Pareto's Law of Income Distribution: Evidence for Germany, the United Kingdom, and the United States

F. Clementi [a,c,*], M. Gallegati [b,c]

[a] *Department of Public Economics, University of Rome 'La Sapienza', Via del Castro Laurenziano 9, 00161 Rome, Italy*

[b] *Department of Economics, Università Politecnica delle Marche, Piazzale Martelli 8, 60121 Ancona, Italy*

[c] *S.I.E.C., Università Politecnica delle Marche, Piazzale Martelli 8, 60121 Ancona, Italy*

Abstract

We analyze three sets of income data: the US Panel Study of Income Dynamics (PSID), the British Household Panel Survey (BHPS), and the German Socio-Economic Panel (GSOEP). It is shown that the empirical income distribution is consistent with a two-parameter lognormal function for the low-middle income group (97%–99% of the population), and with a Pareto or power law function for the high income group (1%–3% of the population). This mixture of two qualitatively different analytical distributions seems stable over the years covered by our data sets, although their parameters significantly change in time. It is also found that the probability density of income growth rates almost has the form of an exponential function.

Key words: Personal income, lognormal distribution, Pareto's law, income growth rate
PACS: 02.60.Ed, 89.75.Da, 89.65.Gh

* Corresponding author. Tel.: +39–06–49–766–843; fax: +39–06–44–61–964.
Email addresses: `fabio.clementi@uniroma1.it` (F. Clementi), `gallegati@dea.unian.it` (M. Gallegati).

1 Introduction

More than a century ago, the economist Vilfredo Pareto stated in his *Cours d'Économie Politique* that there is a simple law which governs the distribution of income in all countries and at all times. Briefly, if N represents all the number of income-receiving units cumulated from the top above a certain income limit x, and A and α are constants, then:

$$N = \frac{A}{x^{\alpha}} \tag{1}$$

and, therefore, $log\,(N) = log\,(A) - \alpha log\,(x)$. In other words, if the logarithms of the number of persons in receipt of incomes above definite amounts are plotted against the logarithms of the amount of these incomes, the points so obtained will be on a straight line whose slope with the axis on which the values of $log\,(x)$ are given will be α. Pareto examined the statistics of incomes in some countries and concluded that the inclination of the line with the $log\,(x)$ axis differed but little from 1.5.

Very recently, considerable investigations with modern data in capitalist economies have revealed that the upper tail of the income distribution (generally less than 5% of the individuals) indeed follows the above mentioned behaviour, and the variation of the slopes both from time to time and from country to country is large enough not to be negligible. Hence, characterization and understanding of income distribution is still an open problem. The interesting problem that remains to be answered is the functional form more adequate for the majority of population not belonging to the power law part of the income distribution. Using data coming from several parts of the world, a number of recent studies debate whether the low-middle income range of the income distribution may be fitted by an exponential [1–8] or lognormal [9–13] decreasing function. [1]

In this paper we have analyzed three data sets relating to a pool of major industrialized countries for several years in order to add some empirical investigations to the ongoing debate on income distribution. When fits are performed, a two-parameter lognormal distribution is used for the low-middle

[1] Recently, a distribution proposed by [14,15] has the form of a deformed exponential function:

$$P_{\kappa}\,(x) = \left(\sqrt{1 + \kappa^2 x^2} - \kappa x\right)^{\frac{1}{\kappa}}$$

which seems to capture well the behaviour of the income distribution at the low-middle range as well as the power law tail.

part of the distribution (97%–99% of the population), while the upper high-end tail (1%–3% of the population) is found to be consistent with a power law type distribution. Our results show that the parameters of income distribution change in time; furthermore, we find that the probability density of income growth rates almost scales as an exponential function.

The structure of the paper is as follows. Section 2 describes the data used in our study. Section 3 presents and analyzes the shape of the income distribution (Section 3.1) and its time development over the years covered by our data sets (Section 3.2). Section 4 concludes the paper.

2 The Data

We have used income data from the US Panel Study of Income Dynamics (PSID), the British Household Panel Survey (BHPS), and the German Socio-Economic Panel (GSOEP) as released in a cross-nationally comparable format in the Cross-National Equivalent File (CNEF). The CNEF brings together multiple waves of longitudinal data from the surveys above, and therefore provides relatively long panels of information. The current release of the CNEF includes data from 1980 to 2001 for the PSID, from 1991 to 2001 for the BHPS, and from 1984 to 2002 for the GSOEP. Our data refer to the period 1980–2001 for the United States, and to the period 1991–2001 for the United Kingdom. As the eastern states of Germany were reunited with the western states of the Federal Republic of Germany in November 1990, the sample of families in the East Germany was merged with the existing data only at the beginning of the 1990s. Therefore, in order to perform analyses that represent the population of reunited Germany, we chose to refer to the subperiod 1990–2002 for the GSOEP.

A key advantage of the CNEF is that it provides reliable estimates of annual income variables defined in a similar manner for all the countries that are not directly available in the original data sets.[2] It includes pre- and post-government household income, estimates of annual labour income, assets, private and public transfers, and taxes paid at household level. In this paper, the household post-government income variable (equal to the sum of total family income from labour earnings, asset flows, private transfers, private pensions, public transfers, and social security pensions minus total household taxes) serves as the basis for all income calculations. Following a generally accepted methodology, the concept of *equivalent income* will serve as a substitute for personal income, which is unobservable. Equivalent income x is calculated as

[2] Reference [16] offers a detailed description of the CNEF. See also the CNEF web site for details: http://www.human.cornell.edu/pam/gsoep/equivfil.cfm.

follows. In a first step, household income h is adjusted for by household type θ using an equivalence scale $e(\theta)$. [3] This adjusted household income $x = h/e(\theta)$ is then attributed to every member of the given household, which implies that income is distributed equally within households.

In the most recent release, the average sample size varies from about 7,300 households containing approximately 20,200 respondent individuals for the PSID-CNEF to 6,500 households with approximately 16,000 respondent individuals for the BHPS-CNEF; for the GSOEP-CNEF data from 1990 to 2002, we have about 7,800 households containing approximately 20,400 respondent individuals.

All the variables are in current year currency; therefore, we use the consumer price indices to convert into constant figures for all the countries. The base year is 1995.

3 Empirical Findings

3.1 The Shape of the Distribution

The main panel of the pictures illustrated in Fig. 1 presents the empirical cumulative distribution of the equivalent income from our data sets for some randomly selected years in the log-log scale. [4]

[Fig. 1 about here.]

As shown in the lower insets, the upper income tail (about 1%–3% of the population) follows the Pareto's law:

$$1 - F(x) = P(X \geq x) = C_\alpha x^{-\alpha} \tag{2}$$

where $C_\alpha = k^\alpha$, $k, \alpha > 0$, and $k \leq x < \infty$. Since the values of x above some value x_R can not be observed due to tail truncation, to fit the (logarithm of the) data for the majority of the population (until the 97^{th}–99^{th} percentiles of

[3] We use the so-called "modified OECD" equivalence scale, which is defined for each household as equal to $1 + 0.5 \times (\#\text{adults} - 1) + 0.3 \times (\#\text{children})$.

[4] To treat each wave of the surveys at hand as a cross-section, and to obtain population-based statistics, all calculations used sample weights which compensate for unequal probabilities of selection and sample attrition. Furthermore, to eliminate the influence of outliers, the data were trimmed. We also dropped observations with zero and negative incomes from all samples.

the income distribution) we use a right-truncated normal probability density function:

$$f(y) = \begin{cases} \dfrac{f(y)}{\int_{-\infty}^{y_R} f(y)dy} \quad , & -\infty < y \leq y_R \\ 0 \quad , & y_R \leq y < \infty \end{cases} \tag{3}$$

where $y = log\,(x)$, and $y_R = log\,(x_R)$. The fit to Equation (3) is shown by the top insets of the pictures.

To select a suitable threshold or cutoff value x_R separating the lognormal part from the Pareto power law tail of the empirical income distribution, we use visually oriented statistical techniques such as the *quantile-quantile* (Q-Q) and *mean excess* plots. Figure 2 gives an example of these graphical tools for the countries at hand.

[Fig. 2 about here.]

The left pictures in the figure are the plots of the quantile function for the standard exponential distribution (i.e., a distribution with a medium-sized tail) against its empirical counterpart. If the sample comes from the hypothesized distribution, or a linear transformation of it, the Q-Q plot is linear. The concave presence in the plots is an indication of a fat-tailed distribution. Since a log-transformed Pareto random variable is exponentially distributed, we conduct experimental analysis on the log-transformed data by excluding some of the lower sample points to investigate the concave departure region on the plots and obtain a fit closer to the straight line. The results are shown by the insets of the left pictures in the figure. The right pictures plot the empirical average of the data that are larger than or equal to x_R, $E\,(X|X \geq x_R)$, against x_R. If the plot is a linear curve, then it may be either a power type or an exponential type distribution. If the slope of the linear curve is greater than zero, then it suggests a power type (as in the main panels); otherwise, if the slope is equal to zero, it suggests an exponential type (as in the insets for the log-transformed data).

3.2 Temporal Change of the Distribution

The two-part structure of the empirical income distribution seems to hold all over the time span covered by our data sets. The distribution for all the years and countries are shown in Fig. 3.

[Fig. 3 about here.]

As one can easily recognize, the distribution shifts over the years covered by our data sets. It is conceivable to assume that the origin of this shift consists in the growth of the countries. To confirm this assumption, we study the fluctuations in the output and equivalent income growth rate, and try to show that the evolution of both these quantities is governed by similar mechanisms, pointing in this way to the existence of a correlation between them as one would expect. We calculate the growth rates using the monthly series of the Index of Industrial Production (IIP) from [17] for output and connecting individual respondents' incomes over time for the equivalent income,[5] and express them in terms of their logarithm.[6] To account for the fact that the variance of the growth rates varies, we scale each growth rate by dividing by the corresponding estimated standard deviation. In Fig. 4 we graph the empirical probability density function for these scaled growth rates, where the data points for the equivalent income in the main panels are the average over the entire period covered by the CNEF surveys.

[Fig. 4 about here.]

As one can easily recognize, after scaling the resulting empirical probability density functions appear identical for observations drawn from different populations. Remarkably, both curves display a simple "tent-shaped" form; hence, the probability density functions are consistent with an exponential decay [18]:

$$f(r) = \frac{1}{\sigma\sqrt{2}} exp\left(-\frac{|r-\bar{r}|}{\sigma}\right) \tag{4}$$

where $-\infty < r < \infty$, $-\infty < \bar{r} < \infty$, and $\sigma > 0$. We test the hypothesis that the two growth rate distributions have the same continuous distribution by using the two-sample Kolmogorov-Smirnov (K-S) test; the results shown in Table 1 mean that the test is not significant at the 5% level.

[Table 1 about here.]

These findings are in quantitative agreement with results reported on the growth of firms and countries [19–26], leading us to the conclusion that the data are consistent with the assumption that a common empirical law might describe the growth dynamics of both countries and individuals.

Even if the functional form of the income distribution expressed as lognormal with power law tail seems stable, its parameters fluctuate within narrow bounds over the years for the same country. For example, the power law slope

[5] To properly weight the sample of individuals represented in all the years of the CNEF surveys, we use the individual's longitudinal sample weights.
[6] All the data have been adjusted to 1995 prices and detrended by the average growth rate, so values for different years are comparable.

has a value $\alpha = [1.1, 3.34]$ for the US between 1980 and 2001, while the curvature of the lognormal fit, as measured by the Gibrat index $\beta = 1/\left(\sigma\sqrt{2}\right)$, ranges between approximately $\beta = 1$ and $\beta = 1.65$; for the UK between 1991 and 2001, $\alpha = [3.47, 5.76]$ and $\beta = [2.18, 2.73]$; for Germany between 1990 and 2002, $\alpha = [2.42, 3.96]$ and $\beta = [1.63, 2.14]$. The time pattern of these parameters is shown by the main panels of Fig. 5, which also reports in one of the insets the temporal change of inequality as measured by the Gini coefficient.

[Fig. 5 about here.]

As one can easily recognize, the information about inequality provided by the Gibrat index seems near enough to those provided by the Gini coefficient, which is a further confirmation of the fact that the lognormal law is a good model for the low-middle incomes of the distribution. The Pareto index is a rather strongly changing index. Among others, the definition of income we use in the context of our analysis contains asset flows. It is conceivable to assume that for the top 1% to 3% of the population returns on capital gains rather than labour earnings account for the majority share of the total income. This suggests that the stock market fluctuations might be an important factor behind the trend of income inequality among the richest, and that capital income plays an important role in determining the Pareto functional form of the observed empirical income distribution at the high income range [27]. The other insets of the pictures also show the time evolution of various parameters characterizing income distribution, such as the income separating the lognormal and Pareto regimes (selected as explained in Section 3.1), the fraction of population in the upper tail of the distribution, and the share of total income which this fraction accounts for.[7] One can observe that the fraction of population and the share of income in the Pareto tail move together in the opposite direction with respect to the cutoff value separating the body of the distribution from its tail, and the latter seems to track the temporal evolution of the Pareto index. This fact means that a decrease (increase) of the power law slope and the accompanying decrease (increase) of the threshold value x_R imply a greater (smaller) fraction of the population in the tail and a greater (smaller) share of the total income which this population account for, as well as a greater (smaller) level of inequality among high income population.

[7] The share of total income in the tail of the distribution is calculated as μ_α/μ, where μ_α is the average income of the population in the Pareto tail and μ is the average income of the whole population.

4 Summary and Conclusions

Our analysis of the data for the US, the UK, and Germany shows that there are two regimes in the income distribution. For the low-middle class up to approximately 97%–99% of the total population the incomes are well described by a two-parameter lognormal distribution, while the incomes of the top 1%–3% are described by a power law (Pareto) distribution.

This structure has been observed in our analysis for different years. However, the distribution shows a rightward shift in time. Therefore, we analyze the output and individual income growth rate distribution from which we observe that, after scaling, the resulting empirical probability density functions appear similar for observations coming from different populations. This effect, which is statistically tested by means of a two-sample Kolmogorov-Smirnov test, raises the intriguing possibility that a common mechanism might characterize the growth dynamics of both output and individual income, pointing in this way to the existence of a correlation between these quantities. Furthermore, from the analysis of the temporal change of the parameters specifying the distribution, we find that these quantities do not necessarily correlate to each other. This means that different mechanisms are working in the distribution of the low-middle income range and that of the high income range. Since earnings from financial or other assets play an important role in the high income section of the distribution, one possible origin of this behaviour might be the change of the asset price, which mainly affects the level of inequality at the very top of the income distribution and is likely to be responsible for the power law nature of high incomes.

Acknowledgements

The authors wish to thank the organizers of the *International Workshop on* `Econophysics of Wealth Distributions` (ECONOPHYS - KOLKATA I, 15–19 March, 2005, SINP, Kolkata) for the very nice hospitality and all the invited participants for the useful interactions.

References

[1] Nirei M, Souma W (2004), Two Factors Model of Income Distribution Dynamics. SFI Working Paper: http://www.santafe.edu/research/publications/workingpapers/04-10-029.pdf

[2] Drăgulescu AA, Yakovenko VM (2001), Evidence for the Exponential Distribution of Income in the USA. The European Physical Journal B 20:585–589

[3] Drăgulescu AA, Yakovenko VM (2001), Exponential and Power-Law Probability Distributions of Wealth and Income in the United Kingdom and the United States. Physica A 299:213–221

[4] Drăgulescu AA (2003), Applications of Physics to Economics and Finance: Money, Income, Wealth, and the Stock Market. E-print: http://arXiv.org/cond-mat/0307341

[5] Silva AC, Yakovenko VM (2005), Temporal Evolution of the "Thermal" and "Superthermal" Income Classes in the USA During 1983–2001. Europhysics Letters 69:304–310

[6] Willis G, Mimkes J (2004), Evidence for the Independence of Waged and Unwaged Income, Evidence for Boltzmann Distributions in Waged Income, and the Outlines of a Coherent Theory of Income Distribution. E-print: http://arXiv.org/cond-mat/0406694

[7] Chatterjee A, Chakrabarti BK, Manna SS (2003), Money in Gas-Like Markets: Gibbs and Pareto Laws. E-print: http://arXiv.org/cond-mat/0311227

[8] Sinha S (2005) Evidence for Power-law Tail of the Wealth Distribution in India. E-print: http://arXiv.org/cond-mat/0502166

[9] Montroll EW, Shlesinger MF (1982), Maximum Entropy Formalism, Fractals, Scaling Phenomena, and $1/f$ Noise: A Tale of Tails. Journal of Statistical Physics 32:209–230

[10] Souma W (2001), Universal Structure of the Personal Income Distribution. Fractals 9:463–470

[11] Souma W (2002), Physics of Personal Income. E-print: http://arXiv.org/cond-mat/0202388

[12] Di Matteo T, Aste T, Hyde ST (2003), Exchanges in Complex Networks: Income and Wealth Distributions. E-print: http://arXiv.org/cond-mat/0310544

[13] Clementi F, Gallegati M (2005), Power Law Tails in the Italian Personal Income Distribution. Physica A 350:427–438

[14] Kaniadakis G (2001), Non-linear Kinetics Underlying Generalized Statistics. Physica A 296:405–425

[15] Kaniadakis G (2002), Statistical Mechanics in the Context of Special Relativity. E-print: http://arXiv.org/cond-mat/0210467

[16] Burkhauser RV, Butrica BA, Daly MC, Lillard DR (2001) The Cross-National Equivalent File: A Product of Cross-national Research. In: Becker I, Ott N, Rolf G (eds) Soziale Sicherung in einer dynamischen Gesellschaft. Festschrift für Richard Hauser zum 65. Geburtstag. Campus, Frankfurt/New York

[17] OECD Statistical Compendium 2003#2

[18] Kim K, Yoon SM (2004), Power Law Distributions in Korean Household Incomes. E-print: http://arXiv.org/cond-mat/0403161

[19] Sutton J (2001) The Variance of Firm Growth Rates: The Scaling Puzzle. STICERD – Economics of Industry Papers 27, Suntory and Toyota International Centres for Economics and Related Disciplines, LSE

[20] Stanley MHR, Amaral LAN, Buldyrev SV, Havlin S, Leschhorn H, Maass P, Salinger MA, Stanley HE (1996) Scaling Behavior in the Growth of Companies. Nature 379:804–806

[21] Amaral LAN, Buldyrev SV, Havlin S, Leschhorn H, Maass P, Salinger MA, Stanley HE, Stanley MHR (1997) Scaling Behavior in Economics: I. Empirical Results for Company Growth. J. Phys. I France 7:621–633

[22] Amaral LAN, Buldyrev SV, Havlin S, Leschhorn H, Maass P, Salinger MA, Stanley HE, Stanley MHR (1997) Scaling Behavior in Economics: The Problem of Quantifying Company Growth. Physica A 244:1–24

[23] Lee Y, Amaral LAN, Canning D, Meyer M, Stanley HE (1998) Universal Features in the Growth Dynamics of Complex Organizations. Physical Review Letters 81:3275–3278

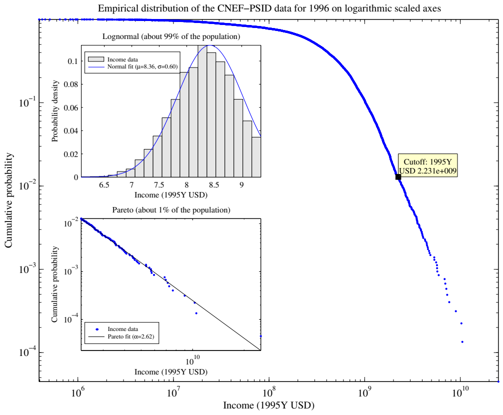
[24] Bottazzi G, Secchi A (2002) On The Laplace Distribution of Firms Growth Rates. LEM Working Paper no. 2002–20

[25] Bottazzi G, Secchi A (2003) A Stochastic Model of Firm Growth. Physica A 324:213–219

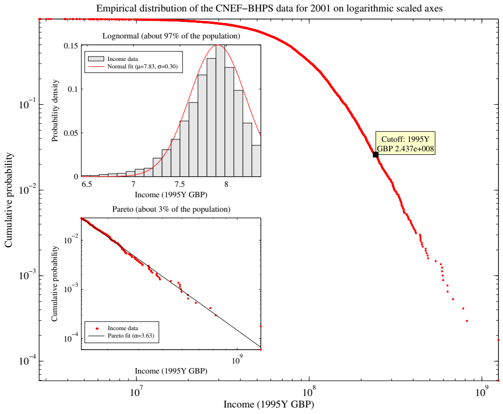
[26] Bottazzi G, Secchi A (2003) Why Are Distributions of Firm Growth Rates Tent-shaped? Economic Letters 80:415–420

[27] Levy M (2003) Are Rich People Smarter? Journal of Economic Theory 110:42–64

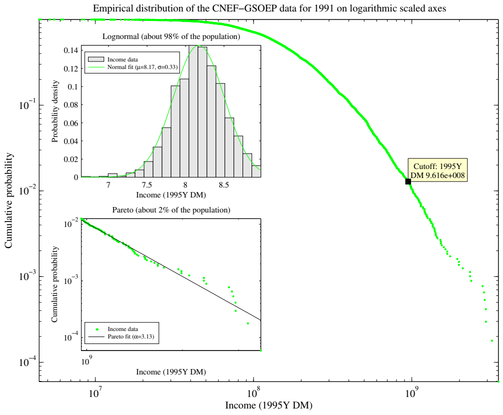
Figures



(a) United States (1996)

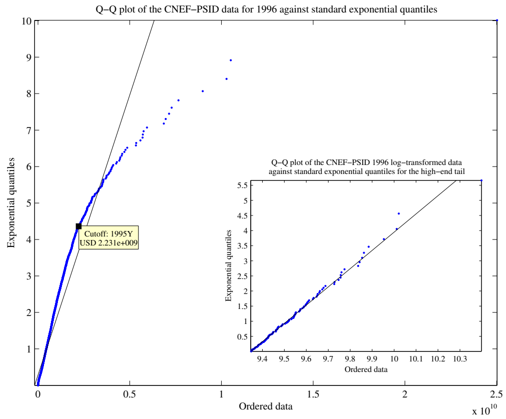


(b) United Kingdom (2001)

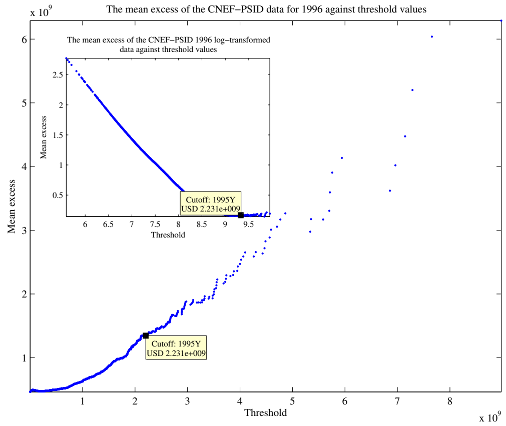


(c) Germany (1991)

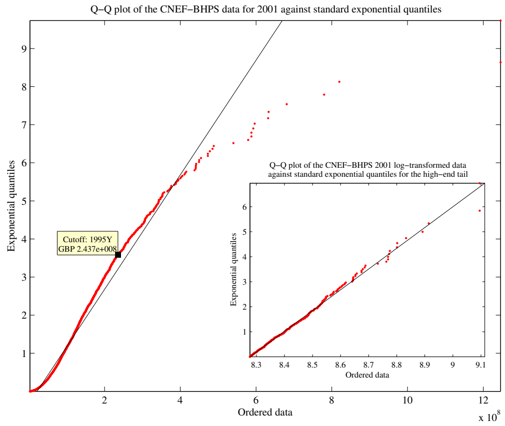
Fig. 1. The cumulative probability distribution of the equivalent income in the log-log scale along with the lognormal (top insets) and Pareto (lower insets) fits for some randomly selected years



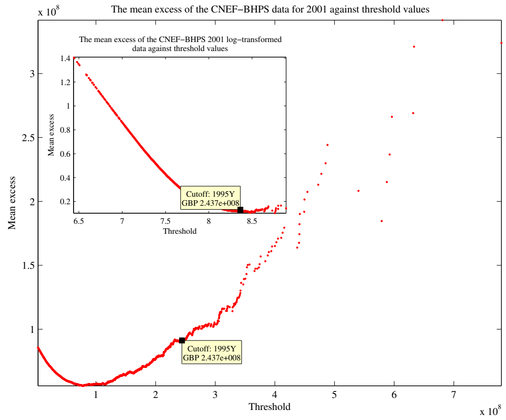
(a) United States (1996)



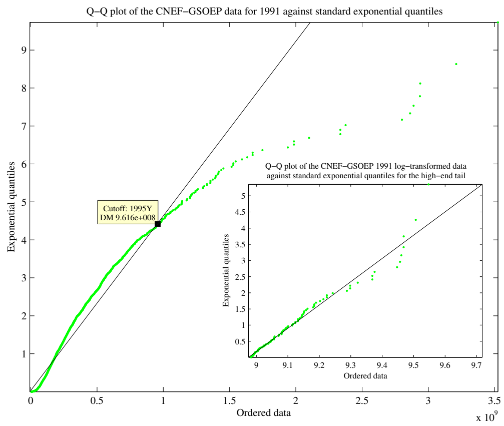
(b) United States (1996)



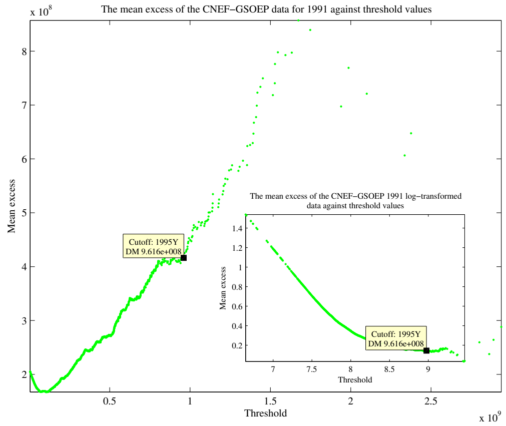
(c) United Kingdom (2001)



(d) United Kingdom (2001)

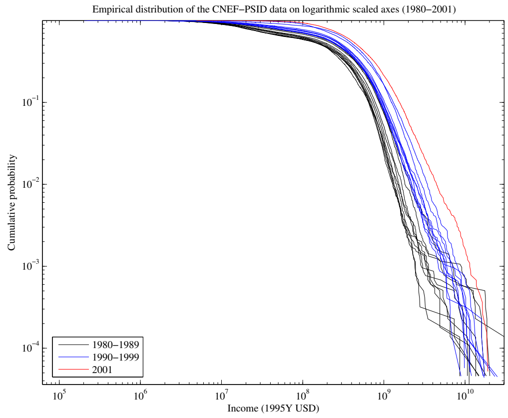


(e) Germany (1991)

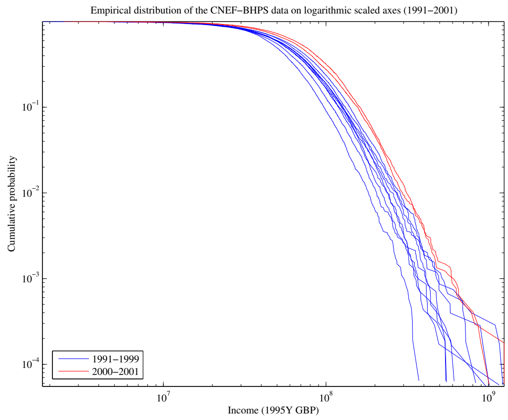


(f) Germany (1991)

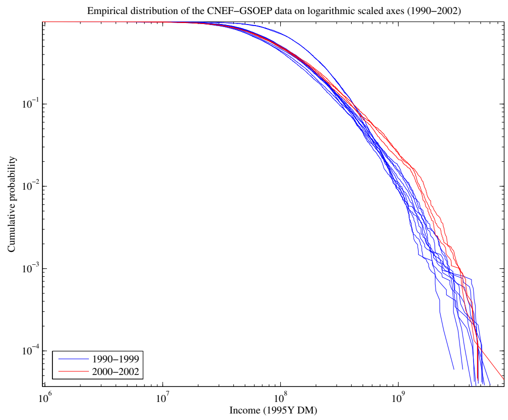
Fig. 2. Q-Q plots (left pictures) against standard exponential quantiles and mean excess plots (right pictures) against threshold values for some randomly selected years. A concave departure from the straight line in the Q-Q plot (as in the left main panels) or an upward sloping mean excess function (as in the right main panels) indicate a heavy tail in the sample distribution. The insets in the pictures apply the same graphical tools to the log-transformed data



(a) United States (1980–2001)

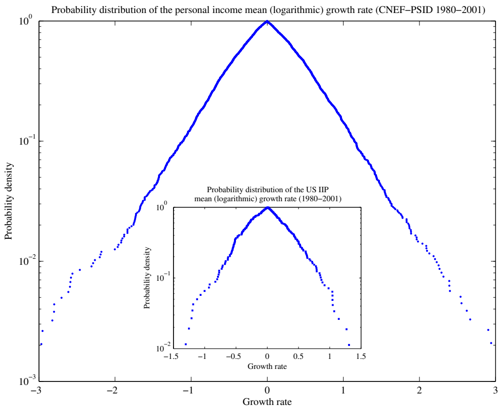


(b) United Kingdom (1991–2001)

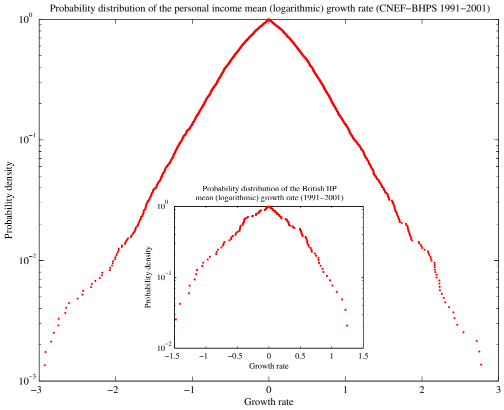


(c) Germany (1990–2002)

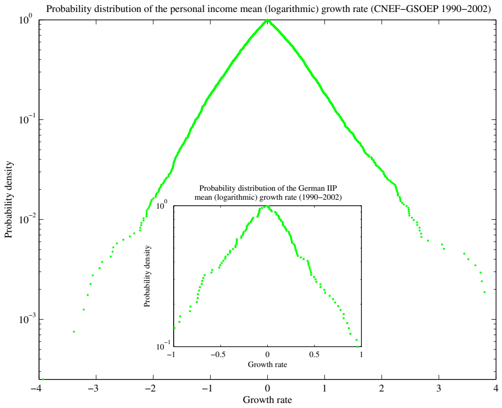
Fig. 3. Time development of the income distribution for all the countries and years



(a) United States (1980–2001)

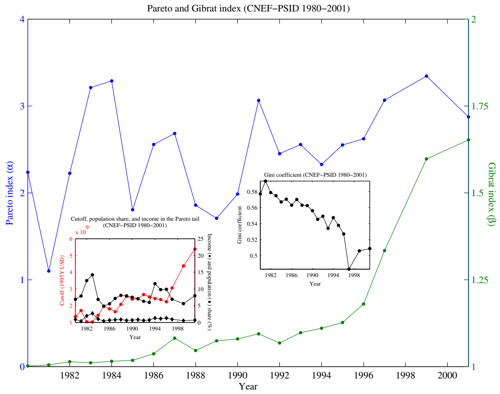


(b) United Kingdom (1991–2001)

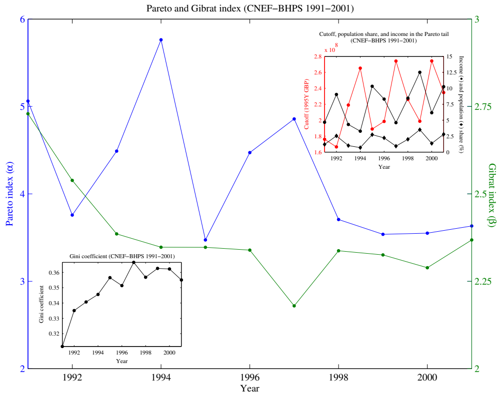


(c) Germany (1990–2002)

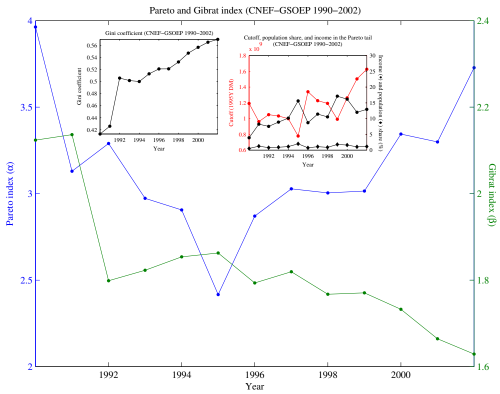
Fig. 4. The probability distribution of equivalent income (main panels) and IIP (insets) growth rate for all the countries and years



(a) United States (1980–2001)



(b) United Kingdom (1991–2001)



(c) Germany (1990–2002)

Fig. 5. Temporal evolution of various parameters characterizing the income distribution

Tables

Table 1
Two-sample Kolmogorov-Smirnov test statistics and p-values for both output and equivalent income growth rate data for all the countries

Country	K-S test statistic	p-value
United States	0.0761	0.1133
United Kingdom	0.0646	0.6464
Germany	0.0865	0.2050