LIS
Working Paper Series

No. 568

Is the Pareto-Lévy Law a Good Representation of Income Distributions?

John K. Dagsvik, Zhiyang Jia, Bjørn H. Vatne and Weizhen Zhu

August 2011

CROSS-NATIONAL
DATA CENTER
in Luxembourg

Luxembourg Income Study (LIS), asbl

Is the Pareto-Lévy Law a Good Representation of Income Distributions?

by

John K. Dagsvik, Zhiyang Jia, Bjørn H. Vatne and Weizhen Zhu

Abstract

Mandelbrot (1960) proposed using the so-called Pareto-Lévy class of distributions as a framework for representing income distributions. We argue in this paper that the Pareto-Lévy distribution is an interesting candidate for representing income distribution because its parameters are easy to interpret and it satisfies a specific invariance-under-aggregation property. We also demonstrate that the Gini coefficient can be expressed as a simple formula of the parameters of the Pareto-Lévy distribution. We subsequently use wage and income data for Norway and seven other OECD countries to fit the Pareto-Lévy distribution as well as the Generalized Beta type II (GB2) distribution. The results show that the Pareto-Lévy distribution fits the data better than the GB2 distribution for most countries, despite the fact that GB2 distribution has four parameters whereas the Pareto-Lévy distribution has only three.

Keywords: Stable distributions, Pareto-Lévy distribution, income distributions, invariance principles, Generalized Beta distributions

JEL classification: C21, C46, C52, D31

John K. Dagsvik. Research Department, Statistics Norway, john.dagsvik@ssb.no

Zhiyang Jia, Research Department, Statistics Norway, zhiyang.jia@ssb.no

Bjørn H. Vatne, The Central Bank of Norway, Bjorn-Helge.Vatne@Norges-Bank.no

Weizhen Zhu, The Financial Supervisory Authority of Norway, weizhen.zhu@finanstisynet.no

1. Introduction

For analysing and representing data on income distributions it may often be convenient to apply a parametric class of distribution functions, for the following reasons. First, a parametric distribution allows researchers to estimate the income distribution in question even if data are scarce. Second, a parametric framework may be convenient for summary representations of the distribution by means of associated parameters, provided that these parameters have a clear and intuitive interpretation and capture central aspects of the distribution. Third, the parameters of a parametric framework may in some cases relate in a simple way to commonly used inequality measures, such as the Gini coefficient, thus facilitating the interpretation of changes in the Gini coefficient in terms of changes in the parameters of the corresponding distribution. This paper discusses the potential of the Pareto-Lévy distribution as a framework for representing income distributions.

In recent years, several researchers have proposed or discussed parametric representations for the size distribution of incomes. These include Champernowne (1953), Singh and Maddala (1976), Dagum (1977), Kloek and van Dijk (1978), McDonald and Ransom (1979), van Dijk and Kloek (1980), McDonald (1984), Esteban (1986), Majumder and Chakravarty (1990), McDonald and Mantrala (1995), McDonald and Xu (1995), Bordley et al. (1996), Parker (1999), Bandourian et al. (2002) and Chotikapanich (2008). The distributions proposed include the log-normal distribution, the Gamma and Generalized Gamma distributions, the Burr 3 and Burr 12 distributions, and the GB1 and GB2 distributions. As discussed by McDonald and Mantrala (1995) and Bordley et al. (1996), many distributions that have been proposed in the literature, such as the Generalized Gamma, the Burr 3 and the Burr 12, are special cases of the GB2 distribution. Bordley et al. (1996) demonstrate that among several four-parameter distributions, the GB2 distribution yields the best fit to US data.

Some of these representations are derived from theoretical principles. For example, Singh and Maddala (1976) and Dagum (1977) derived their functional form as the solution of a differential equation specified to capture the characteristic properties of empirical income distributions, whereas Esteban (1986) proposed characterizing the Generalized Gamma distribution as an income distribution in terms of what he defined as the ‘income share elasticity’. Parker (1999) presents a neoclassical model of optimizing firm behaviour which predicts that income distribution will follow the GB2 distribution.

Unfortunately, the parameters of many of the distributions proposed in the literature do not in general have a simple interpretation. Also, except for Champernowne (1953), Singh and Maddala (1976), Dagum (1977), Esteban (1986) and Parker (1999), the only selection criterion of the parametric distributions seem to be a consideration of goodness of fit. Mandelbrot has emphasized in several of his papers that although goodness of fit is a necessary requirement, it should not be the only one. This is because there is no general agreement on what is the best goodness of fit measure. Since different goodness of fit measures emphasize different parts of the distributions under investigation, different measures may result in different rankings of the distributions. A more fundamental point is that the central focus of a scientific approach should be on the relevant *qualitative* properties of the model in question, not just on its flexibility with respect to fitting the data and the application of sophisticated inference methods.

In the 1960s and 1970s Mandelbrot wrote a number of papers in which he discussed the problem of justifying the stochastic properties of economic variables such as stock market prices and incomes (Mandelbrot, 1960, 1961, 1962, 1963). His argument was that there are certain aggregation operations that take place and have important bearings on the structure of the probability distribution of the variables under investigation. In a similar way to what had happened in the physical sciences, he was led to postulate specific invariance principles which imply that income distributions should belong to the class of stable distributions. The class of stable distributions has the property that a linear combination of two independent stable random variables with the same index of stability is a stable random variable. The class of stable distributions contains the normal distribution as a special case and also follows from a general version of the Central Limit Theorem without the condition of bounded variance. Note that the condition of bounded variance is essential in the classic Central Limit Theorem. More precisely, in the context of income distributions Mandelbrot restricted attention to a subclass of stable distributions; namely, what he called the Pareto-Lévy law. A Pareto-Lévy distribution has the properties that the mean exists, that it is skewed to the right and that, for a suitable choice of location, the probability mass of negative values becomes negligible (within the stable class). The rationale for the emphasis on the Pareto-Lévy law is the need to take into account the fact that ‘income’ is a non-negative variable.

Like Mandelbrot, we argue that the Pareto-Lévy distribution offers an attractive framework for representing income distributions, for a number of reasons. First, the three parameters of the Pareto-Lévy distribution have a clear and intuitive interpretation. Second, the right tail is asymptotically Pareto-distributed, a property which is considered as a typical feature of empirical income distributions. Third, the Pareto-Lévy distribution possesses the linear aggregation property mentioned above. Specifically, since a linear combination of independent Pareto-Lévy distributed variables, which are distributed according to a Pareto-Lévy distribution with given index of stability, has a Pareto-Lévy distribution with the same index, a model using this framework will consequently not depend critically on whether the income concept is based on, say, monthly or yearly income.

A novel contribution of this paper is the demonstration that the Gini coefficient of the Pareto-Lévy distribution can be expressed as a simple closed-form formula of its parameters. This is of considerable interest because it facilitates the interpretation of changes in income distribution in terms of changes in inequality, as measured by the Gini coefficient.

Apart from the work of Van Dijk and Kloek (1980), the Pareto-Lévy distributions have not received much attention in this context. In the theoretical part of this paper, we discuss in detail possible reasons why many are reluctant to use the Pareto-Lévy class to model income distributions and argue that they are no longer valid objections.

In the empirical part of this paper, we estimate Pareto-Lévy distributions using micro-data on wage and income for Norway and grouped income data for seven other OECD countries from the Luxembourg Income Study (LIS) database (2003). This seems to be the first time micro-data on incomes have been used to estimate a Pareto-Lévy distribution. For comparison, we also fit the income data to a GB2 distribution, comparing the fit to the estimated Pareto-Lévy distribution. In contrast to what van Dijk and Kloek (1980) found, our results suggest that the Pareto-Lévy distribution is flexible enough to fit typical empirical income distributions. It even appears to fit the data *better* than the popular GB2 distribution for most countries.

The paper is organized as follows: in the next section we discuss properties of stable distributions and review Mandelbrot's invariance arguments that support the Pareto-Lévy distribution; in Section 3 we derive a closed-form formula for the Gini coefficient as a function of the parameters of the Pareto-Lévy distribution; in Section 4 we review different estimation methods; and in Section 5 we describe the data and report empirical results.

2. Theoretical considerations

In this section we first review some of the properties of the Pareto-Lévy law before we explain why we are using it to represent income distribution. Our reason for this is that although stable distribution is well known among specialists in mathematical statistics, it appears to be less known among economists and even many econometricians.

The stable class follows from an extended version of the Central Limit Theorem, under rather general conditions. Specifically, a random variable, X, is said to have a *stable* distribution if for any positive numbers, a_1 and a_2 there exists a positive number b_1 and a real number b_2 such that

(2.1) $$a_1X_1 + a_2X_2 \overset{d}{=} b_1X + b_2 ,$$

where X_1 and X_2 are independent copies of X, and where $\overset{d}{=}$ denotes equality in distribution. It is well known that the normal distribution has this property, but it is less known that this property holds for a much wider class of distribution functions: namely, the so-called stable class. The distribution of X is called *strictly stable* if (2.1) holds with $b_2 = 0$. The stable class was thoroughly investigated by Paul

Lévy in the 1920s and 1930s (see Lévy, 1925, 1937; see also Gnedenko and Kolmogorov, 1954). Except for a few cases, the probability distributions in the stable class cannot be expressed in closed form. Their characteristic function, however, can be expressed as

$$\phi(\lambda) = \mathrm{E}e^{i\lambda X} = \exp\left(i\lambda\delta - c^{\alpha}\left|\lambda\right|^{\alpha}\left(1 - i\beta \,\mathrm{sign}(\lambda)\tan\left(\frac{\alpha\pi}{2}\right)\right)\right), \tag{2.2}$$

where, $i = \sqrt{-1}$, X is a stable random variable, c > 0 and δ are scale and location parameters, $\beta \in [-1,1]$ is a parameter that represents the skewness of the distribution, α (called the index of stability) represents the tail fatness of the distribution and the formula (2.2) holds for all $\alpha \in (0,2]$ except for $\alpha = 1$. When $\alpha = 1$, the characteristic function is given by a similar formula (see Samorodnitsky and Taqqu, 1994). The parameter c is a scale parameter similar to the standard deviation and represents the 'spread' of the distribution, so that when X is multiplied by a constant, k, the scale parameter changes from c to kc. By tail fatness representation we mean that asymptotically

$$\alpha \cong \frac{-d\log(1-F(x))}{d\log x}$$

when *x* is 'large'. This property means that the distribution has a right Pareto tail (see Pareto, 1897). When $\beta = 0$ the distribution is symmetric, whereas it is said to be totally skewed to the right (left) when $\beta = 1$ $(\beta = -1)$. When $\alpha = 2$ we get the normal distribution, the parameter β vanishes and $c^2 = 2\sigma^2$ where σ^2 is the variance. When $\alpha \in (1,2)$ the variance of the distribution is infinite and the expectation equals δ, whereas the expectation does not exist and the variance is infinite when $\alpha \in (0,1]$. Thus within the stable class the only member that possesses a finite variance is the normal distribution.

For our purpose, we are only interested in a subclass of the stable distributions, namely the Pareto-Lévy class. This class consists of the subclass of stable distributions with $\alpha \in (1,2)$ and $\beta = 1$ and has the property that the probability mass of negative values will be negligible for sufficiently large δ.[1] Thus, a Pareto-Lévy distribution is characterized by only three parameters, the measure of tail fatness α, the mean δ, and the measure of distribution spread c.

[1] There is also another special case within the stable class with support defined on the positive part of the real line. When the index of stability is less than 1, the skewness parameter is equal to 1 and the location parameter is zero, then the corresponding stable distribution is positive only on the positive part of the real line. However, this special case seems not to be of interest in the context of income distribution because the mean does not exist.

The first person to discuss theoretical properties of income distributions seems to have been Pareto (1897). He found that the right tail of income/wealth distribution follows an inverse-power law and he therefore proposed the Pareto distribution to describe the distribution of income. However, it was subsequently recognized that the Pareto distribution gives a poor fit of other parts of the distributions. In several papers Mandelbrot proposed using a generalization of the Pareto distribution – namely, the Pareto-Lévy distribution mentioned above – as a framework for representing incomes and other economic variables (see Mandelbrot, 1960, 1961, 1962, 1963).

Not many researchers have taken Mandelbrot's idea of applying the Pareto-Lévy class as a framework for analysing income distribution very seriously. We suggest that this may be for the following reasons. First, the c.d.f. and the corresponding density of the Pareto-Lévy distributions cannot be expressed in closed form. Second, since we never observe empirical distributions with infinite variances, it may initially seem rather awkward to apply theoretical distributions with infinite variance. Third, the theoretical arguments provided by Mandelbrot to support his choice of the Pareto-Lévy distribution are not entirely convincing. Finally, the Pareto-Lévy distribution is not flexible enough to fit typical empirical income distributions.

Taking a closer look at these arguments, the first one is no longer a serious objection because several estimation and simulation methods now exist that can be used to estimate and produce graphs of stable distributions. One particular method (McCulloch, 1986) is in fact extremely simple to use.

The Second argument against the Pareto-Lévy distribution is not valid either. The variance is just a mathematical expression that should not be taken literally in every case. The situation is similar to the following example: the normal distribution has infinite support, whereas empirical distributions have finite support, but this does not prevent researchers from usefully applying this distribution in a large number of cases. Moreover, the existence of second-order moments implies that, if disturbances hitting economic agents are only idiosyncratic, then aggregate fluctuations disappear as the number of agents grows large. In contrast, distributions with infinite variance do not need aggregate shocks to generate aggregate fluctuations and are good candidates for explaining aggregate large fluctuations in time periods characterized by small aggregate shocks.

As regards the third objection, Mandelbrot (1960) argued as follows: if income is considered the variable of interest, this variable may be broken down into different kinds of income, such as income from waged work, self-employment, capital income, etc. If the distribution of each of the income components is compared with the distribution of total income, it appears that these distributions have more or less the 'same shape'.[2] From this, Mandelbrot (1960) postulated that the distribution of the sum of (independent) income components should also belong to the same class as

[2] The notion of the 'same shape' is, of course, not very precise. Here it will be interpreted as follows: two distributions have the same shape if after a scale transformation they become roughly the same.

the distributions of the components. Hence, provided that the income components are skew to the right, stochastically independent, and have negligible negative left tails, the Pareto-Lévy class of distributions follows. Of course, in practice the income components may not be strictly independently distributed, so Mandelbrot's argument will be valid only for an idealized case and the resulting class of distributions may hold only in an approximate sense. It is also of interest to note that although the independence condition is necessary for (2.1) to imply a Pareto-Lévy distribution, it is sufficient for this equation to hold that (X_1, X_2) has a joint Pareto-Lévy distribution. A similar argument can be applied to aggregation over time.

As mentioned in the introduction, we believe that in addition to the invariance-under-aggregation property explained above, the most interesting feature about the Pareto-Lévy distribution function is that its parameters have a clear and intuitive interpretation as they represent key features (location, scale and tail fatness) of the underlying income distribution, as reviewed in Section 2. This is in contrast to many other parametric distributions that have been applied in this context.

A crucial point, therefore, is whether the Pareto-Lévy law is sufficiently flexible to yield a good fit of empirical income distributions compared to other parametric distributions in use. Our empirical results on income data from eight OECD countries suggest the Pareto-Lévy law is quite flexible. Specifically, we demonstrate in the present paper that in our chosen empirical application the Pareto-Lévy distribution does indeed fit the data rather well.

As indicated in the introduction, we would like to emphasize that model selection criteria based solely on goodness of fit measures are treacherous. The reason is that different and perfectly reasonable goodness of fit measures may yield different model selection results, as will be illustrated in the empirical application below. More fundamentally, scientific modelling should not primarily be a matter of curve fitting. What is needed are alternative qualitative selection criteria that combine goodness of fit properties with theoretical features. With these criteria in mind, the Pareto-Lévy law has clear advantages over other competing parametric distributions proposed in the literature, due to the appealing theoretical properties discussed above.

3. The Gini coefficient as a closed-form function of the parameters of the Pareto-Lévy distribution

In this section we investigate an additional advantage of the Pareto-Lévy law, namely, that it allows us to express the Gini coefficient as a simple formula of the underlying parameters of the distribution. The Pareto-Lévy law is not the only distribution that leads to a closed-form formula for the Gini coefficient. However, the Gini coefficient for the most general family of distributions proposed in this context – namely, the GB2 distribution – is quite complicated (see McDonald, 1984).

In assessing the significance of changes in an income distribution, say F(y), aggregate measures of income inequality are often employed. The most common measure is the Gini coefficient, which can be expressed as

(3.1)
$$G = \frac{E\left|Y_1 - Y_2\right|}{2EY_1},$$

where Y_1 and Y_2 are independent random variables with c.d.f. F(y). In this section we consider the properties of G when F(y) is a Pareto-Lévy distribution. Specifically, we demonstrate how G depends in a simple way on the parameters of the Pareto-Lévy distribution.

When F is assumed to be a Pareto-Lévy distribution it follows that $Y_1 - Y_2$ is stable, symmetric with zero mean and dispersion parameter equal to $c2^{1/\alpha}$ (see Samorodnitsky and Taqqu, 1994). From Samorodnitsky and Taqqu (1994, p.18), we thus find that

(3.2)
$$E\left|Y_1 - Y_2\right| = \frac{2c\Gamma\left(1 - \frac{1}{\alpha}\right)2^{1/\alpha}}{\pi},$$

from which it follows that G can be expressed as

(3.3)
$$G = \frac{c\Gamma\left(1 - \frac{1}{\alpha}\right)2^{1/\alpha}}{\delta\pi}.$$

Let

(3.4)
$$\kappa = \frac{\Gamma(1 - 1/\alpha)2^{1/\alpha}}{\pi}.$$

Since the Gamma function $\Gamma(x)$ is strictly decreasing when x is positive and less than 1, it follows that κ is decreasing as a function of α. Since κ is a monotone mapping of α, it is qualitatively an equivalent representation of tail fatness. Thus κ increases with increasing fatness of the right tail of the income distribution, so we can write the Gini coefficient as

(3.5)
$$G = \frac{c\kappa}{\delta}.$$

The relation in (3.5) asserts that the Gini coefficient is proportionate to the scale parameter and the tail fatness index, and inversely proportionate to the mean of the distribution. Evidently, since incomes are positive, the parameters must fulfil certain restrictions: namely, that

$$P(\delta + cV > 0) \geqq 1 - \varepsilon$$

where ε is small and V denotes a Pareto-Lévy random variable with scale and location parameters equal to 1 and zero respectively. For example, if $\varepsilon = 0.001$ and $\alpha = 1.6$, then it follows that $\delta \geq 4c$.

The formula (3.3), or alternatively (3.5), is of considerable interest because it enables us to assess the impact on inequality (measured by G) from changes in the parameters of the income distribution: namely, c, δ and α. In practice, the tail fatness index α will in most cases belong to the interval [1.5, 1.7]. This corresponds to a κ decreasing from 1.35 to about 1. This means that an increase of the index of stability from 1.5 to 1.7 will yield a substantial decrease in inequality, as measured by the Gini coefficient.

For the sake of computational simplicity, the following may be of interest. Note first that within the interval $x \in [1.3, 1.5]$, $\Gamma(x)$ is practically constant and approximately equal to 0.89. Remember, furthermore, that the Gamma function satisfies the recursive relation $\Gamma(1+x) = x\Gamma(x)$. Hence we obtain that when α belongs to the interval [1.5, 2], we have

$$\Gamma(1-1/\alpha) = \frac{\Gamma(2-1/\alpha)}{1-1/\alpha} \cong \frac{0.89}{1-1/\alpha},$$

which implies that

(3.6) $$\kappa \cong \frac{0.89 \cdot 2^{1/\alpha}}{(1-1/\alpha)\pi}.$$

There are several non-parametric estimators for the Gini coefficient. For example, we realize that the empirical analogue to the formula in (3.1) is given by

(3.7) $$\hat{G} = \frac{\sum_{j=1}^{N}\sum_{i \neq j}^{N} |Y_i - Y_j|}{2(N-1)\sum_{i=1}^{N} Y_i},$$

where N is the sample size and $Y_i, i = 1,2,..., N$ are the individual incomes. Therefore the formula in (3.7) can be applied to estimate the Gini coefficient non-parametrically.

4. Estimation issues

The literature suggests many methods for estimating the parameters of stable distributions. In this section we briefly review a few of these in order to show that although no closed-form analytic expression for Pareto-Lévy distribution function exists, it is not difficult to estimate its parameters. However, here we are not primarily concerned with optimal inference methods. As indicated above, we feel that, with some notable exceptions, the literature on fitting parametric distributions to incomes has focused too narrowly on inference issues and statistical goodness of fit criteria.

4.1. The McCulloch method

An interesting estimation method, attractive because of its simplicity, was proposed by McCulloch (1986). It is based on selected fractiles of the empirical distribution. Specifically, he proposed a method based on the 5 per cent, 25 per cent, 50 per cent, 75 per cent and 95 per cent fractiles of the empirical distribution function. Thus this method requires the knowledge of only these fractiles and access to a set of tables given in McCulloch's paper. The corresponding asymptotic standard errors can also be computed by means of these tables. Below we illustrate how this method performs in comparison with other estimation methods.

4.2. The Koutrouvelis method

The method proposed by Koutrouvelis (1980) is based on properties of the characteristic function of the stable distribution. This method has also been extensively discussed by Kogon and Williams (1998), so we give only a very brief discussion here.

Koutrouvelis noted that (2.2) implies that

(4.1) $$\ln(-\ln(|\phi(\lambda)|^2)) = \ln(2c^\alpha) + \alpha \ln|\lambda|$$

This relation shows that a regression type of estimation procedure may be possible (for estimating c and α). To carry out this procedure, $\phi(\lambda)$ on the left-hand side of (4.1) is replaced by the corresponding empirical characteristic function, given by

(4.2)
$$\hat{\phi}(\lambda) = \frac{1}{N}\sum_{k=1}^{N} e^{i\lambda Y_k} ,$$

and (4.1) is regressed against $\ln|\lambda|$ for suitable chosen values of λ. Furthermore, Koutrouvelis shows that $|\hat{\phi}(\lambda)|^2$ can be expressed as

(4.3)
$$|\hat{\phi}(\lambda)|^2 = \frac{1}{N^2}\sum_{r,q} \cos\left(\lambda\left(Y_r - Y_q\right)\right).$$

Koutrouvelis (1980, 1981) and Kogon and Williams (1998) have studied refinements of the estimation procedure above and have also done simulation experiments to demonstrate that this approach works well and is quite efficient, provided that the λ-values are carefully selected and the data are suitably normalized. Koutrouvelis (1980) indicates how to select appropriate λ-values. Thus by this method both α and c can be estimated. For further discussion on this method, we refer to Kogon and Williams (1998). Moreover, since the assumption of stability implies that the relation expressed in (4.1) is linear, we can obtain an informal test of the stability assumption by plotting $\{\ln(-\ln(|\hat{\phi}(\lambda)|^2))\}$ against $\{\ln|\lambda|\}$. If this plot is approximately linear it indicates that the underlying distribution is stable.

Since $\beta = 1$ in the Pareto-Lévy class, the expectation δ is the only parameter that remains to be estimated. Koutrouvelis shows that (2.2) also implies the relation

(4.4)
$$\text{Arc}\tan\left(\frac{\text{Im}\,\phi(\lambda)}{\text{Re}\,\phi(\lambda)}\right) = \delta\lambda + c^{\alpha}|\lambda|^{\alpha}\,\text{sgn}(\lambda)\tan\left(\frac{\alpha\pi}{2}\right)$$

where Im(.) and Re(.) denote the respective imaginary and real parts of the characteristic function. When c and α have been estimated we can, by replacing $\phi(\lambda)$ with $\hat{\phi}(\lambda)$ use (4.4) to estimate the mean δ. This parameter can also be estimated by the corresponding sample mean. However, unless the sample is large, the sample mean is not a very good estimator for the population mean.

4.3. Least-squares method based on the cumulative distribution function

This method is based on minimizing

$$\sum_{k}\left(F(Y_k) - \hat{F}(Y_k)\right)^2,$$

where $\hat{F}(y)$ is the empirical cumulative income distribution function and $\{Y_k\}$ the income observations. The least-squares approach as applied in this paper relies on computing the cumulative Pareto-Lévy distribution function by a method developed by Nolan (1997). Nolan's method is based on a reparametrized version of the stable distribution provided by Zolotarev (1986). For $\beta = 1$ (totally skewed to the right) we have found that Nolan's procedure works well for computing c.d.f., but does not work well for computing the corresponding probability density.

The least-squares method was originally proposed by Bohman (1975). However, he put forward a different method for computing the Pareto-Lévy distribution. The least-squares method can also be used on grouped data.

4.4. Limited-information maximum-likelihood estimation

Since we have individual observations, it is in principle possible to apply the maximum-likelihood method (full information). We have found that for Pareto-Lévy distributions, existing numerical procedures for density functions, such as those suggested by Nolan (1997) and McCulloch and Panton (1997), do not always work well. However, Nolan's procedure seems to work well for computing stable c.d.f. We have therefore applied a limited-information maximum-likelihood procedure based on grouped data. In this case the log-likelihood function has the form

$$\sum_{k=1}^{m} N_k \ln\left(F\left(\bar{y}_k\right) - F\left(\bar{y}_{k-1}\right)\right)$$

where $\bar{y}_1 < \bar{y}_2 < ... < \bar{y}_m$, is the partition that defines the groups and N_k is the number of observations within $\left(y_{k-1}, y_k\right)$ (group k). In the actual estimations we have chosen the number of points, *m*, equal to 30. This method can also be used on grouped data. In fact, several researchers have used this method with different types of parametric distributions.

5. Data, estimation results and goodness of fit

We used two different data sources for our empirical illustration. The Norwegian individual data are based on the Norwegian income and wealth data for 1994 gathered by Statistics Norway and described by Pedersen (1997). Income data for other countries used in this paper have been obtained from the Luxembourg Income Study (LIS) database. Specifically, we use data for seven countries: Australia (1994), Belgium (1997), Canada (1997), France (1994), Germany (1994), Italy (1995) and the United States (1997). The LIS datasets contain variables on market incomes, public transfers and taxes, household- and person-level characteristics, labor market outcomes, and, in some datasets, expenditures.

5.1 Results for Norway

For the Norwegian data, defining income as the sum of wage income, income from self-employment and capital income, we have excluded those individuals whose income is less than the minimum pension benefit for single people. In 1994 this amount was 60,700 kroner.[3] The sub-population obtained by this censoring consists of 5,618 people. Without such censoring the income distribution in Norway is bimodal. The estimates of the parameters of the Pareto-Lévy distribution are shown in Table 1.

(Table 1 and Figure 1 inserted here)

For the Pareto-Lévy distribution the limited-information maximum-likelihood method is based solely on 30 equally distributed selected data points, as discussed in Section 4.4. Two versions of the least-squares method are applied. The first uses all the income data of the selected sub-population and the second is based solely on the same 30 data points as were used for the maximum-likelihood method. The second is applied for the sake of comparison with the maximum-likelihood method. The precision of the Koutrouvelis method will depend on the choice of $\{\lambda_r\}$. The calculation of standard errors of the parameters of the Koutrouvelis method can be made using the procedure of Koutrouvelis and Bauer (1982). The calculation of the standard errors of the least-squares method is, however, not straightforward. We have estimated standard errors by applying the bootstrap approach for the Koutrouvelis and the least-squares methods. Bootstrap estimates are obtained by simulating 25 replications. As indicated in Section 4.2, it is possible to obtain an informal test of stability by plotting the empirical counterpart of the left-hand side of (4.1) against $\ln|\lambda|$. This plot is displayed in Figure 1, where the data are scaled as suggested in Koutrouvelis (1980). Figure 1 shows that the plot is fairly linear up to $\ln|\lambda| \approx 0.7$. Table 1 shows that the different methods yield modest variations in the estimates. For example, the estimates of the index of stability do not differ significantly according to confidence intervals based on twice the standard deviation on both sides. According to Table 1, it seems, not surprisingly, that McCulloch's method produces considerably higher standard errors of the

[3] The corresponding amount for married people was about 10,000 kroner less than for single people. People with incomes below 60,700 kroner occur in the data. Our interpretation is that people declaring incomes below the subsistence level received undeclared income from other sources: for example, their spouses. The exchange rate in 1994 was 1 US dollar to around 7.05 kroner.

estimates of α and c than the other methods. Apart from the estimates obtained by McCulloch's method, the Gini coefficients calculated by (3.3) vary little from the estimation method and are close to the empirical one, calculated by applying the estimator in (3.7) on all observations.

Remember that our sub-population can in fact be viewed as censored, as we have removed the information about those with incomes below 60,700 kroner. In one estimation procedure we have therefore also applied the corresponding censored theoretical distribution in a limited-information maximum-likelihood estimation procedure.

The Pareto-Lévy c.d.f. has been simulated by means of tables provided by McCulloch and Panton (1997). To assess how well the Pareto-Lévy distribution fits the data, we have displayed the cumulative empirical and the fitted cumulative Pareto-Lévy distribution estimated by the least-squares method. In Figure 2 we display the empirical c.d.f. together with the corresponding fitted Pareto-Lévy c.d.f. by the least-squares and the censored least-squares procedures respectively. We note that the respective plots are almost indistinguishable.

(Figure 2, 3, 4 inserted here)

In Figures 3 and 4 we show Quantile–Quantile (Q–Q) and Percentile–Percentile (P–P) plots of the cumulative empirical distribution functions against the fitted Pareto-Lévy distribution. We note that they are close to straight lines, which indicates a good fit.

For the sake of comparison we have also fitted the GB2 distribution to the income data. The parameter estimates of the GB2 distribution are given in Table 2. We note that these estimates vary considerably across different estimation methods. This may indicate that the parameters of the GB2 distribution do not each represent different and separate qualitative properties of the distribution, as a result of which small changes in the data or estimation method may result in large changes in parameter estimates. This contrasts with the estimates of the Pareto-Lévy distribution, which seem to be fairly stable across different estimation methods. We believe that this is because the parameters α, c and δ capture different and 'orthogonal' qualitative properties of income distribution.

(Table 2, 3 inserted here)

Table 3 shows how selected goodness of fit measures vary depending on distributions and estimation methods. Based on 30 equally distributed points of the empirical c.d.f., we have used the log-likelihood, root of the mean square error and the mean absolute deviation as goodness of fit measures. These measures have been calculated as follows: after the unknown parameters of the Pareto-Lévy income distribution were estimated, we calculated the mean square from the formula in Section 4.3 and the mean absolute deviation as the mean absolute difference between the empirical c.d.f. and the corresponding estimated Pareto-Lévy c.d.f. When the least-squares method is applied, both the Pareto-Lévy and the GB2 distributions give an excellent fit, with the latter slightly better than the former. However, when the limited-information maximum-likelihood method is used, the Pareto-Lévy distribution seems to perform better than the GB2 distribution, despite the fact that the former has only three parameters while the latter has four. The corresponding likelihood functions are almost equal, with the likelihood for the Pareto-Lévy distribution slightly higher than the likelihood for the GB2 distribution.

5.2 Results for the other selected countries

Household income data for the other seven OECD countries have been taken from LIS. Many of the LIS data providers do not allow direct access to their microdata. However, access to the LIS micro databases can be achieved through a remote-execution system, which supports only the standard statistical packages SPSS, STATA and SAS. We have generated grouped data through the remote-execution system ourselves. The grouped data we have used are the 5th, 10th, ..., and the 95th percentiles generated from the LIS household income data for each country.

Following Bandourian et al. (2002), we define the income variable for the dataset as gross wage and salaries, farm income and any self-employment income. As no direct access to the micro-database of LIS is permitted and because of the difficulty of fitting Paréto-Lévy distribution using the standard statistical packages utilized by LIS, we have resorted to grouped data instead. We obtained data grouped into 20 equal-probability intervals, corresponding to the fifth to the ninety-fifth percentiles.

The estimates of the distribution parameters were obtained by applying the limited-information maximum-likelihood estimator proposed in Section 4.4. The estimates of the parameters of the Pareto-Lévy distribution and the GB2 distribution are presented in Table 4. Goodness of fit measures are reported in Table 5.

[Table 4-5 here.]

Table 5 shows that the fit of these two distribution functions is quite close, except perhaps for the US data.[4] This means that the Pareto-Lévy distribution is flexible enough to fit the empirical income data. Pareto-Lévy distribution fits the data better than the GB2 distribution for four of the seven countries based on the limited-information maximum-likelihood method and using the log-likelihood as a goodness of fit measure.

6. Conclusions

In this paper we have discussed the Pareto-Lévy distribution as a framework for analysing income distributions. We have reviewed the theoretical properties of this distribution and we have argued that its parameters have a clear and intuitive interpretation. Specifically, the three parameters of this distribution – namely, α, δ and c– can be interpreted as tail fatness, mean and scale respectively. Although the variance of the Pareto-Lévy distribution is infinite, the interpretation of the scale parameter is entirely similar to the conventional standard deviation as a measure of spread of distribution. Furthermore, we have shown how the Gini coefficient can be expressed as a simple formula in terms of these parameters. We have also argued that the selection of a parametric framework for fitting empirical income distributions should not be based solely on goodness of fit criteria or on the method of estimation. Qualitative properties such as interpretation of parameters and the invariance-under-aggregation property have an essential role to play.

We have gone on to fit the Pareto-Lévy distribution from a sample of Norwegian micro-data on income, as well as on grouped income data from seven other selected countries taken from the LIS database. As regards estimation methods, we have applied several methods in order to illustrate their benefits and drawbacks.

For the sake of comparison, we have also fitted the data to the GB2 distribution, which is a rather flexible four-parameter distribution. Even though the Pareto-Lévy distribution has only three parameters, we have demonstrated in our empirical applications that in terms of goodness of fit it can compete well with the four-parameter-distribution GB2. Unfortunately, the GB2 distribution, as well as other types of distributions that have been proposed in the literature, does not have parameters that can be given such an intuitive interpretation as is the case with the Pareto-Lévy law. Moreover, the parameter estimates of the GB2 distribution do not seem to be stable with respect to different estimation methods.

References

Bandourian R, McDonald JB, Turley R (2002) A Comparison of parametric models of income distribution across countries and over time. Luxembourg Income Study working paper no. 305

Bohman H (1975) Numerical inversions of characteristic functions. Scandinavian Actuarial Journal 121–124

[4] Further investigations (not reported here) show that the estimated Pareto-Lévy distribution does not reproduce the corresponding empirical distribution as well as the GB2 distribution does.

Bordley RF, McDonald JB, Mantrala, A (1996) Something New, Something Old: Parametric Models for the Size Distribution of Income. Journal of income distribution 6: 91–103

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

Chotikapanich D (ed) (2008) Modeling income distributions and Lorenz curves. Springer Verlag, Berlin

Dagum C (1977) A new model of personal income distribution: specification and estimation. Economie Appliquée 30: 413–437

Esteban J (1986) Income share elasticity and the size distribution of income. International Economic Review 27: 439–444

Gnedenko BV, Kolmogorov AN (1954) Limit distributions for sums of independent random variables. Addison Wesley, Reading MA (English translation by K.L. Chung)

Kloek T, Van Dijk HK (1978) Efficient estimation of income distribution parameters. Journal of Econometrics 8: 61–74

Kogon SM, Williams DB (1998) Characteristic function based estimation of Stable distribution parameters. In: Adler RJ, Feldman RE, Taqqu MS (eds) A practical guide to heavy tails. Birkhäuser, Boston MA, pp 311-335

Koutrouvelis IA (1980) Regression-type estimation of the parameters of Stable laws. Journal of the American Statistical Association 75: 918–928

Koutrouvelis IA (1981) An interactive procedure for the estimation of the parameters of Stable laws. Communications in Statistics: Simulations and Computation 10: 17–28

Koutrouvelis IA, Bauer DF (1982) Asymptotic distribution of regression type estimators of parameters of Stable laws. Communications in Statistics: Theory and Methods 11: 2715–2730

Lévy P (1925) Calcul des probabilités.Gauthier-Villars, Paris

Lévy P (1937) Théorie de l'addition des variable aléatoires. Gauthiers-Villars, Paris

Luxembourg Income Study (LIS) Database (2003). On-line (http://www.lisproject.org/techdoc.htm). Microdata runs completed May 2003.

McCulloch JH (1986) Simple consistent estimators of Stable distribution parameters. Communications in Statistics: Simulation and Computation 15: 1109–1136

McCulloch JH, Panton DB (1997) Precise tabulations of the maximally-skewed Stable distributions and densities. Computational Statistics & Data Analysis 23: 307–320

McDonald JB (1984) Some generalized functions for the size distribution of income. Econometrica 52: 647–663

McDonald JB, Mantrala A (1995) The distribution of personal income: Revisited. Journal of Applied Econometrics 10: 201–204

McDonald JB, Ransom MR (1979) Functional forms, estimation techniques and the distribution of income. Econometrica 47: 1513–1525

McDonald JB, Xu YJ (1995) A Generalization of the Beta distribution with applications. Journal of Econometrics 66: 133–152

Majumder A, Chakravarty SR (1990) Distribution of personal income: Development of a new model and its applications to U.S. income data. Journal of Applied Economics 5: 189–196

Mandelbrot B (1960) The Pareto-Lévy law and the distribution of income. International Economic Review 1: 79–106

Mandelbrot B (1961) Stable Paretian random functions and the multiplicative variation of income. Econometrica 29: 517–543

Mandelbrot B (1962) Paretian distributions and income maximization. Quarterly Journal of Economics 76: 57–85

Mandelbrot B (1963) New methods in statistical economies. Journal of Political Economy 71: 421–440

Nolan JP (1997) Numerical calculations of Stable densities and distribution functions. Communications in Statistics: Stochastic Models 13: 759–774

Parker SC (1999) The Generalized Beta as a model for the distribution of earnings. Economic Letters 62: 197–200

Pareto V (1897) Cours d'economie politique. Rouge et Cie, Lausanne and Paris

Pedersen V (1997) The survey of income and wealth 1994. Notes no. 14, Statistics Norway (In Norwegian)

Samorodnitsky G, Taqqu MS (1994) Stable non-Gaussian random processes. Chapman and Hall, New York

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

Van Dijk HK, Kloek T (1980) Inferential procedures in Stable distributions for class frequency data on incomes. Econometrica 48: 1139–1148

Zolotarev VM (1986) One-dimensional Stable distributions. American Mathematical Society; translations of Mathematical Monographs 65, Providence RI

Appendix A

The Generalized Beta type II distribution (GB2)

Several researchers have applied the GB2 distribution to represent the distribution of income data. This distribution is fairly flexible and has density function

$$g(y;a,b,p,q)=\frac{a\,y^{ap-1}}{b^{ap}B(p,q)\left(1+(y/b)^{a}\right)^{p+q}}$$

where a,b,p,q are unknown parameters and B(p,q) is the Beta function

$$B(p,q)=\frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}.$$

The mean in this distribution is given by

$$\frac{b\Gamma(p+1/a)\Gamma(q-1/a)}{\Gamma(p)\Gamma(q)}.$$

Thus the GB2 distribution has four parameters, but unfortunately their interpretation is not so clear. Also the formula for the Gini coefficient is quite complicated. For more details we refer to McDonald (1984) and McDonald and Xu (1995).

Tables and Figures

Table 1 Parameter estimates by different methods for Norwegian data, 1994*

Method	Estimates (Standard deviation)			Theoretical Gini	Empirical Gini
	α	*c* (1,000)	δ (1,000)		
Least-squares, all observations	1.67 (0.02)	56.17 (0.6)	214.08 (2.5)	0.280	
Least-squares, 30 points	1.65 (0.04)	56.02 (0.9)	214.45 (3.2)	0.282	0.275
Koutrouvelis	1.60 (0.02)	54.50 (0.6)	214.80 (2.5)	0.295	

McCulloch	1.57 (0.05)	54.60 (1.4)	214.80 (2.5)	0.308
Maximum-likelihood, 30 points	1.61 (0.02)	52.70 (0.6)	215.60 (2.0)	0.281
Maximum-likelihood censored, 30 points	1.67 (0.02)	49.58 (0.6)	211.76 (1.7)	
# Observations			5,618	

* For the Koutrouvelis and the McCulloch estimation procedures δ is estimated by the sample mean.

Table 2 Parameter estimates of the GB2 distribution, Norwegian data 1994

Method	Estimates (Standard deviation)			
	a	b	p	q
Least-squares, all observations	8.05 (0.4)	213.96 (5.8)	0.30 (0.2)	0.44 (0.2)
Least-squares, 30 points	4.45 (0.4)	290.52 (25.4)	0.58 (1.8)	2.54 (1.8)
Maximum-likelihood, all observations	3.15 (0.4)	191.50 (5.1)	1.24 (0.3)	1.28 (0.2)
Maximum-likelihood, 30 points	1.13 (0.3)	281.17 (25.7)	6.32 (1.9)	9.74 (3.6)

Table 3 Goodness of fit measures

Distribution and estimation method	Root of mean square error	Mean absolute error	Log-likelihood
Pareto-Lévy, least-squares	$0.854 \cdot 10^{-2}$	$0.703 \cdot 10^{-2}$	
GB2, least–squares	$0.709 \cdot 10^{-2}$	$0.682 \cdot 10^{-2}$	
Pareto-Lévy, maximum-likelihood, 30 points	$1.324 \cdot 10^{-2}$	$1.162 \cdot 10^{-2}$	-19057.4
GB2, maximum-likelihood, 30 points	$1.784 \cdot 10^{-2}$	$1.501 \cdot 10^{-2}$	-19070.7

Table 4 Parameter estimates for income distribution in seven OECD countries, LIS data

	Australia	Belgium	Canada	France	Germany	Italy	US
Pareto-Lévy distribution							
α	1.65	1.58	1.61	1.52	1.68	1.55	1.43
	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
δ	44,489.31	1,781,278.00	51,336.04	173,445.43	69,816.83	36,282.59	55,389.86
	(192.38)	(658.50)	(200.40)	(83.55)	(547.09)	(24.30)	(145.60)
c	16,668.57	555,667.10	19,699.72	58,544.26	24,247.63	11,239.83	18,311.78
	(75.34)	(753.20)	(56.20)	(313.82)	(179.82)	(17.30)	(49.24)
GB2							
a	4.57	3.00	4.72	3.91	6.29	3.45	3.65
	(0.29)	(0.26)	(0.16)	(0.31)	(0.62)	(0.18)	(0.01)
b	64,723.92	2,265,405.00	76,520.87	206,761.73	84,662.93	40,039.00	65,792.73
	(806.05)	(62,107.00)	(506.74)	(4,458.50)	(690.40)	(285.80)	(164.36)
p	0.26	0.63	0.22	0.34	0.23	0.53	0.30
	(0.02)	(0.08)	(0.01)	(0.03)	(0.03)	(0.05)	(0.01)
q	0.93	1.56	0.87	0.85	0.56	1.05	0.84
	(0.07)	(0.19)	(0.04)	(0.10)	(0.06)	(0.09)	(0.03)

Table 5 Goodness of fit measures, LIS data

Goodness of fit measure	Australia	Belgium	Canada	France	Germany	Italy	US
Pareto-Lévy distribution							
Log-likelihood	-16,102.1	-8,735.3	-75,244.7	-23,656.1	-13,830.4	-16,555.9	-119,656.3
GB2							
Log-likelihood	-16,152.3	-8,734.8	-74,981.4	-23,675.4	-13,851.5	-16,559.8	-119,377.2

Figure 1 The regression line under the Koutrouvelis estimation procedure

$$x = \ln\lambda, \ \ y = \ln\left(-\ln\left|\hat{\varphi}(\lambda)\right|^2\right)$$

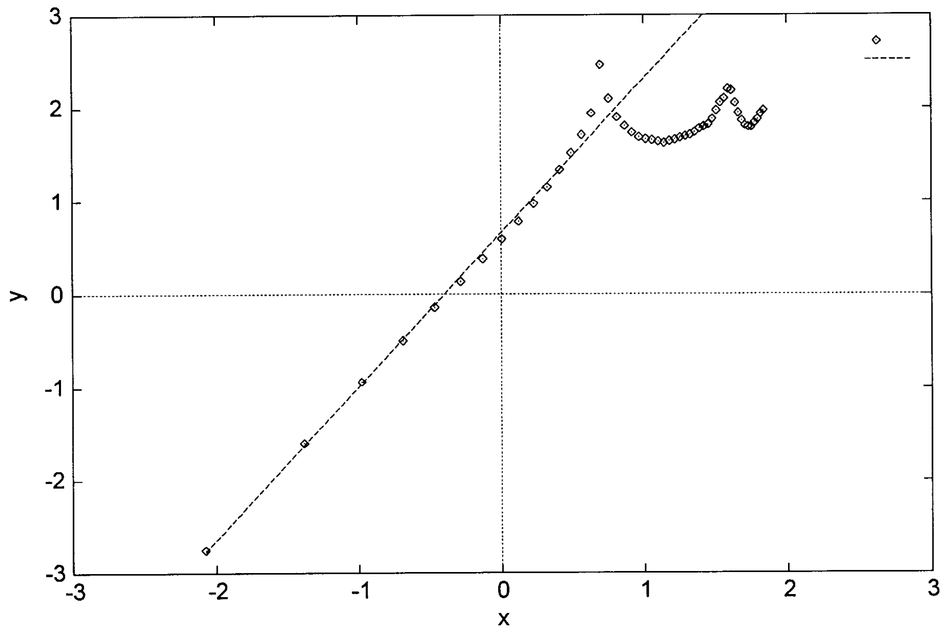

Figure 2 Estimated Pareto-Lévy cdf and the empirical cdf, Norwegian data, 1994

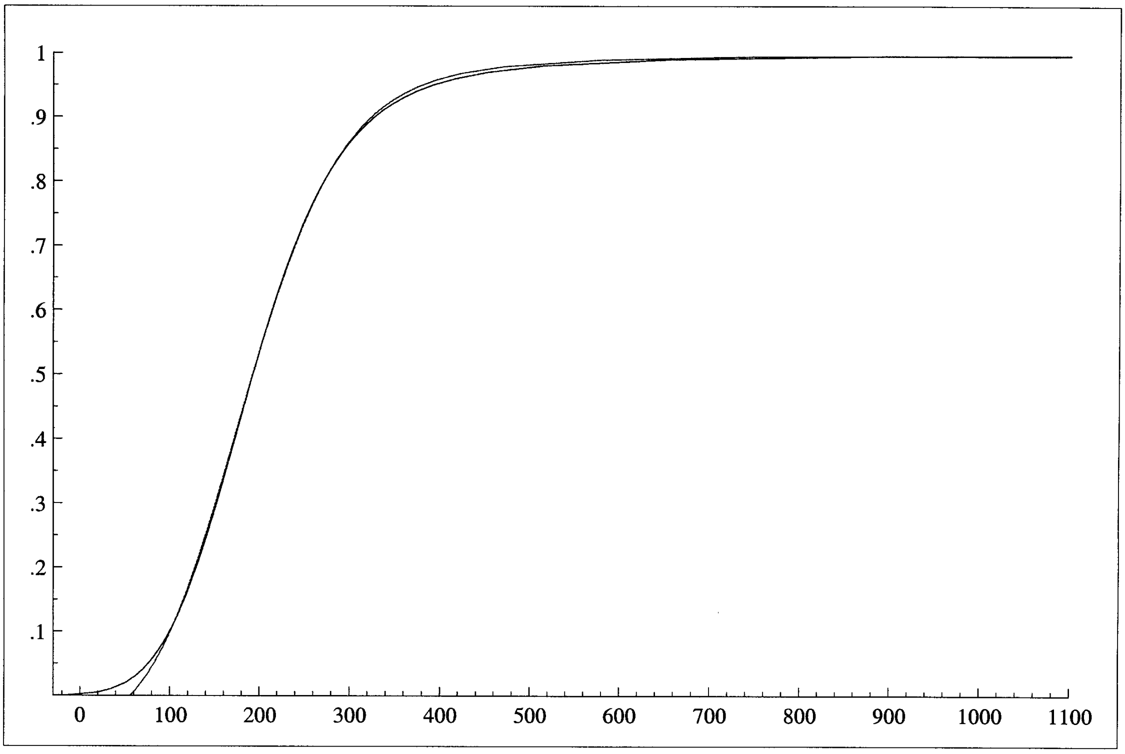

Figure 3 Q–Q plots for two estimation procedures, Norwegian data, 1994

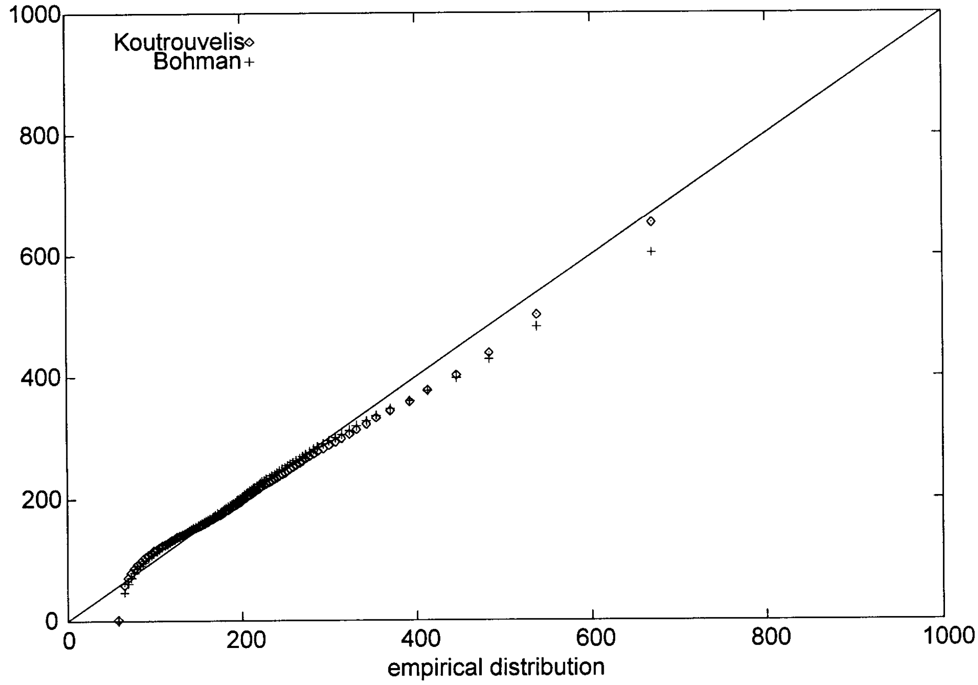

Figure 4 P–P plots for two estimation methods, Norwegian data, 1994