LIS
Working Paper Series

No. 600

Extension of the κ-generalized distribution: new four-parameter models for the size distribution of income and consumption

Masato Okamoto

December 2013

Luxembourg Income Study (LIS), asbl

Extension of the κ-generalized distribution: new four-parameter models for the size distribution of income and consumption

Masato Okamoto

Statistical Survey Researcher

The Ministry of Internal Affairs and Communications

19-1 Wakamatsu-cho, Shinjuku-ku, Tokyo 162-8668   PFC02714@nifty.com

Abstract This paper studies a new kind of generalized beta distribution that is different from the GB1 and GB2 of McDonald (1984). This new four-parameter statistical distribution, the extended κ-generalized distribution of the second kind, abbreviated EκG2, is derived as one of two kinds of generalizations from the κ-generalized distribution of Clementi *et al*. (2007). By empirical comparison with the GB2 using the LIS income/consumption data, the EκG2 is found to be an overall better fit in terms of both frequency-based (FB) evaluation criteria, such as the likelihood, and money-amount-based (MAB) evaluation criteria, such as the accuracy of the estimated Lorentz curve. The EκG2 also overall outperforms the double Pareto-lognormal distribution (dPLN) of Reed (2003) in terms of FB criteria. Although not necessarily superior to the dPLN in terms of MAB criteria, the EκG2 is judged to be an overall better fit to the empirical distributions relative to the dPLN by a combined evaluation using both FB and MAB criteria.

This paper also discusses similarities and differences in characteristics between the EκG2 and GB2, including the shapes of the distributions.

1 Introduction

Parametric income distribution models (PIDMs) are frequently used to approximately recover the original size distributions from grouped data for estimation of income inequality and poverty when the microdata are unavailable. Furthermore, PIDMs that can represent income distributions using only a few parameters are also indispensable when studying determinants for income level and income inequality. One example of such a study is the Mincer-type equation (typically using the lognormal distribution as the error term). Many PIDMs have been proposed and studied. In addition to the lognormal distribution (LN), other PIDMs, such as the Singh-Maddala distribution [19], the Dagum distribution [5], and the generalized beta distribution of the second kind (GB2) [11], are well-known; however, attempts to identify new PIDMs continue. The double Pareto-lognormal distribution (dPLN) proposed by Reed [16] was found to be a better fit than the GB2 to income distributions for several countries by Reed and Wu [17] and Okamoto [13, 15]. Furthermore, Okamoto [13] showed that the Gini index for the overall income distribution can be analytically expressed by parameters of dPLNs (as well as LNs) fitted to the distributions in subgroups (e.g., age groups and regions). By the analytic expression of the Gini index for the mixture distributions, the LN/dPLN enables us to analyze contributions of different subgroup characteristics to the Gini index for the overall income distribution. The κ-generalized distribution (κG) proposed by Clementi *et al*. [3] is a better fit for some countries than the existing three-parameter PIDMs ([3, 13]) and tends to yield better estimates of income inequality even when the goodness-of-fit is inferior to the existing PIDMs in terms of the likelihood value ([13]).

This study, motivated by the κG's tendency to yield a better inequality estimation, extends the κG to four-parameter PIDMs to attain a stronger goodness-of-fit relative to that of the existing four-parameter PIDMs in terms of both frequency-based (FB) evaluation criteria, such as the likelihood, and money-amount-based (MAB) evaluation criteria, such as the accuracy of the estimated Lorentz curve and inequality indices. One's choice in PIDM may vary depending on how one intends to use it. For example, a PIDM that is a best fit to an income distribution for a specific country based on a specific evaluation criterion may be an appropriate choice in some cases. However, the purpose of this study is to derive PIDMs that would be well fitted to income/consumption distributions for many countries in terms of both FB and MAB measures, assuming general/multi-purpose use. For this reason, the new and existing four-parameter PIDMs are fitted to the size distributions of six variables, i.e., gross income, disposable income, consumption and their equivalized variables, in about twenty countries from each of waves 4-6 from the LIS database [10]. The empirical comparisons show that, in the overall evaluation, the extended κ-generalized distribution of the second kind (EκG2), one of the two kinds of generalizations of the κG distribution, is a better fit to those size distributions (of non-equivalized variables in particular) than the GB2 in terms of both FB and MAB criteria. The EκG2 also outperforms the dPLN overall in terms of FB measures (especially in the cases of non-equivalized variables). Although not necessarily superior to the dPLN in terms of MAB measures, the EκG2 is judged to be a better fit to the size distributions (of non-equivalized variables in particular) relative to the dPLN by a combined evaluation using both FB and MAB measures. The extended κ-generalized distribution of the first kind (EκG1), another kind of generalization of the κG, is inferior to other PIDMs in terms of FB measures; that said, in terms of MAB measures, the EκG1 is a better fit than the GB2 and a slightly better fit than the EκG2 to the size distributions of equivalized/non-equivalized gross and disposable income.

This paper proceeds as follows: The next section introduces four-parameter PIDMs to be considered and presents related characteristics, such as the analytic expressions of the distribution function, Lorentz curve and inequality indices. The choice of PIDMs includes the new PIDMs, EκG1 and EκG2, and their inverse distributions denoted by IEκG1 and IEκG2. Section 3 discusses the shape of the probability density functions (PDFs) of the new PIDMs. The EκG1 and EκG2 are shown to have unimodal density functions in typical cases in which the density can be regarded as zero at null income/consumption. Several methods for evaluating goodness-of-fit are introduced in section 4 and then applied to assess and summarize the empirical results obtained by fitting the PIDMs to the LIS datasets from waves 4-6 in section 5. Finally, the last section concludes the discussion. The regularity of the EκG2 in terms of maximum likelihood estimation is proved, and its Fisher information matrix is presented in appendices.

2 Statistical distributions to be compared

Among four-parameter PIDMs, the GB2 proposed by McDonald [11] is probably the most popular model. The cumulative distribution function (CDF) of the GB2 is expressed as follows:

$$F_{\mathrm{GB2}}(x;a,b,p,q)=\frac{1}{B(p,q)}\int_0^z w^{p-1}(1-w)^{q-1}dw=I_z(p,q),\qquad z=\frac{\left(\frac{x}{b}\right)^a}{1+\left(\frac{x}{b}\right)^a},$$

$$\text{where } 0<x<\infty;\ a,b,p,q>0.$$

$I_z(p,q)$ denotes the incomplete beta function. The GB2 is equivalent to the Singh-Maddala distribution [19] when $p=1$ and the Dagum distribution [5] when $q=1$. McDonald [11] also proposed the generalized beta distribution of the first kind (GB1), which has a finite domain.

$$F_{\mathrm{GB1}}(x;a,b,p,q) = I_z(p,q), \qquad z = \left(\frac{x}{b}\right)^a, \qquad \text{where } 0 < x < b;\ a,b,p,q > 0.$$

Reed [16] derived the dPLN by log-transforming the normal Laplace distribution, which is defined as a sum of two independent random variables that follow a normal distribution and an asymmetric Laplace distribution, respectively. The dPLN attains better goodness-of-fit than the GB2 to income distributions in several countries ([13, 15, 17]).

$$F_{\mathrm{dPLN}}(x;\mu,\sigma,\alpha,\beta)$$
$$= \frac{\alpha\beta}{\alpha+\beta}\left[\frac{1}{\beta}x^{\beta}e^{-\beta\mu+\beta^2\sigma^2/2}\Phi^c\left(\frac{\log x - \mu + \beta\sigma^2}{\sigma}\right) + \frac{1}{\beta}\Phi\left(\frac{\log x - \mu}{\sigma}\right)\right.$$
$$\left. - \frac{1}{\alpha}x^{-\alpha}e^{\alpha\mu+\alpha^2\sigma^2/2}\Phi\left(\frac{\log x - \mu - \alpha\sigma^2}{\sigma}\right) + \frac{1}{\alpha}\Phi\left(\frac{\log x - \mu}{\sigma}\right)\right],$$
$$\text{where } 0 < x < \infty;\ \sigma,\alpha,\beta > 0.$$

In the above formula, Φ denotes the CDF of the standard normal distribution, and Φ^c denotes the complementary function of the CDF, defined as $1 - \Phi$.

The κG, a three-parameter PIDM proposed by Clementi *et al.* [3], tends to yield better estimates of the Lorentz curves and income inequalities, although the likelihood values are not necessarily higher than those from the existing three-parameter PIDMs ([13]). The CDF of the κG is expressed as follows:

$$F_{\kappa\mathrm{G}}(x;\alpha,\beta,\kappa) = 1 - \exp_\kappa\left(-\left(\frac{x}{\beta}\right)^\alpha\right) = 1 - \left[\sqrt{1+\kappa^2\left(\frac{x}{\beta}\right)^{2\alpha}} - \kappa\left(\frac{x}{\beta}\right)^\alpha\right]^{1/\kappa},$$
$$\text{where } 0 < x < \infty;\ \alpha,\beta,\kappa > 0.$$

Taking note of the tendency, two kinds of generalizations of the κG are introduced below to try to produce a better fit to empirical income/consumption distributions relative to the existing four-parameter PIDMs in terms of both FB measures, such as the likelihood value, and MAB measures, such as the accuracy of the estimated Lorentz curve.

The κG is derived by 'Weibullizing' the deformed exponential function $\exp_\kappa x \coloneqq \left[\sqrt{1+(\kappa x)^2} + \kappa x\right]^{1/\kappa}$ of Kanidakis *et al*. [7]. Because the deformed logarithmic function $\log_\kappa \pi = \frac{\pi^\kappa - \pi^{-\kappa}}{2\kappa}$, the inverse of $\exp_\kappa x$, is generalized to the two-parameter deformed logarithmic function $\log_{\kappa,r}\pi = \pi^r\,\frac{\pi^\kappa - \pi^{-\kappa}}{2\kappa}$ by Kaniadakis *et al*. [8], it is natural to create a new PIDM by Weibullizing the two-parameter deformed exponential function, implicitly defined as the inverse of $\log_{\kappa,r}\pi$. Hereafter, the new PIDM shall be called the extended κ-generalized distribution of the first kind (EκG1). The inverse of the CDF of the EκG1 is expressed as follows:

$$F_{\mathrm{E\kappa G1}}^{-1}(\pi;a,b,q,r) = b\left[-(1-\pi)^r\,\frac{(1-\pi)^{\frac{1}{2q}} - (1-\pi)^{-\frac{1}{2q}}}{1/q}\right]^{\frac{1}{a}},$$
$$\text{where } 0 < \pi < 1;\ a,b,q > 0, r < \frac{1}{2q}.$$

The CDF of the EκG1 does not allow an explicit expression. When $r = 0$ (and $a = \alpha,\ b = \beta,\ q = \frac{1}{2\kappa}$), the EκG1 is equivalent to the κG, and when $r = -\frac{1}{2q}$, the EκG1 is equivalent to the Singh-Maddala distribution.

Another type of generalization is based on an (implicit) analytic expression of the Lorentz curve of the κG and GB2.[1]

$$L_{\kappa\mathrm{G}}(\pi;\alpha,\kappa)=I_z\left(1+\frac{1}{\alpha},\frac{1}{2\kappa}-\frac{1}{2\alpha}\right),\qquad z=1-(1-\pi)^{2\kappa}=I_\pi^{-1}\left(1,\frac{1}{2\kappa}\right),\tag{1}$$

$$L_{\mathrm{GB2}}(\pi;a,p,q)=I_z\left(p+\frac{1}{a},q-\frac{1}{a}\right),\qquad z=I_\pi^{-1}(p,q).\tag{2}$$

Let parameters α and κ be replaced by a and $\frac{1}{2q}$, respectively, in the κG's Lorentz curve (1); then, its comparison with the GB2's Lorentz curve (2) leads to a natural extension of the Lorentz curve (1), as follows:

$$L_{\mathrm{E\kappa G2}}(\pi;a,p,q)=I_z\left(p+\frac{1}{a},q-\frac{1}{2a}\right),\qquad z=I_\pi^{-1}(p,q).\tag{3}$$

The new PIDM corresponding to the Lorentz curve (3) shall be called the generalized κ-generalized distribution of the second kind (EκG2). The new model has the following CDF:

$$F_{\mathrm{E\kappa G2}}(x;a,b,p,q)=I_z(p,q),\qquad z=\left(\frac{x}{b}\right)^a\left[\sqrt{1+\frac{1}{4}\left(\frac{x}{b}\right)^{2a}}-\frac{1}{2}\left(\frac{x}{b}\right)^a\right]=2\Bigg/\left[\sqrt{1+4\left(\frac{x}{b}\right)^{-2a}}+1\right],$$

where $0<x<\infty$; $a,b,p,q>0$.

The above formula indicates that the EκG2 is a new kind of generalized beta distribution, different from the GB1 and GB2. When $p=1$ (and $a=\alpha$, $b=\beta$, $q=\frac{1}{2\kappa}$), the EκG2 is equivalent to the κG. When a random variable X follows the GB2 with parameters $(1,1,p,q)$, then, $Y=X/\sqrt{1+X}$ follows the EκG2 with parameters $(1,1,p,q)$.[2] A reciprocal of a random variable from the GB2 with parameters (a,b,p,q) follows the GB2 with parameters $(a,1/b,q,p)$. In contrast, a reciprocal of a random variable from the EκG1/EκG2 does not follow the EκG1/EκG2. A reciprocal of a variable from the EκG1 with parameters (a,b,q,r) follows the IEκG1, listed in Table 1, with parameters $(a,1/b,q,r)$. Similarly, the inverse distribution of the EκG2 with (a,b,p,q) is the IEκG2, listed in Table 1, with $(a,1/b,q,p)$. Note that parameters (a,b,p,r) of the IEκG1 must be in a domain defined as $a,b,p>0,r<\frac{1}{2p}$, and parameters (a,b,p,q) of the IEκG2 must be in a domain defined as $a,b,p,q>0$.

The moments, mean log deviation (MLD), Theil index and coefficients of variation (CV) of the new four PIDMs can be expressed analytically. Analytic expressions for the Gini indices can also be derived in the same way as that of the GB2 devised by McDonald [11]. Those formulas, together with those of the PDFs and Lorentz curves, are listed in Table 1.

Some formulas in Table 1 are found in other literature. In particular, those formulas are for moments of the dPLN ([17]), the Gini index and Lorentz curve of the dPLN ([13, 14]) and the MLD and Theil index of the GB2 ([6]).

For the EκG2, which plays the leading role in this paper, a procedure for the maximum-likelihood parameter estimation (when fitted to microdata) is given in Appendix 1 in addition to proof of the regularity in terms of the maximum likelihood estimation in Appendix 2 and the Fisher information matrix in Appendix 3. The parameter estimation procedure and information matrix are similar to those of the GB2 described by Kleiber and Kotz [9]. Because PIDMs are fitted to grouped data in this paper, the estimation procedure in Appendix 1 is inapplicable. The procedure actually employed is explained in section 5.

---

1 Clementi *et al.* [4] give a different expression of the κG's Lorentz curve. Their expression is equivalent to (1).

2 $Z=X/1+X$ follows the GB1 with parameters $(1,1,p,q)$, which is equivalent to the beta distribution. The above EκG2 variable can be expressed as $Y=\sqrt{XZ}$.

Table 1 Distributions and population characteristics

PIDM	Inverse CDF $F^{-1}(\pi)$	CDF $F(x)$	PDF $f(x)$
κG (α,β,κ)	$\beta\left[-\frac{(1-\pi)^{\kappa}-(1-\pi)^{-\kappa}}{2\kappa}\right]^{\frac{1}{\alpha}}$	$1-\left[\sqrt{1+\kappa^2\left(\frac{x}{\beta}\right)^{2\alpha}}-\kappa\left(\frac{x}{\beta}\right)^{\alpha}\right]^{1/\kappa}$	$\frac{\alpha x^{\alpha-1}}{\beta^{\alpha}}\frac{\left[\sqrt{1+\kappa^2\left(\frac{x}{\beta}\right)^{2\alpha}}-\kappa\left(\frac{x}{\beta}\right)^{\alpha}\right]^{1/\kappa}}{\sqrt{1+\kappa^2\left(\frac{x}{\beta}\right)^{2\alpha}}}$
EκG1 (a,b,q,r)	$b\left[-(1-\pi)^{r}\frac{(1-\pi)^{\frac{1}{2q}}-(1-\pi)^{-\frac{1}{2q}}}{1/q}\right]^{\frac{1}{a}}$		$\frac{a\left[-(1-\pi)^{r}\frac{(1-\pi)^{\frac{1}{2q}}-(1-\pi)^{-\frac{1}{2q}}}{1/q}\right]^{-\frac{1}{a}+1}}{b\left[\left(qr+\frac{1}{2}\right)(1-\pi)^{r+\frac{1}{2q}-1}-\left(qr-\frac{1}{2}\right)(1-\pi)^{r-\frac{1}{2q}-1}\right]}$, where $\pi=F_{\mathrm{E\kappa G1}}^{-1}(x;a,b,q,r)$ (*implicit*)
EκG2 (a,b,p,q)	$bz^{\frac{1}{a}}(1-z)^{-\frac{1}{2a}}$, where $z=I_{\pi}^{-1}(p,q)$	$I_z(p,q)$, where $z=\left(\frac{x}{b}\right)^{a}\left[\sqrt{1+\frac{1}{4}\left(\frac{x}{b}\right)^{2a}}-\frac{1}{2}\left(\frac{x}{b}\right)^{a}\right]$	$\frac{a}{bB(p,q)}\frac{z^{p-\frac{1}{a}}(1-z)^{q+\frac{1}{2a}}}{1-\frac{1}{2}z}$, where $z=\left(\frac{x}{b}\right)^{a}\left[\sqrt{1+\frac{1}{4}\left(\frac{x}{b}\right)^{2a}}-\frac{1}{2}\left(\frac{x}{b}\right)^{a}\right]$
IEκG1 (a,b,p,r)	$b\left[-\pi^{r}\frac{\pi^{\frac{1}{2p}}-\pi^{-\frac{1}{2p}}}{1/p}\right]^{-\frac{1}{a}}$		$\frac{a\left[-\pi^{r}\frac{\pi^{\frac{1}{2p}}-\pi^{-\frac{1}{2p}}}{1/p}\right]^{\frac{1}{a}+1}}{b\left[\left(pr+\frac{1}{2}\right)\pi^{r+\frac{1}{2p}-1}-\left(pr-\frac{1}{2}\right)\pi^{r-\frac{1}{2p}-1}\right]}$, where $\pi=F_{\mathrm{IE\kappa G1}}^{-1}(x;a,b,p,r)$ (*implicit*)
IEκG2 (a,b,p,q)	$bz^{\frac{1}{2a}}(1-z)^{-\frac{1}{a}}$, where $z=I_{\pi}^{-1}(p,q)$	$I_z(p,q)$, where $z=1-\left(\frac{x}{b}\right)^{-a}\left[\sqrt{1+\frac{1}{4}\left(\frac{x}{b}\right)^{-2a}}-\frac{1}{2}\left(\frac{x}{b}\right)^{-a}\right]=\left[\sqrt{1+\frac{1}{4}\left(\frac{x}{b}\right)^{-2a}}-\frac{1}{2}\left(\frac{x}{b}\right)^{-a}\right]^{2}$	$\frac{a}{bB(p,q)}\frac{z^{p-\frac{1}{2a}}(1-z)^{q+\frac{1}{a}}}{2(1+z)}$, where $z=1-\left(\frac{x}{b}\right)^{-a}\left[\sqrt{1+\frac{1}{4}\left(\frac{x}{b}\right)^{-2a}}-\frac{1}{2}\left(\frac{x}{b}\right)^{-a}\right]=\left[\sqrt{1+\frac{1}{4}\left(\frac{x}{b}\right)^{-2a}}-\frac{1}{2}\left(\frac{x}{b}\right)^{-a}\right]^{2}$
dPLN $(\mu,\sigma^2,\alpha,\beta)$		$\frac{\alpha\beta}{\alpha+\beta}\left[\frac{1}{\beta}x^{\beta}e^{-\beta\mu+\beta^2\sigma^2/2}\Phi^{c}\left(\frac{\log x-\mu+\beta\sigma^2}{\sigma}\right)+\frac{1}{\beta}\Phi\left(\frac{\log x-\mu}{\sigma}\right)-\frac{1}{\alpha}x^{-\alpha}e^{\alpha\mu+\alpha^2\sigma^2/2}\Phi\left(\frac{\log x-\mu-\alpha\sigma^2}{\sigma}\right)+\frac{1}{\alpha}\Phi\left(\frac{\log x-\mu}{\sigma}\right)\right]$	$\frac{\alpha\beta}{\alpha+\beta}\left[x^{\beta-1}e^{-\beta\mu+\beta^2\sigma^2/2}\Phi^{c}\left(\frac{\log x-\mu+\beta\sigma^2}{\sigma}\right)+x^{-\alpha-1}e^{\alpha\mu+\alpha^2\sigma^2/2}\Phi\left(\frac{\log x-\mu-\alpha\sigma^2}{\sigma}\right)\right]$
GB1 (a,b,p,q)	$bz^{\frac{1}{a}}$, where $z=I_{\pi}^{-1}(p,q)$	$I_z(p,q)$, where $z=\left(\frac{x}{b}\right)^{a}$	$\frac{a}{bB(p,q)}z^{p-\frac{1}{a}}(1-z)^{q-1}=\frac{ax^{ap-1}\left[1-\left(\frac{x}{b}\right)^{a}\right]^{q-1}}{b^{ap}B(p,q)}$, where $z=\left(\frac{x}{b}\right)^{a}$
GB2 (a,b,p,q)	$bz^{\frac{1}{a}}(1-z)^{-\frac{1}{a}}$, where $z=I_{\pi}^{-1}(p,q)$	$I_z(p,q)$, where $z=\frac{\left(\frac{x}{b}\right)^{a}}{1+\left(\frac{x}{b}\right)^{a}}$	$\frac{a}{bB(p,q)}z^{p-\frac{1}{a}}(1-z)^{q+\frac{1}{a}}=\frac{ax^{ap-1}}{b^{ap}B(p,q)\left[1+\left(\frac{x}{b}\right)^{a}\right]^{p+q}}$, where $z=\frac{\left(\frac{x}{b}\right)^{a}}{1+\left(\frac{x}{b}\right)^{a}}$

Table 1 Distributions and population characteristics (*continued*)

PIDM	Lorentz curve	Moments	Gini index
κG (α,β,κ)	$I_z\left(1+\frac{1}{\alpha},\frac{1}{2\kappa}-\frac{1}{2\alpha}\right)$, where $z=1-(1-\pi)^{2\kappa}$	$\frac{b^h}{(2\kappa)^{\frac{h}{\alpha}+1}}B\left(1+\frac{h}{\alpha},\frac{1}{2\kappa}-\frac{h}{2\alpha}\right)$	$1-2\frac{B\left(1+\frac{1}{\alpha},\frac{1}{\kappa}-\frac{1}{2\alpha}\right)}{B\left(1+\frac{1}{\alpha},\frac{1}{2\kappa}-\frac{1}{2\alpha}\right)}$
EκG1 (a,b,q,r)	$I_z\left(1+\frac{1}{a},q-\frac{1}{2a}+\frac{2qr}{a}\right)$, where $z=1-(1-\pi)^{1/q}$	$b^h q^{\frac{h}{a}+1}\cdot$ $B\left(1+\frac{h}{a},q-\frac{h}{2a}+\frac{hqr}{a}\right)$	$1-2\frac{B\left(1+\frac{1}{a},2q-\frac{1}{2a}+\frac{qr}{a}\right)}{B\left(1+\frac{1}{a},q-\frac{1}{2a}+\frac{qr}{a}\right)}$
EκG2 (a,b,p,q)	$I_z\left(p+\frac{1}{a},q-\frac{1}{2a}\right)$, where $z=I_\pi^{-1}(p,q)$	$b^h\frac{B\left(p+\frac{h}{a},q-\frac{h}{2a}\right)}{B(p,q)}$	$\frac{B\left(2p+\frac{1}{a},2q-\frac{1}{2a}\right)}{B(p,q)B\left(p+\frac{1}{a},q-\frac{1}{2a}\right)}\Big[\frac{1}{p}\ {}_3F_2\left(1,p+q,2p+\frac{1}{a};p+1,2p+2q+\frac{1}{2a}\,\Big\vert\,1\right)-$ $\frac{1}{p+\frac{1}{a}}\ {}_3F_2\left(1,p+q+\frac{1}{2a},2p+\frac{1}{a};p+\frac{1}{a}+1,2p+2q+\frac{1}{2a}\,\Big\vert\,1\right)\Big]$
IEκG1 (a,b,p,r)	$I_z\left(p+\frac{1}{2a}-\frac{2pr}{a},1-\frac{1}{a}\right)$, where $z=\pi^{1/p}$	$b^h p^{\frac{h}{a}-1}\cdot$ $B\left(p+\frac{h}{2a}-\frac{hpr}{a},1-\frac{h}{a}\right)$	$2\frac{B\left(2p+\frac{1}{2a}-\frac{pr}{a},1-\frac{1}{a}\right)}{B\left(p+\frac{1}{2a}-\frac{pr}{a},1-\frac{1}{a}\right)}-1$
IEκG2 (a,b,p,q)	$I_z\left(p+\frac{1}{2a},q-\frac{1}{a}\right)$, where $z=I_\pi^{-1}(p,q)$	$b^h\frac{B\left(p+\frac{h}{2a},q-\frac{h}{a}\right)}{B(p,q)}$	$\frac{B\left(2p+\frac{1}{2a},2q-\frac{1}{a}\right)}{B(p,q)B\left(p+\frac{1}{2a},q-\frac{1}{a}\right)}\Big[\frac{1}{p}\ {}_3F_2\left(1,p+q,2p+\frac{1}{2a};p+1,2p+2q-\frac{1}{2a}\,\Big\vert\,1\right)-$ $\frac{1}{p+\frac{1}{2a}}\ {}_3F_2\left(1,p+q-\frac{1}{2a},2p+\frac{1}{2a};p+\frac{1}{2a}+1,2p+2q-\frac{1}{2a}\,\Big\vert\,1\right)\Big]$
dPLN $\begin{pmatrix}\mu,\sigma^2,\\ \alpha,\beta\end{pmatrix}$	$\Phi\left(\frac{\log x-\sigma^2}{\sigma}\right)-$ $\frac{\beta+1}{\alpha+\beta}x^{-\alpha+1}e^{(\alpha^2-1)\sigma^2/2}\Phi\left(\frac{\log x-\alpha\sigma^2}{\sigma}\right)+$ $\frac{\alpha-1}{\alpha+\beta}x^{\beta+1}e^{(\beta^2-1)\sigma^2/2}\Phi^c\left(\frac{\log x+\beta\sigma^2}{\sigma}\right)$, where $\pi=F_{\mathrm{dPLN}}(x;0,\sigma^2,\alpha,\beta)$ (*implicit*)	$\frac{\alpha\beta}{(\alpha-h)(\beta+h)}e^{h\mu+h^2\sigma^2/2}$	$\left[2\Phi(\sigma/\sqrt{2})-1\right]+R$, where $R=2\frac{(\alpha-1)(\beta+1)}{(\alpha+\beta)(\alpha-\beta-1)}\begin{bmatrix}-\frac{\beta}{(\alpha-1)(2\alpha-1)}e^{\alpha(\alpha-1)\sigma^2}\Phi\left(-\frac{2\alpha-1}{\sqrt{2}}\sigma\right)\\ +\frac{\alpha}{(\beta+1)(2\beta+1)}e^{\beta(\beta+1)\sigma^2}\Phi\left(-\frac{2\beta+1}{\sqrt{2}}\sigma\right)\end{bmatrix}$ if $\alpha\neq\beta+1$, $2\frac{\alpha(\alpha-1)}{(2\alpha-1)^2}e^{\alpha(\alpha-1)\sigma^2}\begin{bmatrix}\left(\frac{1}{\alpha}+\frac{1}{\alpha-1}+\frac{2}{2\alpha-1}-(2\alpha-1)\sigma^2\right)\Phi\left(-\frac{2\alpha-1}{\sqrt{2}}\sigma\right)\\ +\sqrt{2}\sigma\phi\left(-\frac{2\alpha-1}{\sqrt{2}}\sigma\right)\end{bmatrix}$ if $\alpha=\beta+1$
GB1 (a,b,p,q)	$I_z\left(p+\frac{1}{a},q\right)$, where $z=I_\pi^{-1}(p,q)$	$b^h\frac{B\left(p+\frac{h}{a},q\right)}{B(p,q)}$	$\frac{B\left(2p+\frac{1}{a},q\right)}{B(p,q)B\left(p+\frac{1}{a},q\right)}\Big[\frac{1}{p}\ {}_3F_2\left(2p+\frac{1}{a},p,1-q;2p+q+\frac{1}{a},p+1\,\Big\vert\,1\right)-$ $\frac{1}{p+\frac{1}{a}}\ {}_3F_2\left(2p+\frac{1}{a},p+\frac{1}{a},1-q;2p+q+\frac{1}{a},p+\frac{1}{a}+1\,\Big\vert\,1\right)\Big]$
GB2 (a,b,p,q)	$I_z\left(p+\frac{1}{a},q-\frac{1}{a}\right)$, where $z=I_\pi^{-1}(p,q)$	$b^h\frac{B\left(p+\frac{h}{a},q-\frac{h}{a}\right)}{B(p,q)}$	$\frac{B\left(2p+\frac{1}{a},2q-\frac{1}{a}\right)}{B(p,q)B\left(p+\frac{1}{a},q-\frac{1}{a}\right)}\Big[\frac{1}{p}\ {}_3F_2\left(1,p+q,2p+\frac{1}{a};p+1,2p+2q\,\Big\vert\,1\right)-\frac{1}{p+\frac{1}{a}}\ {}_3F_2\left(1,p+q,2p+\frac{1}{a};p+\frac{1}{a}+1,2p+2q\,\Big\vert\,1\right)\Big]$

${}_3F_2(\theta_1,\theta_2,\theta_3;\theta_4,\theta_5\vert\cdot)$ and $\phi(\cdot)$ denote the generalized hypergeometric function and the PDF of the standard normal distribution, respectively.

Table 1 Distributions and population characteristics (*continued*)

PIDM	CV	MLD	Theil index
κG (α,β,κ)	$\sqrt{2\kappa\frac{B\left(1+\frac{2}{\alpha},\frac{1}{2\kappa}-\frac{1}{\alpha}\right)}{B\left(1+\frac{1}{\alpha},\frac{1}{2\kappa}-\frac{1}{2\alpha}\right)^2}-1}$	$\log B\left(1+\frac{1}{\alpha},\frac{1}{2\kappa}-\frac{1}{2\alpha}\right)-\log(2\kappa)-\frac{1}{\alpha}\psi(1)+\frac{1}{2\alpha}\psi\left(\frac{1}{2\kappa}\right)+\frac{1}{2\alpha}\psi\left(1+\frac{1}{2\kappa}\right)$	$\frac{1}{\alpha}\psi\left(1+\frac{1}{\alpha}\right)-\frac{1}{2\alpha}\psi\left(\frac{1}{2\kappa}-\frac{1}{2\alpha}\right)-\frac{1}{2\alpha}\psi\left(1+\frac{1}{2\kappa}+\frac{1}{2\alpha}\right)-\log B\left(1+\frac{1}{\alpha},\frac{1}{2\kappa}-\frac{1}{2\alpha}\right)+\log(2\kappa)$
EκG1 (a,b,q,r)	$\sqrt{\frac{B\left(1+\frac{2}{a},q-\frac{1}{a}+\frac{2qr}{a}\right)}{qB\left(1+\frac{1}{a},q-\frac{1}{2a}+\frac{qr}{a}\right)^2}-1}$	$\log B\left(1+\frac{1}{a},q-\frac{1}{2a}+\frac{qr}{a}\right)+\log q-\frac{1}{a}\psi(1)+\left(\frac{1}{2a}-\frac{qr}{a}\right)\psi(q)+\left(\frac{1}{2a}+\frac{qr}{a}\right)\psi(1+q)$	$\frac{1}{a}\psi\left(1+\frac{1}{a}\right)-\left(\frac{1}{2a}-\frac{qr}{a}\right)\psi\left(q-\frac{1}{2a}+\frac{qr}{a}\right)-\left(\frac{1}{2a}+\frac{qr}{a}\right)\psi\left(1+q+\frac{1}{2a}+\frac{qr}{a}\right)-\log B\left(1+\frac{1}{a},q-\frac{1}{2a}+\frac{qr}{a}\right)-\log q$
EκG2 (a,b,p,q)	$\sqrt{\frac{B(p,q)B\left(p+\frac{2}{a},q-\frac{1}{a}\right)}{B\left(p+\frac{1}{a},q-\frac{1}{2a}\right)^2}-1}$	$\log B\left(p+\frac{1}{a},q-\frac{1}{2a}\right)-\log B(p,q)-\frac{1}{a}\psi(p)+\frac{1}{2a}\psi(q)+\frac{1}{2a}\psi(p+q)$	$\frac{1}{a}\psi\left(p+\frac{1}{a}\right)-\frac{1}{2a}\psi\left(q-\frac{1}{2a}\right)-\frac{1}{2a}\psi\left(p+q+\frac{1}{2a}\right)-\log B\left(p+\frac{1}{a},q-\frac{1}{2a}\right)+\log B(p,q)$
IEκG1 (a,b,p,r)	$\sqrt{\frac{pB\left(p+\frac{1}{a}+\frac{2pr}{a},1-\frac{2}{a}\right)}{B\left(p+\frac{1}{2a}-\frac{pr}{a},1-\frac{1}{a}\right)^2}-1}$	$\log B\left(p+\frac{1}{2a}-\frac{pr}{a},1-\frac{1}{a}\right)-\log p+\frac{1}{a}\psi(1)-\left(\frac{1}{2a}-\frac{pr}{a}\right)\psi(p)-\left(\frac{1}{2a}+\frac{pr}{a}\right)\psi(1+p)$	$-\frac{1}{a}\psi\left(1-\frac{1}{a}\right)+\left(\frac{1}{2a}-\frac{pr}{a}\right)\psi\left(p+\frac{1}{2a}-\frac{pr}{a}\right)+\left(\frac{1}{2a}+\frac{pr}{a}\right)\psi\left(1+p-\frac{1}{2a}-\frac{pr}{a}\right)-\log B\left(p+\frac{1}{2a}-\frac{pr}{a},1-\frac{1}{a}\right)+\log p$
IEκG2 (a,b,p,q)	$\sqrt{\frac{B(p,q)B\left(p+\frac{1}{2a},q-\frac{2}{a}\right)}{B\left(p+\frac{1}{2a},q-\frac{1}{a}\right)^2}-1}$	$\log B\left(p+\frac{1}{2a},q-\frac{1}{a}\right)-\log B(p,q)-\frac{1}{2a}\psi(p)+\frac{1}{a}\psi(q)-\frac{1}{2a}\psi(p+q)$	$\frac{1}{2a}\psi\left(p+\frac{1}{2a}\right)-\frac{1}{a}\psi\left(q-\frac{1}{a}\right)+\frac{1}{2a}\psi\left(p+q-\frac{1}{2a}\right)-\log B\left(p+\frac{1}{2a},q-\frac{1}{a}\right)+\log B(p,q)$
dPLN $(\mu,\sigma^2,\alpha,\beta)$	$\sqrt{\frac{(\alpha-1)^2(\beta+1)^2}{\alpha\beta(\alpha-2)(\beta+2)}e^{\sigma^2}-1}$	$\frac{\sigma^2}{2}+\log\left[\frac{\alpha\beta}{(\alpha-1)(\beta+1)}\right]-\frac{\beta-\alpha}{\alpha\beta}$	$\frac{\sigma^2}{2}-\log\left[\frac{\alpha\beta}{(\alpha-1)(\beta+1)}\right]+\frac{\beta-\alpha+2}{(\alpha-1)(\beta+1)}$
GB1 (a,b,p,q)	$\sqrt{\frac{B(p,q)B\left(p+\frac{2}{a},q\right)}{B\left(p+\frac{1}{a},q\right)^2}-1}$	$\log B\left(p+\frac{1}{a},q\right)-\log B(p,q)-\frac{1}{a}\psi(p)+\frac{1}{a}\psi(p+q)$	$\frac{1}{a}\psi\left(p+\frac{1}{a}\right)-\frac{1}{a}\psi\left(p+q+\frac{1}{a}\right)-\log B\left(p+\frac{1}{a},q\right)+\log B(p,q)$
GB2 (a,b,p,q)	$\sqrt{\frac{B(p,q)B\left(p+\frac{2}{a},q-\frac{2}{a}\right)}{B\left(p+\frac{1}{a},q-\frac{1}{a}\right)^2}-1}$	$\log B\left(p+\frac{1}{a},q-\frac{1}{a}\right)-\log B(p,q)-\frac{1}{a}\psi(p)+\frac{1}{a}\psi(q)$	$\frac{1}{a}\psi\left(p+\frac{1}{a}\right)-\frac{1}{a}\psi\left(q-\frac{1}{a}\right)-\log B\left(p+\frac{1}{a},q-\frac{1}{a}\right)+\log B(p,q)$

$\psi(\cdot)$ denotes the digamma function $[\log\Gamma(\cdot)]'=\Gamma'/\Gamma$.

3 Shape of the probability density functions for the extended κG and their inverse distributions

3.1 Shape of the extended κG distribution of the second kind

The PDF of the EκG2 is expressed as follows:

$$f_{\mathrm{E\kappa G2}}(x;a,b,p,q)=\frac{a}{bB(p,q)}\frac{z^{p-\frac{1}{a}}(1-z)^{q+\frac{1}{2a}}}{1-\frac{1}{2}z}$$

$$=\frac{a}{b^{ap}B(p,q)}\frac{x^{ap-1}}{\left(\frac{1}{2}\left(\frac{x}{b}\right)^{a}+\sqrt{1+\frac{1}{4}\left(\frac{x}{b}\right)^{2a}}\right)^{p+2q-1}\sqrt{1+\frac{1}{4}\left(\frac{x}{b}\right)^{2a}}},$$

$$z=\left(\frac{x}{b}\right)^{a}\left[\sqrt{1+\frac{1}{4}\left(\frac{x}{b}\right)^{2a}}-\frac{1}{2}\left(\frac{x}{b}\right)^{a}\right]=2\Bigg/\left[\sqrt{1+4\left(\frac{x}{b}\right)^{-2a}}+1\right].$$

The EκG2 has power tails s.t. $f_{\mathrm{E\kappa G2}}(x)\sim c_1x^{ap-1}$ when $x\to 0$ and $f_{\mathrm{E\kappa G2}}(x)\sim c_2x^{-2aq-1}$ when $x\to\infty$. Because the sign of $\frac{d}{dz}\log f_{\mathrm{E\kappa G2}}(x)$, the derivative of the log-density function with respect to z, corresponds to the sign of a quadratic function of z shown below and $z(x)$ is a strictly monotonic increasing function for x with a positive derivative, the conditions for existence of the mode or local maximum of the PDF can be derived.

$$Az^2+Bz+C,\qquad \text{where } A=-\frac{1}{2}\left(p-\frac{1}{a}\right)+\frac{1}{2}\left(q+\frac{1}{2a}\right)-1,$$

$$B=-\frac{3}{2}\left(p-\frac{1}{a}\right)-\left(q+\frac{1}{2a}\right)+1,\qquad C=p-\frac{1}{a}.$$

In the case $p>\frac{1}{a}$, $f_{\mathrm{E\kappa G2}}(x)\to 0$ when $x\to 0$, and the PDF has a single peak. The peak is located at the following point:

$$x=b\left(\frac{z}{\sqrt{1-z}}\right)^{1/a},$$

$$z=\frac{3p+2q-\frac{2}{a}-1-\sqrt{(p+2q)^2+2\left(p-2q-\frac{2}{a}\right)+1}}{2\left(p+q-\frac{1}{2a}-1\right)}\quad \text{if } p+q-\frac{1}{2a}-1>0,$$

$$z=\frac{2p-\frac{2}{a}}{3p+2q-\frac{2}{a}-1}\quad \text{if } p+q-\frac{1}{2a}-1=0,$$

$$z=\frac{3p+2q-\frac{2}{a}-1+\sqrt{(p+2q)^2+2\left(p-2q-\frac{2}{a}\right)+1}}{2\left(p+q-\frac{1}{2a}-1\right)}\quad \text{if } p+q-\frac{1}{2a}-1<0.$$

In the case $p=\frac{1}{a}$, the PDF takes a finite positive value at the left limit s.t. $f_{\mathrm{E\kappa G2}}(x)\to\frac{a}{bB(p,q)}$ when $x\to 0$. In addition, if the following inequality is satisfied, the PDF has a single peak; otherwise, the PDF is monotonic decreasing.

$$q+\frac{1}{2a}<\frac{1}{2}.$$

The peak is located at the following point:

$$x = b\left(\frac{z}{\sqrt{1-z}}\right)^{1/a}, \qquad z = \frac{2q + \frac{1}{a} - 1}{q + \frac{1}{2a} - 1}.$$

The PDF is monotonic increasing on the left of the peak and monotonic decreasing on the right. In the case $p < \frac{1}{a}$, the PDF is infinite at the left limit s.t. $f_{\mathrm{E\kappa G2}}(x) \to \infty$ when $x \to 0$. In addition, if the following three inequalities are satisfied, the PDF has a local maximum and minimum; otherwise, the PDF is monotonic decreasing.

$$(p+2q)^2 + 2\left(p - 2q - \frac{2}{a}\right) + 1 > 0;\; p + q - \frac{1}{2a} - 1 < 0;\; 3p + 2q - \frac{2}{a} - 1 < 0.$$

The maximum and minimum points are at the following locations:

$$x = b\left(\frac{z}{\sqrt{1-z}}\right)^{1/a},$$

$$z = \frac{3p + 2q - \frac{2}{a} - 1 - \sqrt{(p+2q)^2 + 2\left(p - 2q - \frac{2}{a}\right) + 1}}{2\left(p + q - \frac{1}{2a} - 1\right)} \quad \text{(Maximum point)},$$

$$z = \frac{3p + 2q - \frac{2}{a} - 1 + \sqrt{(p+2q)^2 + 2\left(p - 2q - \frac{2}{a}\right) + 1}}{2\left(p + q - \frac{1}{2a} - 1\right)} \quad \text{(Minimum point)}.$$

Figure 1 charts the PDFs for various parameter values while Figure 2 compares the PDFs of the EκG2 and GB2 for the same parameter values except the scale parameter b. The scale parameters of the EκG2 are adjusted to equalize the mean to that of the GB2 with the unity scale parameter. The density of the EκG2 is thicker than that of the GB2 around the peak and left tail, while the former is thinner than the latter around the right tail.[3] The same condition holds for the left limit of both PDFs, i.e., both PDFs approach zero or a finite positive value or diverge to infinity when $x \to 0$ depending on whether $p > \frac{1}{a}$, $p = \frac{1}{a}$ or $p < \frac{1}{a}$. Unimodality is also common for the two PIDMs in the case $p > \frac{1}{a}$, whereas the possible existence of the mode or local maximum is a different characteristic of the EκG2 from the GB2 in the case $p = \frac{1}{a}$ and $p < \frac{1}{a}$.

3.2 Shape of the extended κG distribution of the first kind

The PDF of the EκG1 is implicitly expressed as follows:

$$f_{\mathrm{E\kappa G1}}(x; a, b, q, r) = \frac{a\left[-(1-\pi)^r \dfrac{(1-\pi)^{\frac{1}{2q}} - (1-\pi)^{-\frac{1}{2q}}}{1/q}\right]^{-\frac{1}{a}+1}}{b\left[\left(qr + \frac{1}{2}\right)(1-\pi)^{r+\frac{1}{2q}-1} - \left(qr - \frac{1}{2}\right)(1-\pi)^{r-\frac{1}{2q}-1}\right]},$$
$$\pi = F_{\mathrm{E\kappa G1}}(x; a, b, q, r).$$

---

[3] The peak of the PDF of the GB2 is located at $x = b[(ap-1)/(aq-1)]^{1/a}$ in the case $p > \frac{1}{a}$.

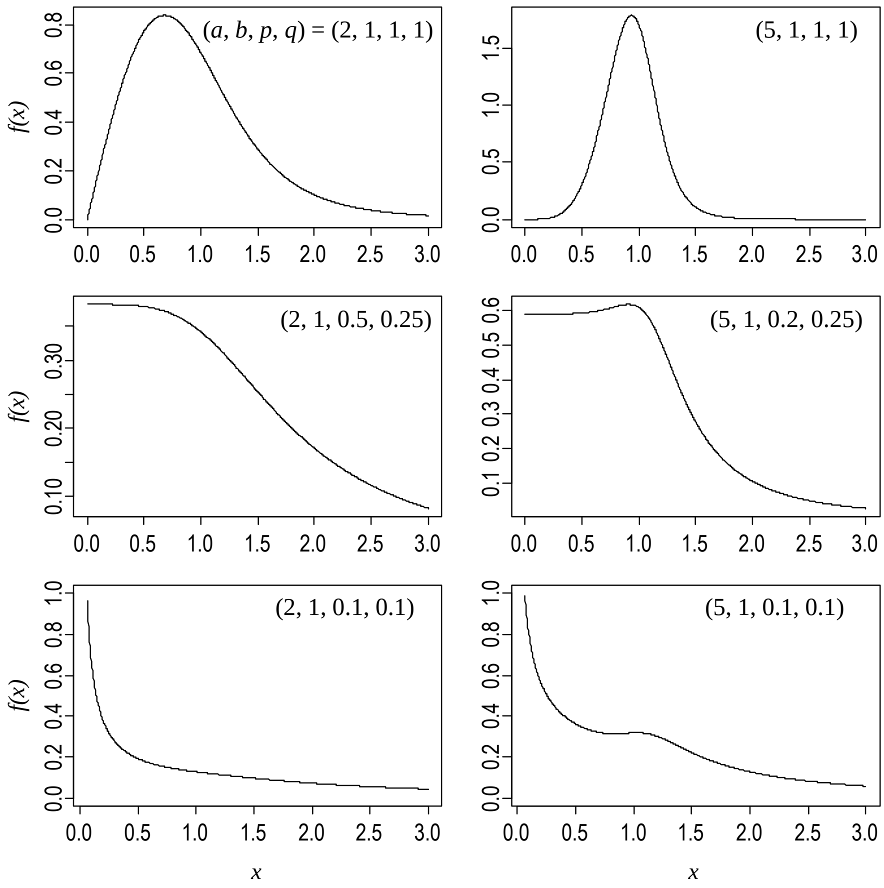

Fig. 1 PDF of the EκG2

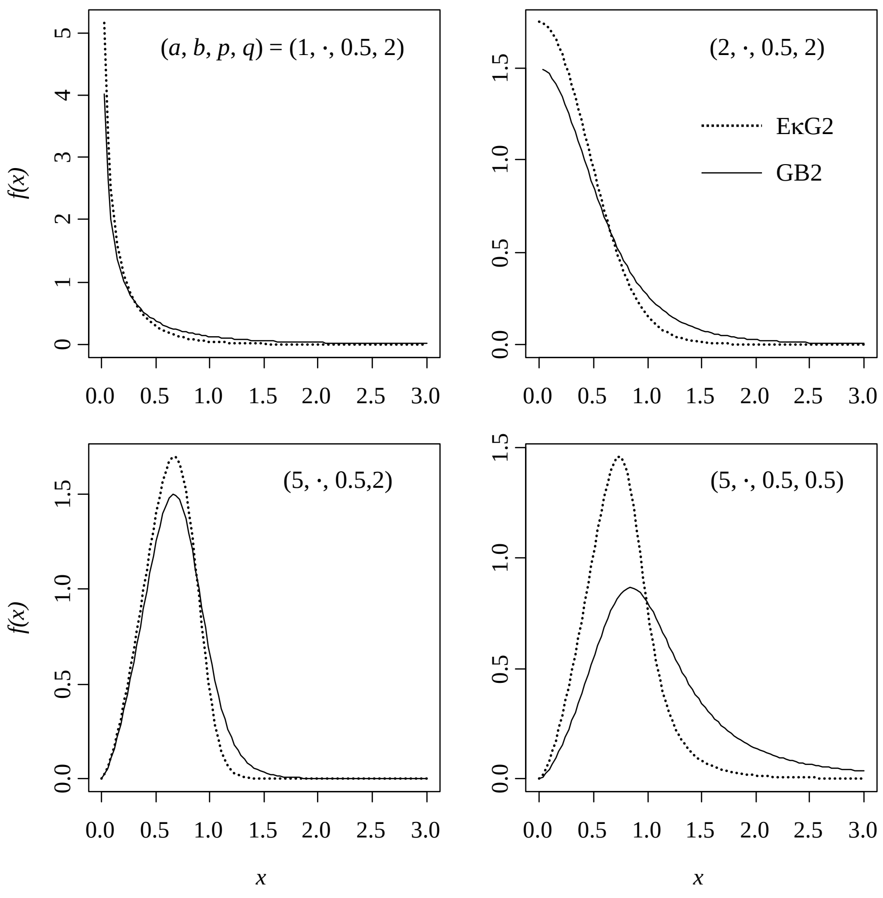

Fig. 2 PDFs of the EκG2 and GB2 with the same mean and parameter values except the scale parameters

The PDF has power tails s.t. $f_{\text{E}\kappa\text{G1}}(x) \sim c_3 x^{a-1}$ when $x \to 0$ and $f_{\text{E}\kappa\text{G1}}(x) \sim c_4 x^{-a/(1/2q-r)-1}$ when $x \to \infty$. Because the sign of $\frac{d}{d\pi}\log f_{\text{E}\kappa\text{G1}}(x)$ corresponds to the sign of the quadratic function below and $\pi(x)$ is a strictly monotonic increasing function for *x* with a positive derivative, the conditions for the existence of the mode or local maximum of the PDF can be derived.

$$AX^2 + BX + C,$$

$$\text{where } X = (1-\pi)^{\frac{1}{q}};\ \ A = \frac{2q(a-r)-1}{a}\left(r + \frac{1}{2q}\right), \qquad C = \frac{2q(a-r)+1}{a}\left(r - \frac{1}{2q}\right),$$

$$B = -A - C + \frac{2}{q}\frac{(a-1)}{a}.$$

In the case $a > 1$, $f_{\text{E}\kappa\text{G1}}(x) \to 0$ when $x \to 0$, and the PDF has a single peak. The peak is located at the following point:

$$x = F^{-1}_{\text{E}\kappa\text{G1}}(1 - X^q),$$

$$X = \frac{-B - \sqrt{B^2 - 4AC}}{2A} \quad \text{if } q(a-r) > \frac{1}{2},$$

$$X = -\frac{C}{B} \qquad \text{if } q(a-r) = \frac{1}{2},$$

$$X = \frac{-B + \sqrt{B^2 - 4AC}}{2A} \quad \text{if } q(a-r) < \frac{1}{2}.$$

In the case $a = 1$, the PDF takes a finite positive value at the left limit s.t. $f_{\text{E}\kappa\text{G1}}(x) \to \frac{a}{b}$ when $x \to 0$. In addition, if the following inequality is satisfied, the PDF has a single peak; otherwise, the PDF is monotonic decreasing.

$$r > \frac{1}{2}.$$

The peak is located at the following point:

$$x = F^{-1}_{\text{E}\kappa\text{G1}}(1 - X^q), \qquad X = \frac{C}{A}.$$

The PDF is monotonic increasing on the left of the peak and monotonic decreasing on the right. In the case $a < 1$, the PDF is infinite at the left limit s.t. $f_{\text{E}\kappa\text{G2}}(x) \to \infty$ when $x \to 0$. In addition, if the following three inequalities are satisfied, the PDF has a local maximum and minimum; otherwise, the PDF is monotonic decreasing.

$$q(a-r) < \frac{1}{2}\ (\Leftrightarrow A < 0);\ \ 0 < B < -2A;\ \ B^2 > 4AC.$$

The maximum and minimum points are at the following locations:

$$x = F^{-1}_{\text{E}\kappa\text{G1}}(1 - X^q),$$

$$X = \frac{-B + \sqrt{B^2 - 4AC}}{2A} \quad \text{(Maximum point)},$$

$$X = \frac{-B + \sqrt{B^2 - 4AC}}{2A} \quad \text{(Minimum point)}.$$

Figure 3 charts the PDFs for various parameter values.

3.3 Shape of the inverse distributions of the extended κG distributions

The IEκG1 also has power tails s.t. $f_{\mathrm{IE\kappa G1}}(x) \sim c_5 x^{a/(1/2p-r)-1}$ when $x \to 0$ and $f_{\mathrm{IE\kappa G1}}(x) \sim c_6 x^{-a-1}$ when $x \to \infty$. The PDF approaches zero or a finite positive value or diverges to infinity when $x \to 0$ depending on whether $a > \frac{1}{2p} - r$, $a = \frac{1}{2p} - r$ or $a < \frac{1}{2p} - r$. In the case $a > \frac{1}{2p} - r$, the PDF is unimodal; otherwise, the PDF is always monotonic decreasing without further restrictions on the parameters, unlike that of the EκG1.

Likewise, the IEκG2 has power tails s.t. $f_{\mathrm{IE\kappa G2}}(x) \sim c_7 x^{2ap-1}$ when $x \to 0$ and $f_{\mathrm{IE\kappa G2}}(x) \sim c_8 x^{-aq-1}$ when $x \to \infty$. The PDF approaches zero or a finite positive value or diverges to infinity when $x \to 0$ according to whether $p > \frac{1}{2a}$, $p = \frac{1}{2a}$ or $p < \frac{1}{2a}$. In the case $p > \frac{1}{2a}$, the PDF is unimodal; otherwise, the PDF is always monotonic decreasing, unlike that of the EκG2.

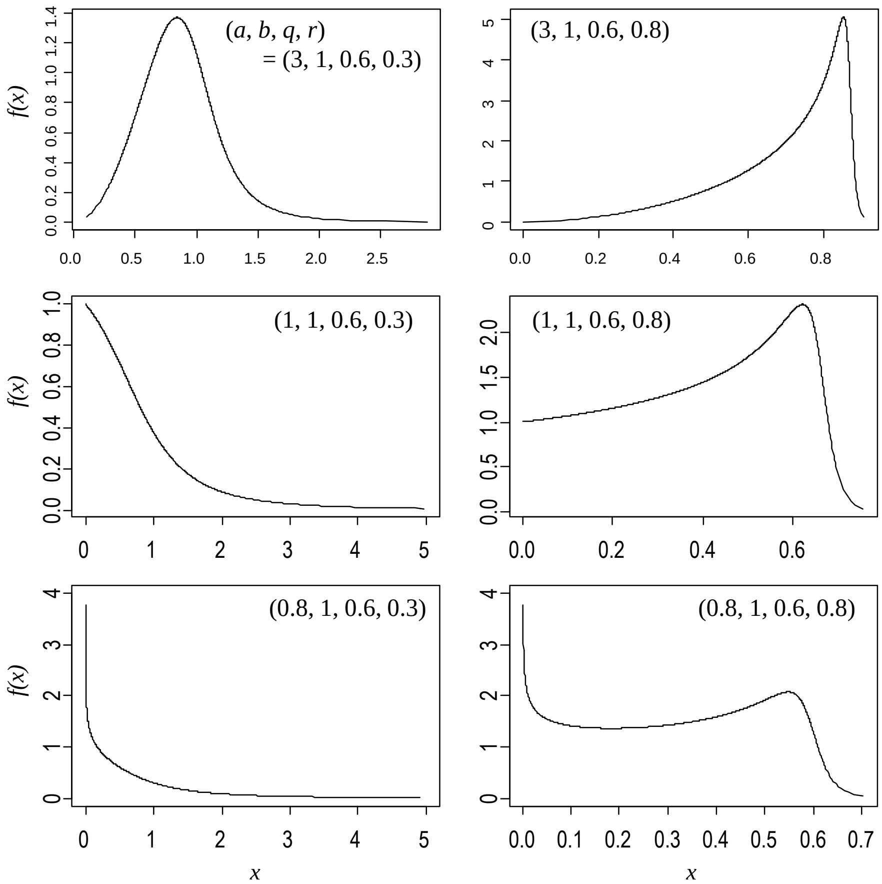

Fig. 3 PDF of the EκG1

4 Methods for evaluating the goodness-of-fit

4.1 Criteria for the goodness-of-fit to individual datasets

As shown by Okamoto [13, 15], a better evaluation based on FB measures such as the likelihood value does not necessarily imply more accurate inequality estimates. Thus, in this study, the goodness-of-fit shall be evaluated using not only FB measures but also MAB measures such as the accuracy of the estimated Lorentz curve which are expected to more closely reflect the accuracy of inequality estimates.

Four-parameter PIDMs are fitted to grouped data using the maximum likelihood (ML) method in the next section. The grouped data were obtained by tabulating income/consumption data in the LIS database by 22 classes defined as ventile groups, with equal subdivision of the lowest and highest ventile groups. The log-likelihood shall be computed using the following formula (its constant term is omitted):

$$\ell = N\left\{\pi_1 \log F(x_1;\hat{\boldsymbol{\theta}}) + \sum_{i=2}^{21} \pi_i \log\left[F(x_i;\hat{\boldsymbol{\theta}}) - F(x_{i-1};\hat{\boldsymbol{\theta}})\right] + \pi_{22} \log\left[1 - F(x_{21};\hat{\boldsymbol{\theta}})\right]\right\} + \sum_{i=1}^{21} \log f(x_i;\hat{\boldsymbol{\theta}}), \tag{4}$$

where $\pi_1 = \pi_2 = \pi_{21} = \pi_{22} = 1/40$; $\pi_i = 1/20$ for $i = 3,\cdots,20$; x_i denotes the upper bound of the income/consumption in group i for $i = 1,\cdots,21$; N denotes sample size; $\hat{\boldsymbol{\theta}}$ denotes the ML parameters, i.e., $\hat{\boldsymbol{\theta}}$ maximizes the log-likelihood in (4). In this study, the log-likelihood shall be mainly employed as a FB measure supplemented with three other FB measures: the squared root of the sum of squared errors (RSSE), the sum of absolute errors (SAE) and the chi-square (χ^2) statistic between observed and estimated frequencies.

$$\text{RSSE} = \sqrt{\sum_{i=1}^{n} \left[\pi_i - \left(F(x_i;\hat{\boldsymbol{\theta}}) - F(x_{i-1};\hat{\boldsymbol{\theta}})\right)\right]^2}, \tag{5}$$

$$\text{SAE} = \sqrt{\sum_{i=1}^{n} \left|\pi_i - \left(F(x_i;\hat{\boldsymbol{\theta}}) - F(x_{i-1};\hat{\boldsymbol{\theta}})\right)\right|}, \tag{6}$$

$$\chi^2 = N\sqrt{\sum_{i=1}^{n} \frac{\left[\pi_i - \left(F(x_i;\hat{\boldsymbol{\theta}}) - F(x_{i-1};\hat{\boldsymbol{\theta}})\right)\right]^2}{F(x_i;\hat{\boldsymbol{\theta}}) - F(x_{i-1};\hat{\boldsymbol{\theta}})}}, \tag{7}$$

where $n = 22$; $x_0 = 0$.

The RSSE between the observed and estimated Lorentz curves (LRSSE) shall be employed as a MAB measure ([13, 15]).

$$\text{LRSSE} = \sqrt{\sum_{i=1}^{n} \left(L_i - L(\lambda_i;\hat{\boldsymbol{\theta}})\right)^2}, \tag{8}$$

where $\lambda_i = \sum_{j\le i} \pi_j / \sum_{j=1}^{n} \pi_j$ expresses the cumulative population share up to group i; L_i expresses the cumulative money share up to group i, i.e., a point (λ_i, L_i) is on the empirical Lorentz curve. Because popular inequality indices including the Gini index are sensitive to distribution tails, the LRSSE in (8) shall be supplemented with its modified version which puts more importance on the accuracy near both ends.

$$\text{LwRSSE} = \sqrt{\sum_{i=1}^{n} w_i \left(L_i - L(\lambda_i;\hat{\boldsymbol{\theta}})\right)^2}, \qquad w_i = \frac{1}{\lambda_i(1-\lambda_i)} \Big/ \sum_{i=1}^{n} \frac{1}{\lambda_i(1-\lambda_i)}. \tag{9}$$

In addition to the LRSSE, the absolute error between the observed and estimated Gini index (AEG) shall also be used as another primary MAB measure supplemented with the absolute error of the mean log deviation (MLD) and the Theil index.

In this study, the overall evaluation of PIDMs shall be performed by summarizing pairwise comparisons of the goodness-of-fit to individual datasets, as explained in the next subsection. The goodness-of-fit of two PIDMs to each dataset shall be compared not only by each single measure but also by combinations of those measures, such as a combination of the likelihood value and LRSSE. When one of the two models is unanimously judged to be superior to another by the selected measures, the former shall be regarded to be a better fit than the latter, while the pairwise comparison shall be regarded as invalid in the case of inconsistent evaluation among combined measures. Unfortunately, there rarely exists a PIDM among the existing PIDMs and new ones introduced in this paper that performs the best in comparison across all PIDMs, by both the likelihood value and the LRSSE or other combinations of FB and MAB measures, although it is ideal to choose the best model based on such across-the-board comparisons. Thus, the search for the 'best' PIDM that overwhelmingly outperforms any other in terms of both FB and MAB criteria remains a future research task.

4.2 Methods for the overall evaluation

The overall evaluation shall be performed to choose the most 'suitable' PIDM by summarizing pairwise comparisons of the goodness-of-fit to individual datasets among four-parameter PIDMs. The PIDMs are fitted to the empirical size distributions of six variables, i.e., household gross income, disposable income and consumption and their equivalized variables in countries from waves 4-6 of the LIS database. Thus, the datasets can be divided into eighteen groups according to wave, variable and equivalization/non-equivalization. The eighteen dataset groups shall hereafter be called 'categories.' The evaluations of the goodness-of-fit of a pair of PIDMs to individual datasets are aggregated using two types of scores. One is the ratio of datasets gained (RoDG) defined below. When the RoDG of a PIDM over its counterpart PIDM is higher than 50%, the former is identified as more frequently a better fit to the datasets than the latter in terms of a given measure or combination of measures introduced in the previous subsection.

$$\mathrm{RoDG} = \frac{\text{No. of datasets gained}}{\text{No. of dataset gained} + \text{No. of datasets lost}} \times 100\ (\%), \qquad (10)$$

where 'datasets gained/lost' is defined as datasets to which a PIDM is a better or worse fit than its counterpart in terms of a given measure or combination of measures. In instances where the two PIDMs happen to tie for the goodness-of-fit to a dataset, both PIDMs are regarded gaining a half. Another type of score is the number of categories gained (NoCG) defined below. When the NoCG of a PIDM is more than 9 (categories), i.e., greater than that of its counterpart, it means that the former is dominant in more categories in terms of number of datasets gained.

$$\mathrm{NoCG} = \sum_i I(\text{No. of datasets gained in category } i > \text{No. of datasets lost in category } i), \qquad (11)$$

where $I(\quad)$ denotes an indicator function, i.e., the function takes a value of one if the inequality in the argument is true, a value of zero if the opposite inequality holds, or a value of 0.5 if equality holds. The reason for employing the NoCG as well as the RoDG is that there is a possibility that a PIDM could be dominant over its counterpart only in specific categories, such as categories of equivalized variables, even if the RoDG is relatively higher. When combined measures (such as a combination of the likelihood value and LRSSE) are applied to evaluate the goodness-of-fit to individual datasets, invalid cases will exist. Because a higher ratio of valid cases is also desirable for suitable PIDMs, the following validity ratio (VR) shall also be computed.

$$\begin{aligned} \mathrm{VR} &= \frac{\text{No. of valid relevant pairwise comparisons}}{\text{Total no. of relevant pairwise comparisons}} \times 100 \\ &= \frac{\text{No. of datasets gained} + \text{No. of datasets lost}}{\text{Total no. of datasets}} \times 100\ (\%) \end{aligned} \qquad (12)$$

5 Empirical comparisons in goodness-of-fit among four-parameter statistical distributions

5.1 Data and estimation procedure

Four-parameter PIDMs introduced in section 2 shall be fitted to empirical size distributions of gross income, disposable income and consumption in many countries included in the LIS database. The consumption distribution is included among the target distributions in this study in consideration of views in the literature that the level of household consumption (during sufficiently long period of time) more accurately reflects the standard-of-living than household income. To make the number of datasets in each category (classified according to wave and variables) as equal as possible, countries for which all three variables are available were selected from waves 4-6. As listed in Table 2, data for approximately 20 countries are available in each wave. As for the consumption data, one or two countries were excluded in waves 5 and 6 because, for those countries, the MLE procedure either does not converge or results in a very poor fit in most PIDMs. Equivalized variables are computed by dividing the respective variables by the square root of the number of household members.

Table 2 LIS datasets used for the empirical evaluation of PIDMs

Country	Country Code	Wave 4	Wave 5	Wave 6
Austrailia	AU	1995	2001	2003
Austria	AT	1995	2000	2003
Belgium	BE	1995	2000	–
Canada	CA	1994	2000	2004*
Czech Rep.	CZ	–	–	2004
Denmark	DK	1995	2000	2004
Estonia	EE	–	2000	2004
France	FR	1994	2000	2005
Germany	DE	1994	2000	2004
Greece	GR	1995	2000	2004
Hungary	HU	1994	1999	2005
Ireland	IE	1995	2000	2004
Israel	IL	1997	2001	2005
Italy	IT	1995	2000	2004
Luxembourg	LU	1997	2000	2004
Netherlands	NL	–	–	2004
Romania	RO	1995	–	–
Russia	RU	–	2000	–
Slovenia	SI	1997	1999	2004
Spain	ES	1995	2000	2004
Sweden	SE	–	2000	2005
Switzerland	CH	–	2000	2004
Taiwan	TW	1995	2000	2005
United Kingdom	UK	1995	1999*	2004*
United States	US	1997	–	–
No. of countries		19	21 (20)	21 (19)

* Consumption data are not used for the evaluation. See explanation in the text.
The reference years for the data in waves 4-6 are listed in columns of 'wave 4,' 'wave 5' and 'wave 6,' respectively.

In a manner similar to that in the studies by Bandourian *et al*. [1] and Reed and Wu [17], PIDMs are fitted to grouped data. Although ventile-grouped data are used in the literature, the LIS data are tabulated into 22 groups, defined as ventile groups with equal subdivisions of the lowest and highest ventile groups in this study, in consideration that popular inequality indices are sensitive to distribution tails. The slightly more detailed grouped data are expected to make the fitting results, especially inequality estimates, closer to those obtained from the microdata while restraining the increase of computational burden, similar to ventile grouping. The tail subdivisions also provide another advantage in that the ML estimation empirically becomes more stable. In certain exceptions, some of the 22 groups are collapsed for small sample data. The tabulations are made using population weights (the product of household weights and the number of household members) for equivalized variables, while household weights are used for tabulating non-equivalized variables.

Seven four-parameter PIDMs introduced in section 2, i.e., the dPLN, GB1, GB2, EκG1, EκG2, IEκG1 and IEκG2, shall be fitted to the grouped data using the MLE procedure, which maximizes the log-likelihood in (4). The simplex Nelder-Mead method ([12]), implemented in function 'optim' of statistical computer package R, is used to solve the maximization problem. When fitting four-parameter PIDMs to grouped data, the maximization is sometimes sensitive to the initial values. To address this issue, two or more sets of initial values are used for the maximization, and the obtained parameters that attain the largest likelihood values are chosen as the final estimates. The sets of initial values are obtained by fitting three-parameter PIDMs that correspond to special cases of the respective PIDM. Empirically, the initial value-setting strategy can obtain sufficiently accurate estimates. For example, to fit the GB2 by the MLE procedure, the Dagum and Singh-Maddala distributions are fitted in advance to obtain the initial values. The κG and Singh-Maddala distributions are used to obtain the initial values for the MLE fitting of the EκG1. Similarly, the κG, Dagum and Singh-Maddala distributions are employed to fit the EκG2. Either the Dagum or Singh-Maddala distribution is not a special case of the EκG2; that said, because the EκG2 is close to the GB2 in form, the ML parameters of both models are used as initial values.

5.2 Goodness of fit to individual datasets and categories in wave 6

Tables 3-1 through 3-6 list the log-likelihood value, LRSSE and AEG (absolute error of the Gini index) for each of the seven four-parameter PIDMs fitted to each dataset by 6 categories (equivalized/non-equivalized gross income, disposable income and consumption) in wave 6. Two types of scores for PIDMs are listed in the bottom two rows. These scores quantify the overall evaluation of the goodness-of-fit to each dataset. The scores in the second row from the bottom indicate the number of datasets to which the respective PIDM was the best fit among the seven PIDMs. For cases where two/three PIDMs were equally superior, value '1' is equally split into the two or three PIDMs. Thus, the scores took on values ranging from 0 to 19 for consumption data and from 0 to 21 for gross and disposable income data. Another type of score found in the bottom row indicates the number of pairwise comparisons in which the respective PIDM was a better fit to the dataset relative to its counterpart PIDM more frequently than the other way around. In cases where a pairwise comparison ended in a draw, value '1' was equally split into the respective two PIDMs. Thus, those scores took on values ranging from 0 to 6.

As for the first type of scores based on across-the-board comparisons, the EκG2 marks the highest in the categories of non-equivalized disposable income and equivalized/non-equivalized gross income in terms of the likelihood value, while the IEκG1 marks the highest in the categories of equivalized/non-equivalized gross and disposable income in terms of the LRSSE and AEG. No PIDM clearly attains a high score in terms of both the likelihood value and LRSSE/AEG. For example, in the categories of non-equivalized disposable income, the EκG2 marks the highest in terms of the likelihood value; nevertheless the EκG2 receives the lowest score (zero) in terms of the LRSSE and AEG. In contrast, the IEκG1 earns the highest mark in terms of the LRSSE and AEG although the IEκG1 received no score in terms of the likelihood value.

The second type of scores based on pairwise comparisons also indicate that the EκG2 is the best among the seven PIDMs in the categories of non-equivalized disposable income and equivalized/non-equivalized gross income in terms of the likelihood value. Furthermore, the EκG2 attains better scores in the categories of consumption and in terms of the LRSSE and AEG relative to the first type of scores, whereas the EκG1 replaces the IEκG1 as the best PIDM in the categories of equivalized/non-equivalized gross income in terms of the LRSSE and AEG. In particular, the IEκG1's score decreases substantially in the category of non-equivalized gross income. In summary, the overall evaluation using the two types of scores implies that the goodness-of-fit of the IEκG1 tends to vary substantially among countries, and the model is unsuitable for general use. Thus, the above example provides a justification for the overall evaluation based on pairwise comparisons.

Figure 4-1 shows the PDFs of the dPLN, GB2, EκG1 and EκG2 fitted to the empirical size distribution of equivalized gross income in Sweden for 1995. Figure 4-2 shows the PDFs of the same four PIDMs fitted to the non-equivalized gross income in Canada for 1994. In both charts, the income levels are proportionally adjusted to make the scale parameter *b* of the GB2 to unity, and close-ups of the PDFs around the peaks and parts of the right distribution tails are also presented. The EκG2 is the best fit to equivalized gross income in Sweden among the four PIDMs in terms of all three criteria – the likelihood value, LRSSE and AEG. The goodness-of-fit of the EκG1 is close to that of the EκG2 in terms of all three criteria. As for non-equivalized gross income in Canada, the EκG2 is the best and the dPLN and EκG1 are much inferior to the EκG2 in terms of the likelihood value, whereas the dPLN is the best and the EκG1 is also better than the EκG2 in terms of the LRSSE and AEG. From Figures 4-1 and 4-2, one can notice that the goodness-of-fit of the PIDMs is related to similarities in the shape of the PDFs. The EκG2's density is thinner than the GB2's density around the right tail, which is similar to the comparisons with the same mean and same parameter values except the scale parameters charted in Figure 2 (although the difference in density is substantially small relative to that in Figure 2). It should also be noted that the EκG2's density around the peak is slightly thinner than the GB2's density, unlike the comparisons in Figure 2.

5.3 The overall evaluation of the goodness-of-fit based on a single criterion

Tables 4-1 through 4-3 contain the NoCGs defined as (11) and RoDGs defined as (10) that summarize all pairwise comparisons regarding goodness-of-fit to datasets from waves 4-6. For example, Table 4-1 summarizes all pairwise comparisons in terms of the likelihood value. The cells at the intersection of column 'EκG2' and row 'dPLN' in the panel for non-equivalized variables contain a score of 9 for the NoCG and 85.0 for the RoDG. The scores indicate that the EκG2 was more frequently a better fit to the datasets in each of the nine categories (gross income, disposable income and consumption data in three waves) and the former was better fitted to 85.0% of all the datasets in the nine categories relative to the dPLN in terms of the likelihood value. The overall evaluation of Tables 4-1 through 4-3 reveals that the EκG2 outperforms the GB1, GB2 and IEκG1 in terms of the LRSSE and AEG as well as the likelihood value. The EκG2 also outperforms the IEκG2 with the exception of achieving equivalent goodness-of-fit in the categories of equivalized variables evaluated in terms of the likelihood value. As for the dPLN and EκG1, the EκG2 is superior to each in terms of the likelihood value in the overall evaluation, whereas the EκG2 is inferior to the dPLN in terms of the AEG and inferior to the EκG1 in terms of the LRSSE and AEG in the categories of non-equivalized variables in the overall evaluation.

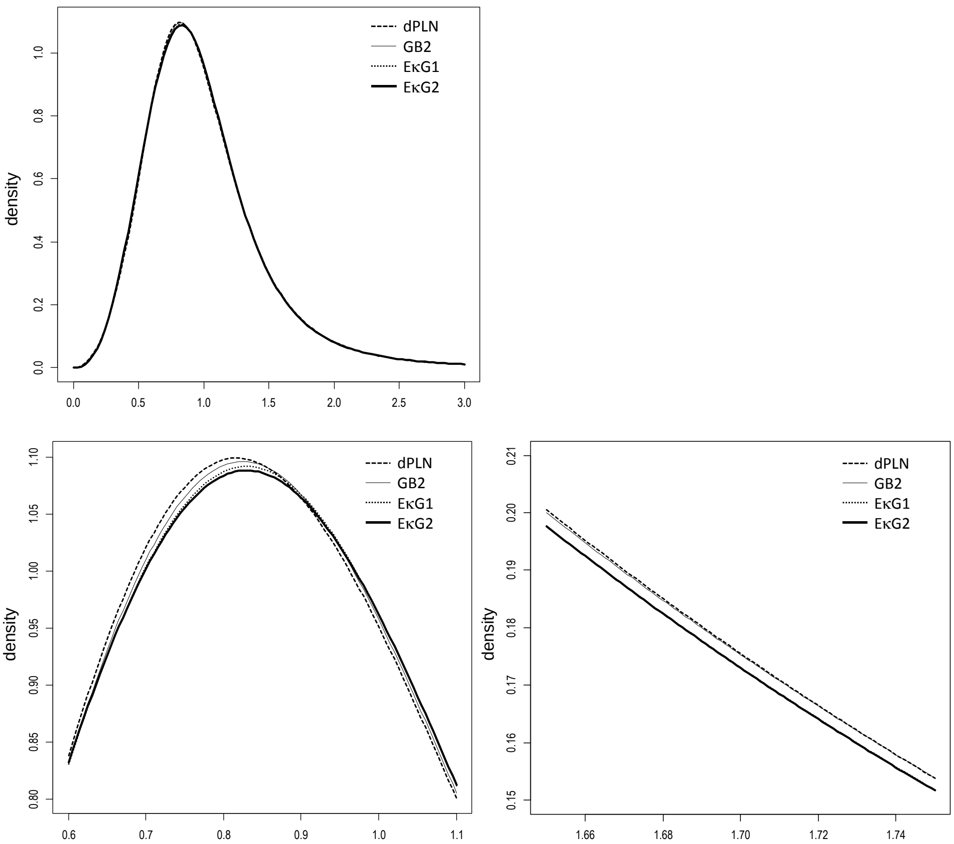

Fig. 4-1 Fitted PIDMs to equivalized gross income in Sweden, 2005

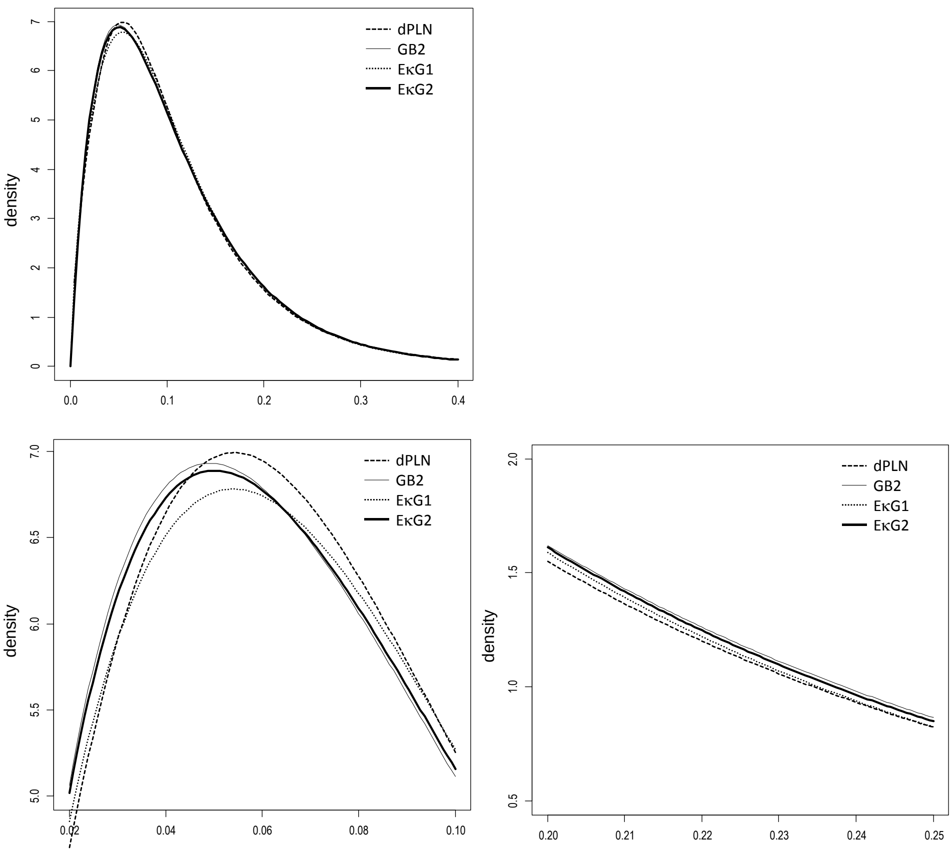

Fig. 4-2 Fitted PIDMs to non-equivalized gross income in Canada, 2004

Tables 5-1 through 5-4 detail the overall evaluation for the pairwise comparisons among the four selected

PIDMs: the dPLN, GB2, EκG1 and EκG2. To save space, detailed results for the GB1, IEκG1 and IEκG2 are omitted because of their inferior goodness-of-fit in terms of both FB and MAB measures. Because the GB2 generally produces a better fit than the GB1 in addition to its being popular, its scores are kept for reference in Tables 5-1 through 5-4. For example, Table 5-1 details the overall evaluation in terms of the likelihood value. The cells at the intersection of column 'GB2 vs. EκG2' and row 'All categories' in Table 5-1 contain a score of 15 for the NoCG and 63.2 for the RoDG. These scores indicate that the EκG2 is more frequently a better fit to the datasets in fifteen of the eighteen categories and a better fit to 63.2% of all the datasets in the eighteen categories relative to the GB2 in terms of the likelihood value. A breakdown of the overall scores into types of categories shows that the EκG2 outperforms the GB2 in all nine categories of non-equivalized variables, marking a higher RoDG, 70.3%, whereas the overall evaluations indicate its lower superiority in the categories of equivalized variables (NoCG 6 and RoDG 56.1%) and, in particular, equivalized consumption. Tables 5-2 and 5-4 show that, overall, the EκG2 outperforms the GB2 in terms of the LRSSE and AEG as well as the likelihood value. The former is dominant over the latter in each category for non-equivalized variables. The RoDGs are above 55%, although lower relative to the corresponding ratios attained when evaluated by the likelihood value.

As for the pairwise comparisons between the EκG1 and EκG2, the latter is dominant over the former in all eighteen categories in terms of the likelihood value, marking a high RoDG (78.2%), whereas the NoCG falls to 7.5 under half of all categories and the RoDG decreases to 47.5% slightly below the neutral rate of 50% when evaluated by the LRSSE. The EκG2 receives the same NoCG score as the EκG1 but the RoDG falls to 47.8% when evaluated by the AEG, indicating that the EκG2 is slightly inferior to the EκG1 in terms of MAB criteria.

The pairwise comparisons between the dPLN and EκG2 yield results similar to those between the EκG1 and EκG2, but accompanied by some indication of possible disagreement among the MAB criteria. In terms of the likelihood value, the EκG2 clearly outperforms the dPLN such that the former is dominant over the latter in sixteen of eighteen categories and especially dominant in all nine categories of non-equivalized variables, marking the RoDG at 71.1% overall and 85.0% in the categories of non-equivalized variables. The LRSSE, one of the MAB measures, yields an overall evaluation that slightly favors the EκG2 in that the EκG2 is dominant over the dPLN in eleven of eighteen categories and tied with the dPLN in one category, along with a slightly-more-than-half rate of the overall RoDG (53.3%). In contrast, the AEG, a different MAB measure, tends to favor the dPLN such that the EκG2 is dominant over the dPLN in only four categories and ties the dPLN in three categories, along with an overall RoDG of 46.9%, slightly lower than the neutral rate of 50%. Thus, the LRSSE does not agree with the AEG in the overall evaluation, although it should also be noted that both evaluations are subtle. For this reason, Table 5-3 is added to present the overall evaluation based on the LwRSSE defined as (9), which places importance on the accuracy of the estimated Lorentz curve near both ends; that said, the EκG2 is still dominant over the dPLN in ten of eighteen categories and its overall RoDG over the dPLN is 51.1%, still above the neutral rate. Thus, the overall evaluation remains favorable to the EκG2 although somewhat closer to that of the AEG, relative to the LRSSE's. Those results imply that evaluation of goodness-of-fit should not rely on a single MAB measure.

The EκG1 is inferior to the dPLN, GB2 and EκG2 in all eighteen categories in terms of the likelihood value, but superior to the GB2 and EκG2 in terms of the LRSSE and AEG in the overall evaluation. As for the pairwise comparisons between the dPLN and EκG1 in terms of MAB measures, the former appears to outperform the latter in terms of the AEG except for equivalized gross and disposable income. On the whole, a definitive judgment is difficult to make.

The (omitted) overall evaluations based on other FB measures, i.e., RSSE, ASE and χ^2 in (5) through (7), are similar to those based on the likelihood value. The EκG2 outperforms other PIDMs in terms of those FB measures as well as the likelihood value. Inconsistent evaluations can occur between FB and MAB measures (and/or among MAB measures). Thus, in the next subsection, the overall evaluation shall be performed again according to combinations of the two types of criteria.

5.4 The overall evaluation of the goodness-of-fit based on combined criteria

Tables 6-1 through 6-6 summarize all pairwise comparisons among the dPLN, GB2, EκG1 and EκG2 by combined measures. In addition to a combination of the likelihood and LRSSE and that of the likelihood and AEG, other combinations are also used for the overall evaluation. These include 1) a combination of the likelihood and absolute error of the MLD; 2) the likelihood and absolute error of the Theil index; and 3) all four FB measures (the likelihood value, RSSE, ASE and χ^2) with LRSSE/AEG.

No matter how the measures are combined, the EκG2 outperforms other PIDMs. In particular, the EκG2 is dominant over its counterparts in almost all nine categories (although tied with the GB2 in some categories) for non-equivalized variables, making the RoDGs higher than or about the same as 70% over the GB2 and higher than or about the same as 80% over the dPLN and EκG1. Overall, the EκG2 is also superior to the dPLN, GB2 and EκG1 in the categories of equivalized variables, although inferior to the dPLN in the categories of equivalized disposable income. It should be noted, however, that the VRs defined as (12) are generally not so high. The VR indicates the rate of valid cases on which all respective measures agree regarding which PIDM is a better fit to a given dataset. The rates are below 50% in the pairwise comparisons between the dPLN and EκG2. Exceptions include instances of applying a combination of the likelihood value and LRSSE, as well instances within categories of equivalized variables when applying a combination of the likelihood value and the accuracy of the Theil index. In the pairwise comparisons between the GB2 and EκG2, the rates are generally above 50% when applying combinations of two measures, including one from FB measures and another from MAB measures. When combining all FB measures with LRSSE/AEG, rates fall below 50%.

The EκG1 is judged to be inferior to the dPLN, GB2 and EκG2 by any combination of FB and MAB measures.

6 Conclusion

This paper studies two new four-parameter PIDMs derived by generalizing the κG distribution, which has an empirical tendency to estimate inequality indices more accurately than the existing three-parameter PIDMs. Empirical comparisons using the LIS datasets indicate that the EκG1, one of the two kinds of generalizations, tends to be a good fit to empirical income/consumption distributions in terms of MAB criteria, such as the accuracy of the estimated Lorentz curve and the Gini index. Thus, the EκG1 appears to inherit the features from the κG. The EκG2, another kind of generalization, can also be viewed as a new variant of a generalized beta distribution that is different from the GB1 and GB2 but closer to the GB2 in form. The latter new model tends to be a good fit to empirical income/consumption distributions in terms of both MAB criteria and FB criteria, such as the likelihood, leading to superiority over the GB1 and GB2 in both terms. Thus, the EκG2 appears to create positive synergetic effect between the κG and GB2.

The search continues for the ideal PIDM that will be a best fit to empirical income/consumption distributions in all aspects. The research most likely requires finding new statistical distributions with at least five or more parameters. Hopefully, the new four-parameter PIDMs in this paper may contribute to the future research for creating better-fitting models.

Table 3-1 Goodness-of-fit to individual datasets from wave 6 – non-equivalized consumption

Country & Year	Log-likelihood*							LRSSE							Absolute error of the Gini index						
	dPLN	GB2	EκG1	IEκG1	EκG2	IEκG2	GB1	dPLN	GB2	EκG1	IEκG1	EκG2	IEκG2	GB1	dPLN	GB2	EκG1	IEκG1	EκG2	IEκG2	GB1
AU03	-6.20	0.00	-13.26	-200.71	-11.50	37.45	-1169.83	1.488	1.530	1.471	0.684	1.458	1.505	0.162	0.437	0.451	0.433	0.191	0.391	0.437	0.039
AT04	-1.19	0.00	2.12	-1.06	-1.61	1.63	-15.18	0.017	0.018	0.013	0.026	0.019	0.016	0.011	0.005	0.006	0.004	0.008	0.006	0.005	0.000
CZ04	-0.73	0.00	0.09	-0.09	0.09	-0.14	-42.09	0.008	0.007	0.009	0.005	0.009	0.004	0.021	0.002	0.002	0.003	0.001	0.003	0.001	0.004
DK04	-572.40	0.00	-630.90	-910.20	0.00	0.00	200.80	0.060	0.029	0.033	0.032	0.029	0.029	0.033	0.012	0.005	0.008	0.001	0.005	0.005	0.010
EE04	-0.67	0.00	-0.16	0.37	-0.04	0.37	-44.42	0.289	0.290	0.288	0.293	0.286	0.293	0.336	0.092	0.092	0.092	0.094	0.091	0.094	0.102
FR05	-7.86	0.00	-31.33	-39.19	0.01	-0.01	0.47	0.006	0.007	0.012	0.034	0.008	0.008	0.011	0.000	0.003	0.002	0.009	0.003	0.003	0.004
DE04	4.37	0.00	-4.44	0.81	-5.29	2.19	-116.97	0.033	0.033	0.032	0.031	0.030	0.035	0.058	0.011	0.011	0.010	0.010	0.010	0.011	0.016
GR04	3.52	0.00	-4.66	0.58	-4.80	2.75	-39.95	0.016	0.022	0.022	0.026	0.028	0.019	0.016	0.005	0.007	0.007	0.009	0.009	0.006	0.002
HU05	-0.08	0.00	-3.73	-3.36	0.03	-0.02	-0.17	0.006	0.005	0.021	0.036	0.006	0.005	0.004	0.001	0.001	0.006	0.011	0.001	0.001	0.000
IE04	0.60	0.00	-0.07	8.70	0.00	0.71	-209.06	0.036	0.036	0.035	0.169	0.035	0.043	0.136	0.009	0.010	0.009	0.058	0.009	0.013	0.028
IL05	-2.66	0.00	-10.72	-15.56	0.00	0.00	-0.22	0.003	0.007	0.010	0.032	0.007	0.007	0.009	0.000	0.002	0.003	0.009	0.002	0.002	0.002
IT04	-1.31	0.00	-4.79	-4.08	-0.01	0.02	-32.61	0.007	0.004	0.015	0.029	0.005	0.003	0.047	0.002	0.001	0.005	0.009	0.001	0.000	0.013
LU04	0.00	0.00	3.91	-1.07	0.69	-0.73	-33.30	0.066	0.065	0.055	0.046	0.059	0.049	0.043	0.023	0.023	0.020	0.016	0.021	0.018	0.005
NL04	-3.14	0.00	3.35	-0.28	3.27	0.29	-129.89	0.048	0.037	0.030	0.031	0.031	0.035	0.051	0.016	0.012	0.010	0.011	0.011	0.012	0.012
SI04	-5.94	0.00	-10.57	-12.77	0.00	0.00	0.46	0.009	0.009	0.007	0.030	0.009	0.008	0.011	0.002	0.002	0.001	0.008	0.002	0.002	0.003
ES04	-22.32	0.00	-32.21	-34.26	0.00	0.00	-84.37	0.053	0.033	0.098	0.116	0.033	0.033	0.044	0.014	0.008	0.029	0.035	0.008	0.008	0.012
SE05	-3.41	0.00	-0.11	1.12	0.85	-0.53	-39.42	0.006	0.005	0.006	0.007	0.007	0.004	0.010	0.001	0.001	0.001	0.001	0.002	0.000	0.002
CH04	0.20	0.00	-4.04	-3.25	0.09	-0.12	-3.42	0.005	0.009	0.006	0.018	0.008	0.011	0.029	0.002	0.003	0.001	0.005	0.003	0.003	0.009
TW05	-17.56	0.00	4.32	11.70	5.96	-1.37	-11.04	0.007	0.006	0.005	0.006	0.005	0.007	0.011	0.001	0.001	0.000	0.001	0.000	0.001	0.002
Score based on across-the-board comparisons	3	0.7	3.5	3.5	2.2	3.2	3	4	1	4	0	2	4	4	2	1	3	1	4	3	5
Score based on pairwise comparisons	3	4	2	1	6	5	0	2	3	5	0	4	6	1	2	3	5	0	5	1	1

* Differences from the corresponding value of the GB2 are listed. (The same is true of the subsequent tables.)

Table 3-2 Goodness-of-fit to individual datasets from wave 6 – non-equivalized disposable income

Country & Year	Log-likelihood*							LRSSE							Absolute error of the Gini index						
	dPLN	GB2	EκG1	IEκG1	EκG2	IEκG2	GB1	dPLN	GB2	EκG1	IEκG1	EκG2	IEκG2	GB1	dPLN	GB2	EκG1	IEκG1	EκG2	IEκG2	GB1
AU03	-54.51	0.00	-61.05	-78.89	0.00	0.00	12.93	0.032	0.019	0.020	0.025	0.019	0.019	0.026	0.004	0.007	0.007	0.002	0.007	0.007	0.010
AT04	-2.97	0.00	-2.84	0.38	1.49	-0.68	-6.27	0.011	0.011	0.007	0.008	0.008	0.014	0.022	0.003	0.004	0.002	0.001	0.003	0.004	0.006
CA04	-22.21	0.00	-7.65	-9.82	4.93	-1.07	-31.59	0.027	0.030	0.027	0.011	0.027	0.032	0.038	0.009	0.010	0.009	0.004	0.009	0.010	0.011
CZ04	-2.65	0.00	-22.83	-21.57	0.01	0.01	0.20	0.012	0.020	0.008	0.021	0.020	0.020	0.021	0.004	0.007	0.002	0.005	0.007	0.007	0.007
DK04	-205.60	0.00	-1055.00	-1314.80	0.20	0.20	71.80	0.039	0.030	0.029	0.049	0.030	0.030	0.032	0.003	0.006	0.004	0.010	0.006	0.006	0.009
EE04	0.01	0.00	-32.36	-32.08	0.15	0.14	-1.35	0.030	0.030	0.074	0.127	0.024	0.026	0.014	0.007	0.007	0.021	0.037	0.005	0.005	0.004
FR05	-3.57	0.00	-14.78	-16.99	0.53	-0.13	-1.89	0.016	0.021	0.012	0.010	0.019	0.021	0.025	0.006	0.007	0.004	0.001	0.007	0.007	0.008
DE04	-3.91	0.00	-21.56	-23.88	0.27	-0.04	-0.87	0.020	0.024	0.011	0.012	0.024	0.025	0.028	0.007	0.008	0.004	0.002	0.008	0.008	0.009
GR04	0.17	0.00	-1.99	-3.96	0.09	-0.13	-8.42	0.005	0.004	0.009	0.030	0.005	0.004	0.017	0.001	0.001	0.002	0.009	0.002	0.000	0.003
HU05	0.07	0.00	-2.99	-2.36	0.07	-0.07	-0.81	0.004	0.006	0.015	0.034	0.005	0.008	0.020	0.000	0.001	0.004	0.010	0.001	0.002	0.005
IE04	0.26	0.00	-0.71	-4.84	-0.66	0.14	-150.69	0.349	0.368	0.341	0.151	0.350	0.441	0.160	0.113	0.120	0.112	0.045	0.101	0.146	0.049
IL05	-20.92	0.00	-19.11	-25.27	0.00	0.00	2.92	0.018	0.036	0.030	0.011	0.036	0.036	0.041	0.006	0.012	0.011	0.002	0.012	0.012	0.014
IT04	-0.14	0.00	-14.95	-17.64	-0.01	0.00	-0.94	0.061	0.060	0.034	0.013	0.060	0.061	0.069	0.020	0.020	0.012	0.004	0.020	0.020	0.023
LU04	67.86	0.00	-4.96	-4.10	0.41	2.04	-2.29	0.013	0.005	0.041	0.066	0.007	0.013	0.023	0.004	0.001	0.012	0.020	0.001	0.005	0.007
NL04	-3.26	0.00	-27.52	-33.77	0.28	0.27	0.28	0.036	0.042	0.031	0.016	0.044	0.043	0.044	0.013	0.015	0.010	0.005	0.015	0.015	0.015
SI04	-12.48	0.00	-2.30	-1.98	0.00	0.00	1.40	0.010	0.010	0.009	0.009	0.010	0.010	0.013	0.003	0.003	0.003	0.000	0.003	0.003	0.003
ES04	-37.67	0.00	-42.95	-51.04	0.00	0.00	5.68	0.016	0.011	0.009	0.023	0.011	0.011	0.015	0.002	0.004	0.002	0.005	0.004	0.004	0.005
SE05	-28.77	0.00	-111.13	-135.45	0.03	0.03	7.33	0.029	0.022	0.022	0.037	0.022	0.022	0.024	0.002	0.005	0.002	0.007	0.005	0.005	0.007
CH04	-0.51	0.00	-0.07	-0.53	0.11	-0.10	-6.93	0.011	0.009	0.009	0.003	0.007	0.010	0.019	0.004	0.003	0.003	0.000	0.002	0.004	0.004
TW05	-4.06	0.00	0.05	-0.67	0.23	-0.10	-20.93	0.015	0.011	0.010	0.003	0.007	0.013	0.027	0.005	0.003	0.003	0.001	0.002	0.004	0.006
UK04	0.27	0.00	-147.79	-142.77	2.42	2.33	1.98	0.044	0.043	0.019	0.025	0.051	0.050	0.055	0.015	0.015	0.006	0.005	0.017	0.017	0.019
Score based on across-the-board comparisons	3	0.5	0	0	9.5	0.5	7.5	1	1	7	9	0	2	1	4	1	1	13	0	1	1
Score based on pairwise comparisons	2	4.5	1	0	6	4.5	3	4	2	6	5	3	1	0	5	2	4	6	3	1	0

Table 3-3 Goodness-of-fit to individual datasets from wave 6 – non-equivalized gross income

Country & Year	Log-likelihood*							LRSSE							Absolute error of the Gini index						
	dPLN	GB2	EκG1	IEκG1	EκG2	IEκG2	GB1	dPLN	GB2	EκG1	IEκG1	EκG2	IEκG2	GB1	dPLN	GB2	EκG1	IEκG1	EκG2	IEκG2	GB1
AU03	-51.83	0.00	-73.45	-87.40	0.11	0.11	12.49	0.052	0.022	0.025	0.033	0.022	0.022	0.029	0.010	0.005	0.008	0.003	0.006	0.006	0.011
AT04	-4.96	0.00	0.18	3.80	2.37	-1.04	-9.22	0.009	0.010	0.004	0.005	0.005	0.014	0.028	0.002	0.003	0.001	0.001	0.001	0.004	0.006
CA04	-49.58	0.00	-21.20	-19.65	4.22	-0.22	-9.80	0.032	0.044	0.038	0.018	0.040	0.045	0.049	0.010	0.014	0.013	0.006	0.013	0.014	0.015
CZ04	-3.14	0.00	-37.19	-37.11	0.03	0.03	0.45	0.014	0.024	0.015	0.034	0.025	0.025	0.028	0.003	0.009	0.003	0.007	0.009	0.009	0.010
DK04	-79.30	0.00	-1114.00	-1243.30	0.00	0.00	22.80	0.039	0.029	0.031	0.071	0.029	0.029	0.029	0.003	0.004	0.001	0.017	0.004	0.004	0.007
EE04	0.00	0.00	-42.22	-41.78	0.06	0.10	-0.38	0.046	0.047	0.094	0.168	0.043	0.041	0.026	0.010	0.010	0.025	0.049	0.009	0.008	0.002
FR05	-2.08	0.00	-11.76	-11.14	0.89	-0.38	-5.38	0.024	0.022	0.014	0.009	0.020	0.024	0.030	0.008	0.008	0.005	0.001	0.007	0.008	0.009
DE04	-20.24	0.00	-63.08	-69.80	0.03	0.03	2.66	0.016	0.016	0.013	0.039	0.017	0.017	0.020	0.001	0.005	0.001	0.010	0.006	0.006	0.007
GR04	0.21	0.00	-2.68	-5.12	0.01	-0.06	-6.66	0.008	0.007	0.006	0.027	0.004	0.008	0.022	0.002	0.002	0.001	0.008	0.001	0.002	0.005
HU05	0.07	0.00	-2.98	-2.36	0.07	-0.07	-0.84	0.004	0.006	0.015	0.034	0.005	0.008	0.020	0.000	0.002	0.004	0.010	0.001	0.002	0.006
IE04	1.63	0.00	0.26	-2.69	0.63	-3.72	-234.62	0.510	0.498	0.466	0.585	0.604	0.437	0.200	0.162	0.158	0.158	0.187	0.161	0.137	0.058
IL05	-8.01	0.00	-35.42	-37.83	0.01	0.01	0.85	0.012	0.017	0.018	0.066	0.018	0.018	0.022	0.001	0.006	0.002	0.018	0.006	0.006	0.008
IT04	-0.45	0.00	-19.81	-19.82	0.10	-0.01	-0.16	0.090	0.090	0.054	0.021	0.090	0.092	0.096	0.029	0.028	0.018	0.007	0.029	0.029	0.030
LU04	-0.20	0.00	-6.74	-4.48	0.08	0.08	-3.40	0.044	0.035	0.068	0.099	0.033	0.032	0.017	0.014	0.011	0.022	0.032	0.010	0.010	0.005
NL04	80.15	0.00	-54.27	-55.48	0.83	0.83	2.11	3.350	0.039	0.028	0.017	0.040	0.040	0.042	n.a.	0.013	0.009	0.001	0.013	0.013	0.014
SI04	-12.48	0.00	-2.30	-1.98	0.00	0.00	1.40	0.010	0.010	0.009	0.009	0.010	0.010	0.013	0.003	0.003	0.003	0.000	0.003	0.003	0.003
ES04	-33.30	0.00	-43.75	-49.87	0.03	0.03	4.51	0.017	0.011	0.009	0.025	0.011	0.011	0.014	0.002	0.004	0.002	0.005	0.004	0.004	0.005
SE05	-23.57	0.00	-75.43	-90.87	0.00	0.00	4.16	0.018	0.025	0.019	0.026	0.025	0.025	0.028	0.003	0.009	0.005	0.004	0.009	0.009	0.010
CH04	-1.91	0.00	-1.70	-1.80	0.43	-0.13	-2.19	0.017	0.017	0.013	0.004	0.015	0.018	0.023	0.006	0.006	0.004	0.000	0.005	0.006	0.006
TW05	-4.57	0.00	0.13	0.10	0.55	-0.22	-22.03	0.013	0.009	0.007	0.005	0.005	0.011	0.027	0.004	0.002	0.002	0.002	0.001	0.003	0.006
UK04	0.01	0.00	-269.92	-261.20	0.14	0.17	-3.78	0.026	0.026	0.027	0.076	0.027	0.028	0.038	0.009	0.009	0.004	0.021	0.009	0.010	0.014
Score based on across-the-board comparisons	3	0	0	1	6.5	2.5	8	5	1	3	6	3	0	3	4	0	6	7	1	0	3
Score based on pairwise comparisons	3	4.5	1	0	6	4.5	2	3	4	6	0	5	2	1	5	2	6	3	4	1	0

Table 3-4 Goodness-of-fit to individual datasets from wave 6 –equivalized consumption

Country & Year	Log-likelihood*							LRSSE							Absolute error of the Gini index						
	dPLN	GB2	EκG1	IEκG1	EκG2	IEκG2	GB1	dPLN	GB2	EκG1	IEκG1	EκG2	IEκG2	GB1	dPLN	GB2	EκG1	IEκG1	EκG2	IEκG2	GB1
AU03	18.48	0.00	9.99	46.46	1.50	-11.72	-1502.47	1.274	1.387	0.902	n.a.	1.109	0.816	0.193	0.372	0.409	0.373	n.a.	0.268	0.225	0.050
AT04	-13.23	0.00	-13.77	-15.07	0.02	0.02	-53.21	0.031	0.017	0.056	0.061	0.018	0.018	0.036	0.008	0.004	0.017	0.019	0.004	0.004	0.009
CZ04	-0.11	0.00	0.03	0.12	-0.44	0.17	-36.49	0.008	0.006	0.006	0.006	0.008	0.004	0.028	0.002	0.001	0.001	0.001	0.002	0.001	0.007
DK04	-19.90	0.00	-1403.50	-1505.90	0.10	-0.10	8.70	0.084	0.067	0.103	0.219	0.068	0.068	0.061	0.017	0.011	0.025	0.063	0.011	0.011	0.008
EE04	-0.01	0.00	1.63	-1.09	1.41	-1.29	-67.17	0.262	0.265	0.265	0.268	0.264	0.267	0.288	0.086	0.087	0.087	0.088	0.087	0.088	0.090
FR05	0.25	0.00	-11.58	-14.41	-0.13	0.05	-4.84	0.003	0.003	0.009	0.022	0.002	0.003	0.010	0.001	0.001	0.002	0.007	0.001	0.001	0.003
DE04	-1.63	0.00	2.31	1.06	2.98	-2.92	-78.01	0.039	0.041	0.039	0.043	0.040	0.044	0.056	0.014	0.014	0.014	0.015	0.014	0.015	0.017
GR04	-1.64	0.00	82.59	-21.58	15.43	-4.53	-361.96	0.068	0.070	0.085	0.078	0.080	0.078	0.042	0.023	0.024	0.029	0.026	0.025	0.026	0.005
HU05	-0.22	0.00	-0.18	0.06	0.06	-0.10	-5.94	0.002	0.002	0.003	0.003	0.002	0.004	0.025	0.000	0.000	0.001	0.001	0.000	0.001	0.007
IE04	0.60	0.00	-0.07	8.70	0.00	0.71	-209.06	0.036	0.036	0.035	0.169	0.035	0.043	0.136	0.009	0.010	0.009	0.058	0.009	0.013	0.028
IL05	0.00	0.00	-32.76	-31.63	0.10	2.59	-1.21	0.009	0.010	0.022	0.041	0.008	0.010	0.015	0.004	0.004	0.006	0.012	0.003	0.004	0.006
IT04	-1.55	0.00	0.53	-0.93	-0.79	1.08	-77.15	0.021	0.014	0.014	0.022	0.014	0.010	0.058	0.007	0.005	0.005	0.007	0.005	0.004	0.016
LU04	1.89	0.00	-1.14	0.65	-2.00	0.87	-13.77	0.023	0.031	0.105	0.034	0.038	0.028	0.032	0.009	0.012	0.036	0.012	0.014	0.010	0.003
NL04	-1.78	0.00	-0.85	0.86	1.77	-1.90	-55.83	0.009	0.006	0.008	0.003	0.008	0.003	0.039	0.004	0.002	0.003	0.001	0.003	0.001	0.010
SI04	-0.09	0.00	-5.59	-6.04	-0.05	0.04	-2.09	0.004	0.005	0.013	0.026	0.004	0.005	0.021	0.001	0.001	0.004	0.008	0.001	0.001	0.007
ES04	-24.38	0.00	-15.58	-14.77	0.02	0.08	-176.08	0.100	0.047	0.108	0.114	0.047	0.049	0.065	0.031	0.014	0.034	0.036	0.014	0.015	0.019
SE05	0.41	0.00	-2.38	-4.33	-0.37	0.06	-20.14	0.003	0.003	0.003	0.004	0.003	0.004	0.010	0.001	0.001	0.001	0.001	0.001	0.002	0.003
CH04	0.47	0.00	-2.39	-0.47	0.18	-0.35	-18.06	0.004	0.006	0.005	0.009	0.005	0.009	0.047	0.001	0.001	0.001	0.003	0.001	0.002	0.013
TW05	-1.23	0.00	-0.02	0.05	0.14	-0.55	-74.89	0.010	0.008	0.009	0.008	0.010	0.006	0.014	0.004	0.003	0.003	0.003	0.003	0.002	0.002
Score based on across-the-board comparisons	4	0	2	2	4.5	5.5	1	5	1	2	0	4	4	3	3	2	0	2	5	2	5
Score based on pairwise comparisons	3	4	1	2	5.5	5.5	0	5	4	2	0	6	3	1	5	4	2	0	6	3	1

Table 3-5 Goodness-of-fit to individual datasets from wave 6 – equivalized disposable income

Country & Year	Log-likelihood*							LRSSE							Absolute error of the Gini index						
	dPLN	GB2	EκG1	IEκG1	EκG2	IEκG2	GB1	dPLN	GB2	EκG1	IEκG1	EκG2	IEκG2	GB1	dPLN	GB2	EκG1	IEκG1	EκG2	IEκG2	GB1
AU03	-15.03	0.00	-57.27	-71.34	0.00	0.00	1.48	0.019	0.029	0.020	0.017	0.029	0.029	0.031	0.007	0.011	0.007	0.000	0.011	0.011	0.012
AT04	0.15	0.00	-0.30	0.18	-0.90	0.44	-39.90	0.003	0.003	0.003	0.003	0.004	0.005	0.032	0.000	0.001	0.001	0.001	0.000	0.002	0.009
CA04	-6.36	0.00	-0.98	3.40	3.46	-2.47	-80.88	0.025	0.025	0.024	0.017	0.023	0.028	0.040	0.008	0.009	0.008	0.006	0.008	0.009	0.012
CZ04	0.31	0.00	1.12	0.35	-1.95	1.70	-26.47	0.005	0.006	0.014	0.004	0.005	0.008	0.041	0.001	0.001	0.004	0.001	0.001	0.002	0.011
DK04	-87.80	0.00	-191.60	-156.30	3.20	0.60	-0.90	0.025	0.030	0.026	0.020	0.029	0.030	0.030	0.009	0.011	0.009	0.007	0.011	0.011	0.011
EE04	1.57	0.00	1.34	-1.64	-1.64	1.06	-15.92	0.017	0.017	0.010	0.034	0.019	0.014	0.023	0.005	0.005	0.003	0.011	0.006	0.004	0.005
FR05	-1.36	0.00	0.06	0.50	0.90	-1.10	-51.81	0.002	0.005	0.003	0.005	0.002	0.008	0.030	0.001	0.002	0.001	0.002	0.001	0.003	0.008
DE04	-1.86	0.00	0.43	0.05	1.15	-1.27	-69.61	0.005	0.008	0.006	0.010	0.006	0.012	0.038	0.002	0.003	0.002	0.004	0.002	0.004	0.011
GR04	0.65	0.00	-0.93	-0.45	0.22	-0.45	-16.23	0.005	0.004	0.004	0.011	0.004	0.005	0.023	0.001	0.000	0.000	0.003	0.000	0.001	0.006
HU05	0.16	0.00	0.07	-0.20	0.08	-0.20	-25.30	0.017	0.014	0.015	0.009	0.015	0.010	0.054	0.006	0.004	0.005	0.003	0.005	0.003	0.014
IE04	0.26	0.00	-0.71	-4.84	-0.66	0.14	-150.69	0.349	0.368	0.341	0.151	0.350	0.441	0.160	0.113	0.120	0.112	0.045	0.101	0.146	0.049
IL05	-10.59	0.00	-34.65	-37.29	0.14	0.14	1.45	0.016	0.028	0.019	0.020	0.031	0.031	0.035	0.005	0.010	0.007	0.002	0.011	0.011	0.012
IT04	1.56	0.00	-0.60	1.14	1.02	-1.48	-36.07	0.038	0.042	0.040	0.039	0.040	0.047	0.069	0.013	0.014	0.013	0.013	0.013	0.016	0.021
LU04	0.99	0.00	-3.07	-0.58	0.25	-0.31	-3.99	0.021	0.015	0.022	0.029	0.015	0.014	0.030	0.005	0.002	0.006	0.008	0.002	0.001	0.009
NL04	4.46	0.00	-1.01	0.33	-4.20	1.85	-80.67	0.028	0.029	0.031	0.027	0.027	0.031	0.053	0.010	0.010	0.011	0.010	0.010	0.011	0.016
SI04	0.79	0.00	0.54	-1.70	-1.04	0.54	-10.24	0.008	0.006	0.008	0.003	0.005	0.007	0.017	0.003	0.002	0.003	0.000	0.002	0.003	0.005
ES04	-4.53	0.00	-20.91	-25.67	0.14	0.02	-0.47	0.008	0.013	0.005	0.017	0.013	0.014	0.015	0.003	0.005	0.001	0.004	0.004	0.005	0.005
SE05	-1.02	0.00	-1.12	1.55	1.35	-1.37	-35.71	0.014	0.016	0.015	0.013	0.014	0.017	0.025	0.005	0.006	0.006	0.005	0.006	0.006	0.008
CH04	0.21	0.00	-0.33	-1.55	-0.20	0.06	-6.39	0.019	0.019	0.018	0.010	0.018	0.021	0.030	0.007	0.007	0.007	0.004	0.007	0.008	0.010
TW05	0.42	0.00	4.37	-4.53	-3.78	3.24	-41.59	0.004	0.004	0.004	0.014	0.005	0.003	0.028	0.001	0.000	0.001	0.004	0.001	0.000	0.008
UK04	2.39	0.00	-33.97	-29.69	-0.79	0.76	-59.00	0.044	0.051	0.031	0.018	0.050	0.052	0.090	0.015	0.017	0.011	0.006	0.017	0.018	0.029
Score based on across-the-board comparisons	10	0	1	1	5	2	2	3	1	2	12	1	2	0	1	2	4	10	2	1	1
Score based on pairwise comparisons	6	3	2	1	5	4	0	5	2	4	6	3	1	0	4	2	5	6	3	1	0

Table 3-6 Goodness-of-fit to individual datasets from wave 6 – equivalized gross income

Country & Year	Log-likelihood*							LRSSE							Absolute error of the Gini index						
	dPLN	GB2	EκG1	IEκG1	EκG2	IEκG2	GB1	dPLN	GB2	EκG1	IEκG1	EκG2	IEκG2	GB1	dPLN	GB2	EκG1	IEκG1	EκG2	IEκG2	GB1
AU03	-17.93	0.00	-69.42	-77.70	0.16	0.16	3.07	0.018	0.028	0.021	0.023	0.030	0.030	0.034	0.005	0.010	0.007	0.003	0.011	0.011	0.013
AT04	0.37	0.00	-0.27	0.03	-0.51	-0.05	-45.32	0.009	0.007	0.009	0.006	0.011	0.003	0.032	0.003	0.002	0.003	0.002	0.003	0.001	0.007
CA04	-20.16	0.00	-7.98	3.88	7.80	-2.71	-25.82	0.047	0.045	0.040	0.028	0.041	0.047	0.055	0.015	0.015	0.013	0.010	0.014	0.015	0.017
CZ04	1.15	0.00	0.10	-2.35	-0.86	0.64	-19.77	0.015	0.018	0.015	0.006	0.017	0.020	0.065	0.005	0.006	0.005	0.001	0.005	0.006	0.021
DK04	-66.80	0.00	-266.30	-202.60	11.00	10.90	13.10	0.026	0.031	0.026	0.018	0.032	0.032	0.032	0.009	0.011	0.009	0.006	0.011	0.011	0.011
EE04	1.57	0.00	-1.92	-5.23	-0.67	0.39	-5.01	0.005	0.007	0.021	0.044	0.007	0.006	0.023	0.000	0.001	0.005	0.013	0.001	0.001	0.007
FR05	-0.94	0.00	-0.29	0.28	0.56	-0.89	-55.06	0.002	0.005	0.004	0.005	0.002	0.009	0.045	0.001	0.002	0.002	0.002	0.001	0.003	0.013
DE04	-1.85	0.00	-14.68	-11.79	0.77	-0.45	-4.11	0.018	0.017	0.006	0.014	0.015	0.019	0.026	0.006	0.006	0.002	0.003	0.005	0.006	0.008
GR04	0.66	0.00	-1.24	-0.89	0.16	-0.37	-13.94	0.007	0.004	0.005	0.016	0.005	0.004	0.024	0.002	0.000	0.001	0.005	0.001	0.001	0.006
HU05	0.15	0.00	0.07	-0.20	0.08	-0.20	-24.77	0.017	0.014	0.015	0.009	0.015	0.010	0.051	0.006	0.004	0.005	0.003	0.005	0.003	0.013
IE04	1.63	0.00	0.26	-2.69	0.63	-3.72	-234.62	0.510	0.498	0.466	0.585	0.604	0.437	0.200	0.162	0.158	0.158	0.187	0.161	0.137	0.058
IL05	0.00	0.00	-47.95	-46.52	0.37	0.30	0.12	0.011	0.011	0.033	0.081	0.015	0.014	0.022	0.004	0.003	0.007	0.023	0.006	0.005	0.008
IT04	0.99	0.00	-4.50	1.25	1.74	-1.83	-42.06	0.050	0.060	0.054	0.042	0.055	0.067	0.126	0.016	0.019	0.017	0.014	0.018	0.021	0.039
LU04	0.02	0.00	-4.82	-3.23	0.01	0.01	-4.04	0.013	0.011	0.036	0.054	0.012	0.012	0.044	0.003	0.000	0.011	0.017	0.001	0.001	0.013
NL04	0.02	0.00	-3.07	0.09	1.05	-1.07	-21.40	0.034	0.037	0.036	0.027	0.035	0.040	0.053	0.012	0.013	0.013	0.010	0.012	0.014	0.017
SI04	0.79	0.00	0.54	-1.70	-1.04	0.54	-10.24	0.008	0.006	0.008	0.003	0.005	0.007	0.017	0.003	0.002	0.003	0.000	0.002	0.003	0.005
ES04	-4.59	0.00	-24.18	-28.42	0.14	0.08	-0.07	0.010	0.014	0.007	0.017	0.015	0.015	0.017	0.003	0.005	0.001	0.004	0.005	0.005	0.006
SE05	-1.34	0.00	1.15	0.14	1.60	-1.95	-81.12	0.013	0.015	0.013	0.016	0.013	0.017	0.035	0.005	0.005	0.005	0.006	0.005	0.006	0.010
CH04	-0.02	0.00	-8.14	-7.03	0.05	-0.04	-0.60	0.021	0.022	0.009	0.006	0.021	0.022	0.027	0.007	0.008	0.003	0.000	0.007	0.008	0.009
TW05	0.01	0.00	4.58	-3.27	-3.92	3.46	-47.76	0.008	0.007	0.003	0.018	0.009	0.005	0.030	0.002	0.002	0.000	0.006	0.003	0.001	0.008
UK04	-8.85	0.00	-187.69	-162.46	-0.01	-0.01	-91.30	0.044	0.042	0.013	0.033	0.042	0.042	0.083	0.016	0.014	0.001	0.008	0.015	0.015	0.028
Score based on across-the-board comparisons	8	1	1	0	9	0	2	3	3	4	8	1	1	1	2	3	4	9	1	1	1
Score based on pairwise comparisons	5	4	1	2	6	3	0	3	4	6	5	2	1	0	4	3	6	5	2	1	0

Table 4-1 Summary of pairwise comparisons on the goodness-of-fit to datasets from waves 4-6 – log-likelihood

Cate-gories	PIDM	No. of categories gained (NoCG)							Ratio of datasets gained (RoDG)						
		dPLN	GB2	EκG1	IEκG1	EκG2	IEκG2	GB1	dPLN	GB2	EκG1	IEκG1	EκG2	IEκG2	GB1
Non-equivalized	dPLN		9	0	0	9	9	5		75.0	31.1	26.7	85.0	83.3	47.8
	GB2	0		0	0	9	6	1	25.0		18.3	16.1	70.3	54.4	34.4
	EκG1	9	9		0	9	9	7	68.9	81.7		37.8	86.9	82.5	62.2
	IEκG1	9	9	9		9	9	6	73.3	83.9	62.2		81.7	86.4	61.1
	EκG2	0	0	0	0		1.5	1	15.0	29.7	13.1	18.3		41.4	31.9
	IEκG2	0	3	0	0	7.5		1	16.7	45.6	17.5	13.6	58.6		33.1
	GB1	4	8	2	3	8	8		52.2	65.6	37.8	38.9	68.1	66.9	
Equivalized	dPLN		5.5	0	0	7	5	0		56.7	25.6	30.6	57.2	54.4	13.9
	GB2	3.5		0	0	6	7.5	0	43.3		27.2	30.3	56.1	57.2	8.3
	EκG1	9	9		5	9	9	0	74.4	72.8		50.6	69.4	75.0	31.1
	IEκG1	9	9	4		9	9	0	69.4	69.7	49.4		74.7	71.9	30.0
	EκG2	2	3	0	0		4.5	0	42.8	43.9	30.6	25.3		50.0	7.8
	IEκG2	4	1.5	0	0	4.5		0	45.6	42.8	25.0	28.1	50.0		7.8
	GB1	9	9	9	9	9	9		86.1	91.7	68.9	70.0	92.2	92.2	

Table 4-2 Summary of pairwise comparisons on the goodness-of-fit to datasets from waves 4-6 – LRSSE

Cate-gories	PIDM	No. of categories gained (NoCG)							Ratio of datasets gained (RoDG)						
		dPLN	GB2	EκG1	IEκG1	EκG2	IEκG2	GB1	dPLN	GB2	EκG1	IEκG1	EκG2	IEκG2	GB1
Non-equivalized	dPLN		3	3.5	1	7	2	0		45.0	51.7	40.0	55.6	41.7	25.6
	GB2	6		5.5	3	9	2	0	55.0		56.1	41.1	58.9	36.1	14.4
	EκG1	5.5	3.5		0	3	4.5	0	48.3	43.9		36.7	47.2	44.4	31.1
	IEκG1	8	6	9		6	7	5.5	60.0	58.9	63.3		60.6	59.4	50.6
	EκG2	2	0	6	3		1	0	44.4	41.1	52.8	39.4		37.2	15.0
	IEκG2	7	7	4.5	2	8		0	58.3	63.9	55.6	40.6	62.8		13.9
	GB1	9	9	9	3.5	9	9		74.4	85.6	68.9	49.4	85.0	86.1	
Equivalized	dPLN		2	5.5	2	4.5	4	0		46.1	50.6	40.6	51.1	46.1	12.8
	GB2	7		7	3	7	0.5	0	53.9		55.0	43.9	55.6	33.9	9.4
	EκG1	3.5	2		1	4.5	2.5	0	49.4	45.0		38.3	47.8	42.8	18.3
	IEκG1	7	6	8		5.5	5	2	59.4	56.1	61.7		55.6	53.3	29.4
	EκG2	4.5	2	4.5	3.5		1.5	0	48.9	44.4	52.2	44.4		40.6	10.6
	IEκG2	5	8.5	6.5	4	7.5		0	53.9	66.1	57.2	46.7	59.4		9.4
	GB1	9	9	9	7	9	9		87.2	90.6	81.7	70.6	89.4	90.6	

Table 4-3 Summary of pairwise comparisons on the goodness-of-fit to datasets from waves 4-6 – absolute error of the Gini index

Cate-gories	PIDM	No. of categories gained (NCaG)							Percentage of countries gained (PCoG)						
		dPLN	GB2	EκG1	IEκG1	EκG2	IEκG2	GB1	dPLN	GB2	EκG1	IEκG1	EκG2	IEκG2	GB1
Non-equivalized	dPLN		1	2	2	1.5	1	0		35.0	42.2	43.3	42.2	31.1	22.8
	GB2	8		8	4	9	2	0	65.0		60.6	48.3	59.4	37.2	21.7
	EκG1	7	1		3	4	2.5	1	57.8	39.4		43.3	45.0	41.7	33.3
	IEκG1	7	5	6		6	5	4	56.7	51.7	56.7		52.2	52.2	47.8
	EκG2	7.5	0	5	3		3	0	57.8	40.6	55.0	47.8		40.6	18.9
	IEκG2	8	7	6.5	4	6		0	68.9	62.8	58.3	47.8	59.4		20.6
	GB1	9	9	8	5	9	9		77.2	78.3	66.7	52.2	81.1	79.4	
Equivalized	dPLN		1	4	4.5	3.5	3	0		43.9	50.0	45.6	51.7	43.9	16.7
	GB2	8		8	5	8	0	0	56.1		57.2	48.3	57.8	31.7	13.9
	EκG1	5	1		1	5	1.5	0	50.0	42.8		45.0	50.6	40.0	19.4
	IEκG1	4.5	4	8		4.5	3	2	54.4	51.7	55.0		51.1	45.0	28.9
	EκG2	5.5	1	4	4.5		0	0	48.3	42.2	49.4	48.9		36.1	14.4
	IEκG2	6	9	7.5	6	9		0	56.1	68.3	60.0	55.0	63.9		12.8
	GB1	9	9	9	7	9			83.3	86.1	80.6	71.1	85.6	87.2	

Table 5-1 Detailed summary of the pairwise goodness-of-fit comparisons – log-likelihood

Categories	No. of cate-gories	dPLN vs. GB2		dPLN vs. EκG1		dPLN vs. EκG2		GB2 vs. EκG1		GB2 vs. EκG2		EκG1 vs. EκG2	
		NoCG	RoDG	NoCG	RoDG	NoCG	RoDG	NoCG	RoDG	NoCG	RoDG	NoCG	RoDG
All categories	18	14.5	65.8	0	28.3	16	71.1	0	22.8	15	63.2	18	78.2
Non-equiv.	9	9	75.0	0	31.1	9	85.0	0	18.3	9	70.3	9	86.9
Consumption	3	3	74.1	0	34.5	3	77.6	0	25.9	3	62.9	3	80.2
Disp. income	3	3	74.6	0	31.1	3	88.5	0	13.1	3	70.5	3	90.2
Gross income	3	3	76.2	0	27.9	3	88.5	0	16.4	3	77.0	3	90.2
Equivalized	9	5.5	56.7	0	25.6	7	57.2	0	27.2	6	56.1	9	69.4
Consumption	3	3	66.4	0	39.7	3	62.1	0	32.8	1	50.9	3	62.1
Disp. income	3	0.5	45.1	0	18.0	1	44.3	0	27.9	2	51.6	3	67.2
Gross income	3	2	59.0	0	19.7	3	65.6	0	21.3	3	65.6	3	78.7

Table 5-2 Detailed summary of the pairwise goodness-of-fit comparisons – LRSSE

Categories	No. of cate-gories	dPLN vs. GB2		dPLN vs. EκG1		dPLN vs. EκG2		GB2 vs. EκG1		GB2 vs. EκG2		EκG1 vs. EκG2	
		NoCG	RoDG	NoCG	RoDG	NoCG	RoDG	NoCG	RoDG	NoCG	RoDG	NoCG	RoDG
All categories	18	5	45.6	9	51.1	11.5	53.3	12.5	55.6	16	57.2	7.5	47.5
Non-equiv.	9	3	45.0	3.5	51.7	7	55.6	5.5	56.1	9	58.9	3	47.2
Consumption	3	2	48.3	1.5	51.7	3	63.8	1.5	50.0	3	56.9	2	55.2
Disp. income	3	0	42.6	1	55.7	1	49.2	3	65.6	3	57.4	0	41.0
Gross income	3	1	44.3	1	47.5	3	54.1	1	52.5	3	62.3	1	45.9
Equivalized	9	2	46.1	5.5	50.6	4.5	51.1	7	55.0	7	55.6	4.5	47.8
Consumption	3	1	48.3	1.5	46.6	2.5	55.2	2	48.3	3	55.2	2.5	53.4
Disp. income	3	0	42.6	2	52.5	1	45.9	3	59.0	2	55.7	1	45.9
Gross income	3	1	47.5	2	52.5	1	52.5	2	57.4	2	55.7	1	44.3

Table 5-3 Detailed summary of the pairwise goodness-of-fit comparisons – LwRSSE

Categories	No. of cate-gories	dPLN vs. GB2		dPLN vs. EκG1		dPLN vs. EκG2		GB2 vs. EκG1		GB2 vs. EκG2		EκG1 vs. EκG2	
		NoCG	RoDG	NoCG	RoDG	NoCG	RoDG	NoCG	RoDG	NoCG	RoDG	NoCG	RoDG
All categories	18	6.5	42.5	8.5	51.9	10	51.1	14	57.5	14	55.6	8	46.1
Non-equiv.	9	3.5	41.1	3.5	51.1	6	50.6	7	58.9	8	57.2	4	44.4
Consumption	3	1.5	50.0	1.5	50.0	3	58.6	2	53.4	3	58.6	3	55.2
Disp. income	3	1	36.1	1	55.7	2	47.5	3	65.6	2	55.7	0	37.7
Gross income	3	1	37.7	1	47.5	1	45.9	2	57.4	3	57.4	1	41.0
Equivalized	9	3	43.9	5	52.8	4	51.7	7	56.1	6	53.9	4	47.8
Consumption	3	1	46.6	1	48.3	2	53.4	2	48.3	2	51.7	2	51.7
Disp. income	3	1	42.6	2	57.4	1	50.8	2	60.7	2	55.7	1	47.5
Gross income	3	1	42.6	2	52.5	1	50.8	3	59.0	2	54.1	1	44.3

Table 5-4 Detailed summary of the pairwise goodness-of-fit comparisons – absolute error of the Gini index

Categories	No. of cate-gories	dPLN vs. GB2		dPLN vs. EκG1		dPLN vs. EκG2		GB2 vs. EκG1		GB2 vs. EκG2		EκG1 vs. EκG2	
		NoCG	RoDG	NoCG	RoDG	NoCG	RoDG	NoCG	RoDG	NoCG	RoDG	NoCG	RoDG
All categories	18	2	39.4	6	46.1	5	46.9	16	58.9	17	58.6	9	47.8
Non-equiv.	9	1	35.0	2	42.2	1.5	42.2	8	60.6	9	59.4	4	45.0
Consumption	3	1	44.8	1	44.8	1.5	48.3	3	58.6	3	56.9	3	53.4
Disp. income	3	0	27.9	0	42.6	0	37.7	3	67.2	3	62.3	0	41.0
Gross income	3	0	32.8	1	39.3	0	41.0	2	55.7	3	59.0	1	41.0
Equivalized	9	1	43.9	4	50.0	3.5	51.7	8	57.2	8	57.8	5	50.6
Consumption	3	1	48.3	0	41.4	1.5	55.2	2	50.0	3	55.2	3	60.3
Disp. income	3	0	39.3	2	54.1	1	47.5	3	62.3	3	60.7	1	50.8
Gross income	3	0	44.3	2	54.1	1	52.5	3	59.0	2	57.4	1	41.0

Table 6-1 Detailed summary of the pairwise goodness-of-fit comparisons using combined measures – log-likelihood & LRSSE

Categories	No. of cate-gories	dPLN vs. GB2			dPLN vs. EκG1			dPLN vs. EκG2			GB2 vs. EκG1			GB2 vs. EκG2			EκG1 vs. EκG2		
		NoCG	VR	RoDG	NoCG	VR	RoDG	NoCG	VR	RoDG	NoCG	VR	RoDG	NoCG	VR	RoDG	NoCG	VR	RoDG
All categories	18	13.5	51.1	61.4	0.5	56.7	31.9	14.5	54.4	72.4	0	47.8	27.3	16	61.1	67.3	17	48.9	73.3
Non-equivalized	9	9	50.6	70.3	0.5	58.3	35.2	9	55.0	86.9	0	47.8	23.3	8.5	63.9	73.0	9	49.4	77.5
Consumption	3	3	53.4	71.0	0	58.6	38.2	3	55.2	87.5	0	58.6	29.4	3	63.8	64.9	3	58.6	73.5
Disp. income	3	3	47.5	69.0	0.5	55.7	38.2	3	54.1	84.8	0	37.7	21.7	2.5	62.3	71.1	3	41.0	76.0
Gross income	3	3	50.8	71.0	0	60.7	29.7	3	55.7	88.2	0	47.5	17.2	3	65.6	82.5	3	49.2	83.3
Equivalized	9	4.5	51.7	52.7	0	55.0	28.3	5.5	53.9	57.7	0	47.8	31.4	7.5	58.3	61.0	8	48.3	69.0
Consumption	3	2.5	65.5	60.5	0	55.2	37.5	3	62.1	63.9	0	60.3	34.3	2	62.1	55.6	3	51.7	76.7
Disp. income	3	0.5	41.0	36.0	0	52.5	21.9	1	45.9	39.3	0	36.1	31.8	2.5	52.5	59.4	2	47.5	55.2
Gross income	3	1.5	49.2	56.7	0	57.4	25.7	1.5	54.1	66.7	0	47.5	27.6	3	60.7	67.6	3	45.9	75.0

Table 6-2 Detailed summary of the pairwise goodness-of-fit comparisons using combined measures – log-likelihood & absolute error of the Gini index

Categories	No. of cate-gories	dPLN vs. GB2			dPLN vs. EκG1			dPLN vs. EκG2			GB2 vs. EκG1			GB2 vs. EκG2			EκG1 vs. EκG2		
		NoCG	VR	RoDG	NoCG	VR	RoDG	NoCG	VR	RoDG	NoCG	VR	RoDG	NoCG	VR	RoDG	NoCG	VR	RoDG
All categories	18	11.5	45.6	56.1	0.5	55.0	26.8	15	43.1	71.0	1	43.9	29.1	17.5	60.8	68.5	18	44.7	78.9
Non-equivalized	9	7	39.4	63.4	0.5	55.6	26.0	9	39.4	84.5	0	42.2	25.0	9	64.4	73.3	9	42.8	87.0
Consumption	3	3	46.6	70.4	0.5	51.7	30.0	3	39.7	82.6	0	50.0	34.5	3	62.1	66.7	3	56.9	78.8
Disp. income	3	2	32.8	55.0	0	59.0	27.8	3	39.3	83.3	0	36.1	22.7	3	65.6	75.0	3	37.7	91.3
Gross income	3	2	39.3	62.5	0	55.7	20.6	3	39.3	87.5	0	41.0	16.0	3	65.6	77.5	3	34.4	95.2
Equivalized	9	4.5	51.7	50.5	0	54.4	27.6	6	46.7	59.5	1	45.6	32.9	8.5	57.2	63.1	9	46.7	71.4
Consumption	3	3	65.5	60.5	0	60.3	34.3	3	55.2	65.6	0.5	62.1	36.1	2.5	65.5	55.3	3	56.9	69.7
Disp. income	3	0.5	44.3	33.3	0	47.5	20.7	1	37.7	39.1	0.5	32.8	35.0	3	50.8	64.5	3	41.0	72.0
Gross income	3	1	45.9	53.6	0	55.7	26.5	2	47.5	69.0	0	42.6	26.9	3	55.7	70.6	3	42.6	73.1

Table 6-3 Detailed summary of the pairwise goodness-of-fit comparisons using combined measures – log-likelihood & absolute error of the MLD

Categories	No. of cate-gories	dPLN vs. GB2			dPLN vs. EκG1			dPLN vs. EκG2			GB2 vs. EκG1			GB2 vs. EκG2			EκG1 vs. EκG2		
		NoCG	VR	RoDG	NoCG	VR	RoDG	NoCG	VR	RoDG	NoCG	VR	RoDG	NoCG	VR	RoDG	NoCG	VR	RoDG
All categories	18	8.5	45.3	46.6	1	53.1	28.8	12	44.7	62.1	2.5	39.4	32.4	16.5	55.8	70.1	17	34.4	76.6
Non-equivalized	9	5	40.6	53.4	1	52.8	32.6	9	40.6	80.8	1	37.2	28.4	9	60.6	69.7	9	35.6	82.8
Consumption	3	2.5	51.7	63.3	0	62.1	27.8	3	46.6	85.2	0	55.2	28.1	3	62.1	61.1	3	58.6	82.4
Disp. income	3	1.5	32.8	45.0	1	44.3	48.1	3	36.1	77.3	1	27.9	35.3	3	60.7	75.7	3	24.6	80.0
Gross income	3	1	37.7	47.8	0	52.5	25.0	3	39.3	79.2	0	29.5	22.2	3	59.0	72.2	3	24.6	86.7
Equivalized	9	3.5	50.0	41.1	0	53.3	25.0	3	48.9	46.6	1.5	41.7	36.0	7.5	51.1	70.7	8	33.3	70.0
Consumption	3	2.5	62.1	61.1	0	67.2	33.3	2	58.6	58.8	0	60.3	31.4	1.5	56.9	48.5	2.5	50.0	69.0
Disp. income	3	0	47.5	13.8	0	47.5	17.2	0	47.5	24.1	0.5	31.1	42.1	3	44.3	85.2	3	23.0	71.4
Gross income	3	1	41.0	44.0	0	45.9	21.4	1	41.0	56.0	1	34.4	38.1	3	52.5	81.3	2.5	27.9	70.6

Table 6-4 Detailed summary of the pairwise goodness-of-fit comparisons using combined measures – log-likelihood & absolute error of the Theil index

Categories	No. of cate-gories	dPLN vs. GB2			dPLN vs. EκG1			dPLN vs. EκG2			GB2 vs. EκG1			GB2 vs. EκG2			EκG1 vs. EκG2		
		NoCG	VR	RoDG	NoCG	VR	RoDG	NoCG	VR	RoDG	NoCG	VR	RoDG	NoCG	VR	RoDG	NoCG	VR	RoDG
All categories	18	11.5	46.7	54.8	0	53.3	29.7	14	47.2	67.6	0	43.1	30.3	16	59.7	66.5	17.5	42.8	74.0
Non-equivalized	9	6	42.2	57.9	0	53.3	28.1	9	40.0	81.9	0	38.9	25.7	9	63.9	72.2	9	37.8	85.3
Consumption	3	3	53.4	67.7	0	51.7	26.7	3	41.4	83.3	0	50.0	31.0	3	63.8	64.9	3	53.4	77.4
Disp. income	3	0.5	36.1	45.5	0	54.1	33.3	3	41.0	80.0	0	31.1	26.3	3	63.9	74.4	3	29.5	88.9
Gross income	3	2.5	37.7	56.5	0	54.1	24.2	3	37.7	82.6	0	36.1	18.2	3	63.9	76.9	3	31.1	94.7
Equivalized	9	5.5	51.1	52.2	0	53.3	31.3	5	54.4	57.1	0	47.2	34.1	7	55.6	60.0	8.5	47.8	65.1
Consumption	3	3	62.1	61.1	0	67.2	35.9	2	62.1	61.1	0	62.1	36.1	1.5	58.6	52.9	3	53.4	64.5
Disp. income	3	1	44.3	37.0	0	42.6	26.9	1	49.2	40.0	0	36.1	36.4	2.5	52.5	59.4	2.5	47.5	62.1
Gross income	3	1.5	47.5	55.2	0	50.8	29.0	2	52.5	68.8	0	44.3	29.6	3	55.7	67.6	3	42.6	69.2

Table 6-5 Detailed summary of the pairwise goodness-of-fit comparisons using combined measures –four FB measures & LRSSE

Categories	No. of cate-gories	dPLN vs. GB2			dPLN vs. EκG1			dPLN vs. EκG2			GB2 vs. EκG1			GB2 vs. EκG2			EκG1 vs. EκG2		
		NoCG	VR	RoDG	NoCG	VR	RoDG	NoCG	VR	RoDG	NoCG	VR	RoDG	NoCG	VR	RoDG	NoCG	VR	RoDG
All categories	18	14	36.7	64.4	1	49.2	33.9	14	42.8	71.4	1	37.8	23.5	14.5	38.6	68.3	18	36.7	81.1
Non-equivalized	9	9	37.8	79.4	0	51.7	34.4	9	44.4	88.8	0	43.9	20.3	8	33.9	78.7	9	36.1	93.8
Consumption	3	3	48.3	67.9	0	44.8	34.6	3	43.1	92.0	0	51.7	26.7	2.5	32.8	78.9	3	41.4	87.5
Disp. income	3	3	32.8	90.0	0	52.5	37.5	3	44.3	85.2	0	36.1	18.2	2.5	36.1	77.3	3	34.4	95.2
Gross income	3	3	32.8	85.0	0	57.4	31.4	3	45.9	89.3	0	44.3	14.8	3	32.8	80.0	3	32.8	100.0
Equivalized	9	5	35.6	48.4	1	46.7	33.3	5	41.1	52.7	1	31.7	28.1	6.5	43.3	60.3	9	37.2	68.7
Consumption	3	3	43.1	64.0	1	43.1	48.0	2.5	44.8	57.7	1	32.8	36.8	1.5	43.1	52.0	3	39.7	65.2
Disp. income	3	0	29.5	22.2	0	44.3	25.9	1	37.7	39.1	0	24.6	26.7	2	42.6	57.7	3	34.4	61.9
Gross income	3	2	34.4	52.4	0	52.5	28.1	1.5	41.0	60.0	0	37.7	21.7	3	44.3	70.4	3	37.7	78.3

Table 6-6 Detailed summary of the pairwise goodness-of-fit comparisons using combined measures – four FB measures & absoluter error of the Gini index (AEG)

Categories	No. of cate-gories	dPLN vs. GB2			dPLN vs. EκG1			dPLN vs. EκG2			GB2 vs. EκG1			GB2 vs. EκG2			EκG1 vs. EκG2		
		NoCG	VR	RoDG	NoCG	VR	RoDG	NoCG	VR	RoDG	NoCG	VR	RoDG	NoCG	VR	RoDG	NoCG	VR	RoDG
All categories	18	12.5	30.8	57.7	1	46.9	29.6	12.5	31.7	64.0	1.5	32.5	25.6	15	37.2	68.7	17	32.5	82.1
Non-equivalized	9	8	26.7	72.9	0.5	47.2	27.1	9	29.4	83.0	0	36.7	22.7	8	32.8	78.0	9	31.1	92.9
Consumption	3	3	41.4	66.7	0.5	41.4	29.2	3	29.3	88.2	0	43.1	32.0	2.5	31.0	77.8	3	37.9	86.4
Disp. income	3	3	18.0	81.8	0	50.8	29.0	3	29.5	77.8	0	31.1	21.1	2.5	36.1	77.3	3	29.5	94.4
Gross income	3	2	21.3	76.9	0	49.2	23.3	3	29.5	83.3	0	36.1	13.6	3	31.1	78.9	3	26.2	100.0
Equivalized	9	4.5	35.0	46.0	0.5	46.7	32.1	3.5	33.9	47.5	1.5	28.3	29.4	7	41.7	61.3	8	33.9	72.1
Consumption	3	3	43.1	64.0	0.5	46.6	44.4	1	36.2	52.4	1	31.0	38.9	1.5	44.8	50.0	2.5	39.7	69.6
Disp. income	3	0	31.1	21.1	0	41.0	24.0	1	31.1	31.6	0.5	21.3	30.8	2.5	39.3	62.5	2.5	27.9	70.6
Gross income	3	1.5	31.1	47.4	0	52.5	28.1	1.5	34.4	57.1	0	32.8	20.0	3	41.0	72.0	3	34.4	76.2

Appendix 1. An estimation procedure for the maximum likelihood parameters of the EκG2

The logarithm of the density function $f_{\text{EκG2}}(x; a, b, p, q)$ equals

$$\log f_{\text{EκG2}} = \log\left(\frac{a}{b}\right) - \log B(p,q) + \left(p - \frac{1}{a}\right)\log z + \left(q + \frac{1}{2a}\right)\log(1-z) - \log\left(1 - \frac{1}{2}z\right),$$

where $z = z(x; a, b) := \frac{2}{\sqrt{1+4\left(\frac{x}{b}\right)^{-2a}}+1} = \left(\frac{x}{b}\right)^a\left[\sqrt{1+\frac{1}{4}\left(\frac{x}{b}\right)^{2a}} - \frac{1}{2}\left(\frac{x}{b}\right)^a\right]$. Its partial derivatives with respect to parameters equal

$$\begin{aligned}
\frac{\partial \log f_{\text{EκG2}}}{\partial a} &= \frac{1}{a} + \left[\left(p-\frac{1}{a}\right)\frac{1}{z} - \left(q+\frac{1}{2a}\right)\frac{1}{1-z} + \frac{1}{2-z}\right]\frac{\partial z}{\partial a} + \frac{1}{a^2}\log z - \frac{1}{2a^2}\log(1-z),\\
\frac{\partial \log f_{\text{EκG2}}}{\partial b} &= -\frac{1}{b} + \left[\left(p-\frac{1}{a}\right)\frac{1}{z} - \left(q+\frac{1}{2a}\right)\frac{1}{1-z} + \frac{1}{2-z}\right]\frac{\partial z}{\partial b},\\
\frac{\partial \log f_{\text{EκG2}}}{\partial p} &= \log z - \psi(p) + \psi(p+q),\\
\frac{\partial \log f_{\text{EκG2}}}{\partial q} &= \log(1-z) - \psi(q) + \psi(p+q),
\end{aligned} \tag{A1}$$

where $\frac{\partial z}{\partial a} = -\frac{1}{a}\frac{z(1-z)}{2-z}\log\frac{1-z}{z^2}$; $\frac{\partial z}{\partial b} = -2\frac{a}{b}\frac{z(1-z)}{2-z}$; $\psi(\cdot)$ denotes the digamma function. The log-likelihood of an I.I.D. sample of size n from an EκG2 distribution equals

$$\log L = n\log\left(\frac{a}{b}\right) - n\log B(p,q) + \left(p-\frac{1}{a}\right)\sum_{i=1}^{n}\log z_i + \left(q+\frac{1}{2a}\right)\sum_{i=1}^{n}\log(1-z_i) - \sum_{i=1}^{n}\log\left(1-\frac{1}{2}z_i\right),$$

where $z_i = z(x_i; a, b)$. Because the partial derivatives of the log-likelihood with respect to parameters should be zero at the maximum likelihood parameters, the following four simultaneous equations are obtained:

$$\begin{aligned}
&p\sum_{i=1}^{n}\frac{1-z_i}{2-z_i}\log\frac{z_i^2}{1-z_i} - q\sum_{i=1}^{n}\frac{z_i}{2-z_i}\log\frac{z_i^2}{1-z_i} = -n - \sum_{i=1}^{n}\frac{z_i(1-z_i)}{(2-z_i)^2}\log\frac{z_i^2}{1-z_i},\\
&p\sum_{i=1}^{n}\frac{1-z_i}{2-z_i} - q\sum_{i=1}^{n}\frac{z_i}{2-z_i} = -\sum_{i=1}^{n}\frac{z_i(1-z_i)}{(2-z_i)^2},\\
&n\psi(p) - n\psi(p+q) = \sum_{i=1}^{n}\log z_i,\\
&n\psi(q) - n\psi(p+q) = \sum_{i=1}^{n}\log(1-z_i).
\end{aligned}$$

Similarly to the MLE procedure for the GB2 provided by Venter [20], because the first and second equations are linear in p and q, the two simultaneous equations can be easily solved to represent p and q as functions of a and b. Thus, the four-parameter MLE procedure reduces to two simultaneous nonlinear equations with two unknowns.

Appendix 2. Proof of the regularity of the EκG2 in terms of maximum likelihood estimation

The EκG2 satisfies the following regularity conditions in terms of MLE:

Regularity Conditions (*cf.* Serfling [18]): Assume that a family of distributions has a density function $f(x; \boldsymbol{\theta})$ with parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_m)$. Consider $\boldsymbol{\theta}$ to be in an open set (not necessarily finite) $\boldsymbol{\Theta}$ of R^m.

(R1) For each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the derivatives

$$\frac{\partial \log f}{\partial \theta_i}, \quad \frac{\partial^2 \log f}{\partial \theta_i \partial \theta_j}, \quad \frac{\partial^3 \log f}{\partial \theta_i \partial \theta_j \partial \theta_k} \quad (i, j, k = 1, 2, \cdots, m)$$

exist, all x;

(R2) For each $\boldsymbol{\theta_0} \in \boldsymbol{\Theta}$, there exist functions $g(x)$, $h(x)$, $H(x)$ (possibly depending on $\boldsymbol{\theta_0}$) such that, for $\boldsymbol{\theta}$ in a neighborhood $\mathrm{N}(\boldsymbol{\theta_0})$, the relations

$$\left|\frac{\partial f}{\partial \theta_i}\right| \le g(x), \quad \left|\frac{\partial^2 f}{\partial \theta_i \partial \theta_j}\right| \le h(x), \quad \left|\frac{\partial^3 \log f}{\partial \theta_i \partial \theta_j \partial \theta_k}\right| \le H(x) \quad (i, j, k = 1, 2, \cdots, m)$$

hold, all x, and

$$\int g(x) dx < \infty, \quad \int h(x) dx < \infty, \quad E_{\boldsymbol{\theta}}[H(x)] < \infty;$$

(R3) For each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the following Fisher information matrix is finite and positive definite:

$$\left[\mathrm{E}_{\boldsymbol{\theta}}\left(\frac{\partial \log f}{\partial \theta_i}\frac{\partial \log f}{\partial \theta_j}\right)\right].$$

The parameters (a, b, p, q) of the EκG2 can hereafter be regarded in the same light as $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)$. The parameter space $\boldsymbol{\Theta}$ for the EκG2 is defined as the Cartesian product of four positive half intervals, i.e., $\boldsymbol{\Theta} \coloneqq (0, \infty) \times (0, \infty) \times (0, \infty) \times (0, \infty)$.

(R1) is obviously satisfied.

Proof of (R2):

$\frac{\partial f_{\text{EκG2}}}{\partial \theta_i}$ and $\frac{\partial^2 f_{\text{EκG2}}}{\partial \theta_i \partial \theta_j}$ can be expressed as sums of terms in the following form:

$$c(\boldsymbol{\theta})[\log z(x; a, b)]^k \{\log[1 - z(x; a, b)]\}^l \frac{[z(x; a, b)]^{p+\nu-\frac{1}{a}}[1 - z(x; a, b)]^{q+\xi+\frac{1}{2a}}}{[2 - z(x; a, b)]^m}, \tag{A2}$$

where $c(\boldsymbol{\theta})$ is continuous on $\boldsymbol{\Theta}$, and $k, l, m, \nu, \xi = 0, 1, 2, \cdots$. Take a bonded neighborhood $\mathrm{N}(\boldsymbol{\theta}_0)$ such that $\inf_{\boldsymbol{\theta} \in \mathrm{N}(\boldsymbol{\theta_0})} \theta_i > 0$ and $\sup_{\boldsymbol{\theta} \in \mathrm{N}(\boldsymbol{\theta_0})} \theta_i < \infty$ for $i = 1, 2, \cdots, 4$. Let $\underline{a}$, $\underline{b}$, $\underline{p}$, $\underline{q}$ denote the infimums and $\overline{a}$, $\overline{b}$, $\overline{p}$, $\overline{q}$ denote the supremums of the parameters in $\mathrm{N}(\boldsymbol{\theta}_0)$. Because a positive value $K (> 1)$ exist such that $K^{-1} \le \frac{z(x; a', b')^{1/a'}}{z(x; a, b)^{1/a}} \le K$ and $K^{-1} \le \frac{[1 - z(x; a', b')]^{1/a'}}{[1 - z(x; a, b)]^{1/a}} \le K$ for $0 < \forall x < \infty$, $\underline{a} \le \forall a, a' \le \overline{a}$ and $\underline{b} \le \forall b, b' \le \overline{b}$, the absolute value of the term in (A2) is bounded by

$$\varphi(x) \coloneqq M K^{\overline{a}(\overline{p}+\overline{q}+\nu+\xi)-\frac{1}{2}} \left|\log z(x; \underline{a}, \overline{b})\right|^k \left|\frac{\overline{a}}{\underline{a}} \log[1 - z(x; \underline{a}, \overline{b})] - \overline{a} \log K\right|^l$$

$$\cdot \left[z(x; \underline{a}, \overline{b})\right]^{\underline{p}+\nu-\frac{1}{\overline{a}}} \left[1 - z(x; \underline{a}, \overline{b})\right]^{\underline{q}+\xi+\frac{1}{2\underline{a}}}$$

for $0 < \forall x < \infty$ and $\forall \boldsymbol{\theta} \in \mathrm{N}(\boldsymbol{\Theta}_0)$, where $M = \sup_{\boldsymbol{\theta} \in \mathrm{N}(\boldsymbol{\Theta}_0)} c(\boldsymbol{\theta})$. The function $\varphi(x)$ is integrable, as follows:

$$\int_0^\infty \varphi(x) dx = \int_0^1 \varphi[x^{-1}(z)] \frac{\partial x}{\partial z} dz =$$

$$\frac{\overline{b}}{2\underline{a}} M K^{\overline{a}(\overline{p}+\overline{q}+\nu+\xi)-1/2} \int_0^1 |\log z|^k \left|\frac{\overline{a}}{\underline{a}} \log(1 - z) - \overline{a} \log K\right|^l z^{\underline{p}+\nu-1} (1 - z)^{\underline{q}+\xi-1} (2 - z) dz < \infty.$$

Note that $\frac{\partial}{\partial x} z(x; a, b) = 2 \frac{a}{b} \frac{z^{1-\frac{1}{a}}(1-z)^{1+\frac{1}{2a}}}{2-z}$. $g(x)$ and $h(x)$ are obtained by summing up the functions $\varphi(x)$s corresponding to the terms that constitute $\frac{\partial f_{\mathrm{EKG2}}}{\partial \theta_i}$ and $\frac{\partial^2 f_{\mathrm{EKG2}}}{\partial \theta_i \partial \theta_j}$, respectively.

$\frac{\partial^3 \log f_{\mathrm{EKG2}}}{\partial \theta_i \partial \theta_j \partial \theta_k}$ can be expressed as a sum of terms in the following form:

$$c(\boldsymbol{\theta})[\log z(x; a, b)]^k \{\log[1 - z(x; a, b)]\}^l \frac{[z(x; a, b)]^\nu [1 - z(x; a, b)]^\xi}{[2 - z(x; a, b)]^m}, \tag{A3}$$

where $c(\boldsymbol{\theta})$ is continuous on $\boldsymbol{\Theta}$, and $k, l, m, \nu, \xi = 0, 1, 2, \cdots$. In a manner similar to that applied above, it can be shown that the absolute value of the term in (A3) is bounded by

$$\phi(x) \coloneqq$$

$$MK^{\overline{a}(\nu+\xi)} \left|\log z\left(x; \underline{a}, \overline{b}\right)\right|^k \left|\frac{\overline{a}}{\underline{a}} \log\left[1 - z\left(x; \underline{a}, \overline{b}\right)\right] - \overline{a} \log K\right|^l \left[z\left(x; \underline{a}, \overline{b}\right)\right]^\nu \left[1 - z\left(x; \underline{a}, \overline{b}\right)\right]^\xi$$

for $0 < \forall x < \infty$ and $\forall \boldsymbol{\theta} \in \mathrm{N}(\boldsymbol{\Theta}_0)$, where $M = \sup_{\boldsymbol{\theta} \in \mathrm{N}(\boldsymbol{\Theta}_0)} c(\boldsymbol{\theta})$. The function $\phi(x)$ has a finite expectation as follows:

$$E_{\boldsymbol{\theta}}[\phi(x)] = \int_0^\infty \phi(x) f_{\mathrm{EKG2}}(x; a, b, p, q) dx$$

$$\leq \frac{1}{B(p, q)} MK^{(\overline{a}+\underline{a})(\nu+\xi)} \int_0^1 \left|\frac{\underline{a}}{\overline{a}} \log z - \underline{a} \log K\right|^k \left|\frac{\overline{a}}{\underline{a}} \log(1 - z) - 2\overline{a} \log K\right|^l z^{p+\frac{\underline{a}}{\overline{a}}\nu - 1} (1 - z)^{q+\frac{\underline{a}}{\overline{a}}\xi - 1} dz < \infty.$$

$H(x)$ is obtained by summing up the functions $\phi(x)$s corresponding to the terms that constitute $\frac{\partial^3 \log f_{\mathrm{EKG2}}}{\partial \theta_i \partial \theta_j \partial \theta_k}$.

Proof of (R3):

Positive definiteness is equivalent to non-existence of some real vector $\boldsymbol{\alpha} = \left(\alpha_a, \alpha_b, \alpha_p, \alpha_q\right) \neq \mathbf{0}$ such that

$$\boldsymbol{\alpha} \cdot \frac{\partial \log f_{\mathrm{EKG2}}}{\partial \boldsymbol{\theta}'} = \alpha_a \frac{\partial \log f_{\mathrm{EKG2}}}{\partial a} + \alpha_b \frac{\partial \log f_{\mathrm{EKG2}}}{\partial b} + \alpha_p \frac{\partial \log f_{\mathrm{EKG2}}}{\partial p} + \alpha_q \frac{\partial \log f_{\mathrm{EKG2}}}{\partial q} = 0 \text{ for } \forall x > 0. \tag{A4}$$

From (A1), it can be seen that $\frac{\partial \log f_{\mathrm{EKG2}}}{\partial \theta_i}$ constitutes terms in the form (A3); suppose there exists $\boldsymbol{\alpha} \neq \mathbf{0}$ that satisfies (A4), then, the sum of the terms correspond to the form (A3) with $k = 1$ and $\nu = 0$ must converge to zero when $z \to 0$ $(x \to 0)$ to make $\boldsymbol{\alpha} \cdot \frac{\partial \log f_{\mathrm{EKG2}}}{\partial \boldsymbol{\theta}'}$ finite. Thus, the following equation must hold:

$$\alpha_a \frac{p}{a} + \alpha_p = 0. \tag{A5}$$

Similarly, the sum of the terms that correspond to the form (A3) with $l = 1$ and $\xi = 0$ must converge to zero when $z \to 1$ $(x \to \infty)$ to make $\boldsymbol{\alpha} \cdot \frac{\partial \log f_{\mathrm{EKG2}}}{\partial \boldsymbol{\theta}'}$ finite. Thus, the following equation must hold:

$$\alpha_a \frac{q}{a} + \alpha_q = 0. \tag{A6}$$

Furthermore, the sum of the terms that correspond to the form (A3) with $k = l = \nu = 0$ must converge to zero when $z \to 0$ to make $\boldsymbol{\alpha} \cdot \frac{\partial \log f_{\mathrm{EKG2}}}{\partial \boldsymbol{\theta}'}$ stay at zero when $z \to 0$. Thus, the following equation must also hold:

$$\alpha_a \frac{1}{a} - \alpha_b \frac{ap}{b} + \alpha_p[\psi(p + q) - \psi(p)] + \alpha_q[\psi(p + q) - \psi(q)] = 0. \tag{A7}$$

Similarly, the sum of the terms that correspond to the form (A3) with $k = l = \xi = 0$ must converge to zero when $z \to 1$ to make $\boldsymbol{\alpha} \cdot \frac{\partial \log f_{\mathrm{EKG2}}}{\partial \boldsymbol{\theta}'}$ stay at zero when $z \to 1$. Thus, the following equation must hold:

$$\alpha_a \frac{1}{a} + \alpha_b \frac{2aq}{b} + \alpha_p[\psi(p + q) - \psi(p)] + \alpha_q[\psi(p + q) - \psi(q)] = 0. \tag{A8}$$

By deducting each side in (A7) from the respective side in (A8), an equation

$$\alpha_b \frac{a}{b}(p + 2q) = 0 \tag{A9}$$

is derived. To satisfy the equation in (A9), α_b must be zero. Because of the equations in (A5) and (A6), α_a must be non-zero, otherwise $\boldsymbol{\alpha} = \mathbf{0}$. However, if $\alpha_a \neq 0$, to make $\boldsymbol{\alpha} \cdot \frac{\partial^2 \log f_{\mathrm{E\kappa G2}}}{\partial z \partial \boldsymbol{\theta}'}$ finite when $z \to 1$, the sum of the terms including $\log(1 - z)$ among the terms consisting of $\boldsymbol{\alpha} \cdot \frac{\partial^2 \log f_{\mathrm{E\kappa G2}}}{\partial z \partial \boldsymbol{\theta}'}$ must converge to zero (the sum of the terms including $1/(1 - z)$ converges to zero because of equation (A6)). Thus, the following equation must hold:

$$p + 2q + 1 = 0. \tag{A10}$$

It contradicts the fact that the left-hand side in (A10) is actually greater than unity. *Q.E.D.*

Appendix 3. The Fisher information matrix of the EκG2

Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)$ be regarded in the same light as parameters (a, b, p, q). Because the EκG2 satisfies the regularity conditions in terms of MLE, the Fisher information matrix can be expressed as follows:

$$I_{\mathrm{E\kappa G2}}(\boldsymbol{\theta}) = \left[E\left(\frac{\partial \log f_{\mathrm{E\kappa G2}}}{\partial \theta_i}\frac{\partial \log f_{\mathrm{E\kappa G2}}}{\partial \theta_j}\right)\right] = -\left[E\left(\frac{\partial^2 \log f_{\mathrm{E\kappa G2}}}{\partial \theta_i \partial \theta_j}\right)\right] \coloneqq \begin{pmatrix} I_{aa} & I_{ab} & I_{ap} & I_{aq} \\ I_{ba} & I_{bb} & I_{bp} & I_{bq} \\ I_{pa} & I_{pb} & I_{pp} & I_{pq} \\ I_{qa} & I_{qb} & I_{qp} & I_{qq} \end{pmatrix}.$$

Let $\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}}$ denote the ML parameters for an I.I.D. sample of size n on $F_{\mathrm{E\kappa G2}}(x; \boldsymbol{\theta})$. Then, $\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}}$ converges to population parameters $\boldsymbol{\theta}$ with probability 1, and $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta})$ converges in distribution to the normal distribution $N(\mathbf{0}, I_{\mathrm{E\kappa G2}}(\boldsymbol{\theta})^{-1})$ when $n \to \infty$. When fitting the EκG2 to grouped data by the 22 categories described in the main text, according to simulation results, sample variances and co-variances of the parameters are larger than $I_{\mathrm{E\kappa G2}}(\boldsymbol{\theta})^{-1}/n$ in absolute value but less than or equal to four times of $|I_{\mathrm{E\kappa G2}}(\boldsymbol{\theta})^{-1}|/n$ if the data are tabulated from a sufficiently large sample, although it should be noted that estimates of b and q may deteriorate when q is relatively large. The elements of $I_{\mathrm{E\kappa G2}}(\boldsymbol{\theta})$ are expressed as follows:

$$I_{aa} = \frac{1}{a^2} + \frac{1}{B(p,q)}\frac{1}{a^2}\left[(p + 2q)\ddot{\Lambda}(p + 1, q + 1,3) - 2\ddot{\Lambda}(p + 1, q + 2,4) + \ddot{\Lambda}(p + 2, q + 1,4)\right],$$

$$\begin{aligned} I_{ab} = I_{ba} = {} & \frac{2}{B(p,q)}\frac{1}{b}[p\Lambda(p, q + 1,1) - q\Lambda(p + 1, q, 1) + \Lambda(p + 1, q + 1,2)] \\ & + \frac{2}{B(p,q)}\frac{1}{b}\left[(p + 2q)\dot{\Lambda}(p + 1, q + 1,3) - 2\dot{\Lambda}(p + 1, q + 2,4)\right. \\ & \left. + \dot{\Lambda}(p + 2, q + 1,4)\right], \end{aligned}$$

$$\begin{aligned} I_{bb} = {} & -\frac{2}{B(p,q)}\frac{a}{b^2}[p\Lambda(p, q + 1,1) - q\Lambda(p + 1, q, 1) + \Lambda(p + 1, q + 1,2)] \\ & + \frac{4}{B(p,q)}\frac{a^2}{b^2}[(p + 2q)\Lambda(p + 1, q + 1,3) - 2\Lambda(p + 1, q + 2,4) \\ & + \Lambda(p + 2, q + 1,4)], \end{aligned}$$

$$I_{bp} = I_{pb} = \frac{2}{B(p,q)}\frac{a}{b}\Lambda(p, q + 1,1),$$

$$I_{bq} = I_{qb} = -\frac{2}{B(p,q)}\frac{a}{b}\Lambda(p + 1, q, 1),$$

$$I_{ap} = I_{pa} = \frac{1}{B(p,q)} \frac{1}{a} \dot{\Lambda}(p, q+1, 1),$$

$$I_{aq} = I_{qa} = -\frac{1}{B(p,q)} \frac{1}{a} \dot{\Lambda}(p+1, q, 1),$$

$$I_{pp} = \psi'(p) - \psi'(p+q),$$
$$I_{pq} = I_{qp} = -\psi'(p+q),$$
$$I_{qq} = \psi'(q) - \psi'(p+q),$$

where $\Lambda(p,q,k) := \int_0^1 \frac{z^{p-1}(1-z)^{q-1}}{(2-z)^k} dz = \frac{1}{2^k} B(p,q)\ {}_2F_1\left(k, p; p+q; \frac{1}{2}\right),$

$$\dot{\Lambda}(p,q,k) := \int_0^1 \frac{z^{p-1}(1-z)^{q-1}}{(2-z)^k} \log\frac{1-z}{z^2} dz = -2\frac{\partial\Lambda(p,q,k)}{\partial p} + \frac{\partial\Lambda(p,q,k)}{\partial q}$$
$$= \frac{1}{2^k}\big(-2\psi(p) + \psi(q) + \psi(p+q)\big) B(p,q)\ {}_2F_1\left(k, p; p+q; \tfrac{1}{2}\right)$$
$$-\frac{1}{2^k} B(p,q)\left[2\frac{\partial}{\partial\theta_2}\ {}_2F_1\left(k,p;p+q;\tfrac{1}{2}\right) + \frac{\partial}{\partial\theta_3}\ {}_2F_1\left(k,p;p+q;\tfrac{1}{2}\right)\right] \text{ and}$$

$$\ddot{\Lambda}(p,q,k) := \int_0^1 \frac{z^{p-1}(1-z)^{q-1}}{(2-z)^k}\left(\log\frac{1-z}{z^2}\right)^2 dz = 4\frac{\partial^2\Lambda(p,q,k)}{\partial p^2} - 4\frac{\partial^2\Lambda(p,q,k)}{\partial p\partial q} + \frac{\partial^2\Lambda(p,q,k)}{\partial q^2}$$
$$= \big(4\psi'(p) + \psi'(q) - \psi'(p+q)\big)\Lambda(p,q,k)$$
$$+\big(-2\psi(p) + \psi(q) + \psi(p+q)\big)\dot{\Lambda}(p,q,k)$$
$$-\frac{1}{2^k}\big(-2\psi(p) + \psi(q) + \psi(p+q)\big) B(p,q)\left[2\frac{\partial}{\partial\vartheta_2}\ {}_2F_1\left(k,p;p+q;\tfrac{1}{2}\right) + \frac{\partial}{\partial\vartheta_3}\ {}_2F_1\left(k,p;p+q;\tfrac{1}{2}\right)\right]$$
$$+\frac{1}{2^k} B(p,q)\left[4\frac{\partial^2}{\partial{\vartheta_2}^2}\ {}_2F_1\left(k,p;p+q;\tfrac{1}{2}\right) + 4\frac{\partial^2}{\partial\vartheta_2\partial\vartheta_3}\ {}_2F_1\left(k,p;p+q;\tfrac{1}{2}\right) + \frac{\partial^2}{\partial{\vartheta_3}^2}\ {}_2F_1\left(k,p;p+q;\tfrac{1}{2}\right)\right].$$

In the above formula, ${}_2F_1(\vartheta_1, \vartheta_2; \vartheta_3; z)$ denotes the hypergeometric function; $\frac{\partial}{\partial\vartheta_i}\ {}_2F_1\left(1, p; p+q; \frac{1}{2}\right)$ stands for $\frac{\partial}{\partial\vartheta_i}\ {}_2F_1(\vartheta_1, \vartheta_2; \vartheta_3; z)\Big|_{(\vartheta_1,\vartheta_2,\vartheta_3,z)=\left(1,p,p+q,\frac{1}{2}\right)}$; $\frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j}\ {}_2F_1\left(1, p; p+q; \frac{1}{2}\right)$ stands for $\frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j}\ {}_2F_1(\vartheta_1, \vartheta_2; \vartheta_3; z)\Big|_{(\vartheta_1,\vartheta_2,\vartheta_3,z)=\left(1,p,p+q,\frac{1}{2}\right)}$. Those partial derivatives of the hypergeometric function must be calculated numerically because routines for those derivatives is not provided by statistical computer packages. Similarly to the Fisher information matrix of the GB2 derived by Brazauskas [2], the second derivatives of the log-likelihood with respect to *p* and *q* are independent of *a* and *b*.

References

1. Bandourian, R., McDonald, J.B., Turley, R.S.: A comparison of parametric models of income distribution across countries over time. LIS working paper 305 (2002)
2. Brazauskas, V.: Fisher information matrix for the Feller-Pareto distribution. Stat. Probabil. Lett. **59**, 159–167 (2002)
3. Clementi, F., Gallegati, M., Kaniadakis, G.: κ-generalized statistics in personal income distribution. Eur. Phys. J. B **52**, 187–193 (2007)
4. Clementi, F., Gallegati, M., Kaniadakis, G.: A κ-generalized statistical mechanics approach to income analysis. J. Stat. Mech–Theory E., P02037 (2009)
5. Dagum, C.: A new model of personal income distribution: specification and estimation. Economie Appliquée **30**, 413–437

(1977)

6. Jenkins, S.P.: Inequality and the GB2 income distribution. ISER working paper 2007-12 (2007)
7. Kaniadakis, G.: Non-linear kinetics underlying generalized statistics. Physica A **296**, 405–425 (2001)
8. Kaniadakis, G., Lissia, M., Scarfone, A.M.: Deformed logarithms and entropies. Physica A **340**, 41–49 (2004)
9. Kleiber, C., Kotz, S.: Statistical Size Distributions in Economics and Actuarial Sciences. Wiley, Hoboken (2003)
10. Luxembourg Income Study (LIS) Database: http://www.lisdatacenter.org (multiple countries; data runs completed in March – July 2013), Luxembourg: LIS.
11. McDonald, J.B.: Some generalized functions for the size distribution of income. Econometrica **52**, 647–663 (1984)
12. Nelder, J.A., Mead, R.: A simplex method for function minimization. Comput. J. **7**, 308-313
13. Okamoto, M.: Evaluation of the goodness of fit of new statistical size distributions with consideration of accurate income inequality estimation. Econ. Bull. **32**(4), 2969–2982 (2012)
14. Okamoto, M.: Erratum to "Evaluation of the goodness of fit of new statistical size distributions with consideration of accurate income inequality estimation". Econ. Bull. **33**(3), 2443–2444 (2013)
15. Okamoto, M.: Comparison in goodness of fit between the double-Pareto lognormal distribution and the generalized beta distribution of the second type. Mimeo (2013)
16. Reed, W.J.: The Pareto law of incomes – an explanation and an extension. Physica A **319**, 579–597 (2003)
17. Reed, W.J., Wu, F.: New four- and five-parameter models for income distributions. In: Chotikapanich, D. (ed.) Modeling Income Distributions and Lorentz Curves. Springer, New York (2008)
18. Serfling, R.J.: Approximation Theorems of Mathematical Statistics. Wiley, New York (1980)
19. Singh, S.K., Maddala, G.S.: A function for the size distribution of incomes. Econometrica **44**, 963–970 (1975)
20. Venter, G.: Transformed beta and gamma distributions and aggregate losses. Proceedings of the Causality Actuarial Society **70**, 156–193 (1983)