 2, the optimal number of tail observations used in the estimation of the Pareto distribution showed in that year a decline by approximately 70% with respect to 2008, while in contrast only few significant changes are detected in the preceding years. This is a probable consequence of the economic crisis started in 2008–2009 in the wake of the global financial crisis, which caused a fall in real mean and median income of about 2% between 2008 and 2010 (see Table 1). This hypothesis seems also confirmed by the results of a relative distribution analysis, which allows for a decomposition of the relative income density between 2008 and 2010 so as to isolate changes occurred along the entire income range due to differences in the first moment.[25] Indeed, from inspection of Figure 5 one can see that the mean downshift between 2008 and 2010 impacted the whole range of the income distribution with varying intensity, affecting more negatively the mass of workers above the 2008 median.

[Figure 5 about here.]

More specifically, the figure displays a decline of the mass in the upper tail above the 70^{th} percentile and a relatively small increase in the upper-median range of the share of workers between approximately the 65^{th} and the 70^{th} percentile of the 2008 distribution, thus indicating a clear convergence of higher incomes toward the center.[26]

3.2 Results from inequality decomposition

Having provided strong evidence for the presence of a Pareto tail in the Italian labour income distribution, we now turn to assessing earnings inequality through decomposition exercises. The situation is summarized in Tables 3 and 4.

[Table 3 about here.]

[Table 4 about here.]

[24] Since a log-transformed Pareto random variable is exponentially distributed, the coordinates of the points on a Pareto Q-Q plot follow immediately from the exponential case by taking the transformation $\ln \frac{X}{x_{\min}} \sim \text{Exp}(\alpha)$.

[25] For our purposes, the "relative distribution" is defined as the ratio of the income density in the comparison year (2010) to the income density in the reference year (2008) evaluated at each quantile of the reference distribution, and can be interpreted as the fraction of the comparison population that fall in each quantile of the reference population. This allows us to identify and locate the changes that have occurred in the entire Italian labour income distribution between the two years. In particular, when the fraction of individuals in a quantile is higher (lower) than the fraction in the reference year, the relative distribution will be higher (lower) than 1. Where there is no change, the relative distribution will be flat at the value 1. Furthermore, this approach also allows to decompose the relative density into changes in location and changes in shape, in order to emphasize differences between the comparison and the reference populations that could be attributed to a change in the average (or median) income or to changes of the shape (including differences in variation, skewness and other distributional characteristics). We refer the reader to Handcock and Morris (1998, 1999) for a more formal definition of the relative distribution.

[26] Alternative indices like the median can be considered. The corresponding results do not differ in a significant way and are not reported here.

Table 3 contains for each wave of data standard distributional measures, such as population and income shares and relative means. Standard employees represented in each year around 65% of total population and received on average 60% of total income. From 2005 to 2010, self-employed decreased both their population and income shares, while atypical workers followed a reversed trend until 2008. As regards the relative mean, for standard employees it ranges between 90% in 2005 and 98% in 2010; for self-employed the percentage increased from 148% in 2005 to 155% in 2008, whereas in 2010 it decreased to 132%; finally, the mean of atypical workers relative to that of the whole population was around 68% over the whole period.

By considering the subgroups made up of individuals with income $< \hat{x}^*_{\min}$ ("non-rich") and $\geq \hat{x}^*_{\min}$ ("rich"), we observe that: *i*) the population and income shares of the non-rich decreased between 2005 and 2008 and increased in 2010; *ii*) this evidence is reversed for the rich group; *iii*) the relative mean income of each group was fairly stable until 2008 (slightly over 70% for the non-rich and more than 200% for the rich) and raised in 2010, notably for the rich.

Table 3 also shows the estimates and corresponding standard errors for both the Theil and Gini measures of inequality. The Theil index for total gross income grew from 0.249 in 2005 to 0.269 in 2010, save for a temporary decrease in 2006. The estimated Gini exhibited a similar pattern until 2008, while in 2010 it stayed the same as in 2005. As depicted in Figure 6, the two indices reveal sharp inequality heterogeneity both at the population subgroup and income source levels.

[Figure 6 about here.]

In particular, self-employed and atypical earning distributions are characterized by high levels of income disparities. However, it is worthwhile to underline that either the ranks and changes of the inequality measured by the two indices are always consistent across the years, thus suggesting robustness of our findings.

Table 4 presents the results of the standard and "nested" Theil decomposition by subgroups ("rich" and "non-rich") and labour income sources (standard, self-employed and atypical). For each wave: *i*) the rows "Within" and "Between" indicate how much of the income source contributions (columns) can be imputed to intra- or inter-groups differences; *ii*) the rows labeled "Non-rich" and "Rich" specify how the incomes in the lower and upper parts of the annual distributions affect each of the above two components; *iii*) the "Source dec." row displays the income source contributions resulting from the standard decomposition rule. Because of the additive property of equation (4), the absolute values sum up both vertically and horizontally; the percent values are calculated with respect to total inequality ("Source dec.") as well as "Within" and "Between" components.

The within-group component of labour income inequality increased from more than 47% in 2005 to around 55% in 2008, while it reduced in 2010. The standard decomposition by income sources highlights the fundamental role played by self-employed in shaping total income inequality, even though their relative impact decreased steadily from 112% to less than 68%. The contribution due to income from standard work was slightly negative in 2005 and positive in the following three waves. In particular, at the end of the observed period it reached a significant value of about 37%. Income stemming from atypical work made marked negative contributions in 2005 and 2006, while in the following two waves it decreased on average by 5%.

The contribution to overall inequality of standard incomes shifted from negative to positive as a consequence of a weaker inequality-decreasing effect of the between-group component (from around -22% in 2005 to approximately -1% in 2008) and, only in 2010, because of the strong increase of the within-group inequality share (about 77%). Moreover, the nested procedure allows us to impute most of this result to the inequality-increasing contribution (nearly 23%) arising among the rich. Self-employed inequality contribution fell over time as a result of decreasing positive effects of both within- and between-group components referred to the rich group. Finally, with regards to the atypical workers we observe a nearly generalised negative contribution, apart from a few but significant exceptions. In particular, the rich incomes accounted for positive between-group contributions over the entire period analysed, whereas for the within-group inequality this is true only starting with the 2008 wave.

4 Summary

In this paper we have examined the distribution of labour earnings in Italy using four waves of data from the Participation Labour Unemployment Survey (PLUS), a sample survey on the Italian labour market supply. The main results are twofold.

First, we have found that the shape of the Italian labour income distribution in any one year of the analysis is highly skewed to the right with a “fat” and long upper tail, a feature pointing to the existence of a relatively small number of very well-paid individuals. This has called into question the use of the traditional Pareto model to properly separate the group of the rich from poorer workers. The results of fitting a Pareto function to earnings above some endogenously determined income threshold revealed that this model is a plausible (if not completely satisfactory) hypothesis for our data. Concomitantly, the analysis of the shape of the Italian labour income distribution also showed a remarkable drop of the share of workers in the top quantiles of the 2010 distribution as a result of the current economic crisis.

Second, in order to shed light on the roots of the labour income inequality, we have carried out a nested decomposition of the Theil inequality measure that emphasized the twofold role played by sources of labour income and their distribution among the groups of rich and non-rich earners.

The results highlighted the decisive role played by self-employment income in shaping total inequality through positive effects on both between- and within-rich components of inequality. When viewed alongside the high level of self-employment rate in Italy, this finding suggests the importance of considering the connection between employment structure and labour income distribution as a valuable key of understanding. Earnings from standard employment exhibited positive within-group contributions to overall inequality due to income disparities among non-rich workers, especially in 2010. These contributions reflect their large income and population shares among the non-rich, as well as distribution homogeneity induced by the centralised Italian bargaining system. Atypical earnings affected inequality negatively in each year, although some positive contributions have recently arised from the group of the rich.

Finally, the empirical results seem to suggest a preliminary effect on inequality due to the ongoing economic crisis. Indeed, between 2008 and 2010 self-employed accounted for a definitely lower income share, relative mean and earnings dispersion that reflected in weaker inequality contributions. These did not significantly affect total inequality, which remained almost stable because of the simultaneous increase of the contributions from standard incomes.

Appendix: Derivation of the *nested* decomposition rule

Consider a total distribution of income, Y, and a population of n units (individuals or households) divided into K mutually exclusive and exhaustive groups[27] receiving income from M different sources, Y^m, such that

$$Y = \sum_{i=1}^{n} y_i = \sum_{k=1}^{K} \sum_{i=1}^{n_k} y_{ik} = \sum_{k=1}^{K} \sum_{i=1}^{n_k} \sum_{m=1}^{M} y_{ik}^{m},$$

where y_{ik}^{m} is the amount of Y^m received by the unit i of group k[28]. Given the Theil well-known formula

$$T(Y) = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{\mu} \ln \frac{y_i}{\mu},$$

a *nested decomposition rule* can be derived through the following 3 simple steps (Giammatteo, 2007).

Step 1. The basic source-based decomposition of the Theil is

$$T(Y) = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{\mu} \ln \frac{y_i}{\mu} = \sum_{m=1}^{M} \left(\frac{1}{n} \sum_{i=1}^{n} \frac{y_i^m}{\mu} \ln \frac{y_i}{\mu} \right) = \sum_{m=1}^{M} T(m),$$

where y_i^m is the amount of Y^m received by the unit i and $T(m) = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i^m}{\mu} \ln \frac{y_i}{\mu}$ is the generic *pseudo*-Theil for the income source m.

Step 2. The standard subgroups decomposition of the Theil index is given by

$$T(Y) = \sum_{k=1}^{K} \pi_k s_k \ln \frac{\mu_k}{\mu} + \sum_{k=1}^{K} \pi_k s_k \left(\frac{1}{n_k} \sum_{i=1}^{n_k} \frac{y_{ik}}{\mu_k} \ln \frac{y_{ik}}{\mu_k} \right) = Tb(Y) + Tw(Y),$$

where $\pi_k s_k = \frac{n_k}{n} \frac{\mu_k}{\mu}$ is the income share of group k. Notice that the first term, $Tb(Y)$, contributes nothing only if $s_k = 1$, $\forall k$; in all other cases it will be strictly positive. The second term, $Tw(Y)$, which corresponds to the weighed mean of the K sub-indices $T_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \frac{y_{ik}}{\mu_k} \ln \frac{y_{ik}}{\mu_k}$, is also never negative and reach its minimum (zero) only in the case of equally distributed incomes inside each subgroup of the population.

Step 3. By considering the following additivity in sub-means

$$\mu_k = \sum_{m=1}^{M} \mu_k^m, \tag{A.1}$$

we are able to divide the between-group component of total inequality into M source contributions as

$$Tb = \sum_{m=1}^{M} \left(\sum_{k=1}^{K} \frac{n_k}{n} \frac{\mu_k^m}{\mu} \ln \frac{\mu_k}{\mu} \right) = \sum_{m=1}^{M} Tb(m), \tag{A.2}$$

[27] Each individual only belongs to one group and the overall population is entirely covered by the K groups.

[28] Hereafter, we exclude the trivial case of constant distributions, i.e. $Y \neq e_n \mu$, where $e_n = (1, 1, \ldots, 1)$. Moreover, for each of the sub-income distribution Y^m the following minimum requirement is always satisfied: $y_i^m \geq 0$, and $y_j^m > 0$ at least for one j.

where $\frac{n_k}{n}\frac{\mu_k^m}{\mu}$ is the m source share of total income for the subpopulation k and $Tb\left(m\right) = \sum_{k=1}^{K} \frac{n_k}{n}\frac{\mu_k^m}{\mu} \ln \frac{\mu_k}{\mu}$ is the pseudo-Theil index computed on the K subgroup means. Following a similar procedure, but considering the individual income relations $y_{ik} = \sum_{m=1}^{M} y_{ik}^m$ instead of (A.1), we can decompose the within-group component of the Theil index by income sources as

$$Tw = \sum_{m=1}^{M} \left[\sum_{k=1}^{K} \pi_k s_k \left(\frac{1}{n_k} \sum_{i=1}^{nk} \frac{y_{ik}^m}{\mu_k} \ln \frac{y_{ik}}{\mu_k} \right) \right] = \sum_{m=1}^{M} Tw\left(m\right), \qquad \text{(A.3)}$$

where $Tw\left(m\right) = \sum_{k=1}^{K} \pi_k s_k T_k\left(m\right)$ is a weighted sum of K pseudo-Theil indices $T_k\left(m\right) = \frac{1}{n_k} \sum_{i=1}^{nk} \frac{y_{ik}^m}{\mu_k} \ln \frac{y_{ik}}{\mu_k}$.

Expressions (A.2) and (A.3) allow us to derive the following *subgroup-source nested decomposition* of the Theil index

$$T\left(Y\right) = Tb + Tw = \sum_{m=1}^{M} Tb\left(m\right) + \sum_{m=1}^{M} Tw\left(m\right),$$

where $Tb\left(m\right)$ and $Tw\left(m\right)$ represent, respectively, the contribution to *between*- and *within*-group inequality coming from the m income component.

References

Acemoglu, D. (2002). "Technical Change, Inequality, and the Labor Market", *Journal of Economic Literature*, **40**:7–72.

Alvaredo, F., and Pisano, E. (2010). "Top Incomes in Italy, 1974–2004", in Atkinson, A.B., and Piketty, T. (Eds.), *Top Incomes. A Global Perspective* (pp. 625–663). New York: Oxford University Press.

Anderson, E. (2005). "Openness and Inequality in Developing Countries: A Review of Theory and Recent Evidence", *World Development*, **33**:1045–1063.

Anscombe, F.J., and Glynn, W.J. (1983). "Distribution of the Kurtosis Statistic b_2 for Normal Statistics", *Biometrika*, **70**:227–234.

Arnold, B.C. (1983). *Pareto Distributions.* Fairland: International Co-operative Publishing House.

Barth, E., and Lucifora, C. (2006). *Wage Dispersion, Markets and Institutions: The Effects of the Boom in Education on the Wage Structure*, Discussion Paper 2181, IZA, Bonn. Available at: `http://ftp.iza.org/dp2181.pdf`.

Beblo, M., and Knaus, T. (2001). "Measuring Income Inequality in Euroland", *Review of Income and Wealth*, **47**:301–320.

Beirlant, J., Goegebeur, Y., Teugels, J., and Segers, J. (2004). *Statistics of Extremes: Theory and Applications.* New York: John Wiley & Sons.

Bertola, G., and Ichino, A. (1995). "Wage Inequality and Unemployment: United States vs. Europe", in Bernanke, B.S., and Rotemberg, J. (Eds.), *NBER Macroeconomics Annual 1995* (pp. 13–54). Cambridge and London: MIT Press.

Blau, F.D., and Kahn, L.M. (1996). "International Differences in Male Wage Inequality: Institutions Versus Market Forces", *Journal of Political Economy*, **104**:791–837.

Casadio, P. (2003). "Wage Formation in the Italian Private Sector After the 1992–93 Income Policy Agreements", in Fagan, G., Mongelli, F.P., and Morgan, J. (Eds.), *Institutions and Wage Formation in the New Europe* (pp. 112–133). Cheltenham UK: Edward Elgar Publishing.

Checchi, D., and Pagani, L. (2005). "The Effects of Unions on Wage Inequality. The Italian Case in the 1990s", *Politica Economica*, **1**:41–68.

Clauset, A., Young, M., and Gleditsch, K.S. (2007). "On the Frequency of Severe Terrorist Attacks", *Journal of Conflict Resolution*, **51**:58–88.

Clauset, A., Shalizi, C.R., and Newman, M.E.J. (2009). "Power-Law Distributions in Empirical Data", *SIAM Review*, **51**:661–703.

Corsetti, G., and Mandrone, E. (2010). "Il lavoro: tra forma e sostanza. Una lettura dell'occupazione non standard in Italia", *Economia & Lavoro*, **44**:71–97.

Cowell, F.A. (1980a). "Generalized Entropy and the Measurement of Distributional Change", *European Economic Review*, **13**:147–159.

Cowell, F.A. (1980b). "On the Structure of Additive Inequality Measures", *Review of Economic Studies*, **47**:521–531.

Cutuli, G. (2008). "Lavoro atipico e salari: una discriminazione nascosta nel mercato del lavoro italiano", *Polis*, **22**:403–422.

D'Agostino, R.B. (1970). "Transformation to Normality of the Null Distribution of g_1", *Biometrika*, **57**:679–681.

Dell'Aringa, C., and Pagani, L. (2007). "Collective Bargaining and Wage Dispersion in Europe", *British Journal of Industrial Relations*, **45**:29–54.

Deville, J.C., and Särndal, C.E. (1992). "Calibration Estimators in Survey Sampling", *Journal of the American Statistical Association*, **87**:367–382.

Efron, B., and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Förster, M., Jesuit, D.K., and Smeeding, T.M. (2003). *Regional Poverty and Income Inequality in Central and Eastern Europe: Evidence From the Luxembourg Income Study*, Working Paper 65, UNU-WIDER, New York. Available at: `http://www.wider.unu.edu/stc/repec/pdfs/rp2003/dp2003-65.pdf`.

Giammatteo, M. (2007). *The Bidimensional Decomposition of Inequality: A Nested Theil Approach*, Working Paper 466, LIS, Luxembourg. Available at: `http://www.lisproject.org/publications/liswps/466.pdf`.

Giammatteo, M. (2009). *The Participation Labour Unemployment Survey: Methodology and Implementation.* Paper presented at the first ITAlian COnference on Survey Methodology (ITACOSM09), Siena, Italy.

Gottschalk, P., and Joyce, M. (1998). "Cross-National Differences in the Rise in Earnings Inequality: Market and Institutional Factors", *Review of Economics and Statistics*, **80**:489–502.

Gregg, P., and Manning, A. (1997). "Skill-Biased Change, Unemployment and Wage Inequality", *European Economic Review*, **41**:1173–1200.

Handcock, M.S., and Morris, M. (1998). "Relative Distribution Methods", *Sociological Methodology*, **28**:53–97.

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

Harrison, A. (1981). "Earnings by Size: A Tale of Two Distributions", *Review of Economic Studies*, **68**:621–631.

Hill, B.M. (1975). "A Simple General Approach to Inference About the Tail of a Distribution", *Annals of Statistics*, **3**:1163–1174.

Kahn, L.M. (2000). "Wage Inequality, Collective Bargaining and Relative Employment From 1985 to 1994: Evidence From Fifteen OECD Countries", *Review of Economics and Statistics*, **82**:564–579.

Keane, M.P., and Prasad, E.S. (2002). "Inequality, Transfer, and Growth: New Evidence From the Economic Transition in Poland", *Review of Economics and Statistics*, **84**:324–341.

Leamer, E.E. (1995). *Trade, Wages and Revolving Door Ideas*, Working Paper 4716, NBER, Cambridge MA. Available at: `http://www.nber.org/papers/w4716.pdf`.

Lemieux, T. (2006). "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?", *American Economic Review*, **96**:461–498.

Lemieux, T. (2008). "The Changing Nature of Wage Inequality", *Journal of Population Economics*, **21**:21–48.

Lemieux, T., MacLeod, W.B., and Parent, D. (2007). *Performance Pay and Wage Inequality*, Working Paper 13128, NBER, Cambridge MA. Available at: `http://www.nber.org/papers/w13128.pdf`.

Lerman, R.I. (1999). "How Do Income Sources Affect Income Inequality?", in Silber, J. (Ed.), *Handbook of Income Inequality Measurement* (pp. 341–362). Boston: Kluwer Academic Press.

Lucidi, F., and Raitano, M. (2009). "Molto flessibili, poco sicuri: lavoro atipico e disuguaglianze nel mercato del lavoro italiano", *Economia & Lavoro*, **43**:99–115.

Lux, T. (1996). "The Stable Paretian Hypothesis and the Frequency of Large Returns: An Examination of Major German Stocks", *Applied Financial Economics*, **6**:463–475.

Lydall, H.F. (1968). *The Structure of Earnings*. London: Oxford University Press.

Maasoumi, E., and Theil, H. (1979). "The Effect of the Shape of the Income Distribution on Two Inequality Measures", *Economics Letters*, **4**:289–291.

Machin, S., and Van Reenen, J. (1998). "Technology and Changes in Skill Structure: Evidence From Seven OECD Countries", *Quarterly Journal of Economics*, **113**:1215–1244.

Mandrone, E., and Marocco, M. (2012). "Non Standard and Disguised Employment Relationship in Italy: An Economical and Juridical Approach", in Mandrone, E. (Ed.), *Labour Economics—PLUS Empirical Studies* (forthcoming). Rome: ISFOL.

Mandrone, E., and Radicchia, D. (2006). *PLUS Participation Labour Unemployment Survey. Indagine campionaria nazionale sulle caratteristiche e le aspettative degli individui sul lavoro*. Rome: ISFOL. Available at: `http://www.lavoro.gov.it/Lavoro/Europalavoro/SezioneEuropaLavoro/DGPOF/ProdottiEditoriali/CollaneEditoriali/LibriFSE/Plus.htm`.

Moore, M.P., and Ranjan, P. (2005). "Globalisation vs. Skill-Biased Technological Change: Implications for Unemployment and Wage Inequality", *Economic Journal*, **115**:391–422.

Morduch, J., and Sicular, T. (2002). "Rethinking Inequality Decomposition with Evidence From Rural China", *Economic Journal*, **112**:93–106.

Mussard, S. (2004). "The Bidimensional Decomposition of the Gini Ratio. A Case Study: Italy", *Applied Economics Letters*, **11**:503–505.

OECD (2011). *Divided We Stand: Why Inequality Keeps Rising – Country Note: Italy*. Available at: `http://www.oecd.org/dataoecd/51/35/49177743.pdf`.

Pareto, V. (1895). "La legge della domanda", *Giornale degli Economisti*, **10**:59–68. English translation in *Rivista di Politica Economica*, **87**:691–700, 1997.

Pareto, V. (1896). "La courbe de la répartition de la richesse", reprinted in Busoni, G. (Ed.), *Œeuvres complètes de Vilfredo Pareto, Tome 3: Écrits sur la courbe de la répartition de la richesse* (1965). Geneva: Librairie Droz. English translation in *Rivista di Politica Economica*, **87**:647–700, 1997.

Pareto, V. (1897a). *Course d'économie politique*. London: Macmillan.

Pareto, V. (1897b). "Aggiunta allo studio della curva delle entrate", *Giornale degli Economisti*, **14**:15–26. English translation in *Rivista di Politica Economica*, **87**:645–700, 1997.

Paul, S. (2004). "Income Sources Effects on Inequality", *Journal of Development Economics*, **73**:435–451.

Picchio, M. (2006). *Wage Differentials and Temporary Jobs in Italy*, Discussion Paper 2006-33, Département des Sciences Economiques, Université catholique de Louvain. Available at: `http://sites-final.uclouvain.be/econ/DP/IRES/2006-33.pdf`.

Piketty, T., and Saez, E. (2003). "Income Inequality in the United States, 1913–98", *Quarterly Journal of Economics*, **118**:1–39.

Piketty, T., and Saez, E. (2006). "The Evolution of Top Incomes: A Historical and International Perspective", *American Economic Review*, **96**:200–205.

Podder, N. (1993). "The Disaggregation of the Gini Coefficient by Factor Components and its Applications to Australia", *Review of Income and Wealth*, **39**:51–61.

Rani, U. (2008). *Impact of Changing Work Patterns on Income Inequality*, Discussion Paper 193, ILO, Geneva. Available at: `http://www.ilo.org/public/english/bureau/inst/publications/discussion/dp19308.pdf`.

Schindler, M. (2009). *The Italian Labor Market: Recent Trends, Institutions, and Reform Options*, Working Paper 09/47, IMF, Washington. Available at: `http://www.imf.org/external/pubs/ft/wp/2009/wp0947.pdf`.

Shorrocks, A.F. (1982). "Inequality Decomposition by Factor Component", *Econometrica*, **50**:193–211.

Shorrocks, A.F. (1983). "The Impact of Income Components on the Distribution of Family Incomes", *Quarterly Journal of Economics*, **98**:311–326.

Shorrocks, A.F. (1984). "Inequality Decomposition by Population Subgroups", *Econometrica*, **52**:1369–1385.

Stephens, M.A. (1986). "Tests Based on EDF Statistics", in D'Agostino, R.B., and Stephen, M.A. (Eds.), *Goodness-of-Fit Techniques* (pp. 97–113). New York: Marcel Dekker.

Theil, H. (1967). *Economics and Information Theory.* Amsterdam: North-Holland.

Torrini, R. (2005). "Cross-Country Differences in Self-Employment Rates: The Role of Institutions", *Labour Economics*, **12**:661–683.

Torrini, R. (2006). *Self-Employment Incidence, Overall Income Inequality and Wage Compression.* Paper prepared for the 29th general conference of the International Association for Research in Income and Wealth, Joensuu, Finland, 20–26 August.

Wood, A. (1994). *North-South Trade, Employment and Inequality: Changing Fortunes in a Skill Driven World.* Oxford: Clarendon Press.

Tables

Table 1 – Sample statistics

	Wave			
	2005	2006	2008	2010
Obs	15,868	16,475	15,299	16,587
Pop. ('000)	21,570	22,619	22,970	22,434
Min	472	231	293	286
p25	11,802	11,094	11,131	10,876
Med.	14,612	14,458	15,042	14,698
p75	18,597	18,574	18,953	18,519
Max	236,035	288,906	392,284	383,305
Mean	17,967	17,182	17,403	17,126
St. dev.	16,786	15,195	18,732	18,869
Skewness[a]	5.97	6.87	9.10	10.47
	(<5e-05)†	(<5e-05)†	(<5e-05)†	(<5e-05)†
Kurtosis[b]	55.87	82.81	121.91	160.36
	(<5e-05)†	(<5e-05)†	(<5e-05)†	(<5e-05)†

Source: authors' own calculations using the PLUS data
Notes
[a] Numbers in round brackets: p-values for the D'Agostino (1970) skewness test; the null hypothesis is H_0: normality versus the alternative H_1: nonnormality due to skewness
[b] Numbers in round brackets: p-values for the Anscombe and Glynn (1983) kurtosis test; the null hypothesis is H_0: normality versus the alternative H_1: nonnormality due to excess kurtosis
† Significant at the 0.1% level

Table 2 – Parameter estimates and goodness-of-fit test for the Pareto distribution fit[a]

Wave	m^*	$\hat{x}^*_{\text{min}}$	$\hat{\alpha}^*_{\text{H}}$	D_{min}[b]
2005	3,291	19,925	1.962	0.061
		(4.673)†	(57.706)†	(0.906)
2006	3,345	19,946	2.225	0.069
		(3.898)†	(58.553)†	(0.920)
2008	3,512	18,953	2.239	0.061
		(6.407)†	(58.921)†	(0.926)
2010	1,083	28,612	1.916	0.074
		(4.453)†	(33.034)†	(0.952)

Source: authors' own calculations using the PLUS data
Legend: m^* = optimal number of observations in the upper tail to be used for estimation of the shape parameter; $\hat{x}^*_{\text{min}}$ = optimal estimate of the lower income limit; $\hat{\alpha}^*_{\text{H}}$ = optimal estimate of the shape parameter; D_{min} = minimum value attained by the K-S statistic
Notes
[a] Numbers in round brackets: t-ratios using standard errors estimated by the methods described in Section 2
[b] Numbers in round brackets: p-values computed via 5,000 Monte Carlo replications; the null hypothesis for the test is that the Pareto model is a statistically good approximation to the model generating the data
† Significant at the 0.1% level

Table 3 – Summary statistics and inequality measures by population subgroups and income sources[a]

	Non-rich[b]	Rich[c]	Standard	Self-employed	Atypical	Gross inc.
2005[d]						
Pop. share	0.804	0.196	0.648	0.223	0.129	1.000
	(0.005)	(0.005)	(0.006)	(0.006)	(0.004)	–
Inc. share	0.579	0.421	0.581	0.331	0.088	1.000
	(0.010)	(0.010)	(0.010)	(0.011)	(0.004)	–
Rel. mean	0.721	2.145	0.897	1.482	0.682	1.000
	(0.009)	(0.034)	(0.012)	(0.036)	(0.016)	–
Theil	0.068	0.187	0.090	0.450	0.165	0.249
	(0.002)	(0.017)	(0.005)	(0.025)	(0.012)	(0.014)
Gini	0.189	0.303	0.210	0.498	0.295	0.337
	(0.003)	(0.012)	(0.004)	(0.012)	(0.010)	(0.007)
2006[d]						
Pop. share	0.807	0.193	0.630	0.189	0.181	1.000
	(0.005)	(0.005)	(0.007)	(0.006)	(0.005)	–
Inc. share	0.598	0.402	0.598	0.283	0.119	1.000
	(0.010)	(0.010)	(0.010)	(0.011)	(0.004)	–
Rel. mean	0.742	2.078	0.948	1.497	0.659	1.000
	(0.009)	(0.034)	(0.012)	(0.043)	(0.014)	–
Theil	0.068	0.171	0.090	0.414	0.173	0.225
	(0.002)	(0.021)	(0.005)	(0.033)	(0.015)	(0.015)
Gini	0.193	0.280	0.211	0.477	0.305	0.323
	(0.003)	(0.013)	(0.004)	(0.014)	(0.010)	(0.007)
2008[d]						
Pop. share	0.781	0.219	0.640	0.176	0.184	1.000
	(0.006)	(0.006)	(0.007)	(0.006)	(0.005)	–
Inc. share	0.559	0.441	0.602	0.272	0.126	1.000
	(0.012)	(0.012)	(0.013)	(0.014)	(0.005)	–
Rel. mean	0.716	2.015	0.941	1.545	0.683	1.000
	(0.012)	(0.039)	(0.016)	(0.063)	(0.020)	–
Theil	0.076	0.241	0.097	0.524	0.283	0.270
	(0.003)	(0.028)	(0.007)	(0.044)	(0.050)	(0.021)
Gini	0.199	0.307	0.215	0.518	0.332	0.339
	(0.003)	(0.018)	(0.004)	(0.004)	(0.019)	(0.017)
2010[d]						
Pop. share	0.930	0.070	0.655	0.182	0.163	1.000
	(0.004)	(0.004)	(0.007)	(0.006)	(0.005)	–
Inc. share	0.763	0.237	0.644	0.240	0.116	1.000
	(0.013)	(0.013)	(0.011)	(0.011)	(0.005)	–
Rel. mean	0.820	3.406	0.983	1.320	0.711	1.000
	(0.011)	(0.118)	(0.013)	(0.048)	(0.023)	–
Theil	0.092	0.252	0.173	0.480	0.223	0.269
	(0.002)	(0.027)	(0.022)	(0.041)	(0.039)	(0.019)
Gini	0.242	0.323	0.274	0.506	0.321	0.337
	(0.003)	(0.021)	(0.009)	(0.016)	(0.019)	(0.008)

Source: authors' own calculations using the PLUS data
Notes
[a] Figures might not add up because of rounding
[b] Includes individuals with income $< \hat{x}^*_{\min}$
[c] Includes individuals with income $\geq \hat{x}^*_{\min}$
[d] Numbers in round brackets: estimated standard errors

Table 4 – Standard and nested decomposition of the Theil index by population subgroups and income sources[a]

	Absolute values				Percent values				Subgroup dec.
	Standard	Self-employed	Atypical	Gross inc.	Standard	Self-employed	Atypical	Gross inc.	
2005									
Non-rich[b]	0.050	-0.003	-0.008	0.039	42.2	-2.3	-6.6	33.3	–
Rich[c]	-0.032	0.112	-0.002	0.079	-26.7	94.8	-1.3	66.7	–
Within	0.018	0.109	-0.009	0.118	15.5	92.4	-7.9	100.0	47.3
Non-rich[b]	-0.142	-0.025	-0.023	-0.190	-108.1	-18.9	-17.6	-144.6	–
Rich[c]	0.113	0.195	0.013	0.321	86.0	148.5	10.1	244.6	–
Between	-0.029	0.170	-0.010	0.131	-22.2	129.6	-7.5	100.0	52.7
Source dec.	-0.011	0.279	-0.019	0.249	-4.4	112.0	-7.7	100.0	100.0
2006									
Non-rich[b]	0.055	-0.001	-0.013	0.041	49.9	-0.9	-11.6	37.4	–
Rich[c]	-0.029	0.098	-0.000	0.069	-26.9	89.9	-0.4	62.6	–
Within	0.025	0.097	-0.013	0.110	23.0	89.0	-12.0	100.0	48.8
Non-rich[b]	-0.128	-0.021	-0.030	-0.179	-111.1	-18.5	-25.9	-155.5	–
Rich[c]	0.125	0.155	0.014	0.294	108.3	135.0	12.2	255.5	–
Between	-0.003	0.134	-0.016	0.115	-2.8	116.5	-13.8	100.0	51.2
Source dec.	0.022	0.232	-0.029	0.225	9.8	103.1	-12.9	100.0	100.0
2008									
Non-rich[b]	0.057	-0.004	-0.011	0.042	38.6	-2.6	-7.5	28.5	–
Rich[c]	-0.034	0.126	0.015	0.106	-23.2	84.7	10.0	71.5	–
Within	0.023	0.122	0.004	0.148	15.4	82.1	2.5	100.0	54.9
Non-rich[b]	-0.137	-0.017	-0.034	-0.187	-112.2	-13.7	-27.6	-153.5	–
Rich[c]	0.136	0.155	0.018	0.309	111.3	127.7	14.5	253.5	–
Between	-0.001	0.139	-0.016	0.122	-0.9	114.0	-13.1	100.0	45.1
Source dec.	0.022	0.261	-0.012	0.270	8.1	96.5	-4.6	100.0	100.0
2010									
Non-rich[b]	0.070	0.014	-0.015	0.070	54.1	11.1	-11.4	53.9	–
Rich[c]	0.030	0.026	0.004	0.060	23.1	19.8	3.2	46.1	–
Within	0.100	0.040	-0.011	0.130	77.2	30.9	-8.1	100.0	48.2
Non-rich[b]	-0.110	-0.021	-0.020	-0.152	-78.9	-15.3	-14.6	-108.8	–
Rich[c]	0.111	0.163	0.017	0.291	79.4	117.2	12.2	208.8	–
Between	0.001	0.142	-0.003	0.139	0.4	101.9	-2.3	100.0	51.8
Source dec.	0.101	0.182	-0.014	0.269	37.4	67.7	-5.1	100.0	100.0

Source: authors' own calculations using the PLUS data

Notes

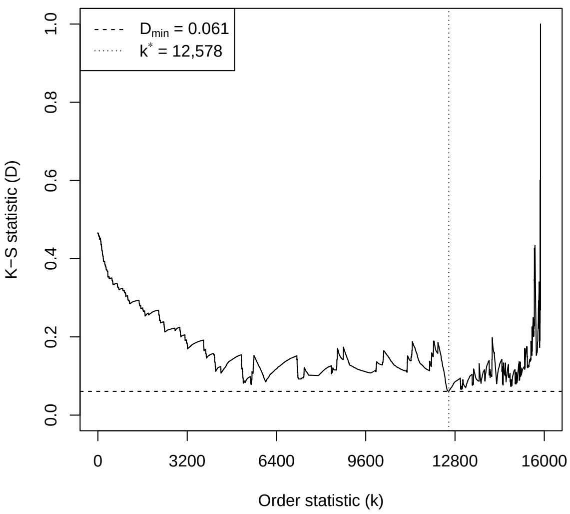
[a] Figures might not add up because of rounding

[b] Includes individuals with income $< \hat{x}^*_{\min}$

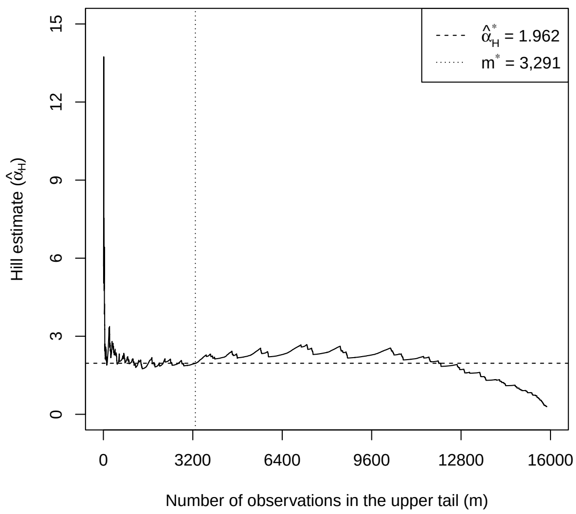
[c] Includes individuals with income $\geq \hat{x}^*_{\min}$

Figures

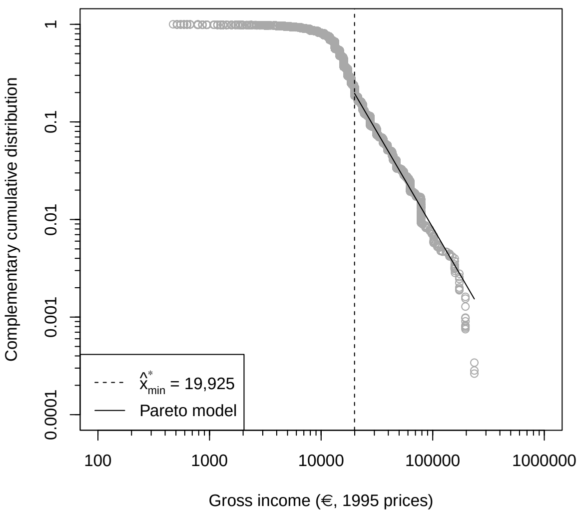
Figure 1 – Pareto distribution fit for the PLUS 2005 wave



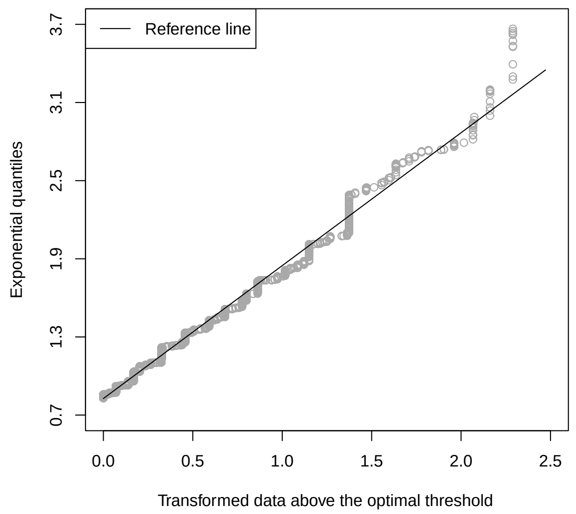
(a) K-S statistic vs. rank of ordered sample values



(b) Hill plot

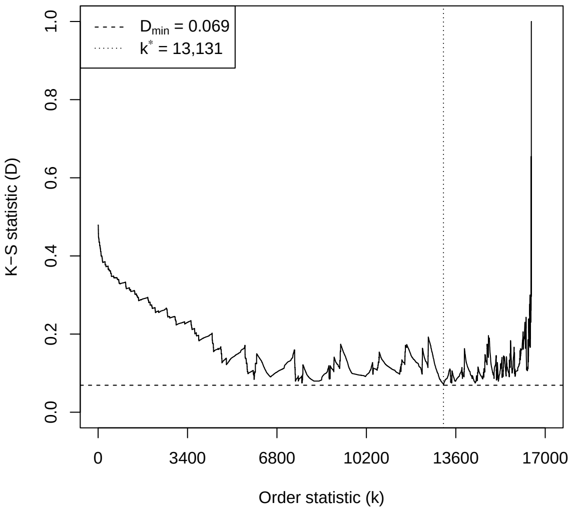


(c) Best-fit Pareto model for the upper tail

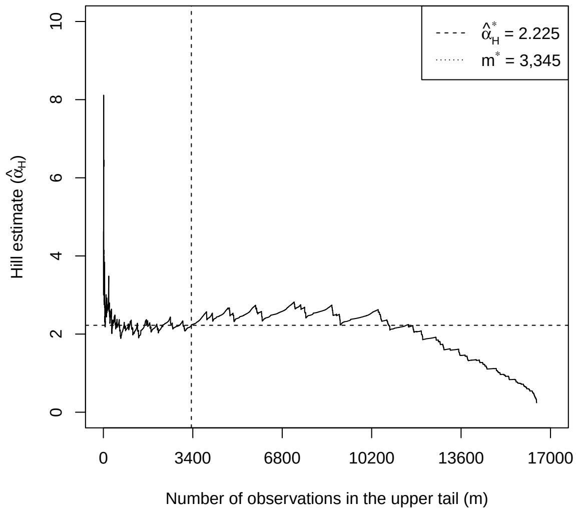


(d) Pareto Q-Q plot

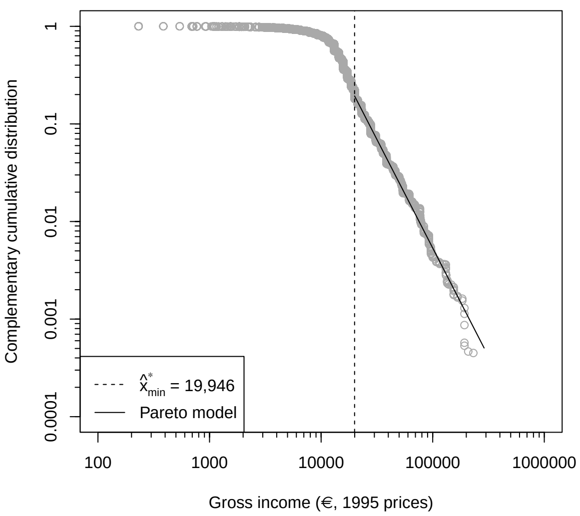
Figure 2 – Pareto distribution fit for the PLUS 2006 wave



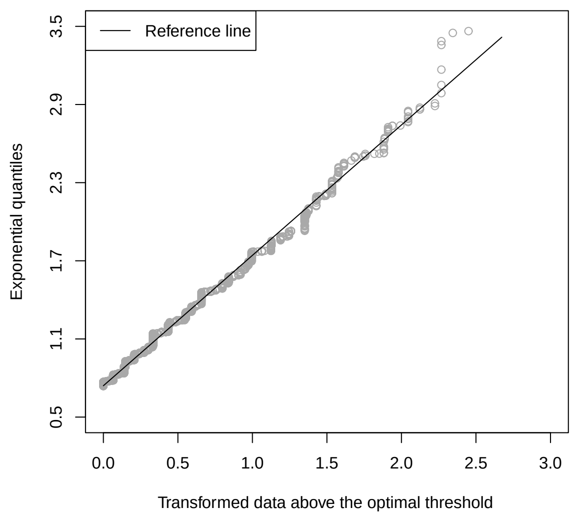
(a) K-S statistic vs. rank of ordered sample values



(b) Hill plot

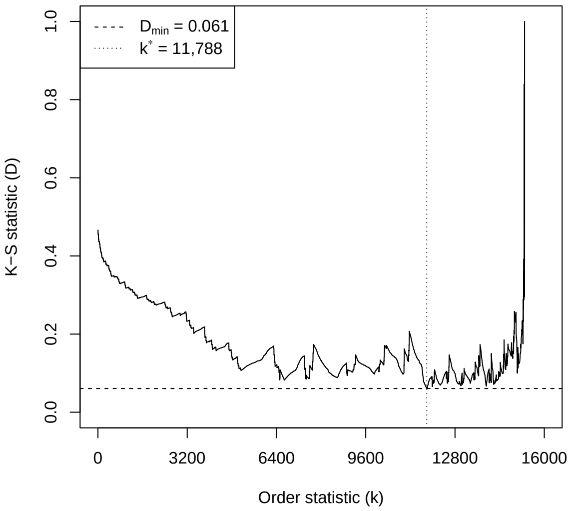


(c) Best-fit Pareto model for the upper tail

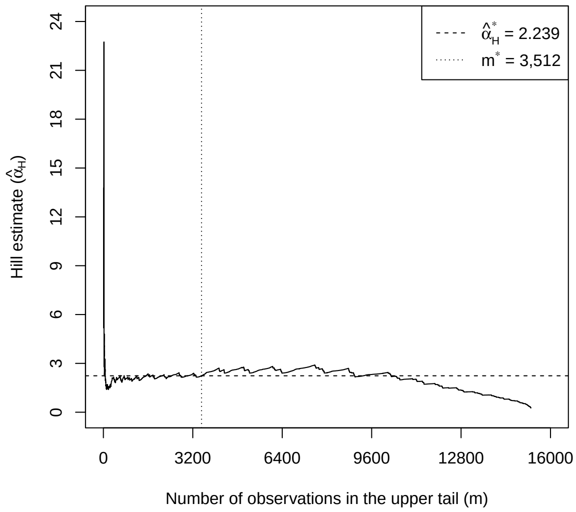


(d) Pareto Q-Q plot

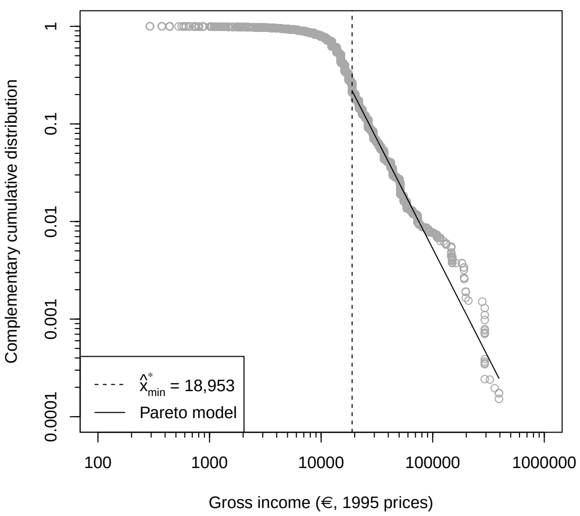
Figure 3 – Pareto distribution fit for the PLUS 2008 wave



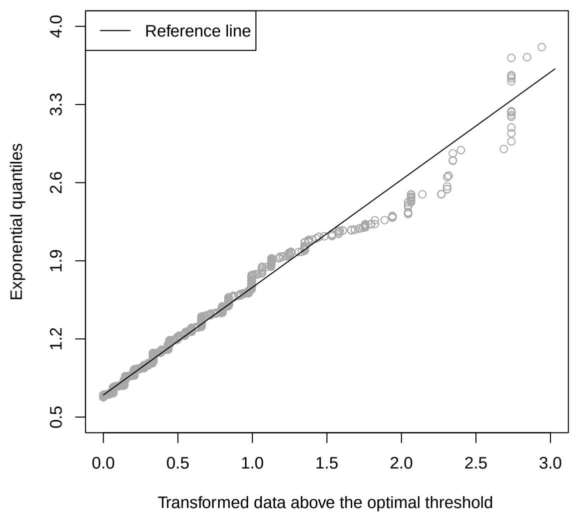
(a) K-S statistic vs. rank of ordered sample values



(b) Hill plot

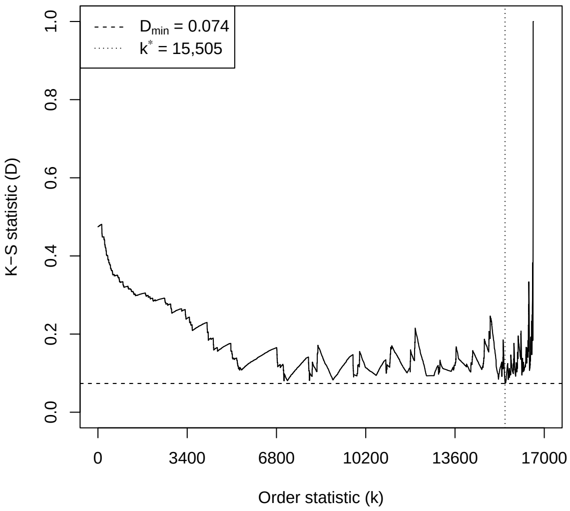


(c) Best-fit Pareto model for the upper tail

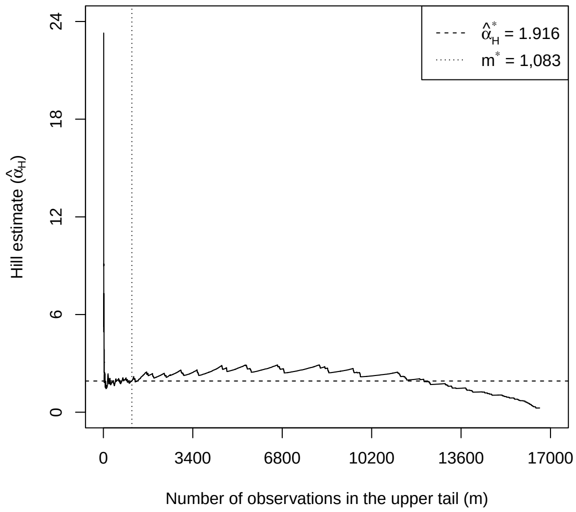


(d) Pareto Q-Q plot

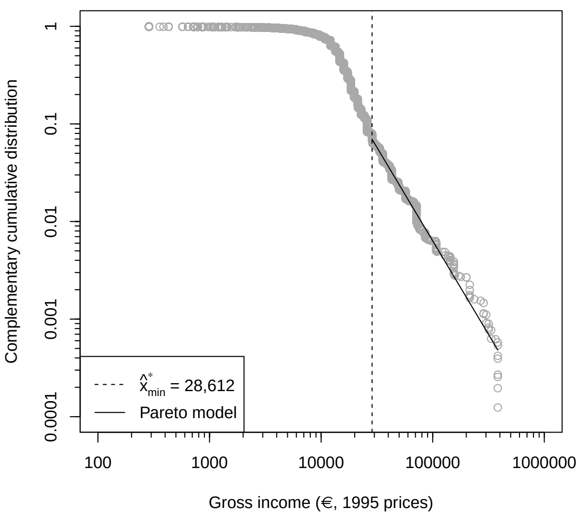
Figure 4 – Pareto distribution fit for the PLUS 2010 wave



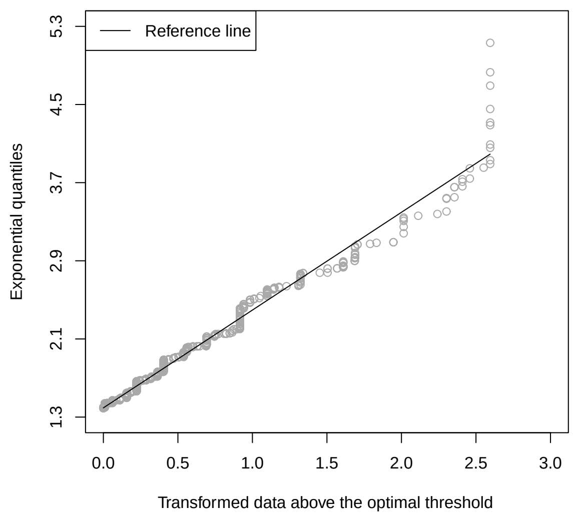
(a) K-S statistic vs. rank of ordered sample values



(b) Hill plot



(c) Best-fit Pareto model for the upper tail



(d) Pareto Q-Q plot

Figure 5 – Comparison between 2008 and 2010 Italian labour income distributions: the mean shift effect

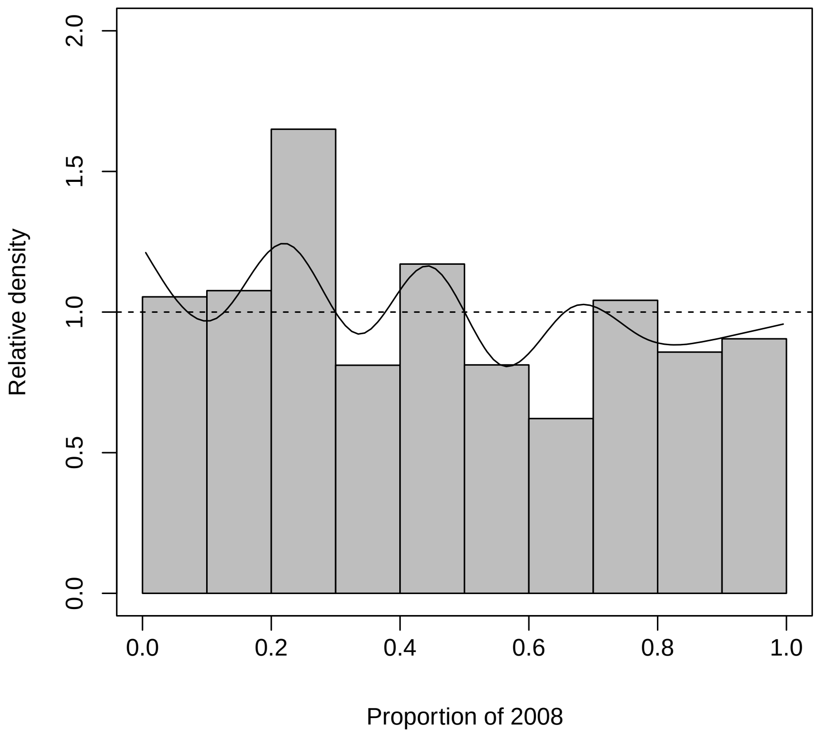
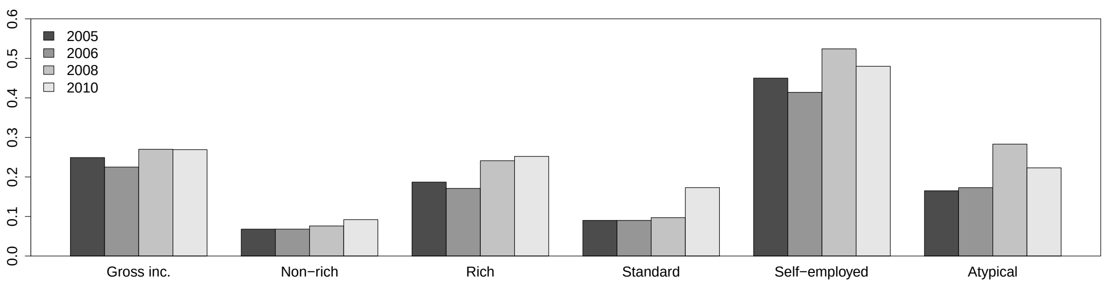
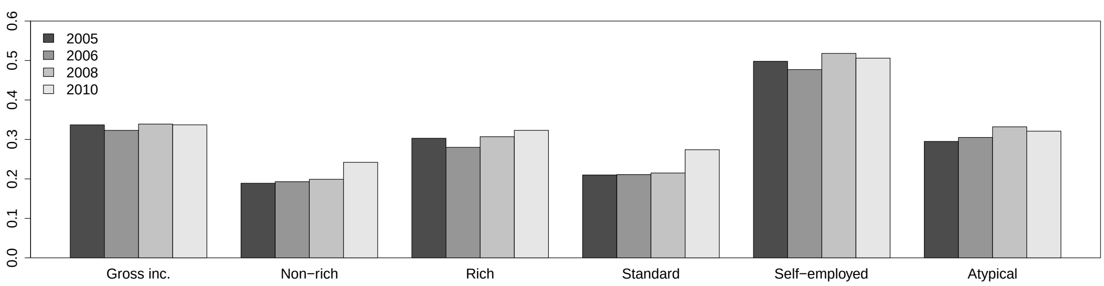


Figure 6 – Measures of the amount of inequality in the Italian labour income distribution by population subgroups and income sources



(a) Theil



(b) Gini

DiSSE Working Papers

- n.41: Cutrini, E., Galeazzi, G. *Can emerging economies decouple from the US business cycle?*
- n.40: Cutrini, E., Micucci, G., Montanaro, P. *I distretti tradizionali di fronte alla globalizzazione: il caso dell'industria calzaturiera marchigiana*
- n.39: Bade, F.-J., Bode, E., Cutrini, E. *Spatial fragmentation of industries by functions*
- n.38: Gentilucci, E., Herrera, R. *Un'analisi critica dei lavori recenti del mainstream sugli effetti economici delle spese militari*
- n.37: Stefani G., Cavicchi A. *Consumer evaluation of a typical Italian salami: an experimental auction approach*
- n.36: Valentini E. *Giving Voice to Employees and Spreading Information within the Firm: the Manner Matters*
- n.35: Cutrini E., Valentini E. *What drives economic specialization in Italian Regions?*
- n.34: Spigarelli F., Goldstein A., Manzetti L. *Italian economic diplomacy at work: catching up the BRICs*
- n.33: Cutrini E., Spigarelli F. *Italian FDI integration with Southeast Europe: country and firm-level evidence*
- n.32: Davino C., Romano R. *Sensitivity Analysis of Composite Indicators through Mixed Model Anova*
- n.31: Rocchi B., Cavicchi A., Baldeschi M. *Consumers' attitude towards farmers' markets: an explorative analysis in Tuscany*
- n.30: Trinchera L., Russolillo G. *On the use of Structural Equation Models and PLS Path Modeling to build composite indicators*
- n.29: Tavoletti E. *The internationalization process of Italian fashion firms: the governance role of the founding team*
- n.28: Croci Angelini E. *Globalization and public administration: a complex relationship*
- n.27: Tavoletti E. *Matching higher education and labour market in the knowledge economy: the much needed reform of university governance in Italy*
- n.26: Ciaschini M., Pretaroli R., Severini F., Socci C. *The economic impact of the Green Certificate market through the Macro Multiplier approach*
- n.25: Ciaschini M., Pretaroli R., Severini F., Socci C. *Environmental tax reform and double dividend evidence*
- n.24: Atkinson A. B. *Poverty and the EU: the New Decade*

- n.23: Cutrini E. *Moving Eastwards while Remaining Embedded: the Case of the Marche Footwear District, Italy*
- n.22: Valentini E. *On the Substitutability between Equal Opportunities and Income Redistribution*
- n.21: Ciaschini M., Pretaroli R., Socci C. *La produzione di servizi sanitari e la variazione delloutput nei principali paesi UE*
- n.20: Cassiani M., Spigarelli F. *Gli hedge fund: caratteristiche, impatto sui mercati e ruolo nelle crisi nanziarie*
- n.19: Cavicchi A. *Regolamentazione e gestione del rischio nel settore agroalimentare. Alcune riflessiioni sull'approccio economico al Principio di Precauzione*
- n.18: Spalletti S. *The History of Manpower Forecasting in Modelling Labour Market*
- n.17: Boffa F., Pingali V. *MIncreasing Market Interconnection: an analysis of the Italian Electricity Spot Market*
- n.16: Scoppola M. *Tariffication of Tariff Rate Quotas under oligopolistic competition: the case of the EU import regimes for bananas*
- n.15: Croci Angelini E., Michelangeli A. *Measuring Well-Being differences across EU Countries. A Multidimensional Analysis of Income, Housing, Health, and Education*
- n.14: Fidanza B. *Quale comparable per la valutazione tramite multipli delle imprese Italiane?*
- n.13: Pera A. *Changing Views of Competition and EC Antitrust Law*
- n.12: Spigarelli F. *Nuovi investitori globali: le imprese cinesi in Italia*
- n.11: Ciaschini M., Pretaroli R., Socci C. *A convenient multi sectoral policy control for ICT in the USA economy*
- n.10: Tavoletti E., te Velde R. *Cutting Porter's last diamond: competitive and comparative (dis)advantages in the Dutch flower industry. Which lessons for Italian SMEs?*
- n.9: Tavoletti E. *The local and regional economic role of universities: the case of the University of Cardiff*
- n.8: Croci Angelini E. *Resisting Globalization: Voting Power Indices and the National Interest in the EU Decision-making*
- n.7: Minervini F., Piacentino D. *Spectrum Management and Regulation: Towards a Full-Fledged Market for Spectrum Bands?*
- n.6: Spalletti S. *Dalle analisi della crescita all'economia dell'istruzione e al capitale umano. Origine e sviluppo*
- n.5: Ciaschini M., Fiorillo F., Pretaroli R., Severini F., Socci C., Valentini E. *Politiche per l'industria: ridurre o abolire l'Irap?*

- n.4: Scoppola M. *Economies of scale and endogenous market structures in international grain trade*
- n.3: De Grauwe P. *What have we learnt about monetary integration since the Maastricht Treaty?*
- n.2: Ciaschini M., Pretaroli R., Socci C. *A convenient policy control through the Macro Multiplier Approach*
- n.1: Cave M. *The Development of Telecommunications in Europe: Regulation and Economic Effects*

Centro eum edizioni Università di Macerata

ISSN: 1971-890X