A new model of income distribution: The κ-generalized distribution

Fabio Clementi · **Mauro Gallegati** · **Giorgio Kaniadakis**

6 June 2011

Abstract This paper proposes a three-parameter statistical model of income distribution by exploiting recent developments on the use of deformed exponential and logarithm functions as suggested by Kaniadakis (2001, 2002, 2005). Formulas for the shape, moments and standard tools for inequality measurement are given. The model is shown to fit remarkably well the personal income data for Great Britain, Germany and the United States in different years, and its empirical performance appears to be competitive with that of other existing distributions.

Keywords Income distribution · Income inequality · κ-generalized distribution

JEL Classification C16 · D31

1 Introduction

In the analysis of income distribution, analysts have found it useful to have a mathematical description for the size distribution of income based on estimates of functional forms. Indeed, a parametric model of the way the empirical distributions look like enables to get an easier grip upon particular features of the income distribution and can be useful in a variety of applications, from the comparison of distributions in different populations and/or over time, to the measurement of inequality and the elaboration of income redistribution policies, up to the characterization of the solution to economic models of the income distribution process.

In applied work, the principal functional forms used as descriptors of the distribution of income have been the lognormal or Pareto densities. Both distributions arise from stochastic models of income growth. For example, Gibrat (1931) demonstrated that if income growth rates are random and independent on the initial size, then lognormality of incomes occurs

The authors acknowledge Springer Verlag. The original publication is available at `http://www.springerlink.com/content/j4875556n6666qu9/`.

F. Clementi(✉)
Department of Studies on Economic Development, University of Macerata,
Piazza G. Oberdan 3, 62100 Macerata, Italy
Tel.: +39-0733-258-3962
Fax: +39-0733-258-3970
E-mail: `fabio.clementi@unimc.it`

M. Gallegati
Department of Economics, Polytechnic University of Marche,
Piazzale R. Martelli 8, 60121 Ancona, Italy

G. Kaniadakis
Department of Physics, Polytechnic University of Turin,
Corso Duca degli Abruzzi 24, 10129 Torino, Italy

irrespective of the initial distribution. Variations on this theme are able to generate Pareto distributions instead (e.g. Champernowne 1953). However, these simple models are evidently limited in the variety of shapes of income distribution that they can be expected to describe. The problem is that neither the lognormal nor the Pareto density provides an adequate fit to actual data. The former does not adequately fit the tails of the distribution, while the latter only fits the very top of the distribution. Moreover, both densities are outperformed in terms of goodness-of-fit by parsimonious alternatives (e.g. the gamma distribution proposed by Salem and Mount 1974). Subsequent work indicated that still further improvements could be made by considering families of densities flexible enough to be able to capture the prominent features of the observed income distributions. For instance, McDonald and Xu (1995) proposed the generalized beta, a quite general family of probability density functions that nests most of the income distributions introduced in the econometrics literature as special or limiting cases. Of these, the Singh-Maddala (1976) and Dagum (1977) distributions have shown them to be a good compromise between parsimony and goodness-of-fit in many instances.

In this paper we introduce a new three-parameter distribution based on the generalization of the exponential and logarithm functions proposed by Kaniadakis (2001, 2002, 2005) and defined as

$$\exp_\kappa(x) = \left(\sqrt{1+\kappa^2 x^2} + \kappa x\right)^{\frac{1}{\kappa}}, \quad x \in \mathbb{R}, \tag{1a}$$

$$\log_\kappa(x) = \frac{x^\kappa - x^{-\kappa}}{2\kappa}, \quad x \in \mathbb{R}_+, \tag{1b}$$

which lead to the standard exponential and logarithm when $\kappa \to 0$ and exhibit a power-law behavior at the boundaries of their support, i.e.

$$\exp_\kappa(x) \underset{x\to\pm\infty}{\sim} \left|2\kappa x\right|^{\pm\frac{1}{|\kappa|}}, \tag{2a}$$

$$\log_\kappa(x) \underset{x\to 0^+}{\sim} -\frac{1}{2|\kappa|} x^{-|\kappa|}, \tag{2b}$$

$$\log_\kappa(x) \underset{x\to +\infty}{\sim} \frac{1}{2|\kappa|} x^{|\kappa|}. \tag{2c}$$

Formally, the distribution can be obtained by maximizing according to Jaynes (1957a,b) maximum entropy principle the Shannon (1948) information measure

$$S \equiv -\int_0^\infty f(x)\log f(x)\,\mathrm{d}x \tag{3}$$

under the natural constraint that normalizes the density,

$$\int_0^\infty f(x)\,\mathrm{d}x = 1, \tag{4}$$

and the three characterizing moments

$$\int_0^\infty \log x f(x)\,\mathrm{d}x = \log\beta - \frac{1}{\alpha}\left[\gamma + \psi\left(\frac{1}{2\kappa}\right) + \log(2\kappa) + \kappa\right], \tag{5a}$$

$$\int_0^\infty \log\left[1+\kappa^2\left(\frac{x}{\beta}\right)^{2\alpha}\right] f(x)\,\mathrm{d}x = 2\kappa - \psi\left(1+\frac{1}{4\kappa}\right) + \psi\left(\frac{1}{2}+\frac{1}{4\kappa}\right), \tag{5b}$$

$$\int_{0}^{\infty} \log\left[\sqrt{1+\kappa^2\left(\frac{x}{\beta}\right)^{2\alpha}}-\kappa\left(\frac{x}{\beta}\right)^{\alpha}\right] f(x)\,\mathrm{d}x = \int_{0}^{\infty} \sinh^{-1}\left[-\kappa\left(\frac{x}{\beta}\right)^{\alpha}\right] f(x)\,\mathrm{d}x = \kappa. \tag{5c}$$

The solution to the variational problem (3)–(5), obtainable using the method of Lagrange multipliers,[1] is given by what we call the κ-generalized distribution

$$f(x;\alpha,\beta,\kappa) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} \frac{\exp_\kappa\left[-(x/\beta)^\alpha\right]}{\sqrt{1+\kappa^2(x/\beta)^{2\alpha}}}, \quad x \geq 0, \tag{6}$$

where $\alpha > 0$ is a shape parameter, $\beta > 0$ is a scale and $\kappa \in [0,1)$ measures the heaviness of the right tail.[2]

In a previous work of us (Clementi et al. 2010) the above functional form was adopted successfully in modelling the personal income distribution in Italy. However, one may imagine that income in other countries is distributed differently. By using sets of data for three developed economies (Great Britain, Germany and the United States) across several years, in this paper we draw conclusions about the quality of the proposed model in describing the distribution of income more generally than just the Italian characteristics. The basic proposition is that the density (6) provides a very good description of the observed income distributions, ranging from the low region to the middle region, and up to the right tail.

The plan of the paper is as follows: Section 2 proposes the new distribution and examines its theoretical properties. Section 3 studies its empirical performance and makes comparisons with other existing distributions. Section 4 summarizes the paper.

2 The distribution and its properties

2.1 Definitions and interrelations

A random variable X is said to have a κ-generalized distribution if it has the probability density function (6). The corresponding cumulative distribution function reads

$$F(x;\alpha,\beta,\kappa) = 1-\exp_\kappa\left[-(x/\beta)^\alpha\right], \tag{7}$$

and the quantile function equals

$$F^{-1}(u;\alpha,\beta,\kappa) = \beta\left[\log_\kappa\left(\frac{1}{1-u}\right)\right]^{\frac{1}{\alpha}}, \quad 0 < u < 1. \tag{8}$$

As $\kappa \to 0$, the distribution tends to the Weibull distribution; it can be easily verified that

$$\lim_{\kappa\to 0} F(x;\alpha,\beta,\kappa) = 1-\exp\left[-(x/\beta)^\alpha\right] \tag{9}$$

and

$$\lim_{\kappa\to 0} f(x;\alpha,\beta,\kappa) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}\exp\left[-(x/\beta)^\alpha\right]. \tag{10}$$

[1] See Appendix A of Clementi et al. (2010).

[2] Notice that the use of the entropy formalism in the analysis of income distribution is not new. For example, Ord et al. (1981), Kapur (1989) and Leipnik (1990) pointed out that several income distributions might be selected if one uses a criterion of maximum entropy. On the inequality side, the entropy-based measure of inequality proposed by Theil (1967) naturally contributed to the development of a general information-theoretic approach to the measurement of inequality (Cowell 1980a,b; and Cowell and Kuga 1981a,b)

For $x \to 0^+$ the distribution behaves similarly to the Weibull model (9) and (10), whereas for large x it approaches a Pareto distribution with scale $\beta(2\kappa)^{-\frac{1}{\alpha}}$ and shape $\frac{\alpha}{\kappa}$, i.e.

$$F(x;\alpha,\beta,\kappa) \underset{x\to+\infty}{\sim} 1-\left[\frac{\beta(2\kappa)^{-\frac{1}{\alpha}}}{x}\right]^{\frac{\alpha}{\kappa}} \tag{11}$$

and

$$f(x;\alpha,\beta,\kappa) \underset{x\to+\infty}{\sim} \frac{\frac{\alpha}{\kappa}\left[\beta(2\kappa)^{-\frac{1}{\alpha}}\right]^{\frac{\alpha}{\kappa}}}{x^{\frac{\alpha}{\kappa}+1}}, \tag{12}$$

thus satisfying the weak Pareto law (Kakwani 1980)

$$\lim_{x\to\infty}\frac{xf(x;\alpha,\beta,\kappa)}{1-F(x;\alpha,\beta,\kappa)}=\frac{\alpha}{\kappa}. \tag{13}$$

From (8) the median is

$$x_{\text{med}}=\beta\left[\log_\kappa(2)\right]^{\frac{1}{\alpha}}. \tag{14}$$

The distribution is unimodal, the mode being at

$$x_{\text{mode}}=\beta\left[\frac{\alpha^2+2\kappa^2(\alpha-1)}{2\kappa^2(\alpha^2-\kappa^2)}\right]^{\frac{1}{2\alpha}}\left\{\sqrt{1+\frac{4\kappa^2(\alpha^2-\kappa^2)(\alpha-1)^2}{\left[\alpha^2+2\kappa^2(\alpha-1)\right]^2}}-1\right\}^{\frac{1}{2\alpha}} \tag{15}$$

if $\alpha>1$; otherwise, the distribution is zero-modal with a pole at the origin.

2.2 Moments and other properties

The r^{th}-order moment about the origin of the κ-generalized distribution equals

$$\mu_r'=\int_0^\infty x^r f(x;\alpha,\beta,\kappa)\,\mathrm{d}x=\beta^r(2\kappa)^{-\frac{r}{\alpha}}\frac{\Gamma\left(1+\frac{r}{\alpha}\right)}{1+\frac{r}{\alpha}\kappa}\frac{\Gamma\left(\frac{1}{2\kappa}-\frac{r}{2\alpha}\right)}{\Gamma\left(\frac{1}{2\kappa}+\frac{r}{2\alpha}\right)}, \tag{16}$$

where $\Gamma(\cdot)$ denotes the gamma function, and exists for $\alpha<r<\frac{\alpha}{\kappa}$. Specifically, $\mu_1'=m$ is the mean of the distribution and

$$\sigma^2=\mu_2'-m^2=\beta^2(2\kappa)^{-\frac{2}{\alpha}}\left\{\frac{\Gamma\left(1+\frac{2}{\alpha}\right)}{1+2\frac{\kappa}{\alpha}}\frac{\Gamma\left(\frac{1}{2\kappa}-\frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{2\kappa}+\frac{1}{\alpha}\right)}-\left[\frac{\Gamma\left(1+\frac{1}{\alpha}\right)}{1+\frac{\kappa}{\alpha}}\frac{\Gamma\left(\frac{1}{2\kappa}-\frac{1}{2\alpha}\right)}{\Gamma\left(\frac{1}{2\kappa}+\frac{1}{2\alpha}\right)}\right]^2\right\} \tag{17}$$

is the variance. Hence, the coefficient of variation equals

$$CV_\kappa=\sqrt{2\frac{(\alpha+\kappa)^2}{\alpha+2\kappa}\frac{\Gamma\left(\frac{2}{\alpha}\right)}{\Gamma^2\left(\frac{1}{\alpha}\right)}\frac{\Gamma\left(\frac{1}{2\kappa}-\frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{2\kappa}+\frac{1}{\alpha}\right)}\frac{\Gamma^2\left(\frac{1}{2\kappa}+\frac{1}{2\alpha}\right)}{\Gamma^2\left(\frac{1}{2\kappa}-\frac{1}{2\alpha}\right)}-1}. \tag{18}$$

It is also possible to define the standardized measures of skewness and kurtosis, respectively given by

$$\gamma_1=\frac{\mu_3}{\sigma^3}=\frac{\mu_3'-3\mu_2'm+2m^3}{\sigma^3} \tag{19}$$

and

$$\gamma_2=\frac{\mu_4}{\sigma^4}=\frac{\mu_4'-4\mu_3'm-6\mu_2'm^2-3m^4}{\sigma^4}, \tag{20}$$

where $\mu_r = \sum_{j=0}^{r} \binom{r}{j} (-1)^{r-j} \mu_j' m^{r-j}$ is the moment about the mean.

2.3 Lorenz curve and inequality measures

The Lorenz (1905) curve of the κ-generalized distribution is available in terms of the first-moment distribution $L(F(x)) = m^{-1} \int_0^x x' p\left(x'\right) \mathrm{d}x'$, namely

$$\begin{aligned} L_\kappa(u) =& 1 - \frac{1+\frac{\kappa}{\alpha}}{2\Gamma\left(\frac{1}{\alpha}\right)} \frac{\Gamma\left(\frac{1}{2\kappa}+\frac{1}{2\alpha}\right)}{\Gamma\left(\frac{1}{2\kappa}-\frac{1}{2\alpha}\right)} \left\{ 2\alpha (2\kappa)^{\frac{1}{\alpha}} (1-u) \left[\log_\kappa\left(\frac{1}{1-u}\right)\right]^{\frac{1}{\alpha}} + B_X\left(\frac{1}{2\kappa}-\frac{1}{2\alpha}, \frac{1}{\alpha}\right) \right. \\ & \left. + B_X\left(\frac{1}{2\kappa}-\frac{1}{2\alpha}+1, \frac{1}{\alpha}\right) \right\}, \quad 0 \le u \le 1, \end{aligned} \tag{21}$$

where $B_X(\cdot,\cdot)$ is the incomplete beta function with $X = (1-u)^{2\kappa}$. Clearly, the curve exists if and only if $\frac{\alpha}{\kappa} > 1$. The use of (21) can be done analytically.

As regards scalar measures of inequality, the well-known Gini (1914) coefficient can be derived using the representation in terms of order statistics due to Arnold and Laguna (1977), i.e. $G = 1 - m^{-1} \int_0^\infty [1-F(x)]^2 \mathrm{d}x$, yielding

$$G_\kappa = 1 - \frac{2\alpha+2\kappa}{2\alpha+\kappa} \frac{\Gamma\left(\frac{1}{\kappa}-\frac{1}{2\alpha}\right)}{\Gamma\left(\frac{1}{\kappa}+\frac{1}{2\alpha}\right)} \frac{\Gamma\left(\frac{1}{2\kappa}+\frac{1}{2\alpha}\right)}{\Gamma\left(\frac{1}{2\kappa}-\frac{1}{2\alpha}\right)}. \tag{22}$$ [3]

Furthermore, the generalized entropy (GE) class of inequality measures (Cowell 1980a,b; and Cowell and Kuga 1981a,b) is defined as

$$GE_\kappa(\theta) = \frac{1}{\theta^2-\theta} \left\{ \left(\frac{\beta}{m}\right)^\theta \left[\frac{(2\kappa)^{-\frac{\theta}{\alpha}}}{1+\frac{\theta}{\alpha}\kappa} \frac{\Gamma\left(\frac{1}{2\kappa}-\frac{\theta}{2\alpha}\right)}{\Gamma\left(\frac{1}{2\kappa}+\frac{\theta}{2\alpha}\right)} \Gamma\left(1+\frac{\theta}{\alpha}\right) \right] - 1 \right\}, \quad \theta \neq 0, 1. \tag{23}$$

The mean logarithmic deviation index is

$$MLD_\kappa = \lim_{\theta\to 0} GE_\kappa(\theta) = \frac{1}{\alpha} \left[\gamma + \psi\left(\frac{1}{2\kappa}\right) + \log(2\kappa) - \alpha \log\left(\frac{\beta}{m}\right) + \kappa \right], \tag{24}$$

where $\gamma = -\psi(1)$ is the Euler-Mascheroni constant and $\psi(z) = \Gamma'(z)/\Gamma(z)$ is the digamma function. The Theil (1967) index is

$$\begin{aligned} T_\kappa = \lim_{\theta\to 1} GE_\kappa(\theta) =& \frac{1}{\alpha} \left[\psi\left(1+\frac{1}{\alpha}\right) - \frac{1}{2}\psi\left(\frac{1}{2\kappa}-\frac{1}{2\alpha}\right) - \frac{1}{2}\psi\left(\frac{1}{2\kappa}+\frac{1}{2\alpha}\right) - \log(2\kappa) \right. \\ & \left. + \alpha \log\left(\frac{\beta}{m}\right) - \frac{\alpha\kappa}{\alpha+\kappa} \right]. \end{aligned} \tag{25}$$

Expressions for each GE index other than for the cases $\theta = 0, 1$ can be derived by straightforward substitution. In particular, the bottom-sensitive index is given by

$$GE_\kappa(-1) = -\frac{1}{2} + \frac{\Gamma\left(1+\frac{1}{\alpha}\right)\Gamma\left(1-\frac{1}{\alpha}\right)}{2\left[1+\left(\frac{\kappa}{\alpha}\right)^2\right]}, \tag{26}$$

[3] Using Stirling approximation for the gamma function, $\Gamma(z) \approx \sqrt{2\pi} z^{z-\frac{1}{2}} \exp(-z)$, and taking the limit as $\kappa \to 0$ in Equation (22), one arrives after some simplification at $G_0 = 1 - 2^{-\frac{1}{\alpha}}$, which is the explicit form of the Gini coefficient for the Weibull distribution (e.g. Kleiber and Kotz 2003, p. 177). Since the exponential distribution is a special case of the Weibull with shape parameter equal to 1, one directly determines that for $\kappa \to 0$ and $\alpha = 1$ the exponential law is also a special limiting case of the κ-generalized distribution, with a "true" value of the Gini coefficient equal to one half (e.g. Drăgulescu and Yakovenko 2001).

whereas the expression for the top-sensitive index is

$$GE_{\kappa}(2)=\frac{1}{2}\left\{2\frac{(\alpha+\kappa)^{2}}{\alpha+2\kappa}\frac{\Gamma\left(\frac{2}{\alpha}\right)}{\Gamma^{2}\left(\frac{1}{\alpha}\right)}\frac{\Gamma\left(\frac{1}{2\kappa}-\frac{1}{\alpha}\right)}{\Gamma\left(\frac{1}{2\kappa}+\frac{1}{\alpha}\right)}\frac{\Gamma^{2}\left(\frac{1}{2\kappa}+\frac{1}{2\alpha}\right)}{\Gamma^{2}\left(\frac{1}{2\kappa}-\frac{1}{2\alpha}\right)}-1\right\}=\frac{1}{2}CV_{\kappa}^{2}. \tag{27}$$

Finally, the Atkinson (1970) class of inequality indices for inequality aversion parameter $\varepsilon=1-\theta$, $\varepsilon>0$ and $\varepsilon\neq 1$, can be easily computed from $GE_{\kappa}(\theta)$ by exploiting the relationship $A_{\kappa}(\varepsilon)=1-[\varepsilon(\varepsilon-1)GE_{\kappa}(1-\varepsilon)+1]^{\frac{1}{1-\varepsilon}}$ (e.g. Cowell 1995); this yields

$$A_{\kappa}(\varepsilon)=1-\left\{\left(\frac{\beta}{m}\right)^{1-\varepsilon}\left[\frac{(2\kappa)^{-\frac{1-\varepsilon}{\alpha}}}{1+\frac{1-\varepsilon}{\alpha}\kappa}\frac{\Gamma\left(\frac{1}{2\kappa}-\frac{1-\varepsilon}{2\alpha}\right)}{\Gamma\left(\frac{1}{2\kappa}+\frac{1-\varepsilon}{2\alpha}\right)}\Gamma\left(1+\frac{1-\varepsilon}{\alpha}\right)\right]\right\}^{\frac{1}{1-\varepsilon}}. \tag{28}$$

The limiting form as $\varepsilon\to 1$ of the equation above is $A_{\kappa}(1)=1-\exp(-MLD_{\kappa})$.[4]

3 Estimation and comparison of alternative distributions

To test the performance of the proposed new statistical distribution we use data from the 2008 release of the Cross-National Equivalent File (CNEF), a commercially available database compiled by researchers at Cornell University (Burkhauser et al. 2001). The CNEF includes data for Great Britain, Germany and the United States, and provides cross-nationally comparable information about income, employment and a number of demographic characteristics.[5] The surveys used to build the CNEF are the British Household Panel Study (BHPS), the German Socio-Economic Panel (GSOEP) and the US Panel Study of Income Dynamics (PSID). The sample period is 1991–2004 for the BHPS, 1984–2007 for the GSOEP[6] and 1980-2005 for the PSID (which switched to biennial data collection after 1997). All calculations are based on the household post-government income expressed in nominal local currency unit. This variable represents the combined income after taxes and government transfers of the head, partner and other family members. Since the PSID stopped estimating household taxes in survey year 1992, from that year onward we use the analogue CNEF variable for which taxes were estimated using the National Bureau of Economic Research TAXSIM model (Feenberg and Coutts 1993). Observations with zero or negative incomes have been removed from the samples of each country. This exclusion has affected only a tiny fraction of the data. Furthermore, incomes have been adjusted for differences in household size using the “modified OECD” equivalence scale (Hagenaars et al. 1994)[7] and weighted by the provided sampling weights.

Estimates by maximum likelihood of the parameters for each country and year are shown in Table 2.[8] Also displayed are the log-likelihood and selected distributional statistics implied by the model parameter estimates, i.e. the mean, Gini coefficient and Theil index, reported to facilitate a comparison with their corresponding empirical estimates in Table 1. The model fit varied slightly across years and countries but was generally excellent. This is demonstrated by the probability plots

[4] Notice that all the measures considered here are functions of the distribution moments, whose existence depends on some conditions guaranteeing the convergence of relevant integrals. As a matter of example, the Gini coefficient (22) exists provided the mean of the distribution $m=\int_0^{\infty}xf(x;\alpha,\beta,\kappa)\,\mathrm{d}x$ converges, which is true when $\frac{\alpha}{\kappa}>1$. As shown by Kleiber (1997), the problem of existence of popular inequality measures is common to various parametric models of income distribution.

[5] The CNEF also includes some data for Australia, Canada and Switzerland. They are not used in this paper.

[6] Before the reunification of Germany, the GSOEP samples only include households living in the western states of the Federal Republic of Germany. The first wave of the East German sample was collected in June 1990, before the currency, economic and social union came into force on July 1.

[7] The “modified OECD” equivalence scale allocates points to each person in a household by taking the first adult as having a weight of 1 point, whereas each additional person who is 14 years or older is allocated 0.5 points, and each child under the age of 14 is allocated 0.3 points. Equivalized household income is derived by dividing total household income by a factor equal to the sum of the equivalence points allocated to the household members. Unlike the old OECD scale, the modified one gives less weight to any additional household member, allowing for higher economies of scale.

[8] The model parameters have been estimated by minimizing the negative of the log-likelihood function via a PORT routine as supplied by the `R` function `nlminb` (R Development Core Team 2011). Convergence was achieved easily within several iterations.

Table 1 Selected standard distributional summary measures for CNEF data, 1980–2007[a]

Wave	Great Britain (BHPS-CNEF) Obs Households	Individuals	Mean (GBP)	Gini	Theil	Germany (GSOEP-CNEF) Obs Households	Individuals	Mean (EUR)	Gini	Theil	United States (PSID-CNEF) Obs Households	Individuals	Mean (USD)	Gini	Theil
1980	—	—	—	—	—	—	—	—	—	—	6,524	18,972	9,788	0.320	0.186
													(122)	(0.006)	(0.017)
1981	—	—	—	—	—	—	—	—	—	—	6,603	18,953	10,848	0.341	0.254
													(215)	(0.012)	(0.043)
1982	—	—	—	—	—	—	—	—	—	—	6,723	19,204	11,473	0.331	0.204
													(149)	(0.007)	(0.020)
1983	—	—	—	—	—	—	—	—	—	—	6,816	19,416	12,401	0.349	0.239
													(212)	(0.009)	(0.037)
1984	—	—	—	—	—	5,603	15,372	10,459	0.256	0.120	6,880	19,498	13,210	0.350	0.216
								(105)	(0.005)	(0.006)			(170)	(0.006)	(0.013)
1985	—	—	—	—	—	5,045	13,865	10,810	0.270	0.141	7,000	19,729	14,520	0.368	0.275
								(123)	(0.006)	(0.011)			(293)	(0.012)	(0.038)
1986	—	—	—	—	—	4,826	13,221	10,969	0.257	0.121	6,984	19,543	15,071	0.362	0.241
								(118)	(0.005)	(0.009)			(219)	(0.007)	(0.018)
1987	—	—	—	—	—	4,764	12,986	11,589	0.255	0.116	7,028	19,583	15,737	0.361	0.236
								(119)	(0.005)	(0.006)			(231)	(0.007)	(0.017)
1988	—	—	—	—	—	4,569	12,392	11,858	0.256	0.116	7,073	19,603	16,924	0.372	0.278
								(118)	(0.005)	(0.005)			(370)	(0.013)	(0.041)
1989	—	—	—	—	—	4,441	11,965	12,348	0.258	0.123	7,091	19,626	18,171	0.381	0.298
								(131)	(0.005)	(0.009)			(408)	(0.012)	(0.040)
1990	—	—	—	—	—	4,396	11,825	12,868	0.269	0.138	7,299	19,883	18,858	0.376	0.268
								(149)	(0.006)	(0.010)			(311)	(0.008)	(0.021)
1991	5,485	13,732	8,691	0.281	0.134	4,423	11,822	13,683	0.266	0.127	7,338	19,905	19,219	0.370	0.248
			(71)	(0.004)	(0.004)			(150)	(0.005)	(0.007)			(261)	(0.006)	(0.012)
1992	5,218	13,139	9,342	0.289	0.147	6,320	16,964	13,198	0.281	0.138	7,529	20,275	18,939	0.359	0.241
			(90)	(0.005)	(0.007)			(131)	(0.005)	(0.007)			(258)	(0.007)	(0.015)
1993	5,069	12,684	9,604	0.290	0.141	6,293	16,684	14,062	0.280	0.139	7,817	21,340	20,148	0.358	0.232
			(87)	(0.004)	(0.004)			(153)	(0.005)	(0.008)			(260)	(0.006)	(0.011)

[a] Numbers in parenthesis: standard errors obtained via 1,000 bootstrap replications

Table 1 continued[a]

Wave	Great Britain (BHPS-CNEF)					Germany (GSOEP-CNEF)					United States (PSID-CNEF)				
	Obs		Mean	Gini	Theil	Obs		Mean	Gini	Theil	Obs		Mean	Gini	Theil
	Households	Individuals	(GBP)			Households	Individuals	(EUR)			Households	Individuals	(USD)		
1994	5,039	12,625	9,879 (90)	0.286 (0.004)	0.139 (0.005)	6,434	16,976	14,590 (164)	0.280 (0.005)	0.139 (0.007)	8,585	23,475	20,291 (314)	0.388 (0.007)	0.293 (0.018)
1995	4,942	12,317	10,407 (112)	0.292 (0.005)	0.153 (0.009)	6,598	17,411	14,860 (179)	0.287 (0.006)	0.150 (0.010)	8,503	23,055	20,918 (297)	0.376 (0.007)	0.271 (0.016)
1996	5,014	12,588	10,804 (99)	0.286 (0.004)	0.139 (0.005)	6,518	17,009	15,224 (176)	0.280 (0.006)	0.140 (0.008)	8,464	22,952	21,933 (302)	0.371 (0.006)	0.258 (0.015)
1997	6,044	14,951	11,431 (110)	0.284 (0.004)	0.138 (0.005)	6,439	16,673	15,458 (183)	0.274 (0.006)	0.134 (0.008)	6,278	17,422	24,174 (314)	0.357 (0.006)	0.232 (0.010)
1998	5,946	14,680	11,878 (131)	0.294 (0.005)	0.161 (0.011)	7,261	18,315	15,622 (209)	0.276 (0.007)	0.136 (0.011)	—	—	—	—	—
1999	8,693	21,313	12,364 (122)	0.284 (0.005)	0.146 (0.008)	7,012	17,605	16,161 (172)	0.274 (0.005)	0.132 (0.006)	6,947	19,553	23,817 (379)	0.395 (0.008)	0.308 (0.035)
2000	8,594	21,230	13,180 (130)	0.279 (0.005)	0.139 (0.008)	12,575	30,771	16,889 (178)	0.275 (0.005)	0.134 (0.006)	—	—	—	—	—
2001	10,354	25,923	13,965 (138)	0.276 (0.005)	0.140 (0.007)	11,338	27,791	17,262 (131)	0.275 (0.004)	0.134 (0.005)	7,357	20,434	29,178 (476)	0.393 (0.008)	0.299 (0.017)
2002	8,706	21,951	13,190 (136)	0.305 (0.004)	0.161 (0.005)	12,050	29,644	17,658 (150)	0.293 (0.004)	0.155 (0.006)	—	—	—	—	—
2003	8,699	21,756	15,126 (147)	0.268 (0.005)	0.130 (0.007)	11,458	27,873	18,151 (147)	0.290 (0.004)	0.144 (0.004)	7,749	21,103	29,260 (456)	0.382 (0.008)	0.280 (0.022)
2004	8,424	21,007	15,810 (184)	0.276 (0.006)	0.141 (0.009)	11,202	27,027	18,041 (164)	0.291 (0.004)	0.146 (0.005)	—	—	—	—	—
2005	—	—	—	—	—	10,868	25,863	18,230 (192)	0.299 (0.005)	0.154 (0.006)	7,960	21,588	32,079 (658)	0.413 (0.011)	0.367 (0.047)
2006	—	—	—	—	—	11,888	27,668	18,345 (243)	0.316 (0.006)	0.177 (0.010)	—	—	—	—	—
2007	—	—	—	—	—	11,119	25,729	18,865 (206)	0.312 (0.005)	0.176 (0.009)	—	—	—	—	—

[a] Numbers in parenthesis: standard errors obtained via 1,000 bootstrap replications

Table 2 Estimated κ-generalized parameters for CNEF data, 1980–2007

Wave	Parameters[a] $\hat{\alpha}$	$\hat{\beta}$	$\hat{\kappa}$	$-\log L$[b]	Mean[c]	Gini[d]	Theil[e]
			Great Britain (BHPS-CNEF)				
1991	2.584 (0.027)	8,971 (45)	0.650 (0.020)	131,812	8,701	0.284	0.139
1992	2.518 (0.026)	9,604 (49)	0.643 (0.020)	128,916	9,324	0.290	0.145
1993	2.400 (0.025)	10,057 (53)	0.565 (0.019)	124,676	9,609	0.292	0.144
1994	2.483 (0.026)	10,259 (53)	0.606 (0.019)	124,508	9,876	0.288	0.142
1995	2.504 (0.027)	10,633 (56)	0.656 (0.020)	122,031	10,367	0.293	0.149
1996	2.525 (0.027)	11,162 (58)	0.635 (0.020)	124,901	10,807	0.288	0.142
1997	2.539 (0.027)	11,822 (61)	0.628 (0.020)	124,295	11,419	0.285	0.140
1998	2.497 (0.026)	12,141 (63)	0.644 (0.019)	123,158	11,800	0.292	0.147
1999	2.622 (0.029)	12,588 (66)	0.686 (0.021)	121,336	12,309	0.285	0.141
2000	2.617 (0.029)	13,533 (70)	0.656 (0.021)	120,871	13,129	0.282	0.136
2001	2.670 (0.029)	14,285 (72)	0.670 (0.020)	119,903	13,881	0.278	0.133
2002	2.046 (0.020)	14,124 (80)	0.399 (0.017)	113,616	13,127	0.313	0.162
2003	2.784 (0.031)	15,481 (77)	0.694 (0.021)	116,036	15,070	0.271	0.126
2004	2.759 (0.031)	16,033 (82)	0.715 (0.021)	113,404	15,709	0.276	0.132
			Germany (GSOEP-CNEF)				
1984	3.276 (0.001)	10,378 (1)	0.894 (3e-04)	596,373,508	10,436	0.257	0.118
1985	2.981 (5e-04)	10,783 (1)	0.800 (3e-04)	604,491,927	10,707	0.268	0.126
1986	3.003 (5e-04)	11,154 (1)	0.743 (3e-04)	610,107,762	10,911	0.258	0.116
1987	3.083 (5e-04)	11,742 (1)	0.779 (3e-04)	615,237,096	11,557	0.256	0.115
1988	3.020 (5e-04)	12,044 (1)	0.765 (3e-04)	622,455,192	11,837	0.260	0.117
1989	2.984 (5e-04)	12,578 (1)	0.733 (3e-04)	629,596,037	12,282	0.258	0.116
1990	2.745 (4e-04)	13,147 (1)	0.671 (3e-04)	643,844,471	12,741	0.271	0.126
1991	2.830 (4e-04)	13,991 (1)	0.705 (3e-04)	655,533,950	13,635	0.268	0.124
1992	2.545 (3e-04)	13,663 (1)	0.618 (3e-04)	827,719,391	13,160	0.283	0.137
1993	2.547 (3e-04)	14,499 (1)	0.628 (2e-04)	837,914,115	14,001	0.285	0.139
1994	2.596 (3e-04)	14,961 (1)	0.662 (3e-04)	848,064,767	14,549	0.284	0.140
1995	2.474 (3e-04)	15,292 (1)	0.617 (2e-04)	850,741,479	14,768	0.291	0.145
1996	2.564 (3e-04)	15,662 (1)	0.640 (2e-04)	853,700,688	15,160	0.285	0.139
1997	2.651 (3e-04)	15,881 (1)	0.663 (2e-04)	859,137,399	15,413	0.279	0.134
1998	2.606 (3e-04)	16,110 (1)	0.638 (2e-04)	861,601,716	15,562	0.280	0.134
1999	2.721 (4e-04)	16,453 (1)	0.715 (3e-04)	865,437,108	16,144	0.280	0.136
2000	2.653 (4e-04)	17,285 (1)	0.677 (3e-04)	869,878,074	16,839	0.281	0.136
2001	2.713 (4e-04)	17,657 (1)	0.691 (3e-04)	868,448,331	17,225	0.277	0.133
2002	2.497 (3e-04)	18,026 (1)	0.658 (3e-04)	874,399,292	17,592	0.294	0.150
2003	2.540 (3e-04)	18,574 (1)	0.672 (3e-04)	876,321,686	18,163	0.292	0.148
2004	2.412 (3e-04)	18,733 (1)	0.591 (2e-04)	875,663,423	18,013	0.294	0.148
2005	2.418 (3e-04)	18,717 (1)	0.639 (3e-04)	875,106,299	18,239	0.301	0.156
2006	2.308 (3e-04)	18,592 (1)	0.651 (2e-04)	874,183,908	18,310	0.316	0.174
2007	2.417 (3e-04)	18,918 (1)	0.706 (3e-04)	878,491,225	18,826	0.311	0.171
			United States (PSID-CNEF)				
1980	2.223 (0.005)	10,022 (12)	0.595 (0.004)	2,920,908	9,750	0.318	0.174
1981	2.200 (0.005)	10,775 (13)	0.627 (0.004)	2,917,591	10,609	0.326	0.186
1982	2.108 (0.004)	11,802 (14)	0.544 (0.004)	2,942,808	11,386	0.325	0.181
1983	2.035 (0.004)	12,488 (17)	0.583 (0.004)	2,952,852	12,277	0.342	0.204
1984	1.945 (0.004)	13,538 (17)	0.532 (0.004)	3,204,164	13,175	0.347	0.208
1985	1.987 (0.004)	14,248 (19)	0.622 (0.004)	3,217,831	14,287	0.357	0.226
1986	1.922 (0.004)	15,143 (20)	0.568 (0.004)	3,184,868	14,972	0.357	0.223
1987	1.928 (0.004)	15,853 (21)	0.567 (0.004)	3,168,692	15,660	0.356	0.222
1988	1.924 (0.004)	16,709 (23)	0.588 (0.004)	3,157,653	16,649	0.361	0.230
1989	1.962 (0.004)	17,463 (23)	0.658 (0.004)	3,581,257	17,841	0.369	0.246
1990	1.956 (0.004)	18,308 (24)	0.656 (0.004)	3,601,127	18,704	0.369	0.247
1991	1.925 (0.004)	18,966 (25)	0.620 (0.004)	3,584,341	19,152	0.367	0.241
1992	1.980 (0.004)	18,875 (23)	0.602 (0.004)	3,623,369	18,795	0.354	0.221
1993	1.912 (0.004)	20,313 (26)	0.559 (0.004)	3,425,653	20,039	0.357	0.223
1994	1.842 (0.004)	19,605 (26)	0.617 (0.003)	3,515,388	20,004	0.382	0.262
1995	1.901 (0.004)	20,387 (26)	0.623 (0.004)	3,665,639	20,677	0.372	0.248
1996	1.956 (0.004)	21,337 (28)	0.655 (0.004)	3,658,714	21,787	0.369	0.246
1997	2.013 (0.005)	23,796 (37)	0.650 (0.005)	2,386,326	24,072	0.359	0.230
1999	1.760 (0.004)	23,101 (37)	0.589 (0.004)	2,787,289	23,563	0.391	0.275
2001	1.978 (0.005)	27,085 (41)	0.760 (0.005)	2,906,628	28,981	0.389	0.288
2003	1.815 (0.004)	28,704 (42)	0.584 (0.004)	3,183,241	28,962	0.380	0.256
2005	1.774 (0.003)	30,029 (38)	0.637 (0.003)	4,440,930	31,299	0.400	0.293

[a] Numbers in parenthesis: estimated standard errors
[b] Negative of the log-likelihood function corresponding to the best set of parameters found
[c] Analytic value obtained by substituting the estimated parameters into Equation (16) with $r = 1$
[d] Analytic value obtained by substituting the estimated parameters into Equation (22)
[e] Analytic value obtained by substituting the estimated parameters into Equation (25)

shown in Figure 1 for the most recent data available (for brevity, we do not report plots for each year and country but they are available from the authors on request). These are plots of the cumulative probabilities of income expected given the estimated κ-generalized parameters against the cumulative probabilities of income observed in the data. Excellent goodness-of-fit is demonstrated by the fact that every plot lies extremely close to the 45°-ray from the origin, and much closer than is typically observed in plots of this type.

For comparison, the results of fitting other existing three-parameter distributions able to accommodate sufficient flexibility to model heterogeneous income data are given in Tables 3 and 4. Namely, these models are the Singh-Maddala (1976) distribution

$$F(x;a,b,q) = 1 - \left[1 + \left(\frac{x}{b}\right)^{a}\right]^{-q}, \quad x > 0, \quad a,b,q > 0, \tag{29}$$

and the Dagum (1977) type I distribution

$$F(x;a,b,p) = \left[1 + \left(\frac{x}{b}\right)^{-a}\right]^{-p}, \quad x > 0, \quad a,b,p > 0, \tag{30}$$

which are closely related (Kleiber 1996). The corresponding densities are

$$f(x;a,b,q) = \frac{aqx^{a-1}}{b^{a}\left[1 + \left(\frac{x}{b}\right)^{a}\right]^{1+q}} \tag{31}$$

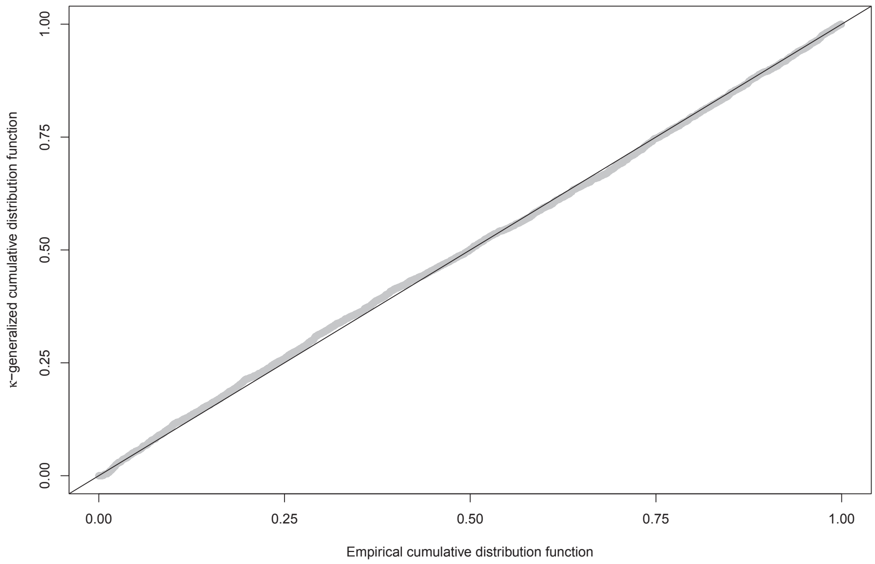
and

$$f(x;a,b,p) = \frac{apx^{ap-1}}{b^{ap}\left[1 + \left(\frac{x}{b}\right)^{a}\right]^{p+1}}. \tag{32}$$

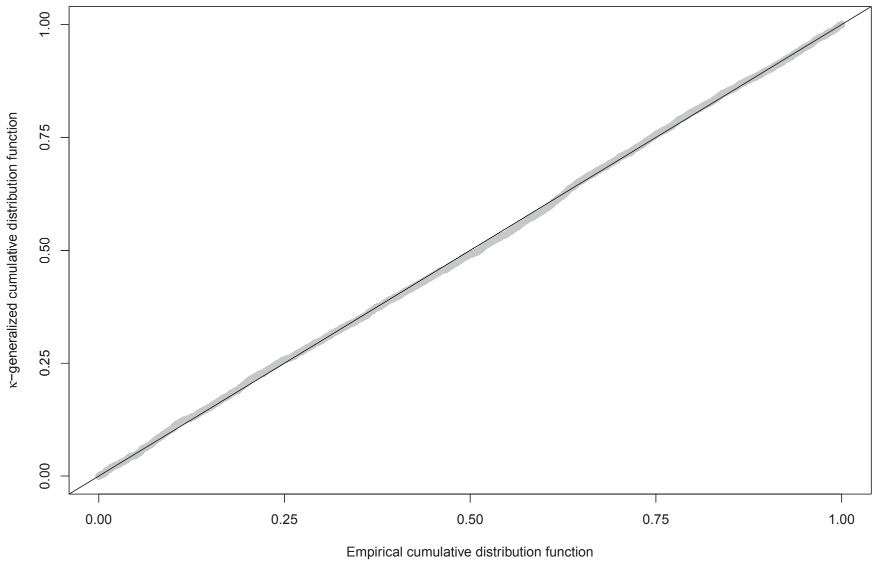
In order to decide which distribution better models the data, we adopt the Vuong (1989) testing approach to model selection for non-nested hypothesis. This approach sets the model selection criterion in a hypothesis testing framework. More specifically, it tests whether the models under consideration are equally close to the "true" model. The null hypothesis states that both models are equivalent against the alternative that H_f is better than H_g or H_g is better than H_f. The proposed statistic is asymptotically normal under the null hypothesis and is quite straightforward to compute. Tables 5 and 6 report the results of the comparison for the three candidate models. As can be seen, if one takes the 5% as the relevant significance level only in two cases (i.e. when comparing to the Dagum type I distribution for Great Britain and Germany) the present distribution is selected by the Vuong test as the worse model a high percentage of times. However, if one lowers the significance level at 1% the test concludes that the competing models are always at least observationally equivalent.

To further explore the performance of the above theoretical densities, we carry out a detailed goodness-of-fit analysis by using six waves of GSOEP-CNEF data for the years 2002 through 2007. Due to the inclusion of a special sample for high-income households ("Sample G"), these waves are likely to offer a more reliable picture of the distribution of income, especially for the top percentiles that are usually under-represented in sample surveys.[9] More specifically, we run a battery of widely-used goodness-of-fit tests based on empirical distribution function (EDF) statistics: Kolmogorv-Smirnov (KS),

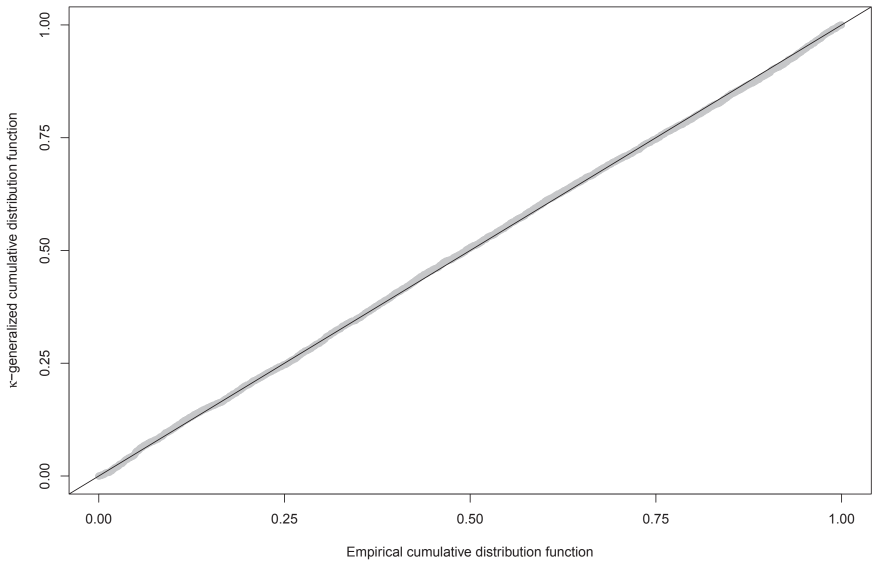
[9] Sample G surveyed since 2002 is the so-called "high-income sample", selected independently from all other subsamples from the population of private households (see Haisken-DeNew and Frick 2005 for a more detailed description of the various subsamples in the GSOEP). The original selection scheme required that the responding household had a monthly income of at least DM 7,500 (EUR 3,835) in 2001. Starting in 2003, the selection scheme was changed in such a way that only households with a net monthly income of at least EUR 4,500 were followed. Sample G represents about 7.3% individuals in private households with the highest income, with a small number of them belonging to the top 1% of the income distribution. Nevertheless, none of these individuals would belong to the "economic elite" as defined in Bach et al. (2009), where the GSOEP data have been integrated by official income tax statistics to shed more light on very high German incomes.



[a] Great Britain (BHPS-CNEF) 2004



[b] Germany (GSOEP-CNEF) 2007



[c] United States (PSID-CNEF) 2005

Fig. 1 Probability plots for κ-generalized estimates

Table 3 Estimated Singh-Maddala parameters for CNEF data, 1980–2007

Wave	Parameters[a] $\hat{a}$	$\hat{b}$	$\hat{q}$	$-\log L$[b]	Mean[c]	Gini[d]	Theil[e]
Great Britain (BHPS-CNEF)							
1991	2.797 (0.033)	10,406 (219)	1.851 (0.073)	131,840	8,679	0.282	0.133
1992	2.729 (0.032)	11,193 (237)	1.859 (0.072)	128,959	9,303	0.288	0.139
1993	2.583 (0.031)	12,759 (328)	2.209 (0.101)	124,695	9,591	0.291	0.140
1994	2.669 (0.032)	12,542 (293)	2.053 (0.087)	124,548	9,854	0.287	0.137
1995	2.717 (0.032)	12,254 (258)	1.808 (0.070)	122,090	10,344	0.292	0.143
1996	2.735 (0.033)	13,106 (294)	1.894 (0.079)	124,927	10,781	0.286	0.137
1997	2.752 (0.033)	13,897 (306)	1.907 (0.078)	124,330	11,396	0.284	0.135
1998	2.712 (0.032)	14,041 (293)	1.828 (0.070)	123,223	11,778	0.291	0.142
1999	2.856 (0.035)	14,055 (282)	1.701 (0.065)	121,388	12,280	0.283	0.135
2000	2.835 (0.034)	15,588 (327)	1.827 (0.073)	120,921	13,101	0.280	0.130
2001	2.894 (0.035)	16,206 (318)	1.771 (0.068)	119,979	13,855	0.277	0.128
2002	2.107 (0.024)	26,666 (1,100)	4.177 (0.269)	113,688	13,123	0.314	0.161
2003	3.010 (0.037)	17,295 (332)	1.718 (0.067)	116,107	15,039	0.269	0.121
2004	3.010 (0.037)	17,398 (319)	1.604 (0.059)	113,483	15,679	0.274	0.127
Germany (GSOEP-CNEF)							
1984	3.674 (0.001)	9,987 (2)	1.170 (0.001)	596,498,624	10,408	0.254	0.111
1985	3.309 (0.001)	10,885 (2)	1.348 (0.001)	604,921,936	10,688	0.266	0.120
1986	3.322 (0.001)	11,660 (3)	1.486 (0.001)	610,341,874	10,894	0.256	0.110
1987	3.404 (0.001)	12,092 (3)	1.421 (0.001)	615,479,419	11,533	0.254	0.109
1988	3.285 (0.001)	12,772 (3)	1.519 (0.001)	622,769,488	11,807	0.257	0.111
1989	3.274 (0.001)	13,350 (3)	1.539 (0.001)	629,954,047	12,263	0.257	0.111
1990	2.957 (5e-04)	15,008 (4)	1.802 (0.001)	644,533,135	12,726	0.270	0.121
1991	3.084 (0.001)	15,316 (4)	1.648 (0.001)	655,882,675	13,610	0.266	0.118
1992	2.768 (4e-04)	16,133 (4)	1.933 (0.001)	827,915,291	13,137	0.282	0.132
1993	2.735 (4e-04)	17,429 (5)	1.986 (0.001)	838,404,571	13,974	0.283	0.133
1994	2.795 (4e-04)	17,435 (5)	1.858 (0.001)	848,448,940	14,513	0.282	0.133
1995	2.645 (4e-04)	18,806 (5)	2.058 (0.001)	851,342,940	14,741	0.289	0.139
1996	2.738 (4e-04)	18,891 (5)	1.991 (0.001)	854,223,939	15,126	0.282	0.132
1997	2.824 (4e-04)	18,790 (5)	1.921 (0.001)	859,740,228	15,376	0.277	0.127
1998	2.773 (4e-04)	19,563 (5)	2.030 (0.001)	862,119,987	15,527	0.278	0.128
1999	2.924 (4e-04)	18,576 (5)	1.729 (0.001)	865,792,915	16,090	0.276	0.127
2000	2.838 (4e-04)	20,179 (5)	1.861 (0.001)	870,375,877	16,793	0.278	0.129
2001	2.937 (4e-04)	19,820 (5)	1.727 (0.001)	868,844,526	17,183	0.275	0.126
2002	2.705 (4e-04)	20,896 (5)	1.824 (0.001)	874,735,648	17,546	0.292	0.143
2003	2.772 (4e-04)	20,960 (5)	1.735 (0.001)	876,470,237	18,113	0.290	0.141
2004	2.602 (4e-04)	23,161 (7)	2.089 (0.001)	875,876,862	17,977	0.293	0.142
2005	2.635 (4e-04)	21,826 (6)	1.846 (0.001)	875,165,918	18,189	0.298	0.149
2006	2.540 (4e-04)	21,095 (6)	1.732 (0.001)	874,325,246	18,262	0.314	0.167
2007	2.668 (4e-04)	20,518 (5)	1.579 (0.001)	878,598,598	18,762	0.309	0.163
United States (PSID-CNEF)							
1980	2.443 (0.006)	11,961 (63)	1.913 (0.017)	2,920,987	9,729	0.317	0.169
1981	2.449 (0.006)	12,162 (58)	1.719 (0.014)	2,918,058	10,593	0.326	0.181
1982	2.308 (0.006)	14,944 (85)	2.115 (0.019)	2,943,232	11,371	0.325	0.177
1983	2.236 (0.006)	15,314 (90)	1.955 (0.017)	2,952,943	12,246	0.341	0.198
1984	2.133 (0.005)	17,629 (110)	2.154 (0.020)	3,204,171	13,153	0.347	0.204
1985	2.209 (0.005)	16,403 (87)	1.740 (0.014)	3,218,064	14,252	0.357	0.220
1986	2.099 (0.005)	19,294 (113)	2.041 (0.017)	3,185,647	14,937	0.357	0.217
1987	2.119 (0.005)	19,791 (117)	1.987 (0.017)	3,169,123	15,629	0.357	0.217
1988	2.126 (0.005)	20,221 (115)	1.882 (0.015)	3,158,121	16,614	0.361	0.224
1989	2.181 (0.005)	19,499 (95)	1.635 (0.012)	3,582,061	17,785	0.368	0.237
1990	2.171 (0.005)	20,584 (103)	1.652 (0.012)	3,601,690	18,637	0.369	0.237
1991	2.123 (0.005)	22,394 (120)	1.792 (0.014)	3,584,913	19,089	0.367	0.232
1992	2.169 (0.005)	22,924 (117)	1.893 (0.014)	3,624,536	18,747	0.354	0.214
1993	2.066 (0.005)	27,022 (158)	2.172 (0.018)	3,427,034	19,986	0.357	0.216
1994	2.002 (0.004)	24,253 (129)	1.887 (0.014)	3,517,721	19,932	0.381	0.251
1995	2.069 (0.005)	24,935 (128)	1.871 (0.013)	3,667,720	20,599	0.371	0.237
1996	2.147 (0.005)	24,807 (126)	1.732 (0.013)	3,659,886	21,685	0.367	0.233
1997	2.187 (0.006)	28,364 (175)	1.810 (0.017)	2,387,432	23,961	0.356	0.218
1999	1.917 (0.005)	29,854 (209)	2.000 (0.018)	2,788,055	23,454	0.389	0.262
2001	2.203 (0.006)	27,915 (143)	1.402 (0.010)	2,907,491	28,780	0.386	0.268
2003	1.963 (0.005)	37,841 (242)	2.079 (0.018)	3,184,520	28,838	0.377	0.244
2005	1.945 (0.004)	35,903 (180)	1.773 (0.012)	4,442,936	31,148	0.398	0.279

[a] Numbers in parenthesis: estimated standard errors
[b] Negative of the log-likelihood function corresponding to the best set of parameters found
[c] Analytic value obtained by substituting the estimated parameters into Equation (6.47) of Kleiber and Kotz (2003, p. 201)
[d] Analytic value obtained by substituting the estimated parameters into Equation (6.69) of Kleiber and Kotz (2003, p. 206)
[e] Analytic value obtained by substituting the estimated parameters into the relevant expression given by Kleiber and Kotz (2003, p. 206)

Table 4 Estimated Dagum type I parameters for CNEF data, 1980–2007

Wave	Parameters[a] $\hat{a}$	$\hat{b}$	$\hat{p}$	$-\log L$[b]	Mean[c]	Gini[d]	Theil[e]
Great Britain (BHPS-CNEF)							
1991	4.171 (0.059)	9,608 (98)	0.584 (0.015)	131,801	8,683	0.283	0.137
1992	4.134 (0.059)	10,405 (107)	0.567 (0.015)	128,903	9,293	0.288	0.142
1993	4.184 (0.065)	11,059 (119)	0.532 (0.015)	124,669	9,618	0.292	0.145
1994	4.240 (0.063)	11,270 (115)	0.538 (0.014)	124,490	9,849	0.287	0.140
1995	4.109 (0.060)	11,584 (121)	0.562 (0.015)	122,015	10,308	0.291	0.145
1996	4.129 (0.062)	12,011 (129)	0.575 (0.016)	124,893	10,792	0.287	0.141
1997	4.209 (0.063)	12,811 (135)	0.562 (0.015)	124,284	11,395	0.284	0.138
1998	4.160 (0.062)	13,306 (138)	0.550 (0.015)	123,143	11,734	0.290	0.144
1999	4.154 (0.061)	13,491 (142)	0.590 (0.016)	121,322	12,241	0.283	0.137
2000	4.275 (0.063)	14,643 (149)	0.566 (0.015)	120,849	13,063	0.279	0.133
2001	4.376 (0.064)	15,557 (152)	0.556 (0.015)	119,878	13,774	0.275	0.129
2002	4.699 (0.080)	17,565 (162)	0.360 (0.009)	113,489	13,031	0.309	0.161
2003	4.468 (0.066)	16,722 (162)	0.571 (0.015)	116,017	14,952	0.267	0.122
2004	4.347 (0.065)	17,254 (174)	0.583 (0.016)	113,394	15,567	0.272	0.126
Germany (GSOEP-CNEF)							
1984	4.218 (0.001)	10,215 (2)	0.796 (3e-04)	596,407,810	10,373	0.253	0.110
1985	4.307 (0.001)	11,223 (2)	0.657 (2e-04)	604,538,125	10,600	0.263	0.119
1986	4.467 (0.001)	11,586 (2)	0.646 (2e-04)	610,066,211	10,856	0.256	0.112
1987	4.461 (0.001)	12,115 (2)	0.668 (3e-04)	615,219,922	11,487	0.253	0.109
1988	4.498 (0.001)	12,643 (2)	0.632 (2e-04)	622,412,402	11,748	0.256	0.112
1989	4.559 (0.001)	13,279 (2)	0.614 (2e-04)	629,558,310	12,198	0.255	0.111
1990	4.644 (0.001)	14,563 (2)	0.522 (2e-04)	643,563,082	12,582	0.267	0.121
1991	4.446 (0.001)	14,935 (2)	0.592 (2e-04)	655,455,299	13,545	0.265	0.120
1992	4.218 (0.001)	14,724 (2)	0.567 (2e-04)	827,593,618	13,140	0.283	0.137
1993	4.350 (0.001)	16,021 (2)	0.529 (2e-04)	837,628,474	13,901	0.282	0.135
1994	4.247 (0.001)	16,248 (2)	0.563 (2e-04)	847,863,227	14,457	0.282	0.136
1995	4.330 (0.001)	17,142 (2)	0.508 (2e-04)	850,369,478	14,635	0.288	0.141
1996	4.362 (0.001)	17,327 (2)	0.529 (2e-04)	853,386,810	15,034	0.281	0.135
1997	4.450 (0.001)	17,520 (2)	0.534 (2e-04)	858,836,472	15,262	0.275	0.129
1998	4.447 (0.001)	17,812 (2)	0.527 (2e-04)	861,260,138	15,432	0.277	0.130
1999	4.248 (0.001)	17,597 (2)	0.596 (2e-04)	865,237,809	16,021	0.276	0.130
2000	4.342 (0.001)	18,854 (2)	0.556 (2e-04)	869,602,555	16,691	0.277	0.131
2001	3.920 (0.001)	17,262 (2)	0.726 (2e-04)	868,608,351	17,089	0.279	0.135
2002	4.069 (0.001)	19,544 (2)	0.570 (2e-04)	874,260,376	17,500	0.292	0.146
2003	3.983 (0.001)	19,655 (3)	0.611 (2e-04)	876,263,599	18,127	0.291	0.146
2004	4.131 (0.001)	20,542 (3)	0.540 (2e-04)	875,539,021	17,989	0.293	0.147
2005	3.869 (0.001)	19,885 (3)	0.601 (2e-04)	875,061,125	18,249	0.300	0.156
2006	3.657 (0.001)	19,768 (3)	0.607 (2e-04)	874,169,722	18,304	0.316	0.173
2007	3.635 (0.001)	19,664 (3)	0.651 (2e-04)	878,459,102	18,791	0.310	0.168
United States (PSID-CNEF)							
1980	3.649 (0.012)	10,725 (29)	0.587 (0.004)	2,921,084	9,793	0.320	0.178
1981	3.523 (0.012)	11,467 (33)	0.603 (0.004)	2,917,761	10,635	0.327	0.188
1982	3.687 (0.013)	13,079 (35)	0.535 (0.003)	2,943,118	11,441	0.327	0.185
1983	3.376 (0.011)	13,500 (40)	0.579 (0.004)	2,953,037	12,348	0.345	0.210
1984	3.369 (0.011)	14,910 (45)	0.550 (0.003)	3,204,765	13,303	0.352	0.217
1985	3.167 (0.010)	15,156 (48)	0.611 (0.004)	3,218,086	14,361	0.360	0.231
1986	3.344 (0.011)	17,063 (48)	0.532 (0.003)	3,185,001	15,002	0.358	0.225
1987	3.289 (0.011)	17,552 (53)	0.553 (0.003)	3,169,086	15,747	0.359	0.227
1988	3.211 (0.011)	18,309 (56)	0.570 (0.003)	3,157,932	16,727	0.363	0.234
1989	3.082 (0.009)	18,690 (56)	0.614 (0.003)	3,581,518	17,836	0.368	0.244
1990	3.059 (0.009)	19,492 (59)	0.620 (0.004)	3,601,377	18,726	0.370	0.246
1991	3.127 (0.010)	20,600 (61)	0.588 (0.003)	3,584,550	19,197	0.368	0.243
1992	3.355 (0.010)	21,068 (55)	0.548 (0.003)	3,623,421	18,772	0.353	0.220
1993	3.442 (0.011)	23,466 (60)	0.501 (0.003)	3,425,215	19,985	0.356	0.221
1994	3.201 (0.009)	22,743 (59)	0.516 (0.003)	3,514,866	19,775	0.376	0.250
1995	3.251 (0.009)	23,292 (59)	0.530 (0.003)	3,665,156	20,488	0.368	0.239
1996	3.146 (0.009)	23,352 (64)	0.585 (0.003)	3,658,546	21,683	0.367	0.240
1997	3.322 (0.012)	26,535 (82)	0.557 (0.003)	2,386,023	23,869	0.354	0.222
1999	3.014 (0.010)	26,203 (87)	0.541 (0.003)	2,787,106	23,541	0.391	0.275
2001	2.860 (0.009)	28,096 (99)	0.678 (0.004)	2,906,850	28,703	0.384	0.271
2003	3.187 (0.010)	33,079 (94)	0.517 (0.003)	3,182,711	28,817	0.377	0.252
2005	2.947 (0.008)	33,959 (91)	0.555 (0.003)	4,440,792	31,052	0.396	0.283

[a] Numbers in parenthesis: estimated standard errors
[b] Negative of the log-likelihood function corresponding to the best set of parameters found
[c] Analytic value obtained by substituting the estimated parameters into Equation (6.91) of Kleiber and Kotz (2003, p. 214)
[d] Analytic value obtained by substituting the estimated parameters into Equation (6.103) of Kleiber and Kotz (2003, p. 217)
[e] Analytic value obtained by substituting the estimated parameters into Equation (11) of Jenkins (2009, p. 395) with $q = 1$

Table 5 Vuong model selection test for CNEF data, 1980–2007[a]

Year	Great Britain (BHPS-CNEF)				Germany (GSOEP-CNEF)				United States (PSID-CNEF)			
	κ-gen vs. SM		κ-gen vs. D		κ-gen vs. SM		κ-gen vs. D		κ-gen vs. SM		κ-gen vs. D	
	Statistic	*p*-value	Statistic	*p*-value	Statistic	*p*-value	Statistic	*p*-value	Statistic	*p*-value	Statistic	*p*-value
1980	—	—	—	—	—	—	—	—	0.258	0.398	1.486	0.069
1981	—	—	—	—	—	—	—	—	1.147	0.126	1.696	0.045*
1982	—	—	—	—	—	—	—	—	1.465	0.071	2.194	0.014*
1983	—	—	—	—	—	—	—	—	0.300	0.382	1.490	0.068
1984	—	—	—	—	1.647	0.050*	0.687	0.246	0.026	0.489	3.263	0.001**
1985	—	—	—	—	4.324	1e-05**	0.702	0.241	0.648	0.258	2.264	0.012*
1986	—	—	—	—	2.423	0.008**	-1.082	0.140	2.614	0.004**	1.028	0.152
1987	—	—	—	—	2.960	0.002**	-0.443	0.329	1.542	0.062	3.125	0.001**
1988	—	—	—	—	3.543	2e-04**	-0.772	0.220	1.292	0.098	2.483	0.007**
1989	—	—	—	—	3.571	2e-04**	-0.645	0.260	2.133	0.016*	3.398	3e-04**
1990	—	—	—	—	5.572	1e-08**	-1.779	0.038*	1.743	0.041*	3.350	4e-04**
1991	2.413	0.008**	-2.429	0.008**	3.560	2e-04**	-1.285	0.099	1.661	0.048*	2.038	0.021*
1992	3.147	0.001**	-2.235	0.013*	1.862	0.031*	-2.661	0.004**	3.636	1e-04**	0.453	0.325
1993	1.917	0.028*	-1.182	0.119	4.155	2e-05**	-2.575	0.005**	4.511	3e-06**	-2.092	0.018*
1994	3.444	3e-04**	-2.576	0.005**	3.376	4e-04**	-2.499	0.006**	6.907	2e-12**	-1.983	0.024*
1995	4.107	2e-05**	-2.224	0.013*	4.424	5e-06**	-2.871	0.002**	5.915	2e-09**	-1.965	0.025*
1996	2.052	0.020*	-1.845	0.033*	3.932	4e-05**	-2.235	0.013*	3.320	5e-04**	-1.178	0.119
1997	2.743	0.003**	-1.616	0.053	4.511	3e-06**	-2.067	0.019*	4.091	2e-05**	-1.581	0.057
1998	4.635	2e-06**	-1.819	0.034*	3.029	0.001**	-2.185	0.014*	—	—	—	—
1999	3.525	2e-04**	-1.862	0.031*	2.893	0.002**	-2.534	0.006**	2.957	0.002**	-1.309	0.095
2000	3.472	3e-04**	-2.505	0.006**	3.500	2e-04**	-1.735	0.041*	—	—	—	—
2001	5.089	2e-07**	-2.151	0.016*	5.020	3e-07**	2.373	0.009**	3.276	0.001**	2.079	0.019*
2002	6.801	1e-11**	-5.649	1e-08**	3.905	5e-05**	-2.667	0.004**	—	—	—	—
2003	4.943	4e-07**	-1.605	0.054	2.126	0.017*	-3.469	3e-04**	4.247	1e-05**	-2.829	0.002**
2004	5.035	2e-07**	-0.890	0.187	2.570	0.005**	-2.301	0.011*	—	—	—	—
2005	—	—	—	—	0.749	0.227	-2.272	0.012*	5.270	1e-07**	-0.772	0.220
2006	—	—	—	—	1.281	0.100	-0.516	0.303	—	—	—	—
2007	—	—	—	—	1.233	0.109	-1.569	0.058	—	—	—	—

[a] κ-gen = κ-generalized; SM = Singh-Maddala; D = Dagum type I. The null hypothesis is that the compared models are equivalent. Star codes for significance: ∗∗ = 1%, ∗ = 5%

Kuiper (KUI), supremum class Anderson-Darling (AD), Cramér-von Mises (CVM) and quadratic class Anderson-Darling (AD2).[10] We also compute the so-called "upper tail" Anderson-Darling statistic, both in its supremum (ADup) and quadratic (AD2up) version, which is convenient to use when it is necessary to test the goodness-of-fit of a distribution in the right tail of the data, while the fit in the left tail or around the median is of less importance (see Chernobai et al. 2005). All *p*-values are derived by making use of a nonparametric bootstrap method (Efron and Tibshirani 1993).[11] That is, given our n-vector of incomes, we generate 1,000 synthetic datasets by drawing new sequences of n observations uniformly at random from the original data. We then fit each synthetic dataset individually to the three distributions and calculate the test statistics for each one relative to its own models. Then we simply count what fraction of the time each resulting statistic is larger than the value for the empirical data. This fraction is the *p*-value for each 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.[12]

Table 7 reports the goodness-of-fit results for the six sets of data. *P*-values are always larger than 0.05, meaning that (if one takes 5% as the relevant significance level) in all cases the data can be statistically described by the three densities.

[10] For more formal definitions see Stephens (1986). Notice, however, that within the class of supremum-type statistics for goodness-of-fit the KS distance tends to be more sensitive near the center of the distribution with respect to the tails, whereas the KUI and AD tests provide equal sensitivity at the tails as at the median. Similarly for the quadratic-type goodness-of-fit statistics, the AD2 distance places more weight on observations in the tails of the distribution than the CVM criterion. Again, the KS test is known to be less powerful than all the other tests in all practical situations (e.g. Thode 2002).

[11] One of the features of the EDF statistics is that their distributions are 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 procedure described here is instead recommended.

[12] Note crucially that for each synthetic dataset we compute the test statistics relative to the best-fit models for that dataset, not relative to the distributions fitted to the original data. In this way we ensure that we are performing for each synthetic dataset the same calculations that we performed for the real data, a crucial requirement if we wish to get unbiased estimates of the *p*-values.

Table 6 Conclusions drawn from Vuong testing approach to model selection[a]

Country	Conclusion	Comparison at the 5% significance level		Comparison at the 1% significance level	
		H_f = κ-gen H_g = SM	H_f = κ-gen H_g = D	H_f = κ-gen H_g = SM	H_f = κ-gen H_g = D
Great Britain (BHPS-CNEF)	H_f better	1.000	0.000	0.857	0.000
	H_g better	0.000	0.714	0.000	0.286
	H_f and H_g equivalent	0.000	0.286	0.143	0.714
Germany (GSOEP-CNEF)	H_f better	0.875	0.042	0.750	0.042
	H_g better	0.000	0.583	0.000	0.292
	H_f and H_g equivalent	0.125	0.375	0.250	0.667
United States (PSID-CNEF)	H_f better	0.636	0.455	0.500	0.227
	H_g better	0.000	0.182	0.000	0.045
	H_f and H_g equivalent	0.364	0.364	0.500	0.727

[a] κ-gen = κ-generalized; SM = Singh-Maddala; D = Dagum type I. The numbers denote the (relative) frequency of times that a given conclusion is reached for each pairwise model comparison. Figures might not add up because of rounding

However, fitting the κ-generalized distribution results both in lower values of the test statistics and higher p-values, thus offering superior performance over the Singh-Maddala and Dagum type I models. In particular, this conclusion is strongly supported by the "upper tail" Anderson-Darling tests.

The above evidence holds vis-à-vis a further check involving a class of fit criteria having a distributional interpretation that is close to the GE inequality indices (Bandyopadhyay et al. 2009). Members of this class are given by

$$J_\alpha(\mathbf{x},\mathbf{y}) = \frac{1}{n(\alpha^2-\alpha)}\sum_{i=1}^{n}\left[\left(\frac{x_i}{\mu_1}\right)^{\alpha}\left(\frac{y_i}{\mu_1}\right)^{1-\alpha}-1\right], \tag{33}$$

where $\mathbf{x}$ is the sample vector of incomes, $\mathbf{y}$ the vector of the corresponding quantiles for the theoretical distribution, μ_1 and μ_2 the means of the marginal distributions of $\mathbf{x}$ and $\mathbf{y}$, and $\alpha \in \mathbb{R}$ a sensitivity parameter that can be calibrated according to which part of the distribution one wants the goodness-of-fit criterion (33) to be particularly sensitive: choosing a large positive value for α would put a lot of weight on discrepancies between the proposed model of income distribution and the data in the upper tail, whereas choosing a substantial negative value would put more weight on lower-tail discrepancies.[13] Formally, we test the hypothesis $H_0 : J_\alpha(\mathbf{x},\mathbf{y}) = 0$ against the alternative $H_1 : J_\alpha(\mathbf{x},\mathbf{y}) \neq 0$, where $\mathbf{y}$ is the vector of quantiles $y_i = F_*^{-1}\left(\frac{i}{n+1}\right)$, $i = 1,\ldots,n$, derived from the three models under scrutiny.[14] Table 8 reports the estimated values of J_α for $\alpha = -1,0,1,2$ along with the associated p-values computed by performing 1,000 bootstrap samplings (see the discussion above). Observe that all the three distributions provide a satisfactory fit to the data (high p-values), but the discrepancy with them is always larger in the case of the Singh-Maddala and Dagum type I models independently of the part of the distribution one is interested in testing for goodness-of-fit. In particular, the κ-generalized density still results in a superior fit at the upper end of the distribution according to the "top-sensitive" goodness-of-fit criterion J_2.

Our statistical findings are also detectable through graphical analysis. Figure 2 presents for all GSOEP-CNEF waves under study the relationship between the income log-rank and log-size. This double-logarithmic framework, known as the Zipf plot, has been used rarely in economics, but is more common in physics (see e.g. Takayasu 1990).[15] In particular, it is

[13] Like the GE class of inequality measures, expression (33) is not defined for $\alpha = 0$ and $\alpha = 1$, as the denominator $n(\alpha^2-\alpha) = 0$ in both cases. Expressions for these values of α (the "middle-sensitive" goodness-of-fit criteria) are therefore calculated by using L'Hôpital rule, by which the limit of an undefined ratio between two functions of the same variable is equal to the limit of the ratio of their first derivatives. Expressions for each J_α index other than for the cases $\alpha = 0,1$ can be, instead, derived by substitution.

[14] Note that we use $\frac{i}{n+1}$ rather than $\frac{i}{n}$. Had we used $\frac{i}{n}$ then it would automatically be set to 1 when $i = n$ and the inversion of F_* would return an infinite value—see e.g. expression (8).

[15] Let $\mathbf{x} = (x_1,\ldots,x_n)$ be a set of n incomes for which the cumulative distribution function is $\hat{F}(x_i) = \frac{i}{n}$, $i = 1,\ldots,n$, and suppose that the observations are ordered from largest to smallest so that the index i is the rank of x_i. The Zipf plot of the sample is the graph of $\log i$ against $\log x_i$. Because of the ranking, $\frac{i}{n} = 1-\hat{F}(x_i)$, so $\log i = \log\left[1-\hat{F}(x_i)\right] + \log n$. Thus, the log of the rank is simply a transformation of the cumulative distribution function.

Table 7 Goodness-of-fit tests based on the EDF for GSOEP-CNEF data, 2002–2007[a]

Wave	Model	KS	KUI	AD	ADup	CVM	AD2	AD2up
	κ-gen	**1.474 (0.594)**	**2.565 (0.719)**	**365.950 (0.486)**	**224.245 (0.667)**	**0.465 (0.615)**	**3.682 (0.633)**	**28.655 (0.693)**
2002	SM	2.066 (0.534)	3.572 (0.653)	1,037.865 (0.491)	882.540 (0.654)	0.919 (0.573)	5.982 (0.576)	81.677 (0.551)
	D	2.132 (0.553)	3.800 (0.650)	500.319 (0.491)	1,221.453 (0.660)	1.113 (0.563)	7.272 (0.572)	134.325 (0.554)
	κ-gen	**1.090 (0.865)**	**2.102 (0.821)**	**567.808 (0.489)**	**357.853 (0.494)**	**0.300 (0.650)**	**2.381 (0.651)**	**23.093 (0.723)**
2003	SM	1.529 (0.597)	2.649 (0.756)	1,710.225 (0.491)	1,552.215 (0.485)	0.527 (0.580)	3.618 (0.596)	33.167 (0.548)
	D	1.513 (0.672)	2.807 (0.701)	836.219 (0.485)	1,712.478 (0.501)	0.569 (0.580)	3.936 (0.577)	43.227 (0.553)
	κ-gen	**1.222 (0.740)**	**2.088 (0.871)**	**6,022.054 (0.491)**	**4,728.169 (0.440)**	**0.349 (0.633)**	**2.953 (0.638)**	**166.113 (0.502)**
2004	SM	1.664 (0.683)	2.941 (0.730)	23,144.977 (0.489)	27,474.108 (0.429)	0.681 (0.570)	4.666 (0.581)	876.692 (0.493)
	D	1.669 (0.690)	3.021 (0.715)	7,356.811 (0.476)	30,271.160 (0.458)	0.719 (0.568)	5.110 (0.569)	976.230 (0.495)
	κ-gen	**0.918 (0.909)**	**1.826 (0.824)**	**144.364 (0.480)**	**2,530.707 (0.469)**	**0.243 (0.660)**	**1.938 (0.669)**	**72.351 (0.501)**
2005	SM	1.241 (0.691)	2.280 (0.717)	382.472 (0.477)	13,912.534 (0.435)	0.313 (0.618)	2.073 (0.652)	303.558 (0.479)
	D	1.116 (0.824)	2.203 (0.743)	220.179 (0.482)	13,292.489 (0.443)	0.315 (0.622)	2.106 (0.639)	294.161 (0.481)
	κ-gen	**0.742 (0.974)**	**1.448 (0.944)**	**65.163 (0.595)**	**3,045.142 (0.443)**	**0.138 (0.727)**	**1.193 (0.742)**	**59.764 (0.526)**
2006	SM	1.181 (0.779)	2.131 (0.809)	145.467 (0.600)	15,625.518 (0.441)	0.324 (0.606)	2.087 (0.654)	285.903 (0.504)
	D	1.121 (0.782)	2.070 (0.803)	88.641 (0.597)	16,889.356 (0.447)	0.289 (0.608)	1.897 (0.648)	315.304 (0.500)
	κ-gen	**1.124 (0.751)**	**2.138 (0.712)**	**86.756 (0.587)**	**376.259 (0.608)**	**0.247 (0.654)**	**1.924 (0.653)**	**25.336 (0.630)**
2007	SM	1.546 (0.607)	2.613 (0.782)	198.295 (0.560)	1,451.253 (0.642)	0.444 (0.603)	3.016 (0.615)	95.369 (0.531)
	D	1.547 (0.586)	2.776 (0.641)	111.898 (0.569)	1,533.487 (0.634)	0.446 (0.598)	3.046 (0.614)	106.879 (0.539)

[a] KS = Kolmogorov-Smirnov; KUI = Kuiper; AD = supremum class Anderson-Darling; ADup = supremum class "upper tail" Anderson-Darling; CVM = Cramér-von Mises; AD2 = quadratic class Anderson-Darling; AD2up = quadratic class "upper tail" Anderson-Darling. The null hypothesis is that data come from the fitted κ-generalized (κ-gen), Singh-Maddala (SM) or Dagum type I (D) distributions. *P*-values (in round brackets) have been computed via 1,000 bootstrap replications. Boldface entries denote the best fitting model

natural to use when focusing on the top part of the distribution because it accentuates the upper tail, making it easier to detect deviations in that part of the distribution from the theoretical prediction of a particular model. The lines shows the predicted Zipf plots obtained from the fit of the models considered. As the figure reveals, all of them are in good agreement with the actual data in the low-middle range of the income distributions. However, there is a systematic departure of empirical observations from the Singh-Maddala and Dagum type I predictions at the top tail, whereas in the same part of the income distributions the theoretical Zipf plot for the κ-generalized distribution lies much closer to the empirical one.

To assess more robustly whether the κ-generalized distribution provides a statistically better fit in the high-income range of German data from 2002 to 2007 as compared to the Singh-Maddala and Dagum type I, we repeat our hypothesis-testing exercises by fitting each model to observations in the richest 50% of the income distribution only (to ensure that model fit is maximized at the top of it)[16] and then running goodness-of-fit tests with appropriate corrections for left-truncation.[17] Specifically, we determine the ultimate best fit on the basis of the ADup, AD2up and J_2 statistics, which assign a higher weight to observations in the upper tail of the distribution. As can be seen from Table 9, these measures still suggest a superior fit of the κ-generalized density in the upper tail of the left-truncated samples for all cases, thus confirming that our model does a better job than the Singh-Maddala and Dagum type I in the top part of the income distribution. This is of particular relevance in the current context, since the upper tail of the three densitites is heavy in that it decays like a power function as income increases.[18]

[16] We chose the 50th percentile as the left-truncation point after experiments balancing goodness-of-fit with ease of estimation, since the numerical optimization routine implemented in `R`'s `nlminb` command did not converge with higher truncation levels. The sample log-likelihood for each year's data was specified as $\log L = \sum_{i=1}^{n} \{\log f_{*}(x_i) - \log[1 - F_{*}(z)]\}$, where $i = 1, \ldots, n$ indexes each individual sample observation and z is the level of income corresponding to the left-truncation point. We do not report estimates for each year but they are available upon request.

[17] See Chernobai et al. (2005). Corrections concerned the cumulative distribution and quantile functions of the three models considered. Nadarajah and Kotz (2006) provide formulas and `R` programs for computing several quantities of interest for the truncated versions of any given distribution.

[18] See Clementi et al. (2010) on the upper tail behaviour of the κ-generalized. For the other distributions, see Kleiber (1996, 2008) and Kleiber and Kotz (2003).

Table 8 J_α statistics for GSOEP-CNEF data, 2002–2007[a]

Wave	Model	$J_\alpha \times 10^2$			
		$\alpha=-1$	$\alpha=0$	$\alpha=1$	$\alpha=2$
2002	κ-gen	**0.149 (0.620)**	**0.134 (0.577)**	**0.137 (0.572)**	**0.145 (0.564)**
	SM	0.286 (0.540)	0.289 (0.513)	0.332 (0.514)	0.398 (0.518)
	D	0.321 (0.541)	0.345 (0.523)	0.406 (0.520)	0.498 (0.518)
2003	κ-gen	**0.104 (0.740)**	**0.087 (0.705)**	**0.089 (0.661)**	**0.094 (0.619)**
	SM	0.146 (0.565)	0.129 (0.524)	0.150 (0.517)	0.186 (0.516)
	D	0.143 (0.549)	0.137 (0.516)	0.161 (0.514)	0.200 (0.512)
2004	κ-gen	**0.681 (0.541)**	**0.710 (0.477)**	**0.971 (0.474)**	**1.463 (0.476)**
	SM	0.918 (0.544)	0.963 (0.469)	1.441 (0.469)	2.482 (0.470)
	D	0.858 (0.521)	0.980 (0.471)	1.479 (0.470)	2.568 (0.467)
2005	κ-gen	**0.338 (0.482)**	**0.414 (0.476)**	**0.548 (0.468)**	**0.784 (0.459)**
	SM	0.435 (0.484)	0.571 (0.479)	0.835 (0.470)	1.375 (0.459)
	D	0.430 (0.484)	0.567 (0.476)	0.830 (0.473)	1.366 (0.464)
2006	κ-gen	**0.346 (0.486)**	**0.430 (0.478)**	**0.577 (0.470)**	**0.842 (0.458)**
	SM	0.520 (0.481)	0.669 (0.478)	0.959 (0.467)	1.560 (0.454)
	D	0.529 (0.480)	0.686 (0.475)	0.987 (0.471)	1.613 (0.457)
2007	κ-gen	**0.153 (0.551)**	**0.154 (0.525)**	**0.170 (0.529)**	**0.193 (0.527)**
	SM	0.278 (0.522)	0.305 (0.504)	0.371 (0.503)	0.474 (0.504)
	D	0.277 (0.521)	0.313 (0.511)	0.384 (0.512)	0.493 (0.514)

[a] The null hypothesis is that the discrepancy between the fitted κ-generalized (κ-gen), Singh-Maddala (SM) or Dagum type I (D) distribution and the data is zero. *P*-values (in round brackets) have been computed via 1,000 bootstrap replications. Boldface entries denote the best fitting model

4 Summary

We have derived a function to describe the size distribution of incomes starting from a generalization of the maximum entropy method that follows from Kaniadakis (2001, 2002, 2005). Expressions for the shape, moments and various tools for inequality measurement that are functions of the parameters in the model have been given. The performance of the distribution has been checked against real data on personal income for Great Britain, Germany and the United States in different years and has been found to fit remarkably well. Furthermore, we have found that in a satisfactory number of cases the model is not to be considered statistically inferior when compared to other existing functional forms that have been considered successful in describing the income size distribution. In particular, the new proposed model suggests a statistically superior fit in the right tail of data with respect to the others in many instances.

References

Arnold BC, Laguna L (1977) On generalized Pareto distributions with applications to income data. Iowa State University Press, Ames

Atkinson AB (1970) On the measurement of inequality. J Econ Theory 2:244–263

Bach S, Corneo G, Steiner V (2009) From bottom to top: The entire income distribution in Germany, 1992–2003. Rev Income Wealth 55:303–330

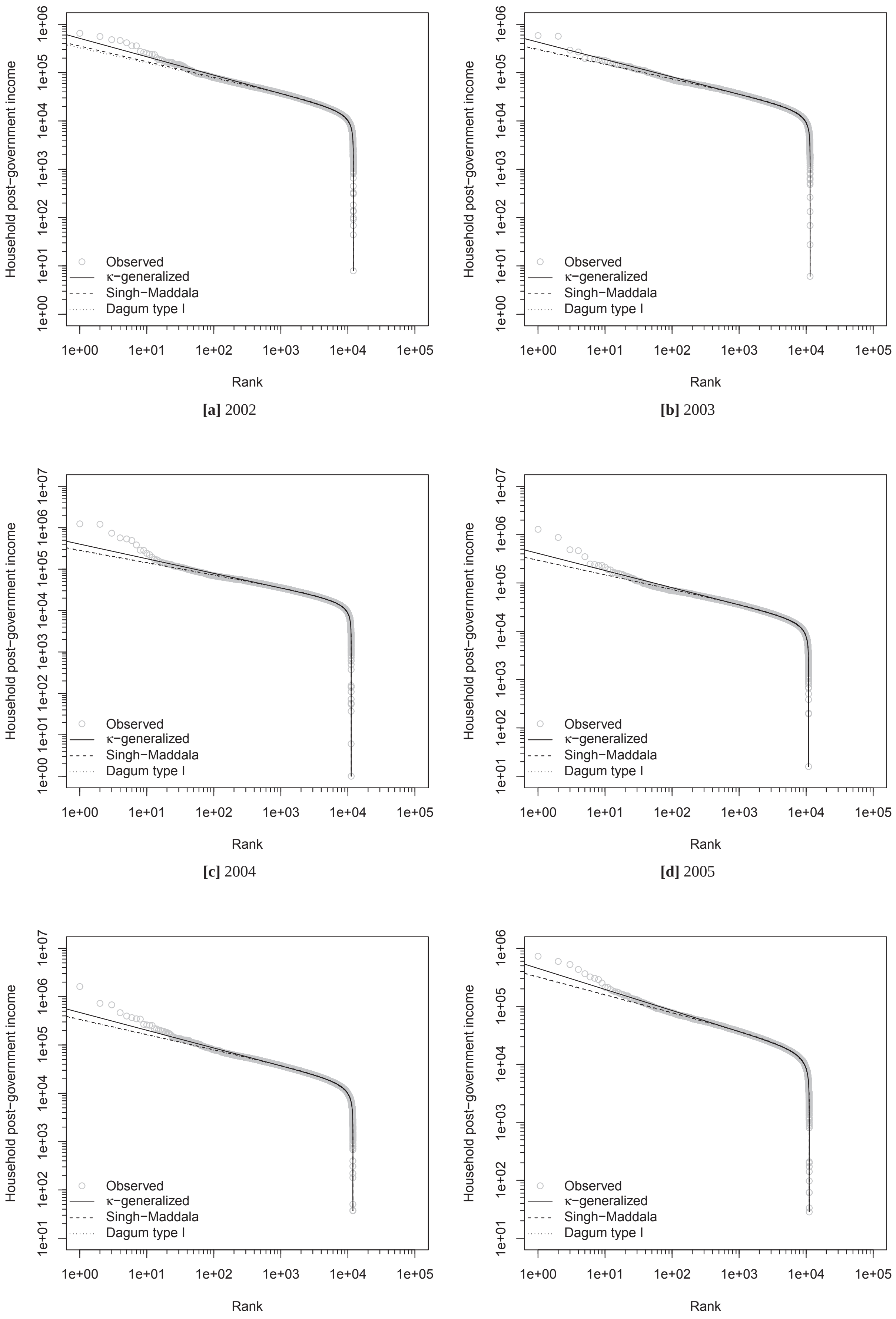
Bandyopadhyay S, Cowell FA, Flachaire E (2009) Goodness-of-Fit: An Economic Approach. STICERD - Distributional Analysis Research Programme Papers 101, Suntory and Toyota International Centres for Economics and Related Disciplines, LSE. `http://sticerd.lse.ac.uk/dps/darp/darp101.pdf`

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 (Social insurance in a dynamic society. Papers in honor of the 65th birthday of Richard Hauser), Geburtstag Campus, Frankfurt and New York, pp 354–376

Champernowne DG (1953) A model of income distribution. Econ J 63:318–351

Chernobai A, Rachev S, Fabozzi F (2005) Composite goodness-of-fit tests for left-truncated loss samples. Technical Report, Department of Statistics and Applied Probability, University of California Santa Barbara. `http://www.pstat.ucsb.edu/research/papers/KSmissing20050604-JFE.pdf`

Clementi F, Gallegati M, Kaniadakis G (2010) A model of personal income distribution with application to Italian data. Empir Econ 39:559–591



[a] 2002 **[b]** 2003

[c] 2004 **[d]** 2005

[e] 2006 **[f]** 2007

Fig. 2 Zipf plot (double-logarithmic plot of income vs. rank) for GSOEP-CNEF data, 2002–2007. The lines are the predicted Zipf plots obtained from the fit of the models under scrutiny

Table 9 Goodness-of-fit tests on sub-samples of GSOEP-CNEF data obtained by truncating the lower 50% of the distributions, 2002–2007[a]

Wave	Model	ADup	AD2up	$J_2 \times 10^2$
	κ-gen	**440.321 (0.664)**	**49.524 (0.555)**	**0.399 (0.518)**
2002	SM	447.491 (0.661)	50.720 (0.558)	0.402 (0.509)
	D	472.309 (0.576)	53.957 (0.504)	0.420 (0.478)
	κ-gen	**853.236 (0.472)**	**21.972 (0.524)**	**0.201 (0.484)**
2003	SM	914.227 (0.477)	23.820 (0.521)	0.209 (0.482)
	D	965.152 (0.347)	25.251 (0.444)	0.218 (0.423)
	κ-gen	**8,069.806 (0.497)**	**374.807 (0.549)**	**2.696 (0.499)**
2004	SM	8,362.094 (0.443)	387.498 (0.483)	2.713 (0.440)
	D	9,134.727 (0.439)	421.817 (0.480)	2.784 (0.440)
	κ-gen	**7,402.94 (0.525)**	**229.861 (0.564)**	**1.746 (0.510)**
2005	SM	7,839.985 (0.458)	242.191 (0.519)	1.763 (0.453)
	D	8,141.219 (0.426)	251.549 (0.469)	1.793 (0.442)
	κ-gen	**4,334.116 (0.454)**	**113.217 (0.540)**	**1.549 (0.477)**
2006	SM	4,535.821 (0.426)	118.336 (0.514)	1.571 (0.456)
	D	5,372.537 (0.444)	139.648 (0.511)	1.666 (0.457)
	κ-gen	**552.449 (0.630)**	**42.479 (0.559)**	**0.432 (0.506)**
2007	SM	580.253 (0.631)	45.263 (0.563)	0.445 (0.494)
	D	613.266 (0.635)	48.426 (0.552)	0.461 (0.496)

[a] ADup = supremum class "upper tail" Anderson-Darling test; AD2up = quadratic class "upper tail" Anderson-Darling test; J_2 = "top-sensitive" Bandopadhyay-Cowell-Flachaire test. The null hypothesis is that data come from the fitted left-truncated κ-generalized (κ-gen), Singh-Maddala (SM) or Dagum type I (D) distributions. *P*-values (in round brackets) have been computed via 1,000 bootstrap replications. Boldface entries denote the best fitting model

Cowell FA (1980a) Generalized entropy and the measurement of distributional change. Europ Econ Rev 13:147–159

Cowell FA (1980b) On the structure of additive inequality measures. Rev Econ Stud 47:521–531

Cowell FA (1995) Measuring inequality. Prentice Hall/Harvester Wheatsheaf, Hemel Hempstead

Cowell FA, Kuga K (1981a) Additivity and the entropy concept: An axiomatic approach to inequality measurement. J Econ Theory 25:131–143

Cowell FA, Kuga K (1981b) Inequality measurement: An axiomatic approach. Europ Econ Rev 15:287–305

Dagum C (1977) A new model of personal income distribution: Specification and estimation. Econ Appl 30:413–436

Drăgulescu AA, Yakovenko VM (2001) Evidence for the exponential distribution of income in the USA. Eur Phys J B 20:585–589

Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman and Hall, New York

Feenberg D, Coutts E (1993) An introduction to the TAXSIM model. J Policy Anal Manag 12:189–194

Gibrat R (1931) Les inégalités économiques. Applications: Aux inégalités des richesses, à la concentration des entreprises, aux population des villes, aux statistiques des familles, etc., d'une loi nouvelle: La loi de l'effet proportionnel. Librairie du Recueil Sirey, Paris

Gini C (1914) Sulla misura della concentrazione e della variabilità dei caratteri. Atti del Reale Istituto veneto di scienze, lettere ed arti 73:1201–1248. English translation: On the measurement of concentration and variability of characters. Metron 63:3–38, 2005

Hagenaars AJM, De Vos K, Zaidi MA (1994) Poverty statistics in the late 1980s: Research based on micro-data. Office for Official Publications of the European Communities, Luxembourg.

Haisken-DeNew JP, Frick JR (2005) Desktop Companion to the German Socio-Economic Panel (SOEP). German Institute for Economic Research (DIW). `http://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.38951.de/dtc.409713.pdf`

Jaynes ET (1957a) Information theory and statistical mechanics. Phys Rev 106:620–630

Jaynes ET (1957b) Information theory and statistical mechanics. II. Phys Rev 108:171–190

Jenkins SP (2009) Distributionally-sensitive inequality indices and the GB2 income distribution. Rev Income Wealth 55:392–398

Kakwani N (1980) Income inequality and poverty: Methods of estimation and policy applications. Oxford University Press, New York

Kaniadakis G (2001) Non-linear kinetics underlying generalized statistics. Physica A 296:405–425

Kaniadakis G (2002) Statistical mechanics in the context of special relativity. Phys Rev E 66:056125

Kaniadakis G (2005) Statistical mechanics in the context of special relativity. II. Phys Rev E 72:036108

Kapur JN (1989) Maximum entropy models in science and engineering. Wiley Eastern, New Delhi

Kleiber C (1996) Dagum vs. Singh-Maddala income distributions. Econ Lett 53:265–268

Kleiber C (1997) The existence of population inequality measures. Econ Lett 57:39–44

Kleiber C (2008) A guide to the Dagum distributions. In: Chotikapanich D (ed), Modeling income distributions and Lorenz curves, Springer Science+Business Media, LLC, New York, pp 97–117

Kleiber C, Kotz S (2003) Statistical size distributions in economics and actuarial sciences. John Wiley & Sons, New York

Leipnik RB (1990) A maximum relative entropy principle for distribution of personal income with derivations of several known income distributions. Commun Stat Theory 19:1003–1036

Lorenz MO (1905) Methods of measuring the concentration of wealth. Pub Am Stat Assn 9:209–219

McDonald JB, Xu YJ (1995) A generalization of the beta distribution with applications. J Econometrics 66:133–152. Errata. J Econometrics 69:427–428

Nadarajah S, Kotz S (2006) R programs for computing truncated distributions. J Stat Softw 16: Code Snippet 2. http://www.jstatsoft.org/v16/c02/paper

Ord JK, Patil GP, Taillie C (1981) The choice of a distribution to describe personal incomes. In: Taillie C, Patil GP, Baldessari BA (eds) Statistical distributions in scientific work, vol. 6, D. Reidel Publishing Company, Dordrecht, pp 193–201

R Development Core Team (2011) R: A Language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. http://www.R-project.org

Salem ABZ, Mount TD (1974) A convenient descriptive model of income distribution: The gamma density. Econometrica 42:1115–1127

Shannon CE (1948) A mathematical theory of communication. Bell Sys Tech J 27:379–423 and 623–657

Singh SK, Maddala GS (1976) A function for size distribution of incomes. Econometrica 44:963–970

Stephens MA (1986) Tests based on EDF statistics. In: D'Agostino RB, Stephens MA (eds) Goodness-of-fit techniques, Marcel Dekker, New York, pp 97–193

Takayasu H (1990) Fractals in the physical sciences. Manchester University Press, Manchester

Theil H (1967) Economics and information theory. North-Holland, Amsterdam

Thode Jr HC (2002) Testing for normality. Marcel Dekker, New York

Vuong QH (1989) Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57:307–333