IOPscience iopscience.iop.org

Home Search Collections Journals About Contact us My IOPscience

A generalized statistical model for the size distribution of wealth

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

J. Stat. Mech. (2012) P12006

(http://iopscience.iop.org/1742-5468/2012/12/P12006)

View the table of contents for this issue, or go to the journal homepage for more

Download details:
IP Address: 95.74.179.217
The article was downloaded on 06/12/2012 at 13:11

Please note that terms and conditions apply.

Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

A generalized statistical model for the size distribution of wealth

F Clementi[1], M Gallegati[2] and G Kaniadakis[3]

[1] Dipartimento di Scienze Politiche, della Comunicazione e delle Relazioni Internazionali, Università degli Studi di Macerata, Piazza G Oberdan 3, I-62100 Macerata, Italy
[2] Dipartimento di Scienze Economiche e Sociali, Università Politecnica delle Marche, Piazzale R Martelli 8, I-60121 Ancona, Italy
[3] Dipartimento Scienza Applicata e Tecnologia, Politecnico di Torino, Corso Duca degli Abruzzi 24, I-10129 Torino, Italy
E-mail: fabio.clementi@unimc.it, mauro.gallegati@univpm.it and giorgio.kaniadakis@polito.it

Received 21 September 2012
Accepted 13 November 2012
Published 6 December 2012

Online at stacks.iop.org/JSTAT/2012/P12006
doi:10.1088/1742-5468/2012/12/P12006

Abstract. In a recent paper in this journal (Clementi *et al* 2009 *J. Stat. Mech.* P02037), we proposed a new, physically motivated, distribution function for modeling individual incomes, having its roots in the framework of the κ-generalized statistical mechanics. The performance of the κ-generalized distribution was checked against real data on personal income for the United States in 2003. In this paper we extend our previous model so as to be able to account for the distribution of wealth. Probabilistic functions and inequality measures of this generalized model for wealth distribution are obtained in closed form. In order to check the validity of the proposed model, we analyze the US household wealth distributions from 1984 to 2009 and conclude an excellent agreement with the data that is superior to any other model already known in the literature.

Keywords: rigorous results in statistical mechanics, applications to game theory and mathematical economics, scaling in socio-economic systems

ArXiv ePrint: 1209.4787

© 2012 IOP Publishing Ltd and SISSA Medialab srl
1742-5468/12/P12006+25$33.00

J. Stat. Mech. (2012) P12006

Contents

1. Introduction **2**

2. The κ-generalized statistical distribution and its properties **4**

3. Specification of finite mixture models for net wealth distribution **5**

4. Moments of finite mixture models for net wealth distribution **6**

5. The Lorenz curve and the Gini ratio of the net wealth distribution models **7**

6. Application **8**

6.1. The US data on household net wealth 8

6.2. Estimation and comparison of finite mixture models for net wealth distribution 13

7. Summary and conclusions **22**

References **23**

1. Introduction

The quantitative and formal development of the personal or size distribution of income and the measurement of income inequality was first introduced by the Italian economist Vilfredo Pareto. He specified his type I model early in 1895 [2], and in 1896 and 1897 his types II and III [3]–[5], and made an inequality interpretation of his shape parameter. Based on Pareto's economic foundations, and on the stochastic foundations afterward developed by other authors [6, 7], the Pareto law (Pareto type I) is now overwhelmingly considered as the income distribution model of high income groups.

Since Pareto's seminal contribution, many probability density functions have been proposed in the literature that are suitable for describing the size distribution of income amongst the population as a whole—see, e.g., the comprehensive survey contained in [8]. Fitting of parametric functional forms has also been common for the distribution of *wealth*[4]. However, the problem for the wealth researcher is that virtually all of the

[4] Income and wealth are commonly used to assess the economic well-being of individuals, families or households. Although some correlation exists between them, the relationship is not perfect: greater income is likely to mean greater wealth, but not always. The two measures, in fact, are not synonymous. Income is a *flow*, since it is meaningful only when defined in relation to a period of time (hourly, weekly, monthly or annual income). Wealth is a *stock*, increasing as new assets are acquired or savings accumulated, and the only time information required is when the stock was measured (no periodicity is necessary). The link between the flow from income and the stock of wealth is obvious: the greater the former, the more rapidly the latter will increase. Accordingly, a high income may be associated with low wealth—this is the case, for example, with young people starting their careers; on the other hand, a low income could accompany high wealth—this is the case with some retirees who have little income but who have accumulated and paid for substantial assets. At a practical level, wealth is distributed much more unequally than income because of life cycle savings and bequest motives [9]. Data on stocks of wealth also present distinctive features in comparison with income data that make empirical analysis non-standard in several

J. Stat. Mech. (2012) P12006

models suggested within the context of the income distribution literature are defined for variables taking only strictly positive values, although published statistical data of wealth distributions give clear evidence of presenting highly significant frequencies of households or individuals with null and/or negative wealth. The early contributions systematically dismissed these frequencies and fitted their respective proposed models to the positive observations only, thus omitting a significant part of the story[5].

To the best of our knowledge, Dagum [15, 16] was the first and only one to specify and test a four-parameter model for wealth distributions (Dagum type II). The fourth parameter in the Dagum model is an estimate of the frequency of economic units with wealth equal to zero. This model is highly relevant to describe total (gross) wealth distribution because of the always large observed percentage of economic units with null total wealth. Dagum [17]–[20] made further developments of his type II model to analyze the distribution of *net* wealth, which is equal to gross wealth minus total debt. The support of the Dagum model of net wealth is the real line $\mathbb{R} = (-\infty, \infty)$, thus allowing us to fit the subset of economic units with null and negative wealth. Furthermore, it contains as particular cases both the Dagum types I and II distributions [15].

In more detail, the Dagum general model of net wealth distribution is a mixture (or a convex combination) of an atomic and two continuous distributions. The atomic distribution concentrates its unit mass of economic agents at zero, and therefore accounts for the economic units with null net wealth. The continuous distribution accounting for the negative net wealth observations is given by a Weibull function. It has a fast left tail convergence to zero, and therefore it has finite moments of all orders. The other continuous distribution, specified as the Dagum type I model, accounts for the positive values of net wealth and presents a heavy right tail, thus having a small number of finite moments of positive order. This different behavior at the two tails of the distribution stems from the fact that, unlike the right tail of income and (gross or net) wealth distributions, which tend slowly to zero when income and wealth tend to infinity, the distribution of the negative values (left tail) of net wealth tends very fast to zero when the variable tends to minus infinity, since economic units face a short term challenge of either moving out of the negative range of net wealth or bankruptcy.

The purpose of the present work is to provide estimates for the 1984–2009 US net wealth distributions of this Dagum general model, partly motivated by the fact that there are no applications other than Dagum's ones [17]–[20] that we are aware of—the only notable exception being represented by [21], who fitted the model to Finnish net wealth data in 1984 and 1989. Furthermore, since other approaches can be entertained and comparative study of their relative merits performed, we also explore the possibility of using alternative distributions to characterize positive net wealth values. That is, we formalize, analyze and fit to our US net wealth data finite mixture models based upon the Singh–Maddala and κ-generalized distributions as specifications for the positive values. The Singh–Maddala distribution [22] is known to be very successful in fitting the empirical

ways (see the ongoing discussion above for details). However, as far as the shape of the particular distribution is concerned, income and wealth share qualitatively the same characteristics: many empirical wealth distributions are indeed positively skewed with 'fat' and long right-hand tails, as are income distributions.

[5] In the 1950s, [10] and [11] proposed the Pareto type I model and the lognormal distribution, respectively. Afterward, other models were proposed: in 1969 the Pareto types I and II by [12]; in 1975, the log–logistic by [13] and the Pearson type V by [14]. All of these models are restricted to describe only the positive range of wealth, since they are not defined for zero and/or negative values.

J. Stat. Mech. (2012) P12006

income distributions. The κ-generalized distribution was proposed in previous works of ours [1], [23]–[26] to describe the distribution of personal income in some developed economies for different years. Positive conclusions were drawn about its ability to provide an accurate description of the observed distributions, ranging from the low to the middle region, and up to the right tail. The empirical success of the κ-generalized distribution was complemented by goodness-of-fit comparisons showing that fitting the distribution to available income data offers superior performance over other existing models (including the Singh–Maddala and Dagum type I) in a significant number of cases.

The content of the paper is organized as follows: section 2 recalls some basic properties of the κ-generalized statistical distribution; section 3 presents the main analytical properties of the net wealth distribution models; section 4 deduces their corresponding moments; section 5 derives the parametric forms of the Lorenz curve and Gini ratio for the distribution of net wealth; section 6 fits the specified models to the US data on household net wealth covering the years 1984–2009 and section 7 presents the conclusions.

2. The κ-generalized statistical distribution and its properties

After 2001, a physical mechanism emerging in the context of special relativity was proposed by one of us [27]–[30], predicting a deformation of the exponential function. According to this mechanism, the classical exponential distribution transforms into a new distribution, which at high energies presents a Pareto fat tail. More precisely, this mechanism deforms the ordinary exponential function $\exp(x)$ into the generalized exponential function $\exp_\kappa(x)$ given by

$$\exp_\kappa(x) = \left(\sqrt{1+\kappa^2 x^2} + \kappa x\right)^{1/\kappa}. \tag{1}$$

The above deformation is generated by the fact that the propagation of the information has a finite speed, and the deformation parameter κ is proportional to the reciprocal of this speed. The κ-generalized exponential has the important properties

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

$$\exp_\kappa(x) \underset{x\to 0}{\sim} \exp(x). \tag{2b}$$

It is remarkable that for classical systems where the information propagates instantaneously $\kappa = 0$ results, so the ordinary exponential emerges naturally after noting that $\exp_0(x) = \exp(x)$. Moreover, in the low energy region $x \to 0$, according to equation (2b), the exponential distribution emerges again, because the system behaves classically. In contrast, in systems where the information propagates with a finite speed—these systems are intrinsically relativistic—$\kappa \neq 0$ results, so the exponential tails become fat according to equation (2a) and the Pareto law emerges.

The generalized exponential represents a very useful and powerful tool to formulate a new statistical theory able to treat systems described by distribution functions exhibiting power-law tails and admitting a stable entropy [31, 32]. Furthermore, non-linear evolution models already known in statistical physics [33]–[35] can be easily adapted or generalized within the new theory.

J. Stat. Mech. (2012) P12006

The function $\exp_\kappa(x)$ was also adopted successfully in the analysis of various nonphysical systems [36]–[38]. In [1], [23]–[26] we used the function $\exp_\kappa(x)$ to model the personal income distribution by defining the cumulative distribution function through

$$F(x) = 1 - \exp_\kappa\left[-(x/\beta)^\alpha\right], \qquad x \geq 0, \qquad \alpha, \beta > 0, \qquad \kappa \in [0, 1). \tag{3}$$

The corresponding probability density function reads

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} \frac{\exp_\kappa\left[-(x/\beta)^\alpha\right]}{\sqrt{1+\kappa^2(x/\beta)^{2\alpha}}}. \tag{4}$$

It follows immediately that for low incomes the distribution function behaves similarly to the Weibull model $F(x) = 1 - \exp\left[-(x/\beta)^\alpha\right]$, whereas for large x it approaches a Pareto distribution with scale $\beta(2\kappa)^{-1/\alpha}$ and shape α/κ, i.e. $F(x) \underset{x\to+\infty}{\sim} 1 - \left[(\beta(2\kappa)^{-1/\alpha})/x\right]^{\alpha/\kappa}$. Similarly, the density function for $x \to 0^+$ behaves as a Weibull distribution $f(x) = (\alpha/\beta)(x/\beta)^{\alpha-1}\exp\left[-(x/\beta)^\alpha\right]$, while for $x \to +\infty$ it reduces to the Pareto law $f(x) = ((\alpha/\kappa)\left[\beta(2\kappa)^{-1/\alpha}\right]^{\alpha/\kappa})/x^{(\alpha/\kappa)+1}$.

3. Specification of finite mixture models for net wealth distribution

The general model of net wealth distribution as a mixture of an atomic and two continuous distributions takes the form

$$f(w) = \sum_{i=1}^{3} \theta_i f_i(w), \qquad -\infty < w < \infty, \qquad \theta_i \geq 0, \qquad \sum_i \theta_i = 1, \tag{5}$$

where w denotes the wealth variable and $\{\theta_i\}_{i=1,\ldots,3}$ are the mixture proportions. The two-parameter Weibull density

$$f_1(w) = \frac{s}{\lambda}\left(\frac{|w|}{\lambda}\right)^{s-1} \exp\left[-\left(\frac{|w|}{\lambda}\right)^s\right], \qquad w < 0, \qquad (s, \lambda) > 0 \tag{6}$$

describes the distribution of economic units with negative net wealth, while the null net wealth observations are accounted for by a distribution that concentrates its unit mass at $w = 0$, i.e.

$$f_2(0) = 1. \tag{7}$$

The other continuous distribution, $f_3(w)$, accounts for the positive values of net wealth, and is alternatively specified by the following three-parameter densities:

(1) the Singh–Maddala

$$f_3^{\mathrm{SM}}(w) = \frac{aqw^{a-1}}{b^a\left[1+(w/b)^a\right]^{1+q}}, \qquad w > 0, \qquad (a, b, q) > 0; \tag{8}$$

(2) the Dagum type I

$$f_3^{\mathrm{D}}(w) = \frac{apw^{ap-1}}{b^{ap}\left[1+(w/b)^a\right]^{p+1}}, \qquad w > 0, \qquad (a, b, p) > 0; \tag{9}$$

J. Stat. Mech. (2012) P12006

(3) the κ generalized given by equation (4).

The corresponding cumulative distribution function reads

$$F(w) = \theta_1 F_1(w) + \theta_2 F_2(w) + \theta_3 F_3(w), \qquad \theta_1 + \theta_2 = \rho, \qquad \theta_3 = 1 - \rho, \tag{10}$$

where

$$F_1(w) = \begin{cases} \exp\left[-\left(\frac{|w|}{\lambda}\right)^s\right], & w < 0; \\ 1, & w \geq 0; \end{cases} \tag{11a}$$

$$F_2(w) = \begin{cases} 0, & w < 0; \\ 1, & w \geq 0; \end{cases} \tag{11b}$$

$$F_3(w) = \begin{cases} 0, & w \leq 0; \\ F_3(w), & w > 0. \end{cases} \tag{11c}$$

Hence

$$F(w) = \begin{cases} \theta_1 \exp\left[-\left(\frac{|w|}{\lambda}\right)^s\right], & w < 0; \\ \rho, & w = 0; \\ \rho + (1-\rho) F_3(w), & w > 0, \end{cases} \tag{12}$$

with $F_3(w)$ having the following alternative mathematical specifications:

$$F_3^{\mathrm{SM}}(w) = 1 - \left[1 + \left(\frac{w}{b}\right)^a\right]^{-q}; \tag{13a}$$

$$F_3^{\mathrm{D}}(w) = \left[1 + \left(\frac{w}{b}\right)^{-a}\right]^{-p}; \tag{13b}$$

$$F_3^{\kappa-\mathrm{gen}}(w) = 1 - \exp_\kappa\left[-\left(\frac{w}{\beta}\right)^\alpha\right]. \tag{13c}$$

4. Moments of finite mixture models for net wealth distribution

It follows from model (5) that the rth-order moment about the origin is[6]

$$\mu_r = E(W^r) = \int_{-\infty}^{\infty} w^r f(w)\,\mathrm{d}w = \theta_1 E_1(W^r) + \theta_2 E_2(W^r) + \theta_3 E_3(W^r), \tag{14}$$

where

$$E_1(W^r) = (-1)^r \lambda^r \Gamma\left(1 + \frac{r}{s}\right) \tag{15}$$

and $E_2(W^r) = 0$.

[6] In what follows, $\Gamma(\cdot)$ stands for the Euler gamma function.

J. Stat. Mech. (2012) P12006

As for $E_3(W^r)$ in the last member of equation (14), according to the alternative distributions considered to characterize positive net wealth values one gets[7]

$$E_3^{\mathrm{SM}}(W^r) = \frac{b^r \Gamma(1+r/a)\Gamma(q-r/a)}{\Gamma(q)}; \tag{16a}$$

$$E_3^{\mathrm{D}}(W^r) = \frac{b^r \Gamma(p+r/a)\Gamma(1-r/a)}{\Gamma(p)}; \tag{16b}$$

$$E_3^{\kappa-\mathrm{gen}}(W^r) = \beta^r (2\kappa)^{-r/\alpha} \frac{\Gamma(1+r/\alpha)}{1+(r/\alpha)\kappa} \frac{\Gamma(1/2\kappa - r/2\alpha)}{\Gamma(1/2\kappa + r/2\alpha)}. \tag{16c}$$

The mean net wealth equals

$$\mu_1 = E(W) = -\theta_1 \lambda \Gamma\left(1+\frac{1}{s}\right) + \theta_3 E_3(W), \tag{17}$$

where $E_3(W)$ is alternatively given by equations (16) with $r=1$.

5. The Lorenz curve and the Gini ratio of the net wealth distribution models

By definition, the Lorenz curve [39] describes a relation between the cumulative distribution function, $F(w)$, and the first cumulative moment distribution function, given by

$$L(u) = \frac{1}{\mu_1}\int_0^w w' f(w')\,\mathrm{d}w' = \frac{1}{\mu_1}\int_0^u w(u')\,\mathrm{d}u', \qquad u \in [0,1], \tag{18}$$

where $u = F(w)$ and $w(u) = F^{-1}(u)$ denotes the quantile function. Given the mathematical structure of the general net wealth distribution model (5) and (10), we have

$$L^{\mathrm{SM}}(u) = \begin{cases} -\dfrac{\lambda\theta_1}{\mu_1}\Gamma\left(1+\dfrac{1}{s}, \log\dfrac{\theta_1}{u}\right), & 0 \le u < \theta_1; \\ -\dfrac{\lambda\theta_1}{\mu_1}\Gamma\left(1+\dfrac{1}{s}\right), & \theta_1 \le u \le \rho; \\ \dfrac{1}{\mu_1}\left\{(1-\rho)\,bq\left[B\left(q-\dfrac{1}{a}, 1+\dfrac{1}{a}\right) - B\left(\left[\dfrac{1-u}{1-\rho}\right]^{1/q}; q-\dfrac{1}{a}, 1+\dfrac{1}{a}\right)\right]\right. \\ \qquad \left. -\lambda\theta_1\Gamma\left(1+\dfrac{1}{s}\right)\right\}, & u > \rho; \end{cases} \tag{19a}$$

$$L^{\mathrm{D}}(u) = \begin{cases} -\dfrac{\lambda\theta_1}{\mu_1}\Gamma\left(1+\dfrac{1}{s}, \log\dfrac{\theta_1}{u}\right), & 0 \le u < \theta_1; \\ -\dfrac{\lambda\theta_1}{\mu_1}\Gamma\left(1+\dfrac{1}{s}\right), & \theta_1 \le u \le \rho; \\ \dfrac{1}{\mu_1}\left\{(1-\rho)\,bpB\left(\left[\dfrac{u-\rho}{1-\rho}\right]^{1/p}; p+\dfrac{1}{a}, 1-\dfrac{1}{a}\right) - \lambda\theta_1\Gamma\left(1+\dfrac{1}{s}\right)\right\}, & u > \rho; \end{cases} \tag{19b}$$

[7] See [8] for relevant expressions. Formulas for the moments of the κ-generalized distribution are given in [1], [24]–[26].

J. Stat. Mech. (2012) P12006

$$L^{\kappa-\text{gen}}(u) = \begin{cases} -\dfrac{\lambda\theta_1}{\mu_1}\Gamma\left(1+\dfrac{1}{s}, \log\dfrac{\theta_1}{u}\right), & 0 \le u < \theta_1; \\ -\dfrac{\lambda\theta_1}{\mu_1}\Gamma\left(1+\dfrac{1}{s}\right), & \theta_1 \le u \le \rho; \\ \dfrac{1}{\mu_1}\left\{\dfrac{(1-\rho)\beta}{(2\kappa)^{1+1/\alpha}}\left[B\left(\dfrac{1}{2\kappa}-\dfrac{1}{2\alpha}, 1+\dfrac{1}{\alpha}\right) - B\left(\left[\dfrac{1-u}{1-\rho}\right]^{2\kappa}; \dfrac{1}{2\kappa}-\dfrac{1}{2\alpha}, 1+\dfrac{1}{\alpha}\right)\right]\right. \\ \quad \left. -\lambda\theta_1\Gamma\left(1+\dfrac{1}{s}\right)\right\}, & u > \rho, \end{cases} \tag{19c}$$

where $B(\cdot,\cdot)$ and $B(\cdot;\cdot,\cdot)$ denote, respectively, the complete and incomplete Euler beta functions. Equations (19) determine the path of the net wealth Lorenz curve $L(u)$ over the closed interval $[0,1]$ for the different specifications of the net wealth finite mixture model. It follows that, for $u = 1$, $L(1) = 1$.

Since the net wealth Lorenz curve presents negative values for all $u < \rho$, it can be proved that the Gini inequality ratio takes the form [19, 20]

$$G = \left\{2\int_0^1 [u - L(u)]\,\mathrm{d}u\right\} \Big/ [1 + \rho\,|L(\theta_1)|] = \left[1 - 2\int_0^1 L(u)\,\mathrm{d}u\right] \Big/ [1 - \rho L(\theta_1)], \tag{20}$$

where

$$\int_0^1 L(u)\,\mathrm{d}u = \int_0^{\theta_1} L(u)\,\mathrm{d}u + \int_{\theta_1}^{\rho} L(u)\,\mathrm{d}u + \int_{\rho}^{1} L(u)\,\mathrm{d}u. \tag{21}$$

Using equations (19), the Gini ratio becomes

$$G^{\text{SM}} = \frac{\mu_1 - 2[(1-\rho)^2 bqB(2q-1/a, 1+1/a) - \lambda\theta_1(1-\theta_1 2^{-1-1/s})\Gamma(1+1/s)]}{\mu_1 + \rho\lambda\theta_1\Gamma(1+1/s)}; \tag{22a}$$

$$G^{\text{D}} = \frac{\mu_1 - 2\{(1-\rho)^2 bp[B(p+\frac{1}{a}, 1-\frac{1}{a}) - B(2p+\frac{1}{a}, 1-\frac{1}{a})] - \lambda\theta_1(1-\theta_1 2^{-1-\frac{1}{s}})\Gamma(1+\frac{1}{s})\}}{\mu_1 + \rho\lambda\theta_1\Gamma(1+\frac{1}{s})}; \tag{22b}$$

$$G^{\kappa-\text{gen}} = \frac{\mu_1 - 2[\frac{(1-\rho)^2\beta}{(2\kappa)^{1+1/\alpha}}B(1/\kappa - 1/2\alpha, 1+1/\alpha) - \lambda\theta_1(1-\theta_1 2^{-1-1/s})\Gamma(1+1/s)]}{\mu_1 + \rho\lambda\theta_1\Gamma(1+1/s)}. \tag{22c}$$

6. Application

6.1. The US data on household net wealth

The empirical analysis is based on data drawn from the Panel Study of Income Dynamics (PSID), a nationally representative household survey collected by the Survey Research Center at the University of Michigan since 1968. The PSID provides detailed information about economic, demographic, sociological and psychological aspects of many US households. Since the focus is on the distribution of wealth, we use all (nine) waves currently available of the special PSID supplement asking information on household wealth holdings. This supplement was added in 1984 and was conducted on a periodic basis prior

J. Stat. Mech. (2012) P12006

to 1999 (in 1984, 1989 and 1994). After 1997 the basic PSID survey switched to biennial data collection, and starting with 1999 the wealth questions have been included in each wave (1999, 2001, 2003, 2005, 2007 and 2009).

As shown in table 1, the number of households participating in the various waves varies between 6 000 and 9 000, providing samples for analysis that are reasonably representative of the 'true' wealth distribution in the US[8]. In particular, we are concerned with the distribution of *net* wealth, which is constructed as the sum of values of several asset types net of debt held by each household[9]. Since net wealth is expressed in nominal local currency units, all figures have been deflated to allow for meaningful comparisons over the period covered by the data. To do so, we have employed the Consumer Price Index deflator (yearly series based on year 2005) provided by the OECD[10]. Furthermore, after a simple adjustment for differences in relative needs of households according to their size[11], net wealth values have been weighted by using appropriate sampling weights provided by the PSID staff in order to produce representative estimates for all households in the target population.

Table 1 also provides a number of summary statistics. Consider first the prevalence of zero and negative values. On the basis of the PSID data, the proportion of households with negative net wealth rose steadily between 1984 and 2009 (from less than 7% to over 14%) whilst the proportion of households with zero net wealth increased somewhat between 1984 and 1994 (from slightly more than 4% to about 5%), followed by a decline toward 3% until 2001. By 2003, the percentage of these households started increasing again to almost 4% and stayed nearly the same in the following two waves (2005 and 2007) before reaching, in 2009, about the same level of as in 1984. Notwithstanding these differences in the proportions of negatives and zeros with regard to time trends and levels, when their joint prevalence is taken into account we find it to be relatively high on average (around

[8] For more on this issue, see for instance [40] and [41]. In particular, measured against the standards set by two prominent American household wealth surveys—the Survey of Consumer Finances (SCF) and the Survey of Income and Program Participation (SIPP)—the PSID does not differ substantially from them when it comes to measuring total wealth and its distribution among the great bulk of the US population. Moreover, its measurement error characteristics look to be consistently better than are those of the SCF and the SIPP: the PSID has indeed a lower item nonresponse rate than these alternative data sets and thus less need to construct imputed values [42].

[9] The PSID asks about eight broad wealth categories: (1) value of farm or business assets; (2) value of checking and savings accounts, money market funds, certificates of deposit, savings bonds, Treasury bills and other individual retirement accounts (IRAs); (3) value of real estate other than main home; (4) value of shares of stock in publicly held corporations, mutual funds or investment trusts, including stocks in IRAs; (5) value of vehicles or other assets 'on wheels'; (6) value of other investments in trusts or estates, bond funds, life insurance policies, special collections; (7) value of private annuities or IRAs; (8) value of home equity (calculated as home value minus remaining mortgage). More complete definitions of the asset and debt categories are available at the PSID web site: http://psidonline.isr.umich.edu/.

[10] Available at http://stats.oecd.org/.

[11] When the distribution of wealth is defined over households and not over individuals, a problem arises with regard to the possibility of comparing wealth holdings of different units. The reason is that households vary in size and thus wealth levels are not a good indicator of their well-being, as households with different numbers of members may have different needs in the use of wealth even when this is of the same order of magnitude across them. In this case, a correction should be made to meaningfully compare different situations. This correction is called an *equivalence scale*. There is a wide range of equivalence scales in use in different countries and by different organizations. All take account of household size: in many scales this is the only factor, whilst in those taking into account other considerations it is the factor with greatest weight. Choices of equivalence scale in recent wealth studies are reviewed in [43]. Here we adopt a simple equivalence scale that is most commonly used in international studies [44] where net household wealth is divided by the square root of the number of household members.

J. Stat. Mech. (2012) P12006

Table 1. Summary statistics for US household net wealth, 1984–2009.

Stats	Wave								
	1984	1989	1994	1999	2001	2003	2005	2007	2009
Obs.	6 918	7 113	7 415	6 851	7 195	7 565	8 002	8 289	8 690
Mean	121 613	135 095	135 885	185 055	189 139	201 991	223 506	256 281	248 753
Median	36 940	38 988	42 390	45 958	50 735	50 295	53 276	56 744	39 143
Skewness	18.340	15.592	13.821	18.234	19.673	15.079	17.883	26.006	31.766
Kurtosis	410.821	364.775	302.102	454.349	598.511	317.070	513.130	909.552	1 193.073
Gini	0.758	0.759	0.751	0.789	0.774	0.788	0.782	0.803	0.850
% with $W < 0$	6.807	8.096	8.636	9.363	9.409	9.439	10.240	11.137	14.385
% with $W = 0$	4.285	4.554	4.751	3.557	3.428	3.872	3.715	3.725	4.484
% with $W > 0$	88.908	87.350	86.614	87.080	87.163	86.689	86.045	85.138	81.130

14% of the sample size). This situation is quite different from that generally faced in the case of income data, where it is often assumed that income can only take on positive values—in practice, there may be non-positive incomes but usually the number of these is so small that one can just ignore them. By contrast, in the case of net wealth data the assumption of dealing with a positive quantity cannot be justified, since it is a matter of fact that many people enter a period of indebtedness at some point in their life. Therefore, net wealth may legitimately take on negative and zero values, and the proportion of such observations could be non-negligible (as in our case) in representative samples of the target population[12].

Results on the time trend in real mean household net wealth show that it rose continuously by some 111% from 1984 to 2007 and then fell by almost 3% between 2007 and 2009, for an overall annual growth rate of about 3% over the entire period. The time trend (although not the magnitude of level changes) in median net wealth appears to mirror that of the mean. Indeed, the PSID data show median net wealth rising in real terms by some 54% from 1984 to 2007—save for a temporary slight decrease by less than 1% between 2001 and 2003—and then quickly reaching the same level as in 1989 by a sharp fall-off of around 31% between 2007 and 2009, for an overall annual growth rate of about 0.2% over the 25 years.

The change over time in the relationship between the mean and median is shown in figure 1. To provide an indication of how the distribution of wealth across households has changed, the evolution of the relative positions of households at the two ends of the distribution (i.e. the bottom and top quintile groups or bottom and top 20%) is also displayed[13]. As noted above, both mean and median net wealth increased from 1984 to 2007, with the mean typically increasing to a greater extent than the median. This suggests that in recent decades wealth became more concentrated among households at the upper end of the distribution, and indeed in those years where the divergence between the mean

[12] For further discussion on this issue, we refer the reader to [21] and [45].

[13] Changes in the aggregates of figure 1 over the 25 year span are measured by *index numbers*. An index number is calculated by dividing the value in the year of interest by the value in the base year—1984 in our case—and then multiplying the result by 100. The base year index is always 100 and the index for each subsequent year will be above or below 100, depending on whether there has been an increase or decrease in the data compared with the base year.

J. Stat. Mech. (2012) P12006

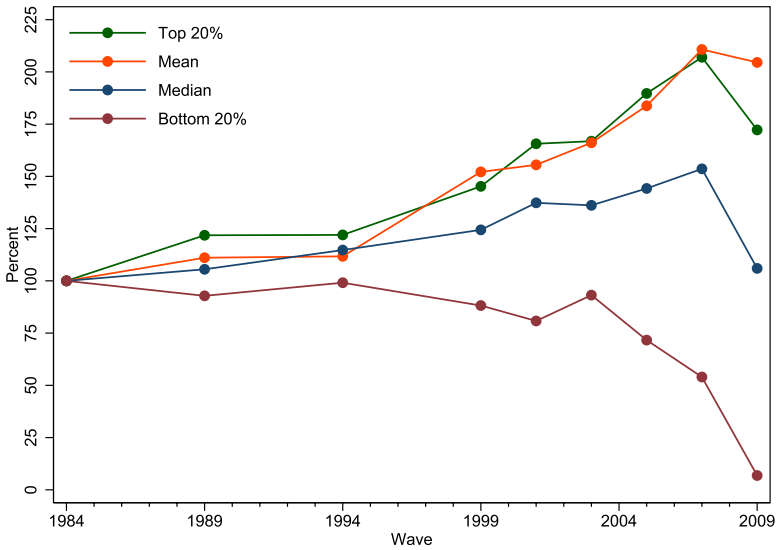

Figure 1. Statistics of US household net wealth distribution, 1984–2009.

and the median became wider—i.e. between 1994 and 2007—the largest changes in net wealth holdings of households in the top of the real distribution were also observed. By contrast, both measures fell during the 2007–09 recession. The relatively greater decline in the median than in the mean suggests that the recession more adversely affected the households at the bottom of the wealth distribution than those further up, as shown by the worsening relative position for the bottom 20% of them.

One might suspect that the differences in the pace of real growth between the mean and median net wealth are partly caused by the presence of long and heavy tails in the distribution of US household net wealth, particularly at the top of the data range. Indeed, the positive skewness values listed in the fourth row of table 1 suggest that the distribution of net wealth in any one year has a long tail toward the upper end, thus indicating a non-trivial prevalence of values that are 'extremes' in relation to the rest of the data. Furthermore, in each of the wave years the level of kurtosis is huge as compared to the normal distribution (fifth row of table 1), meaning that the upper tail of net wealth distribution is inevitably 'fat'—i.e. declines to zero more slowly than exponentially. As the median would not be affected by the extreme values, this results in average net wealth holdings that are consistently larger than median ones in all cases.

Additional information about the fatness of the upper tail of the US net wealth distribution can be obtained from visual examination of the sample mean excess plot shown in figure 2[14]. For a sequence of threshold values $\{w_i\}_{i=1,\ldots,N}$, the mean excess plot reports the mean of exceedances over w_i against w_i itself. Putting it differently, this is a plot of the set of pairs $(w_i, e_n(w_i))_{i=1,\ldots,N-1}$, where $e_n(w_i) = (1/\sum_{j=i+1}^{N}\pi_j)\sum_{j=i+1}^{N}\pi_j(w_j - w_i)$ is the sample *mean excess function* (weighted by household weights $\{\pi_i\}_{i=1,\ldots,N}$) and $\{w_i\}_{i=1,\ldots,N}$ are the sample observations ranked from least to greatest. If the points in the plot show an upward trend, then this is a sign of heavy-tailed behavior. Exponentially

[14] Properties of the mean excess plot are reviewed, for instance, in [46]. We do not report plots for each year but they are available upon request. Since we are interested here in the upper tail behavior of the distribution, the plot has been drawn only for the positive values of net wealth.

J. Stat. Mech. (2012) P12006

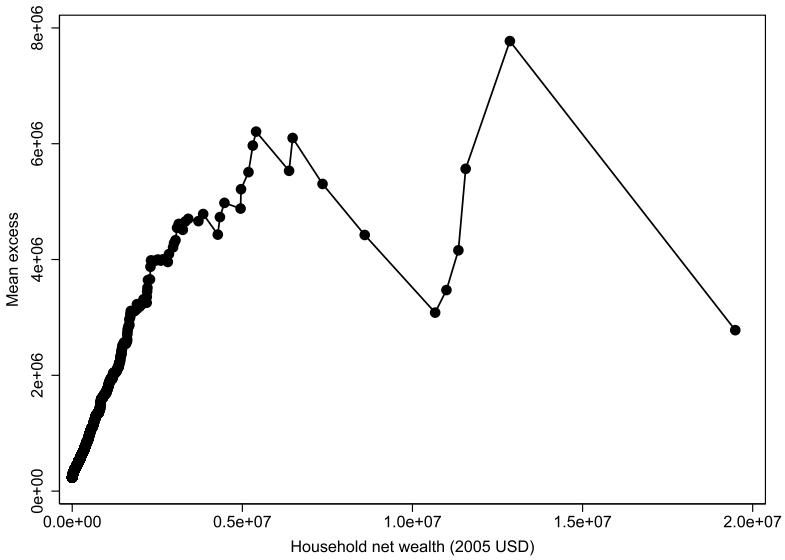

Figure 2. Mean excess plot for the positive values of US household net wealth in 2003.

distributed data would give an approximately horizontal line and data from a short-tailed distribution would show a downward trend. In particular, if the empirical mean excess plot seems to follow a reasonably straight line with positive slope above a certain net wealth value, then this is an indication of Pareto (power-law) behavior in the tail. This is precisely the kind of behavior we observe in the 2003 PSID data. In fact, apart from some noisiness by the most extreme observations, there is evidence for consistent upward trends of the data and straightening out of the plots above some points onwards, hence providing a statistical justification for the emergence of power laws as limiting behavior for the very wealthy.

Does this finding matter when it comes to inequality judgments? Figure 3 displays the pattern of Gini coefficient for the distribution of US household net wealth over the period 1984–2009 (the corresponding values are reported in the fourth-last row of table 1). At least three different sub-periods are shown: from the second half of the 1980s to the first half of the 1990s, from the late 1990s to the first half of the 2000s, and the last time interval (2007–09). According to the PSID, net wealth inequality remained virtually unchanged during the first sub-period. Indeed, the Gini coefficient rose slightly between 1984 and 1989 (from 0.758 to 0.759) and then fell in 1994 to a level below that of 1984 (0.751). By contrast, inequality increased sharply between 1994 and 1999, with the Gini coefficient of net wealth climbing to 0.789. The following years still show almost the same degree of inequality: the Gini coefficient was estimated at 0.788 in 2003 and 0.782 in 2005, except for a temporary decrease to a value of 0.774 in 2001. Finally, between 2007 and 2009 net wealth inequality was up steeply, with the Gini coefficient advancing from 0.803 to 0.850.

Figure 3 also displays the evolution between 1984 and 2009 of the share of total net wealth held by the richest 20% of households, which amounted on average to around 80% of the whole over the period. A noteworthy result is that the observed time pattern of inequality seems to have been driven by the conspicuous wealth holdings at the very top

J. Stat. Mech. (2012) P12006

Figure 3. Gini coefficient and net wealth share of top 20% across years.

end of the distribution. Indeed, as can be seen from the figure, the time profile of net wealth share of the wealthiest 20% is analogous to that of the Gini coefficient: after rising to a peak in 1999, it went down and then started to increase again until 2009[15].

To sum up the above, wealth in the US has become more concentrated in recent decades. Net wealth inequality increased by the mid-1990s, and the increase was not interrupted during the 2007–09 recession. The share of total net wealth held by the top wealth owners has also grown during the same period, whereas at the other end of the wealth distribution there was a sharp increase in the number of households with zero or negative net wealth. Needless to say, this has resulted in a widening gap between the rich and the poor that advocates more attention be paid to the implementation of appropriate and practical policies aimed at reducing inequalities, limiting their negative effects on the socio-economic system and reversing the mechanisms producing them [47].

6.2. Estimation and comparison of finite mixture models for net wealth distribution

Table 2 presents the parameter estimates and other relevant statistics arising from the fitting of the net wealth distribution models previously discussed to the PSID data from 1984 to 2009. The parameters were estimated in all cases by minimizing the negative of the log-likelihood function via a modified Newton–Raphson procedure implemented in Stata's `ml` command [48], with the parameter covariance matrix estimates based on the negative inverse Hessian. Convergence was achieved easily within a few iterations.

The small value of the errors indicates that all the parameters were very precisely estimated. The mixture proportions (the θ) correspond exactly to the sample estimates shown in table 1, and the scale parameters (the b, β and λ) reflect the changes over the period in both the median and the mean among the positive and negative values of real

[15] The correlation coefficient between the two series of Gini coefficient and the net wealth share received by the top 20% is 0.998, which is highly significant (p-value < 0.001).

J. Stat. Mech. (2012) P12006

Table 2. Estimated mixture models for the US household net wealth, 1984–2009. (SM, Singh–Maddala mixture model; D, Dagum mixture model; κ-gen, κ-generalized mixture model.)

		Parameters[a]												
Wave	Model	a (α)	b (β)	q, p, κ	γ[b]	θ_1	θ_2	s	λ	logLik	AIC	BIC	Mean[c]	Gini[d]
1984	SM	0.757	373 565	3.754	2.843	0.068	0.043	0.578	4511	−84 249	168 511	168 579	111 616	0.736
		(0.012)	(56 465)	(0.316)		(0.003)	(0.002)	(0.018)	(382)					
	D	1.614	138 163	0.377	0.608	0.068	0.043	0.578	4511	−84 230	168 474	168 542	121 361	0.753
		(0.042)	(5 720)	(0.015)		(0.003)	(0.002)	(0.018)	(382)					
	κ-gen	0.718	76 514	0.374	1.919	0.068	0.043	0.578	4511	−84 229	168 471	168 539	114 181	0.741
		(0.009)	(1 663)	(0.022)		(0.003)	(0.002)	(0.018)	(382)					
1989	SM	0.743	459 094	3.814	2.833	0.081	0.046	0.619	6639	−87 573	175 161	175 230	130 298	0.749
		(0.012)	(76 568)	(0.348)		(0.003)	(0.002)	(0.018)	(473)					
	D	1.520	152 781	0.402	0.611	0.081	0.046	0.619	6639	−87 583	175 180	175 249	152 050	0.781
		(0.040)	(7 050)	(0.017)		(0.003)	(0.002)	(0.018)	(473)					
	κ-gen	0.702	89 241	0.367	1.912	0.081	0.046	0.619	6639	−87 565	175 143	175 212	133 595	0.754
		(0.009)	(1 998)	(0.023)		(0.003)	(0.002)	(0.018)	(473)					
1994	SM	0.769	449 157	3.713	2.857	0.086	0.048	0.716	10 759	−91 866	183 745	183 816	133 116	0.745
		(0.012)	(72 386)	(0.338)		(0.003)	(0.002)	(0.020)	(629)					
	D	1.528	154 125	0.421	0.644	0.086	0.048	0.716	10 759	−91 879	183 771	183 842	156 209	0.779
		(0.038)	(6 831)	(0.017)		(0.003)	(0.002)	(0.020)	(629)					
	κ-gen	0.727	95 745	0.379	1.919	0.086	0.048	0.716	10 759	−91 861	183 736	183 807	137 029	0.751
		(0.010)	(2 083)	(0.024)		(0.003)	(0.002)	(0.020)	(629)					
1999	SM	0.724	477 324	3.380	2.446	0.094	0.036	0.751	11 529	−86 534	173 082	173 152	173 472	0.775
		(0.012)	(77 394)	(0.288)		(0.004)	(0.002)	(0.022)	(642)					
	D	1.422	181 486	0.420	0.597	0.094	0.036	0.751	11 529	−86 548	173 111	173 181	212 146	0.812
		(0.038)	(9 213)	(0.018)		(0.004)	(0.002)	(0.022)	(642)					
	κ-gen	0.678	107 494	0.389	1.742	0.094	0.036	0.751	11 529	−86 527	173 067	173 137	179 416	0.781
		(0.009)	(2 579)	(0.024)		(0.004)	(0.002)	(0.022)	(642)					

doi:10.1088/1742-5468/2012/12/P12006

Table 2. (Continued.)

Wave	Model	Parameters[a] a (α)	b (β)	q, p, κ	γ[b]	θ_1	θ_2	s	λ	logLik	AIC	BIC	Mean[c]	Gini[d]
2001	SM	0.683	980 104	4.669	3.188	0.094	0.034	0.724	11 083	−91 364	182 742	182 812	181 487	0.765
		(0.010)	(197 710)	(0.486)		(0.003)	(0.002)	(0.020)	(623)					
	D	1.514	229 564	0.366	0.554	0.094	0.034	0.724	11 083	−91 373	182 759	182 829	211 855	0.794
		(0.041)	(10 520)	(0.015)		(0.003)	(0.002)	(0.020)	(623)					
	κ-gen	0.652	118 900	0.326	2.004	0.094	0.034	0.724	11 083	−91 354	182 722	182 792	185 349	0.769
		(0.008)	(2 778)	(0.023)		(0.003)	(0.002)	(0.020)	(623)					
2003	SM	0.703	667 612	3.767	2.649	0.094	0.039	0.682	13 602	−96 151	192 315	192 385	192 121	0.777
		(0.011)	(112 537)	(0.329)		(0.003)	(0.002)	(0.019)	(791)					
	D	1.443	214 742	0.400	0.577	0.094	0.039	0.682	13 602	−96 158	192 329	192 399	232 063	0.812
		(0.037)	(10 101)	(0.016)		(0.003)	(0.002)	(0.019)	(791)					
	κ-gen	0.665	120 624	0.370	1.796	0.094	0.039	0.682	13 602	−96 140	192 294	192 364	198 801	0.784
		(0.008)	(2 775)	(0.022)		(0.003)	(0.002)	(0.019)	(791)					
2005	SM	0.660	1 371 330	4.978	3.285	0.102	0.037	0.728	11 804	−102 470	204 954	205 024	216 342	0.775
		(0.010)	(290 880)	(0.531)		(0.003)	(0.002)	(0.019)	(600)					
	D	1.456	267 209	0.371	0.541	0.102	0.037	0.728	11 804	−102 482	204 978	205 048	264 727	0.813
		(0.037)	(12 103)	(0.015)		(0.003)	(0.002)	(0.019)	(600)					
	κ-gen	0.632	138 871	0.315	2.008	0.102	0.037	0.728	11 804	−102 462	204 937	205 008	221 517	0.780
		(0.008)	(3206)	(0.022)		(0.003)	(0.002)	(0.019)	(600)					
2007	SM	0.645	1 568 538	4.995	3.221	0.111	0.037	0.670	13 715	−106 821	213 657	213 728	239 641	0.787
		(0.009)	(324 452)	(0.508)		(0.003)	(0.002)	(0.015)	(713)					
	D	1.450	303 012	0.360	0.523	0.111	0.037	0.670	13 715	−106 833	213 679	213 750	291 719	0.821
		(0.037)	(13 682)	(0.014)		(0.003)	(0.002)	(0.015)	(713)					
	κ-gen	0.617	150 343	0.307	2.007	0.111	0.037	0.670	13 715	−106 810	213 633	213 704	244 070	0.791
		(0.007)	(3469)	(0.021)		(0.003)	(0.002)	(0.015)	(713)					

doi:10.1088/1742-5468/2012/12/P12006

Table 2. (Continued.)

Wave	Model	Parameters[a]								logLik	AIC	BIC	Mean[c]	Gini[d]
		a (α)	b (β)	q, p, κ	γ[b]	θ_1	θ_2	s	λ					
2009	SM	0.640	930 909	3.988	2.550	0.144	0.045	0.707	17 847	−110 594	221 202	221 273	221 273	0.828
		(0.009)	(171 313)	(0.348)		(0.004)	(0.002)	(0.014)	(756)					
	D	1.334	245 091	0.393	0.524	0.144	0.045	0.707	17 847	−110 605	221 225	221 295	295 004	0.869
		(0.033)	(12 039)	(0.015)		(0.004)	(0.002)	(0.014)	(756)					
	κ-gen	0.605	128 792	0.353	1.713	0.144	0.045	0.707	17 847	−110 583	221 181	221 251	230 967	0.835
		(0.007)	(3120)	(0.022)		(0.004)	(0.002)	(0.014)	(756)					

[a] Numbers in round brackets: estimated standard errors.
[b] Values of the Paretian upper tail index derived from parameter estimates of the Singh–Maddala (SM), Dagum (D) and κ-generalized (κ-gen) mixture models, respectively, as $\gamma^{\mathrm{SM}} = aq$, $\gamma^{\mathrm{D}} = ap$ and $\gamma^{\kappa-\mathrm{gen}} = \alpha/\kappa$.
[c] Analytic values obtained by substituting the estimated parameters into equations (16) and (17) with $r = 1$.
[d] Analytic values obtained by substituting the estimated parameters into equations (22).

doi:10.1088/1742-5468/2012/12/P12006

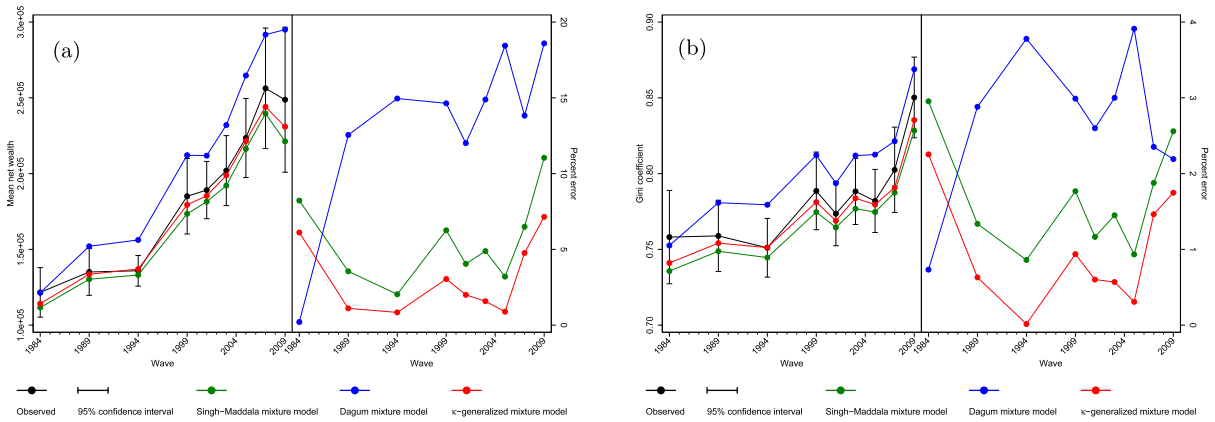

Figure 4. Observed and predicted values for the mean and Gini coefficient of US household net wealth, 1984–2009. The vertical bars denote the symmetric 95% normal-approximation confidence intervals for the empirical values calculated via the bootstrap resampling method based on 100 replications. Percent error is calculated as follows: percent error = $(|\text{predicted} - \text{observed}|/\text{observed}) \times 100$. (a) Mean and (b) Gini.

net wealth[16]. The other parameters (the a, α, p, κ, q and s), characterizing distributional shape, are easiest to interpret by comparing predicted values for key distributional summary measures with their sample counterparts, as the effect of changing one of them is contingent on the value of the other parameters. For example, figure 4 shows that the overall mean net wealth and Gini coefficient as estimated from the mixture models are very close to their sample estimates[17]. However, the agreement (in both magnitude and temporal behavior) between the implied and sample estimates of the mean and Gini coefficient is much closer for the Singh–Maddala and κ-generalized mixture models than for the Dagum one. The mean and Gini coefficient associated with the latter model are in fact above the 95% upper confidence limit of their corresponding sample estimates in six (from 1989 to 2005) and three (1994, 2003 and 2005) cases out of nine, respectively, and their percent error turns out to be relatively large compared to the other models—save for 1994, where both the mean and Gini predictions exhibited the lowest error, and 2009 with respect to the Gini coefficient implied by the Singh–Maddala mixture model. Overall, judging by the percent error of the mean and Gini coefficient of the net wealth distributions estimated from the three mixture models, the performance of the κ-generalized mixture model is appreciably superior to the other ones over most of the time span investigated.

The results summarized in table 2 allow us to emphasize a distinctive feature of wealth distributions, i.e. the concentration of density mass at zero. There is often a marked spike

[16] The correlation coefficients between the Weibull scale parameters (λ) and the two series of the median and mean net wealth values among the negatives are close to unity (0.982 and 0.955, respectively) and highly significant (p-value < 0.001 in both cases). Similarly, the correlation coefficients between the values of the scale parameter of the Singh–Maddala (b), Dagum type I (b) and κ-generalized (β) distributions and the two series of the median and mean net wealth levels among the positives are all significant at the 1% confidence level and equal, respectively, to 0.931, 0.983 and 0.998 for the median and 0.812, 0.925 and 0.935 for the mean.

[17] The analytic values for the mean and Gini coefficients, also reported in the last two columns of table 2, were obtained by substituting the estimated parameters into the relevant expressions given by equations (16) and (17) with $r = 1$ for the mean and equations (22) for the Gini.

J. Stat. Mech. (2012) P12006

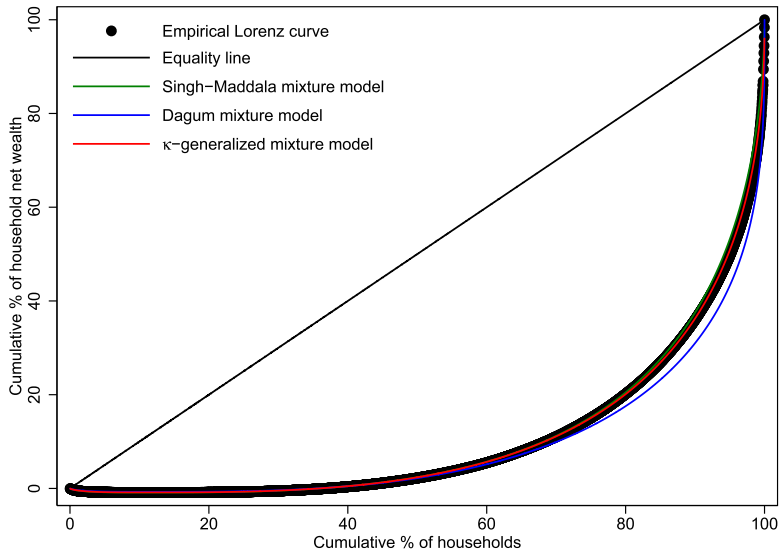

Figure 5. Observed and calculated Lorenz curves for the US household net wealth in 2003.

at zero because a relatively large fraction of the population has no wealth. Similar spikes do not occur with income distributions, where it is often the case that the density mass vanishes when income goes to zero. As the Weibull, Singh–Maddala, Dagum type I and κ-generalized distributions are zero modal with a pole at the origin if, respectively, $s < 1$, $a \leq 1$, $ap \leq 1$ and $\alpha \leq 1$, it is easily verified from the estimates of these parameters that the probability density functions of the three mixture models inevitably transfer some density mass from the neighboring values to the cited spike at zero—i.e. they diverge, rather than vanish, when the argument goes to zero from both the negative and positive ends of the wealth range[18]. This finding of a divergent probability density in the limit of zero wealth is also shared by other studies on the distribution of wealth (e.g. [49]).

The parameter estimates reported in table 2 were also used to build estimated Lorenz curves by applying equations (19). The curves for 2003 are presented in figure 5 together with the empirical Lorenz curve estimate. Even if it is small, one can see a difference between the three predictions, in that the Lorenz curve estimated from the Dagum mixture model lies below the empirical one for approximately the top 30% of the wealthiest households, while the Singh–Maddala and κ-generalized mixture models lead to estimated Lorenz curves exhibiting a degree of inequality that is much more in line with the observed one. In particular, the mean absolute difference between the empirical Lorenz data and the predicted values (averaged from all the survey years) amounts to 0.004, 0.007 and 0.002, respectively, for the Dagum, Singh–Maddala and κ-generalized mixture models, thus indicating once again that the latter model gives a better match to the observed data than the other two.

It is interesting to note that the κ-generalized mixture model provides a better fit to most of the data than any of the alternative models regardless of the criterion used for comparison. For instance, by inspection of AIC and BIC values reported in the fourth-

[18] The behavior around the mode of Weibull, Singh–Maddala and Dagum type I distributions is reviewed, e.g., in [8]. For the κ-generalized distribution see [1], [24]–[26].

J. Stat. Mech. (2012) P12006

Table 3. Vuong test for model selection, 1984–2009. (SM, Singh–Maddala mixture model; D, Dagum mixture model; κ-gen, κ-generalized mixture model. The null hypothesis is that the competing models are equally close to the 'true' data generating process.)

Wave	SM versus κ-gen		D versus κ-gen	
	Statistic	p-value	Statistic	p-value
1984	−3.917	9×10^{-5}[a]	−0.159	0.874
1989	−2.085	0.037[a]	−2.771	0.006[a]
1994	−1.176	0.240	−2.966	0.003[a]
1999	−1.741	0.082	−3.674	2×10^{-4}[a]
2001	−2.665	0.008[a]	−2.481	0.013[a]
2003	−2.486	0.013[a]	−2.529	0.011[a]
2005	−2.121	0.034[a]	−2.336	0.019[a]
2007	−2.434	0.015[a]	−2.462	0.014[a]
2009	−1.964	0.050[a]	−2.675	0.007[a]

[a] Significant at the 5% level.

and third-last columns of table 2, it emerges that the two selection criteria agree on the κ-generalized mixture model as the preferred one for all of the survey waves[19]. To see if these differences in the performance of the alternative specifications are statistically significant, we adopt the Vuong approach to model selection [52]. This approach sets the model selection criterion in a hypothesis testing framework. More specifically, it tests the null hypothesis that the models under consideration are equidistant from a unknown 'true' model against the alternative hypothesis that one model is closer. The test statistic is asymptotically normal under the null hypothesis and is quite straightforward to compute. Table 3 shows the results of the comparisons for the three mixture models. As can be seen, if one takes 5% as the relevant significance level only in three cases (i.e. when comparing to the Singh–Maddala mixture model in the survey years 1994 and 1999 and to the Dagum one in 1984), the test concludes that the κ-generalized mixture model is observationally equivalent to its competitors, while in all the other cases (more than 83% of all cases) its superiority as a descriptive model is found to be statistically significant.

The above evidence holds vis-à-vis a further check involving goodness-of-fit indicators such as the root mean squared error, defined as the square root of the average squared error between the observed and predicted values of the cumulative distribution function. In mathematical terms this is expressed as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[F_*(w_i) - \hat{F}_N(w_i)\right]^2}, \tag{23}$$

where $F_*(w)$ is the distribution function deduced from the fitted mixture models and $\hat{F}_N(w) = \sum_{i=1}^{N}\pi_i \mathbf{1}_A(w) / \sum_{i=1}^{N}\pi_i$ denotes the empirical distribution function of the N

[19] Model selection criteria such as the Akaike [50] and Bayesian [51] information criteria (AIC and BIC) will select, when comparing models with the same number of parameters, the model with the smallest log-likelihood value according to the formula $(2 \times \mathrm{logLik}) + (d \times npar)$, where $npar$ represents the number of parameters in the fitted model, and $d = 2$ for the usual AIC or $d = \ln N$ (N being the number of observations) for the so-called BIC.

J. Stat. Mech. (2012) P12006

Table 4. Goodness-of-fit comparisons for estimated mixture models of US household net wealth, 1984–2009. (Source: authors' own calculations using the PSID supplemental wealth files.)

Wave	Model	RMSE ($\times 10^2$)	Rank	A^2 ($\times 10^2$)	p-value[a]	Rank
1984	SM	1.192	3	0.088	0.594	3
	D	0.986	1	0.054	0.743	1
	κ-gen	1.022	2	0.063	0.673	2
1989	SM	0.981	2	0.057	0.644	2
	D	1.118	3	0.064	0.594	3
	κ-gen	0.911	1	0.045	0.693	1
1994	SM	0.997	2	0.049	0.703	2
	D	1.080	3	0.056	0.693	3
	κ-gen	0.936	1	0.042	0.812	1
1999	SM	0.924	2	0.049	0.713	2
	D	1.058	3	0.055	0.614	3
	κ-gen	0.812	1	0.036	0.782	1
2001	SM	0.916	2	0.053	0.663	3
	D	1.008	3	0.050	0.733	2
	κ-gen	0.798	1	0.038	0.782	1
2003	SM	0.823	2	0.047	0.703	2
	D	0.947	3	0.050	0.673	3
	κ-gen	0.716	1	0.035	0.812	1
2005	SM	0.740	2	0.035	0.703	2
	D	0.882	3	0.041	0.653	3
	κ-gen	0.639	1	0.026	0.802	1
2007	SM	0.871	2	0.046	0.624	2
	D	0.923	3	0.047	0.713	3
	κ-gen	0.747	1	0.034	0.822	1
2009	SM	0.907	2	0.046	0.634	2
	D	0.992	3	0.048	0.663	3
	κ-gen	0.822	1	0.035	0.792	1

[a] Upper tail p-value obtained by 100 bootstrap replications. The null hypothesis is that data come from the fitted Singh–Maddala (SM), Dagum (D) or κ-generalized (κ-gen) mixture model.

sample data ordered from lowest to highest carrying the π_i along ($\mathbf{1}_A$ is the indicator function of the set $A = \{w|w_i \leq w\}$ and π_i refers to the sampling weight of the ith observation). Clearly, lower values of RMSE indicate a better fit. The comparison results between the competing models based on the above criterion are shown in table 4. As can be seen, the κ-generalized mixture model of net wealth ranks first for all years but 1984, where it is outperformed by the Dagum mixture model.

Similar results are obtained by additionally performing an Anderson–Darling goodness-of-fit test that data come from the fitted Singh–Maddala, Dagum or κ-generalized mixture model. This test is known to be more powerful than other tests based on the empirical distribution function, since it provides equal sensitivity at the tails and

J. Stat. Mech. (2012) P12006

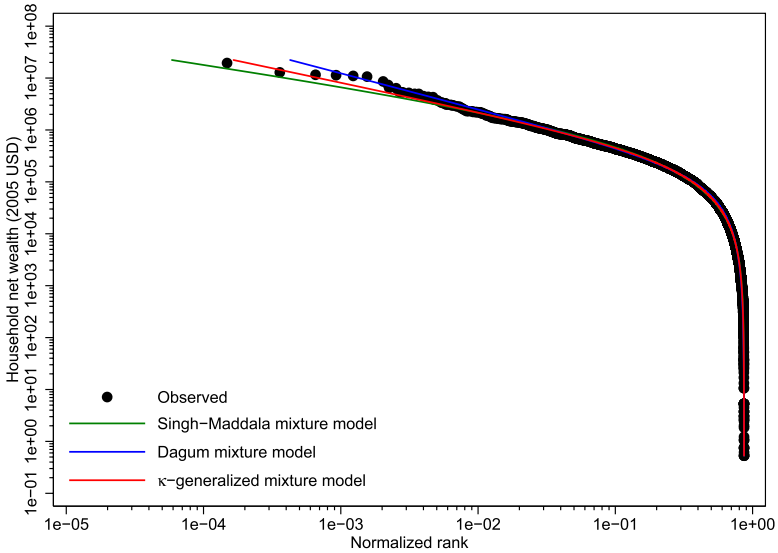

Figure 6. Zipf plot for the positive values of US household net wealth in 2003.

at the median of the distribution [53][20]. The last three columns of table 4 report the test results for the nine 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 models. However, except for 1984, fitting the κ-generalized mixture model results in both lower values of the test statistic and higher p-values, thus offering superior performance over the Singh–Maddala and Dagum mixture models.

Can these findings be ultimately ascribed to the different performances of the alternative densities used to characterize positive net wealth values? Figure 6 presents for the 2003 PSID wave the relationship between log rank and log size along the positive support of the net wealth distribution. This double-logarithmic framework, known as the Zipf plot, is 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[21]. The lines show 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 positive support of net wealth distribution. However, at the top tail there is a systematic departure of empirical observations from the theoretical predictions of the mixtures using the Singh–Maddala and Dagum type I specifications as descriptions of the positive net

[20] The formula used for the test statistic is the one reported by [54], which allows for weighted observations. Since the distribution of the Anderson–Darling test statistic is only known for data sets truly drawn from any given distribution [55], while in our case the underlying distribution is itself determined by fitting to the data and hence varies from one data set to the next, the p-values for the test have been derived by making use of a nonparametric bootstrap method [56]. That is, given our N-vector of net wealth data, we generated 100 synthetic data sets by drawing new sequences of N observations uniformly at random from the original data. We then fitted each synthetic data set individually to the three mixture models and calculated the test statistics for each one relative to its own models. Then we simply counted what fraction of the time each resulting statistic was 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.

[21] For an illustration of basic properties of the Zipf plot, see, e.g., [57].

J. Stat. Mech. (2012) P12006

wealth values, while in the same part of the distributions the theoretical Zipf plot for the κ-generalized mixture model lies much closer to the empirical one.

This point is of particular relevance in the current context, for both the documented presence of long and fat tails toward the upper end of the US net wealth distribution and the fact that all of the three densities accounting for the positive range of wealth obey the weak Pareto law [6]. The weak version of the Pareto law states that the right-hand tail of a distribution behaves in the limit as a simple Pareto model, with an exponent that is a function of the parameters governing the shape of the distribution (see e.g. [58] for an overview). The values of the Pareto index derived from parameter estimates of the Singh–Maddala, Dagum and κ-generalized mixture models are given in the sixth column of table 2[22]. Remarkably, according to the κ-generalized mixture model, the set of values for the index of the Pareto tail is closely in the narrow range $(1, 2]$ that is generally found in empirical studies on the US wealth distribution ([62]–[64], for instance), whereas for the other two models the Paretian upper tail index oscillates systematically above (Singh–Maddala) and below (Dagum) the limits of this range.

7. Summary and conclusions

This paper mainly deals with the specification, analysis and application of models for net wealth distribution with support in the interval $(-\infty, \infty)$. These are mixtures—or, equivalently, convex representations—of three distributions with non-overlapping intervals, which have the advantage of providing a relatively flexible functional form and at the same time retain the advantages of parametric forms that are amenable to inference. The first distribution is a two-parameter Weibull model that describes the distribution of economic units with negative net wealth, i.e. covering the open interval $(-\infty, 0)$; the second is a degenerate distribution with its unit mass concentrated at $w = 0$; and the third is, alternatively, the three-parameter Singh–Maddala, Dagum type I or κ-generalized model that accounts for the distribution of economic units with positive net wealth, hence defined in the open interval $(0, \infty)$.

We have obtained closed formulas for the different probability functions, moments and standard tools for inequality measurement (i.e. the Lorenz curve and Gini concentration ratio). Except for the Dagum general model of net wealth [17]–[20], to the best of our knowledge this is the first time that the analytical properties of finite mixture models for net wealth based on alternative distributions to characterize positive values are fully derived.

The performance of the three mixture models has been checked against real data on US household net wealth for different years. Goodness-of-fit comparisons reveal that all three models are in good agreement with actual data, but the departure of empirical observations from the predictions of the Singh–Maddala and Dagum mixture models is always larger than in the case of the κ-generalized model. In particular, the latter model suggests a superior fit in the right tail of data with respect to the others in many instances.

Finite mixture models deserve further attention in future. A feature of these models is that each of the parameters may be made a function of covariates summarizing household

[22] See table footnote (c) for formulas used to analytically derive the Pareto tail exponent in the three cases. For more details on the upper tail behavior of the κ-generalized distribution, we refer the reader to [1, 25, 26]. For the Singh–Maddala and Dagum distributions see instead [59] and [60, 61].

J. Stat. Mech. (2012) P12006

characteristics. Estimation of 'heterogeneous' wealth distributions such as these, with distributional shape allowed to vary with personal characteristics, provides a route to decomposition analysis of the sources of differences in wealth inequality across years or countries[23]. This could be a good starting point for future research.

References

[1] Clementi F, Gallegati M and Kaniadakis G, *A κ-generalized statistical mechanics approach to income analysis*, 2009 *J. Stat. Mech.* P02037

[2] Pareto V, *La legge della domanda*, 1895 *G. Econ.* **10** 59
Pareto V, 1997 *Riv. Polit. Econ.* **87** 691 (Engl. transl.)

[3] Pareto V, 1896 *La courbe de la répartition de la richesse* (Reprinted in: Busoni G 1965 *Œuvres Complètes de Vilfredo Pareto* (Geneva: Librairie Droz) pp 1–15 Tome 3: *Écrits sur la Courbe de la Répartition de la Richesse*)
Pareto V, 1997 *Riv. Polit. Econ.* **87** 647 (Engl. transl.)

[4] Pareto V, 1897a *Cours d'Économie Politique* (London: Macmillan)

[5] Pareto V, *Aggiunta allo studio della curva delle entrate*, 1897b *G. Econ.* 15 (Engl. transl.)
Pareto V, 1997 *Riv. Polit. Econ.* **87** 645

[6] Mandelbrot B, *The Pareto–Lévy law and the distribution of income*, 1960 *Int. Econ. Rev.* **1** 79

[7] Ord J K, *Statistical models for personal income distribution*, 1975 *Statistical Distributions in Scientific Work* ed G P Patil, S Kotz and J K Ord (Dordrecht: Reidel) pp 151–8 Vol 2 *Model Building and Model Selection*

[8] Kleiber C and Kotz S, 2003 *Statistical Size Distributions in Economics and Actuarial Sciences* (New York: Wiley)

[9] Davies J B, Sandström S, Shorrocks A F and Wolff E N, *The global pattern of household wealth*, 2009 *J. Int. Dev.* **21** 1111

[10] Wold H O A and Whittle P, *A model explaining the Pareto distribution of wealth*, 1957 *Econometrica* **25** 591

[11] Sargan J T, *The distribution of wealth*, 1957 *Econometrica* **24** 568

[12] Stiglitz J E, *Distribution of income and wealth among individuals*, 1969 *Econometrica* **37** 382

[13] Atkinson A B, *The distribution of wealth in Britain in the 1960—the estate duty method rexamined*, 1975 *The Personal Distribution of Income and Wealth* ed J D Smith (New York: Columbia University Press) pp 277–319

[14] Vaughan R N, *A Study of the Distribution of Wealth*, 1975 *PhD Thesis* University of Cambridge, Cambridge

[15] Dagum C, *A new model of personal income distribution: specification and estimation*, 1977 *Econ. Appl.* **30** 413

[16] Dagum C, *The Distribution of Income and Wealth in Canada, the USA and Great Britain*, 1978 *Paper presented at the 4th Annual Convention of the Eastern Economic Association (Washington, DC, April)*

[17] Dagum C, *A model of net wealth distribution specified for negative, null and positive wealth. A case study: Italy*, 1990 *Income and Wealth Distribution, Inequality and Poverty* ed C Dagum and M Zenga (Berlin: Springer) pp 42–56

[18] Dagum C, *Human capital, income and wealth distribution models with applications*, 1994 *Proc. Business and Economic Statistics Section, American Statistical Association, 154th Mtg* pp 91–6

[19] Dagum C, *Income and wealth distributions, dagum system of*, 2006a *Encyclopedia of Statistical Sciences* 2nd edn, vol 5, ed N Balakrishnan, C B Read and B Vidakovic (New York: Wiley) pp 3363–78

[20] Dagum C, *Wealth distribution models: analysis and applications*, 2006b *Statistica* **66** 235

[21] Jenkins S P and Jäntti M, 2005 Methods for Summarizing and Comparing Wealth Distributions *ISER Working Paper 2005-05* University of Essex, Colchester, UK, available at: www.iser.essex.ac.uk/publications/working-papers/iser/2005-05.pdf

[22] Singh S K and Maddala G S, *A function for size distribution of incomes*, 1976 *Econometrica* **4** 963

[23] Clementi F, Gallegati M and Kaniadakis G, *κ-generalized statistics in personal income distribution*, 2007 *Eur. Phys. J.* B **57** 187

[24] Clementi F, Di Matteo T, Gallegati M and Kaniadakis G, *The κ-generalized distribution: a new descriptive model for the size distribution of incomes*, 2008 *Physica* A **387** 3201

[23] For an application of parametric models in this manner, but to income, see [65]–[67].

J. Stat. Mech. (2012) P12006

[25] Clementi F, Gallegati M and Kaniadakis G, *A model of personal income distribution with application to italian data*, 2010 *Empir. Econ.* **39** 559
[26] Clementi F, Gallegati M and Kaniadakis G, *A new model of income distribution: the κ-generalized distribution*, 2012 *J. Econ.* **105** 63
[27] Kaniadakis G, *Non-linear kinetics underlying generalized statistics*, 2001a *Physica* A **296** 405
[28] Kaniadakis G, *H-theorem and generalized entropies within the framework of nonlinear kinetics*, 2001b *Phys. Lett.* A **288** 283
[29] Kaniadakis G, *Statistical mechanics in the context of special relativity*, 2002 *Phys. Rev.* E **66** 056125
[30] Kaniadakis G, *Statistical mechanics in the context of special relativity. II*, 2005 *Phys. Rev.* E **72** 036108
[31] Kaniadakis G, *Maximum entropy principle and power-law tailed distributions*, 2009a *Eur. Phys. J.* B **70** 3
[32] Kaniadakis G, *Relativistic entropy and related Boltzmann kinetics*, 2009b *Eur. Phys. J.* A **40** 275
[33] Kaniadakis G and Delsanto P P, *A local interaction simulation approach to the solution of diffusion problems*, 1993 *Math. Comput. Modelling* **17** 31
[34] Kaniadakis G, Lavagno A and Quarati P, *Kinetic approach to fractional exclusion statistics*, 1996 *Nucl. Phys.* B **466** 527
[35] Kaniadakis G, Lavagno A and Quarati P, *Kinetic model for q-deformed bosons and fermions*, 1997 *Phys. Lett.* A **227** 227
[36] Rajaonarison D, Bolduc D and Jayet H, *The K-deformed multinomial logit model*, 2005 *Econ. Lett.* **86** 13
[37] Rajaonarison D, *Deterministic heterogeneity in tastes and product differentiation in the K-logit model*, 2008 *Econ. Lett.* **100** 396
[38] Bertotti M L and Modanese G, *Exploiting the flexibility of a family of models for taxation and redistribution*, 2012 *Eur. Phys. J.* B **85** 1
[39] Lorenz M O, *Methods of measuring the concentration of wealth*, 1905 *Publ. Am. Stat. Assoc.* **9** 209
[40] Hill M S, 1991 *The Panel Study of Income Dynamics: A User's Guide* vol 2 (Thousand Oaks, CA: SAGE Publications)
[41] Wolff E N, *The size distribution of wealth in the United States: a comparison among recent household surveys*, 1999 *Wealth, Work, and Health: Innovations in Measurement in the Social Sciences* ed J P Smith and R J Willis (Ann Arbor, MI: University of Michigan Press) pp 209–32
[42] Curtin R T, Juster T and Morgan J N, *Survey estimates of wealth: an assessment of quality*, 1989 *The Measurement of Saving, Investment, and Wealth* ed R E Lipsey and H Stone Tice (Chicago, IL: University of Chicago Press) pp 473–552
[43] Sierminska E and Smeeding T, *Measurement Issues: Equivalence Scales, Accounting Framework, and Reference Unit*, paper presented at the LWS Workshop: 'Construction and Usage of Comparable Microdata on Wealth: The LWS', Perugia, Italy, available at: www.lisproject.org/lws/introduction/files/SierminskaSmeedingrev.pdf (January 2005)
[44] Atkinson A B, Rainwater L and Smeeding T M, 1995 *Income Distribution in OECD Countries: Evidence from Luxembourg Income Study* (Paris: Organization for Economic Co-operation and Development)
[45] Cowell F A, *Inequality Among the Wealthy*, 2011 CASE Paper 150, London School of Economics, London, available at: http://sticerd.lse.ac.uk/dps/case/cp/CASEpaper150.pdf
[46] Beirlant J, Goegebeur Y, Teugels J and Segers J, 2004 *Statistics of Extremes: Theory and Applications* (Chichester: Wiley)
[47] Stiglitz J E, 2012 *The Price of Inequality: How Today's Divided Society Endangers Our Future* (New York: W W Norton)
[48] *StataCorp, Stata Statistical Software: Release 12, StataCorp LP, College Station, TX*, 2011
[49] Drăgulescu A and Yakovenko V M, *Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States*, 2001 *Physica* A **299** 213
[50] Akaike H, *Information theory and an extension of the likelihood ratio principle*, 1973 *Proc. 2nd Int. Symp. of Information Theory* ed B N Petrov and F Csaki (Budapest: Akademiai Kiado) pp 267–81
[51] Schwarz G E, *Estimating the dimension of a model*, 1978 *Ann. Stat.* **6** 461
[52] Vuong Q H, *Likelihood ratio tests for model selection and non-nested hypotheses*, 1989 *Econometrica* **57** 307
[53] Thode H C, 2002 *Testing for Normality* (New York: Dekker)
[54] Monahan J F, 2011 *Numerical Methods of Statistics* 2nd edn (New York: Cambridge University Press)
[55] Stephens M A, *Tests based on EDF statistics*, 1986 *Goodness-of-Fit Techniques* ed R B D'Agostino and M A Stephens (New York: Dekker) pp 97–193
[56] Efron B and Tibshirani R J, 1993 *An Introduction to the Bootstrap* (New York: Chapman and Hall)
[57] Stanley M H R, Buldyrev S V, Havlin S, Mantegna R N, Salinger M A and Stanley H E, *Zipf plots and the size distribution of firms*, 1995 *Econ. Lett.* **49** 453
[58] Krämer W and Ziebach T, *The weak Pareto law and regular variation in the tails*, 2004 *Statistica* **64** 505

[59] Wilfling B and Krämer W, *The Lorenz ordering of Singh-Maddala income distributions*, 1993 *Econ. Lett.* **43** 53
[60] Kleiber C, *Dagum versus Singh-Maddala income distributions*, 1996 *Econ. Lett.* **53** 265
[61] Kleiber C, *A guide to the Dagum distributions*, 2008 *Modeling Income Distributions and Lorenz Curves* ed D Chotikapanich (New York: Springer) pp 97–117
[62] Levy M and Solomon S, *New evidence for the power-law distribution of wealth*, 1997 *Physica* A **242** 90
[63] Klass O, Biham O, Levy M, Malcai O and Solomon S, *The Forbes 400 and the Pareto wealth distribution*, 2006 *Econ. Lett.* **90** 290
[64] Klass O S, Biham O, Levy M, Malcai O and Solomon S, *The Forbes 400, the Pareto power-law and efficient markets*, 2007 *Eur. Phys. J.* B **55** 143
[65] Biewen M and Jenkins S P, *A framework for the decomposition of poverty differences with an application to poverty differences between countries*, 2005 *Empir. Econ.* **30** 331
[66] Quintano C and D'Agostino A, *Studying inequality in income distribution of single-person households in four developed countries*, 2006 *Rev. Income Wealth* **52** 525
[67] Betti G, D'Agostino A and Lemmi A, *Fuzzy monetary poverty measures under a Dagum income distributive hypothesis*, 2008 *Modeling Income Distributions and Lorenz Curves* ed D Chotikapanich (New York: Springer) pp 225–40

J. Stat. Mech. (2012) P12006