Discussion Papers No. 246, January 1999
Statistics Norway, Research Department

John K. Dagsvik and Bjørn H. Vatne

Is the Distribution of Income Compatible with a Stable Distribution?

Abstract:
Mandelbrot (1961) proposed to apply the class of Pareto-Levy distributions—which belong to the Stable distributions—as a framework for modelling income distributions. He also presented theoretic arguments in favor of the Pareto-Levy distributions. In this paper we provide additional theoretical justification for this class of distributions. We also use micro data on individual market income to estimate the parameters of a Pareto-Levy distribution. Several estimation methods have been applied. The estimated Pareto-Levy distribution appears to fit the data well.

Keywords: Stable distributions, Pareto-Levy distributions, Income distributions

JEL classification: C13, D31

Acknowledgement: We are grateful for comments by Rolf Aaberge.

Address: John K. Dagsvik, Statistics Norway, Research Department. E-mail: john.dagsvik@ssb.no

Discussion Papers comprise research papers intended for international journals or books. As a preprint a Discussion Paper can be longer and more elaborate than a standard journal article by including intermediate calculation and background material etc.

Abstracts with downloadable PDF files of
Discussion Papers are available on the Internet: http://www.ssb.no

For printed Discussion Papers contact:

Statistics Norway
Sales- and subscription service
P.O. Box 1260
N-2201 Kongsvinger

Telephone: +47 62 88 55 00
Telefax: +47 62 88 55 95
E-mail: Salg-abonnement@ssb.no

1 Introduction

This paper discusses the potential of the so-called class of Pareto-Levy distributions as a framework for analyzing income distributions. In the sixties and seventies Mandelbrot wrote a number of papers where he discussed the problem of justifying the stochastic properties of economic variables such as stock market prices and incomes. His argument was that in addition to the effects from behavior of optimizing agents in the market, there are certain aggregation operations that take place, and which have important bearings on the structure of the probability distribution of the relevant variables under investigation. Similarly to the tradition in the physical sciences he was led to postulate specific invariance principles which imply that, for example, the income distribution should belong to the class of stable distributions. Recall that the class of stable distributions has the property that a linear combination of two independent stable random variables is a stable random variable. The class of stable distributions contains the normal distribution as a special case, and it also follows from a general version of the Central Limit Theorem without the condition of bounded variances. Recall that the condition of bounded variance is essential in the classical Central Limit Theorem. More precisely, Mandelbrot restricted attention to a subclass of stable distributions, namely the Pareto-Levy laws. A Pareto-Levy distribution has the property that the probability mass of large negative values becomes negligeable (within the stable class). The rationale for the focus on the Pareto-Levy law is the need to take into account the fact that "income" is a non-negative variable.

So far the Pareto-Levy distributions have not received much attention in the context of analyzing the distribution of income. There may be several reasons for this. One reason is that the estimation of the parameters of a stable distribution requires the use of nonstandard inference procedures, since one cannot express the probability density of a stable distribution on closed form except in a few special cases. A second reason may be that the results from the few empirical studies that have been carried out have been ambiguous. For example Dijk and Kloek (1980) have concluded that the Pareto-Levy distribution does not produce a particularly good fit to the data. Third, the theoretical arguments given by Mandelbrot (1960), and which we shall discuss shortly, may not have seemed convincing to many researchers.

In this paper we maintain that the Pareto-Levy class of distribution has a number of attractive properties. Specifically, a Pareto-Levy distribution depends on three parameters that have a clear statistical interpretation which we shall discuss further below (Section 6). Since the sum of Pareto-Levy distributed variables has Pareto-Levy distribution, a model based on this framework will consequently not depend critically on whether the income concept is based on monthly or yearly income, provided the parameters that measure the right tail thickness are equal. It is also of interest to note that the variables do not necessarily have to be independent. We also propose a theoretical justification for the Pareto-Levy class of distributions which differs somewhat from the arguments given by Mandelbrot (1960). Subsequently, we apply a large set of micro-data on income from the wage sector to fit a Pareto-Levy distribution. This seems to be the first time micro data on incomes have been used to estimate a Pareto-Levy distribution of income.

From a economic point of view the present analysis is limited in that it does not propose a story of how the parameters of the income distribution relate to underlying economic determinants. However, we claim that since the Pareto-Levy distributions can be represented by three parameters which have a clear statistical interpretation, and these distributions also are supported by plausible invariance properties, this representation may be of substantial interest for assessing the relation between key economic variables and the parameters of the income distribution.

The paper is organized as follows: In the next section we discuss a possible theoretical rationale which supports the class of Pareto-Levy distributions. In Section 3 we describe the data and in Section 4 we discuss different estimation methods. In Section 5 we report empirical results, and in Section 6 we discuss the relationship between the parameters of the Pareto-Levy class and aggregate means of income inequality.

2 Theoretical considerations

Apparently, the first one to put forward theoretical considerations about the properties of the income distribution was Pareto (1897). His concern was to provide a justification for the properties of the right tail of the distribution that corresponds to empirical income distributions. It was subsequently recognized that the so-called Pareto distribution that followed from Pareto's arguments often fitted the right tail of empirical income distributions, but gave a poor fit of other parts of the distributions. In a series of papers, Mandelbrot proposed a generalization of the Pareto distribution, namely what he called the Pareto-Levy distributions, as a framework for analyzing income distributions. The Pareto-Levy distributions constitute a subclass of the class of stable distributions. The stable class follows from the Central Limit Theorem in the most general version. Specifically, this class has the property that if Y_1 and Y_2 are independent stable random variables, then Y_3 given by

$$Y_3 = aY_1 + bY_2, \tag{1}$$

where a and b are constants, is stable. 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 stable class. The stable class was thoroughly investigated by Paul Levy in the thirties. Except for a few cases, the probability distributions in the stable class cannot be expressed on closed form. Their characteristic function, however, can be expressed as

$$\phi(\lambda) = Ee^{i\lambda Y} = \exp\left(i\lambda\mu - \sigma^\alpha |\lambda|^\alpha (1 - i\beta \operatorname{sign}(\lambda)) \tan\left(\frac{\alpha\pi}{2}\right)\right), \tag{2}$$

where Y is a stable random variable, $\sigma > 0$, and μ are scale and location parameters, $\beta \in [-1, 1]$ is a parameter that represents the skewness of the distribution, α represents the thickness of the tail of the distribution and the formulae (2) holds for all $\alpha \in (0, 2]$ except for $\alpha = 1$. When $\alpha = 1$, the characteristic function is given by a similar formulae, cf. Samorodnitsky and Taqqu (1994). When $\beta = 0$ the distribution is symmetric while it is maximally skew to the right (left) when $\beta = 1\,(\beta = -1)$. When $\alpha = 2$, we get the normal distribution and the parameter β vanishes. When $\alpha \in (1, 2)$ the variance of the distribution is infinite and the expectation equals μ, while both the variance and the expectation are infinite when $\alpha \in (0, 1]$. Thus, within the stable class the only member that possesses a finite variance is the normal distribution. In Mandelbrot's definition, the Pareto-Levy class consists of the subclass of stable distributions with $\alpha \in (1, 2)$, and $\beta = 1$. In this case the probability mass of negative values will be negligible for sufficiently large μ.

Apparently, very few researches have taken Mandelbrot's idea of applying the Pareto-Levy class as a framework for analyzing income distributions seriously. The reasons for this are not clear. Our conjecture is that this state of affair may be related to the following arguments:

(i) The stable distributions are difficult to estimate since one cannot, in general, express the stable densities on closed form.

(ii) Since we never observe empirical distributions with infinite variances, it may at first glance seem somewhat awkward to apply theoretical distributions with infinite variance.

(iii) The class of stable distributions is not flexible enough to fit typical empirical income distributions.

(iv) The theoretical arguments provided by Mandelbrot to support the choice of the stable distributions are not entirely convincing.

Let us now take a closer look at these arguments. The first argument (i) is no longer a serious objection because there now exists several estimation methods which work well. The second argument (ii) is not very relevant either. The variance is just a mathematical expression that should not be taken too literally in every case. The situation is somewhat similar to the following example: the normal distribution has infinite support while empirical distributions have finite support, but this does not prevent us from usefully applying this distribution in a large number of cases. It remains to be seen to which extent the third argument is valid; we intend to demonstrate in the present paper that in our chosen empirical application the stable distribution fits the data rather well. The most important argument is perhaps the fourth one. Mandelbrot (1960) argued as follows: If one considers income as the variable of interest, this variable may be decomposed into different kinds of incomes such as income from wage work, self-employment work, capital income, etc. If one considers the distribution of each of the income components, and total income, it appears, according to Mandelbrot, that these distributions have more or less the "same shape". Given this point of departure, Mandelbrot is lead to postulate that the distribution of the sum of (independent) income components (which by assumption have distributions belonging to the "same class"), should also belong to the same class. Hence, provided that the income components are stochastically independent, one obtains the stable class. A difficulty with the argument above is that it is rather vague about what is actually meant by "the same shape". For example, for eq. (1) to hold it is necessary that both the distribution of Y_1 and Y_2 are stable with the same α. In many cases this simply does not seem to be true. For example, our own empirical investigations seem to indicate that the distribution of capital income or income from self-employment have a smaller α than the α associated with the distribution of income from wage work. It therefore seems desirable to establish further theoretical arguments to support the relevance of the stable distributions. This is the topic to be discussed next.

Let $Y(t)$ denote income at time t. By time we shall understand the age of the individual.

Assumption 1 *The income process* $\{Y(t), t \geq 0\}$, *can at each point in time, be expressed as*

$$Y(t) = \mu_t + a_t V(t) \tag{3}$$

where $\mu_t = EY(t)$, $a_t \geq 0$, *is a deterministic function of* t *and* $V(t)$ *is a random variable with (marginal) distribution function that is independent of* t *for each given* t. *Moreover, the distribution of* $\{V(t), t \geq 0\}$ *is independent of* $\{\mu_t, t \geq 0\}$, *and* $\{a_t, t \geq 0\}$.

An example of processes which satisfy Assumption 1 is the class of self-similar processes. Recall that a process $\{X(t), t \geq 0\}$ is self-similar if for any scalar $a > 0$, $X(at)$ has the same distribution as $a^H X(t)$, where H is a positive constant. The intuition is that in the context of income as well as many other cases it may be plausible to assume that apart from a scale transform the properties of the process is invariant with respect to the (ratio) scale representation of time. If for example $Y(t) - \mu_t$ is self-similar if $Y(t) - \mu_t$ has the same distribution as $t^H (Y(1) - \mu_1)$. In this case Assumption 1 holds with $a_t = t^H$.

In Section 6 we report empirical evidence that support Assumption 1. Note that Assumption 1 does not imply that the process $\{V(t), t \geq 0\}$ is stationary since it only concerns the one-dimensional marginal distributions. The scalar a_t will in general depend on the properties of the law P of the stochastic process $\{Y(t) - \mu_t, y > 0\}$. Let P denote this law. Thus, we shall write $a_t = a_t(P)$ to indicate that a_t depends on the law P. Let P^* denote the special case of

P which corresponds to a process with stationary independent increments, and let $\mathcal{D}$ denote be the class of possible laws for the process $\{Y(t) - \mu_t,\ t > 0\}$.

Assumption 2 *$P^* \in \mathcal{D}$.*

Assumption 2 means that a process with stationary independent increments is a possible income process.

Theorem 1 *Under Assumptions 1 and 2, the income process has stable marginal distribution functions.*

Proof. By Assumption 1 we get

$$
\begin{aligned}
Ee^{i\lambda Y(t)} &= e^{i\lambda\mu_t} Ee^{i\lambda a_t(P)V(t)} \\
&= e^{i\lambda t\mu_t}\psi\left(a_t(P)\lambda\right)
\end{aligned}
\tag{4}
$$

for $\lambda \in R$, where

$$\psi(\lambda) \equiv Ee^{i\lambda V(t)} = Ee^{i\lambda V(0)}. \tag{5}$$

Since $\psi(\lambda)$ is the characteristic function of $V(t)$ it is by Assumption 1 independent of t and of $a_t(P)$. By Assumption 2 $\psi(a_t(P^*)\lambda)$ is the characteristic function of a random variable that is the sum of i.i.d. random variables. Let b_t be a norming constant such that

$$\lim_{t\to\infty}\psi\left(\lambda a_t(P^*)/b_t\right)$$

exists. Then since

$$\lim_{t\to\infty}\psi\left(\lambda a_t(P^*)/b_t\right) = \psi\left(\lambda \lim_{t\to\infty} a_t(P^*)/b_t\right)$$

it follows that $\psi(\lambda)$ is the characteristic function of a stable random variable due to Levy's general version of the Central Limit Theorem. (See for example Gnedenko and Kolmogorov (1954).) But then, by Assumption 1, it also follows that $Y(t)$ is stable. ■

As pointed out by Mandelbrot, not every member of the stable class is relevant in the context of income distributions. This is due to the fact that "income" sa a non-negative variable. Thus, to ensure that the probability mass on the negative part of the real line is negligible we shall, as mentioned above, only consider the Pareto-Levy class of distributions, that is the stable distributions for which the mean is finite and which are maximally skew to the right. It follows from the theory of stable laws that β must be equal to one to ensure that the probability mass become negligible on the negative part of the real line (for sufficiently large value of the mean), cf. Samorodnitsky and Taqqu (1994). Thus, in the light of the argument above the Pareto-Levy class should be appealing from a theoretical point of view. It therefore remains to be investigated how well the Pareto-Levy distribution conforms with empirical evidence.

3 The income data

The data used in this paper is based on the Norwegian income and wealth data for 1994 gathered by Statistics Norway described in Pedersen (1997). We have chosen to use market income for individuals, defined as the sum of wage income, income from self-employment and capital income. As population we have chosen the persons that have a total income larger than the minimum benefit from the social security system. In 1994 this amount was 60 701 NOK. The population consists of 5618 persons. The empirical income distribution is given in Figure 1.

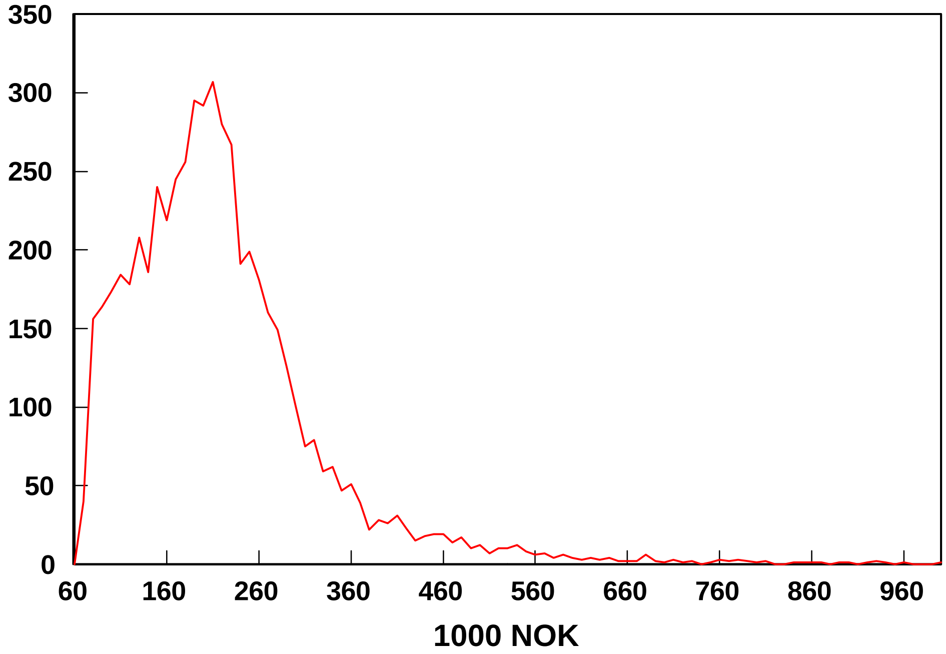
4 Estimation methods

There are many methods suggested in the literature for estimating the parameters of the stable distributions. In this section we shall discuss a few of these methods.

4.1 Estimation of α based on the tails and fractiles of the empirical distribution

There are several versions of this method. The naive approach, suggested by Pareto, consists in plotting $-\ln(1-F(Y_j))$ against $\ln Y_j$, where $Y_j,\ j=1,2,\dots n$, are the observations generated from $F(\cdot)$. If the underlying distribution has a Pareto right tail then this shows up as a straight line in the Pareto quantile plot. The Pareto quantile plot of the income distribution is displayed in Figure 3. The slope of the Pareto quantile plot provides an estimate for $1/\alpha$. The estimate of the slope using the upper 55 per cent of the data is 0.326 which gives a value of α of about 3.1.

Figure 1: Distribution of market income 1994



This estimate is inconsistent with the class of stable distributions. However, this estimate does not "prove" that the underlying distribution is non-stable. According to McCulloch (1997) "tail index estimates greater than 2 are to be expected for stable distributions with α as low as 1.65." On the other hand DuMouchel (1983) states that "When the true distribution is not stable the estimates of α is not a robust measure of the rate of decrease of the tail probabilities".

A more sophisticated method based on specific fractiles of the empirical distribution was derived by McCulloch (1986). McCulloch (1986) proposes a method based on the 5%, 25%, 50 %, 75% and 95% fractiles of the empirical distribution function. This method is remarkably simple to use and requires only the knowledge of the fractiles mentioned above and access to a set of tables given in McCulloch's paper.

4.2 Methods based on the empirical characteristic function

There are several estimation methods that are based on the empirical characteristic function (2). The empirical version of the characteristic function of a distribution is defined as

$$\widehat{\phi}(\lambda) = \frac{1}{n}\sum_{j=1}^{n} e^{i\lambda Y_j}. \tag{6}$$

The empirical characteristic function (6) is an unbiased estimator of the characteristic function, i.e.

$$E\widehat{\phi}(\lambda) = \phi(\lambda) \equiv Ee^{i\lambda Y_1}, \tag{7}$$

and the variance is given by

$$\operatorname{Var}\left[\widehat{\phi}(\lambda)\right] = \frac{1}{n}\operatorname{Var}\left[e^{i\lambda Y_1}\right] = \frac{\phi(2\lambda) - \phi(\lambda)^2}{n}. \tag{8}$$

Equation (8) demonstrates that the empirical characteristic function yields a sharp estimate of $\phi(\lambda)$ when n is large. However, it is not enough to consider convergence for a given λ but rather the properties of the *empirical process* $\left\{\widehat{\phi}(\lambda), \lambda \in R\right\}$, and how the parameters of the mean function $\phi(\lambda)$ can be recovered from $\left\{\widehat{\phi}(\lambda), \lambda \in R\right\}$. One difficulty here is that we observe the process $\left\{\widehat{\phi}(\lambda), \lambda \in R\right\}$ for all real λ and it is difficult to device optimal estimation procedures that take all this information into account. Paulson, Holcomb, and Leitch (1975) have discussed maximum likelihood procedures based on the fact that the process $\left\{\widehat{\phi}(\lambda), \lambda \in R\right\}$ is asymptotically a Gaussian process. Specifically, they base the inference on the likelihood of $\left\{\widehat{\phi}(\lambda_k), k = 1, 2, \ldots, m\right\}$ when the Gaussian approximation is assumed to hold, where $(\lambda_1, \lambda_2, \ldots, \lambda_m)$ are suitably selected real numbers.

In this paper we shall instead review an alternative and remarkably simple method proposed by Koutrouvelis (1980). Koutrouvelis noted that (2) implies that

$$\ln\left(-\ln|\phi(\lambda)|^2\right) = \ln(2\sigma^\alpha) + \alpha\ln|\lambda| \tag{9}$$

which yields

$$\ln\left(-\ln\left|\widehat{\phi}(\lambda)\right|^2\right) = \ln(2\sigma^\alpha) + \alpha\ln|\lambda| + \eta(\lambda) \tag{10}$$

where $\eta(\lambda)$ is a random error term which has approximately zero mean when n is large. By choosing an appropriate set of λ-values, we realize that it is possible to estimate $\ln(2\sigma^\alpha)$ and α by regression analysis with $\{\ln|\lambda_k|\}, k = 1, 2, \ldots, m$, as independent variable and

$$\left\{\ln\left(-\ln\left|\widehat{\phi}(\lambda_k)\right|^2\right)\right\}$$

as dependent variable. It is easy to show that $\left|\widehat{\phi}(\lambda)\right|^2$ can be expressed as

$$\left|\widehat{\phi}(\lambda)\right|^2 = \frac{1}{n^2}\sum_{j,k}\cos(\lambda(Y_j - Y_k)). \tag{11}$$

Koutrouvelis (1980) has done simulation experiments to demonstrate that this approach works well and is quite efficient provided that 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 σ can be estimated. Since $\beta = 1$ in the Pareto-Levy class the expectation μ is the only parameter that remains. This parameter can therefore be estimated by the corresponding sample mean. Moreover, since the assumption of stability implies that the relation expressed in (10) is linear, one obtains an informal test of the stability assumption by plotting $\left\{\ln\left(-\ln\left|\widehat{\phi}(\lambda_k)\right|^2\right)\right\}$ against $\{\ln|\lambda_k|\}$. If this plot is approximately linear this indicates that the underlying distribution is stable.

On the basis of (10) Koutrouvelis estimated the parameters by using the method of ordinary least squares. This method produces consistent estimates but the estimated standard deviations will be biased due to the fact that $\widehat{\phi}(\lambda_k)$ and $\widehat{\phi}(\lambda_j)$ are dependent.

In the present case where the distribution is maximally skew to the right an interesting modification of the Koutrouvelis method applies. This approach is based on the property that the two-sided Laplace transform of $F(y)$ exists and is equal to

$$g(z) \equiv Ee^{-zY} = \exp(cz^\alpha - \mu z) \tag{12}$$

for $z \geq 0$, where

$$c = -\frac{\sigma^\alpha}{\cos\left(\frac{\alpha\pi}{2}\right)}. \tag{13}$$

Note that since $\alpha > 0$, c will be positive. The corresponding empirical process equals

$$\widehat{g}(z) = \frac{1}{n}\sum_{j=1}^{n} e^{-zY_j} \tag{14}$$

which provides a consistent and unbiased estimate of $g(z)$. From (12) it follows that

$$\ln \widehat{g}(z_j) = cz_j^{\alpha} - \mu z_j + \theta(z_j) \tag{15}$$

where $z_1, z_2, \ldots, z_m$, are suitable chosen positive real numbers and $\theta(z_j)$ is a random variable with $E\theta(z_j) \approx 0.$ when n is large. By running nonlinear regression analysis c, α and μ can be estimated. If μ is estimated by the empirical mean one can estimate α and $\ln c$ by linear regression since

$$\ln(\ln \widehat{g}(z_j) - \widehat{\mu}) = \alpha \ln z_j + \ln c + \nu(z_j) \tag{16}$$

where $\nu(z_j)$ is a random term with $E\nu(z_j) \approx 0$, and $\widehat{\mu}$ denotes the empirical mean.

4.3 Estimation by inversion of the empirical characteristic function

Bohman (1975) suggested a discrete version of the Fourier inversion procedure. In this procedure F is approximated by F^* given by

$$F^*(y) = \frac{1}{2} + \frac{\eta y}{2\pi} - \sum_{\nu=1-H,\nu\neq 0}^{H-1} \frac{\phi(\eta\nu)}{2\pi i\nu} e^{-i\eta\nu y} \tag{17}$$

where H and ν are chosen in a suitable manner, and $\phi(\cdot)$ is the characteristic function of $F(\cdot)$. The parameters are estimated by minimizing

$$\sum_j \left[W(Y_j)\left(F^*(Y_j) - \widehat{F}_n(Y_j)\right)\right]^2, \tag{18}$$

where $\widehat{F}_n(\cdot)$ is the empirical cumulative distribution, $Y_1, Y_2 \ldots$, are the observations and $W(\cdot)$ is a suitable weight function. This method has certain advantages compared to the methods based on the empirical characteristic function: It provides a method for estimating the parameters by putting most weight on the part of the distribution that is considered the most important.

Table 1: Parameter estimates by different methods

	Estimates (Standard deviation)					
Method	α		σ		μ	
Bohman	1.69	(0.024)	56.5	(0.577)	214.7	(2.454)
Koutrouvelis	1.60	(0.023)	54.5	(0.558)	214.8	(2.452)
McCulloch	1.57	(0.047)	54.6	(1.395)	214.8	(2.452)

This method can also be used on grouped data. The method is computationally efficient as the number of observations does not enter the computation of $\widehat{F}_n(x)$.

Let us now consider the problem of evaluating the performance of the estimators. The calculation of standard deviation of the parameters is not straight forward for the Koutrouvelis' and Bohman's inversion method. The precision of the Koutrouvelis' method will depend on the choice of $\{\lambda_k\}$. However, conditionally on the choice of $\{\lambda_k\}$ one can find asymptotic estimates of the standard deviation of the parameter estimates by carrying out the procedure of generalized least squares. How this method works in small samples is not known. We have therefore chosen to estimate standard deviations by applying the bootstrap approach. For the inversion method the bootstrap approach is currently the only alternative which is also easy to apply.

5 Estimation results and Goodness of fit

As indicated in Section 4.2 one can obtain an informal test of stability by plotting the left hand side of (11) against $\ln|\lambda|$

This plot is displayed in Figure 2 where the data are scaled as suggested in Koutrouvelis (1980). From the figure we realize that the plot is fairly linear up to $\ln|\lambda| \approx 0.7$. The question is now whether or not it is sufficient to consider λ-values in the interval determined by $\ln|\lambda| < 0.7$.

In the appendix we demonstrate that provided the (true)characteristic function satisfies eq. (9) then it is sufficient to use the information represented by the empirical characteristic function for arguments with absolute value less than 0.7. The estimates of the parameters of the income distribution is displayed in Table 1. The reported standard deviations are obtained from 25

Figure 2: Pareto quantile plot of the income distribution

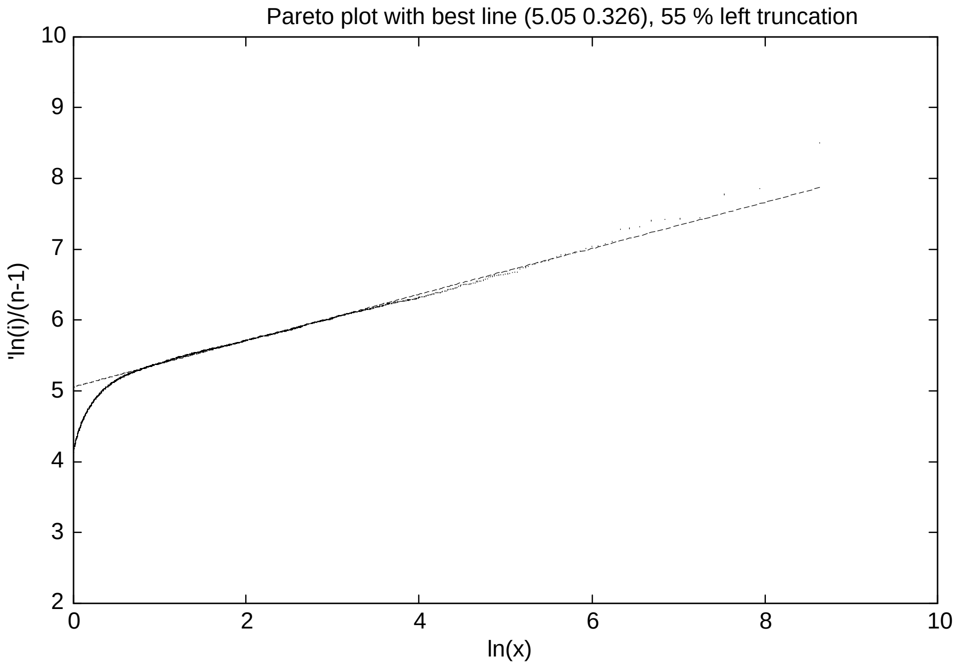


Figure 3: Observed and simulated densities based on the parameter estimates obtained by different estimation procedures

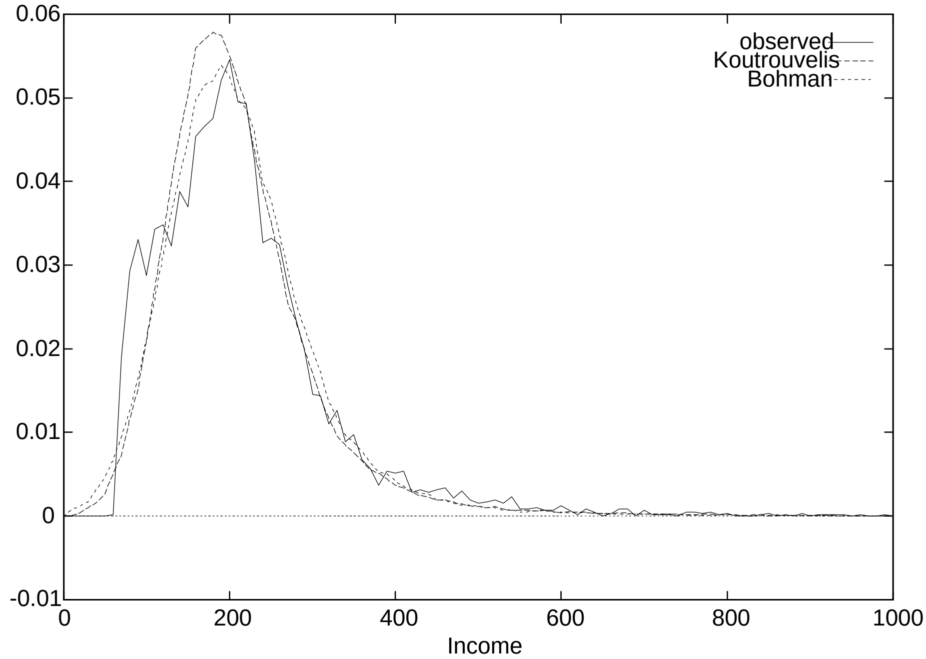


Figure 4: Q-Q plot for two estimation procedures

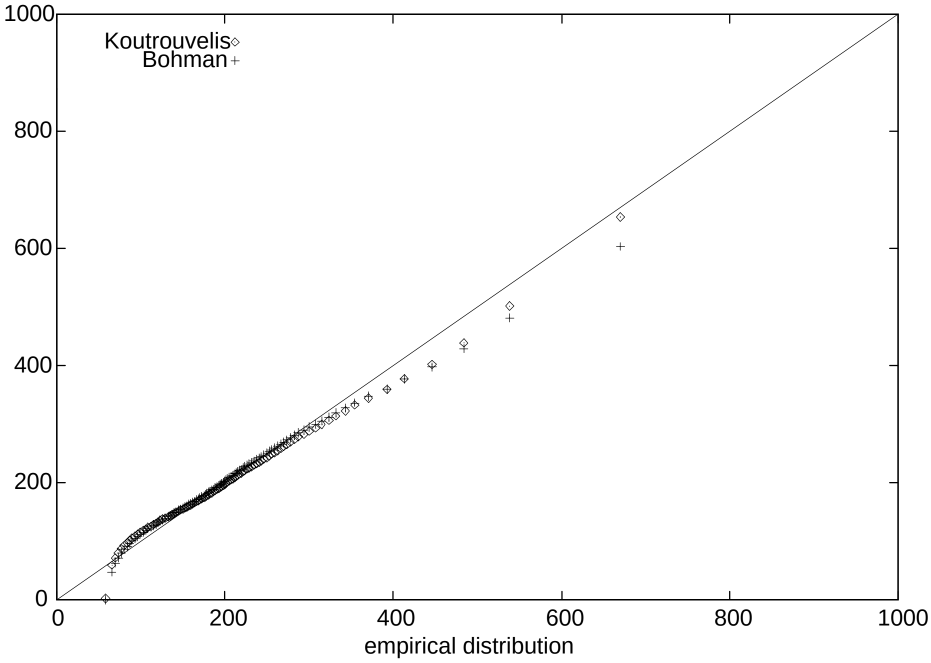


Figure 5: P-P plot for two estimation methods

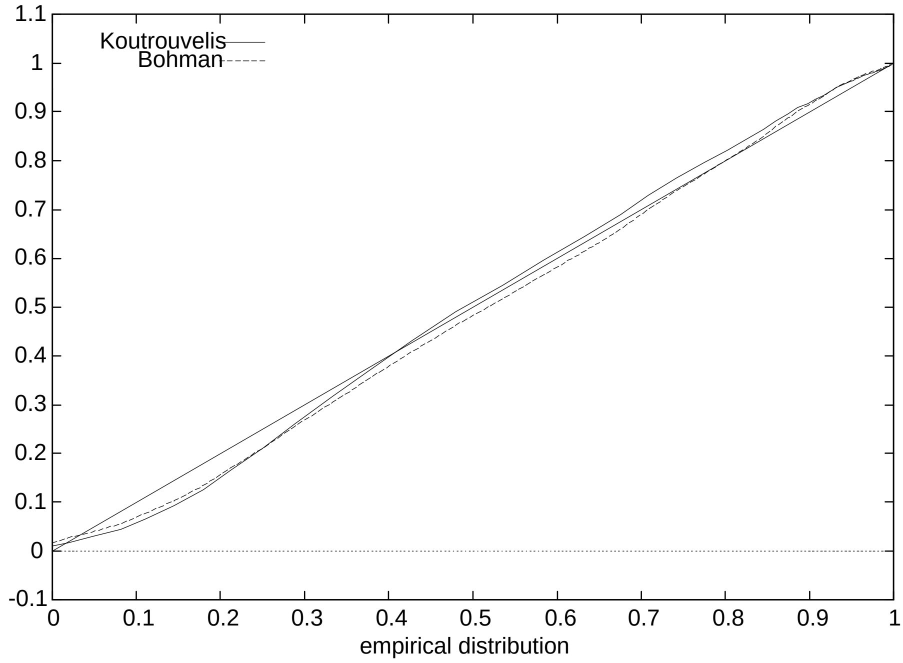
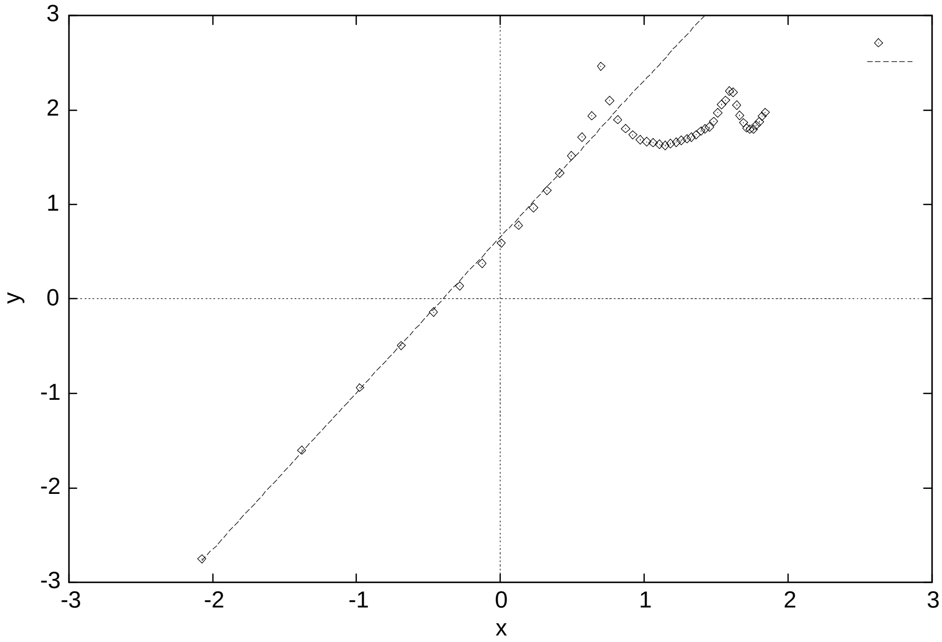


Figure 6: The regression line under the Koutrouvelis estimation procedure

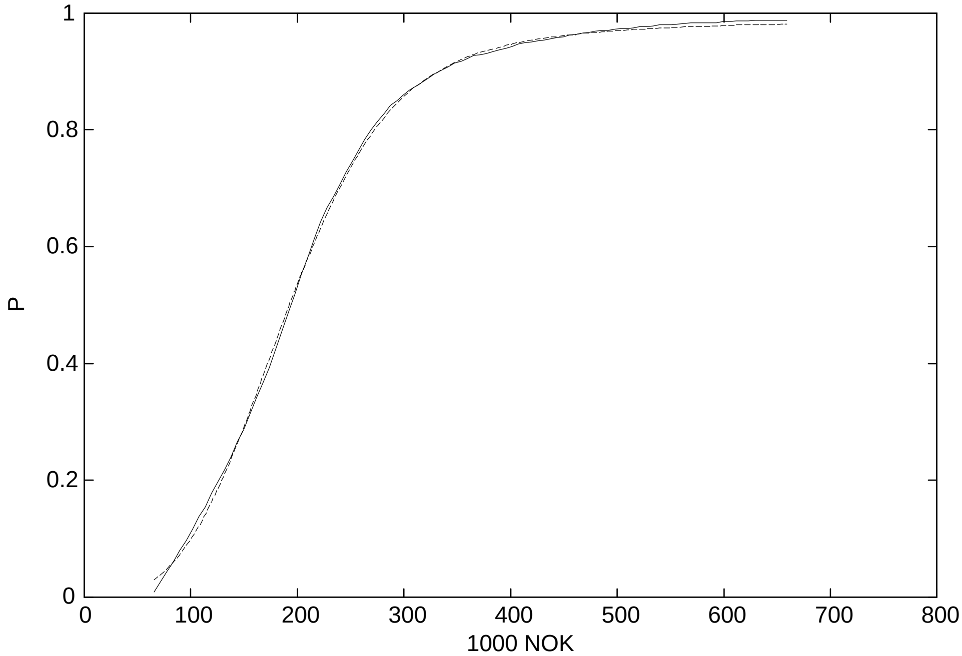


bootstrap replications.

From Table 1 we see that the three estimation methods we have considered in this paper preforms well and yield quite similar estimates. According to Table 1 it seems that McCulloch's method produces considerable higher standard deviation of the estimates of α and σ than the other methods.

In Figure 3 we display the empirical density together with the fitted Pareto-Levy densities estimated by Bohman's and Koutrouvelis' procedure, respectively. To assess how well the Pareto-Levy distribution fit the data we have displayed the cumulative empirical and the fitted cumulative Pareto-Levy distribution estimated by Bohman's inversion method. The Pareto-Levy density and cumulative distribution have been simulated by means of a procedure developed by Chambers, Mallows, and Stuck (1976). In Figure 4 and 5 we show Quantile-Quantile (Q-Q) plot and Percentile-Percentile (P-P) plot of the cumulative empirical distribution functions against the fitted Pareto-Levy distribution. We note that these plots are close to straight lines which indicates a good fit. The corresponding Kolmogorov-Smirnov statistics is displayed

Figure 7: Estimated Pareto-Levy cumulative (solid) and the empirical cumulative distribution function (dashed). Bohman's method

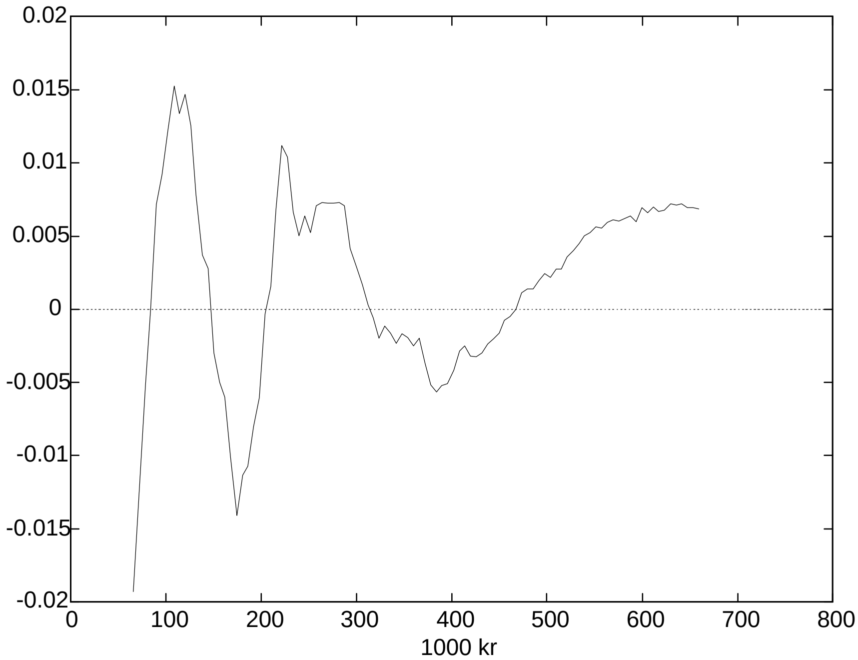


in Figure 8. We note that the Kolmogorov-Smirnov statistics does not exceed 0.015 when the income is above 100 000 NOK. Although we have plotted the corresponding empirical and estimated densities in Figure 3, is not of much interest to use this figure as an indication of the fit because the empirical density provides a rather imprecise estimate of the underlying theoretical density.

6 Aggregate measures of income inequality

In assessing the significance of changes in the income distribution, $F(y)$, aggregate measures of income inequality is often employed. The most common measure is the Gini-coefficient which

Figure 8: The Kolmogorov-Smirnov statistics



can be expressed as

$$G = \frac{\int_0^\infty F(y)(1 - F(y))\, dy}{\int_0^\infty (1 - F(y))\, dy}. \tag{19}$$

Another summary measure is the A-coefficient which can be expressed as

$$A = -\frac{\int_0^\infty F(y) \ln F(y)\, dy}{\int_0^\infty (1 - F(y))\, dy}. \tag{20}$$

We refer to Aaberge (1995) for a discussion of the properties of (20).

In this section we shall consider the properties of G and A when $F(y)$ is a Pareto-Levy distribution. Specifically, we shall demonstrate how G and A depends on the parameters of the Pareto-Levy distribution. Consider first the distribution of the Gini coefficient. From (19) we get

$$\int_0^\infty \left(F(y) - F(y)^2\right) dy = \int_0^\infty \left(1 - F(y)^2\right) dy - \int_0^\infty (1 - F(y))\, dy.$$

Note next that if Y_1 and Y_2 denote two i.i.d. random variables with c.d.f. $F(y)$, then $F(y)^2$ is the distribution of $\max(Y_1, Y_2)$, which implies that

$$\int_0^\infty \left(1 - F(y)^2\right) dy = E \max(Y_1, Y_2).$$

Hence we can express the Gini coefficient as

$$G = \frac{E \max(Y_1, Y_2) - EY_3}{EY_1} = \frac{E(\max(Y_1 - Y_3, Y_2 - Y_3))}{EY_1}, \tag{21}$$

where Y_3 is a random variable with distribution $F(y)$, which is independent of Y_1 and Y_2. Since Y_1, Y_2 and Y_3 are i.i. Pareto-Levy distributed it follows that $Y_1 - Y_3$ and $Y_2 - Y_3$ each have a distribution that is symmetric and stable, cf. Samorodnitsky and Taqqu (1994). We therefore must have that

$$E(\max(Y_1 - Y_3, Y_2 - Y_3)) = \frac{1}{2} E|Y_1 - Y_3|,$$

which yields

$$G = \frac{E|Y_1 - Y_3|}{2EY_1}. \tag{22}$$

Moreover, it follows that $Y_1 - Y_3$ has zero mean and skewness parameter, and dispersion parameter equal to $\sigma 2^{1/\alpha}$. From Samorodnitsky and Taqqu (1994), p.18, we thus find that

$$E|Y_1 - Y_3| = \frac{2\sigma\Gamma\left(1 - \frac{1}{\alpha}\right) 2^{1/\alpha}}{\pi}, \tag{23}$$

from which it follows that G can be expressed as

$$G = \frac{\sigma\Gamma\left(1 - \frac{1}{\alpha}\right) 2^{1/\alpha}}{\mu\pi}. \tag{24}$$

Consider next the A coefficient. Recall that when $F(y)$ is Pareto-Levy, it can be expressed as

$$F(y) = F_0\left(\frac{y - \mu}{\sigma}\right) \tag{25}$$

where $F_0(y)$ is a Pareto-Levy distribution with $\sigma = 1$ and $\mu = 0$. Consequently, if follows from (20) that the A coefficient can be expressed as

$$A = \frac{\sigma}{\mu} h(\alpha) \tag{26}$$

where

$$h(\alpha) \equiv -\int_0^\infty F_0(y) \ln F_0(y)\, dy. \tag{27}$$

Note that $h(\alpha)$ does not depend on σ and μ.

It is also well known that if the income distribution is lognormal then the Gini-coefficient is equal to $\sigma\sqrt{2}/\mu$. We can therefore conclude that provided α remains constant over time the Gini coefficient is proportional to the A coefficient, and also proportional to the Gini coefficient obtained under the lognormal distribution[1]. Consequently, unless α changes, the A and G coefficient will show the same trend.

It is easy to show that this property also holds in the more general case in which Assumption 1 holds. Strøm, Wennemo, and Aaberge (1993) have estimated the A and G coefficient on Norwegian micro data for the period 1973 to 1990 for selected population groups. In Appendix B we report estimates of the ratio G/A. Unfortunately, we are unable to estimate standard errors with the data that are available to us. Appendix B shows that the G/A ratio is remarkably constant over this 17-year time span. Thus, according to the result of Appendix B Assumption 1 is not rejected.

7 Conclusions

In this paper we have discussed the use of the Pareto-Levy class of distributions a framework for analyzing income distributions. We have demonstrated how this class can be justified from theoretical invariance principle. We have subsequently applied different methods to estimate the parameters of a Pareto-Levy distribution from a large sample of Norwegian microdata on income. Several estimation methods have been compared. The resulting estimated Pareto-Levy distribution provides an excellent fit to the empirical distribution.

[1] The A coefficient is sensetive to changes in the lower part of the distribution while the Gini coefficient is more senitive to changes in the upper part.

References

Aaberge, R. (1995). Choosing measures of inequality for empirical applications. Discussion Paper 158, Statistics Norway.

Abramowitz, M. and I. A. Stegun (1972). *Handbook of Mathematical Functions.* New York: Dower publications, Inc.

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

Chambers, J. M., C. L. Mallows, and B. W. Stuck (1976). A method for simulating stable random variables. *Journal of the American Statistical Association 70*, 340–344.

Dijk, H. K. V. and T. Kloek (1980). Inferential procedures in stable distributions for class frequency data on incomes. *Econometrica 48*, 1139–1147.

DuMouchel, W. H. (1983). Estimating the stable index alpha in order to measure tail thickness: A critique. *The Annals of Statistics 11*, 1019–1031.

Gil-Pelaez, J. (1951). Note on the inversion theorem. *Biometrika 38*, 481–482.

Gnedenko, B. V. and A. N. Kolmogorov (1954). *Limit Distributions for Sums of Independent Random Variables.* Reading, Massachusetts: Addison Wesley. (English translation by K. L. Chung).

Koutrouvelis, I. A. (1980). Regression-type estimation of the parameters of stable laws. *Journal of the American Statistical Association 75*, 981–928.

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

McCulloch, J. H. (1986). Simple consistent estimators of stable parameters. *Communications in Statistics- Simulation and Computation 15*, 1109–1136.

McCulloch, J. H. (1997). Measuring tail thickness to estimate the stable index alpha: A critique. *Journal of Business and Economic Statistics 15*, 74–81.

Pareto, V. (1897). *Cours d'Economie Politique.* Lausanne and Paris: Rouge and Cie.

Paulson, A. S., E. W. Holcomb, and R. A. Leitch (1975). The estimation of the parameters of the stable laws. *Biometrika 62*, 163–170.

Pedersen, V. (1997). The survey of income and wealth 1994. Notater 14, Statistisk sentralbyrå. (in norwegian).

Samorodnitsky, G. and M. S. Taqqu (1994). *Stable Non-Gaussian Random Processes*. New York: Chapman and Hall.

Strøm, S., T. Wennemo, and R. Aaberge (1993). Income inequality in Norway 1978-1990. Reports 17, Statistics Norway. (In Norwegian).

Appendix A

An estimate of the approximation error of distribution functions in terms of the approximation error of the corresponding characteristic function

In this appendix we demonstrate that one may utilize the information represented by the empirical characteristic function solely for arguments $|\lambda| < b$, where $b > 0$, is a suitable real number, provided $\widehat{\phi}(\lambda)$ satisfies (9). To this end let

$$\widetilde{\phi}_b(\lambda) = \begin{cases} \widehat{\phi}(\lambda) & \text{for } |\lambda| < b \\ 0 & \text{otherwise.} \end{cases} \tag{28}$$

This means that $\widetilde{\phi}_b(\lambda)$ is the Fourier Stieltjes transform of $\widetilde{F}_b(y)$ where $\widetilde{F}_b(y)$ is defined by

$$\widetilde{F}_b(y) = \frac{1}{2} - \frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{e^{-iy\lambda}\widetilde{\phi}_b(\lambda)\,d\lambda}{\lambda} = \frac{1}{2} + \frac{1}{2\pi i}\int_0^b \frac{\left(e^{iy\lambda}\widehat{\phi}(-\lambda) - e^{-iy\lambda}\widehat{\phi}(\lambda)\right)d\lambda}{\lambda}, \tag{29}$$

(see Gil-Pelaez (1951)). Similarly

$$F(y) = \frac{1}{2} - \frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{e^{-i\lambda y}\phi(\lambda)\,d\lambda}{\lambda} = \frac{1}{2} + \frac{1}{2\pi i}\int_0^{\infty} \frac{\left(e^{iy\lambda}\phi(-\lambda) - e^{-iy\lambda}\phi(\lambda)\right)d\lambda}{\lambda}. \tag{30}$$

From (29) and (30) it follows that

$$\begin{aligned}
\left|F(y) - \widetilde{F}_b(y)\right| &\le \frac{1}{2\pi}\int_b^{\infty} \frac{\left|e^{iy\lambda}\phi(-\lambda) - e^{-iy\lambda}\phi(\lambda)\right|d\lambda}{\lambda} \\
&\quad + \frac{1}{2\pi}\int_0^b \frac{\left|e^{iy\lambda}\left(\phi(-\lambda) - \widehat{\phi}(-\lambda)\right) - e^{-iy\lambda}\left(\phi(\lambda) - \widehat{\phi}(\lambda)\right)\right|d\lambda}{\lambda} \\
&\le \frac{1}{2\pi}\int_b^{\infty} \frac{|\phi(-\lambda)| + |\phi(\lambda)|}{\lambda}d\lambda \\
&\quad + \frac{1}{2\pi}\int_0^b \frac{\left(\left|\phi(-\lambda) - \widehat{\phi}(-\lambda)\right| + \left|\phi(\lambda) - \widehat{\phi}(\lambda)\right|\right)d\lambda}{\lambda} \\
&= \frac{1}{\pi}\int_b^{\infty} \frac{\exp(-\sigma^\alpha\lambda^\alpha)\,d\lambda}{\lambda} + \frac{b}{\pi}\sup_{|\lambda|<b}\left|\frac{\phi(\lambda) - \widehat{\phi}(\lambda)}{\lambda}\right|.
\end{aligned} \tag{31}$$

By change of variable $\sigma^\alpha\lambda^\alpha = x$ then the last integral can be expressed as

$$\frac{1}{\pi}\int_b^{\infty} \frac{e^{-\sigma^\alpha\lambda^\alpha}d\lambda}{\lambda} = \frac{1}{\alpha\pi}\int_{\sigma^\alpha b^\alpha}^{\infty} \frac{e^{-x}dx}{x} = \frac{E_1(\sigma^\alpha b^\alpha)}{\sigma\pi} \tag{32}$$

where $E_1(x)$ denotes the Exponential integral function (cf. Abramowitz and Stegun (1972)).If $\sigma = 1$, $\alpha = 1.6$ and $\log b = 0.7$ we thus get from Table of $E_1(x)$ in Abramowitz and Stegun (1972) p.p. 242, that

$$\frac{1}{\sigma\pi} E_1\left(\sigma^\alpha b^\alpha\right) \simeq 2 \cdot 10^{-3}. \tag{33}$$

Hence

$$\sup_y \left| F(y) - \widetilde{F}_b(y) \right| \leq 2 \cdot 10^{-3} + 0.64 \sup_{|\lambda|<b} \left| \frac{\phi(\lambda) + \phi(-\lambda)}{\lambda} \right|. \tag{34}$$

Since $\phi(0) = \widehat{\phi}(0)$ it follows that the last term on the right hand side exists for all $\lambda \in R$. We realize that the first term on the right hand side of (34) is negligeable and therefore the right hand side of (34) therefore depends on how well $\widehat{\phi}(\lambda)$ approximates $\phi(\lambda)$ for $\lambda \in (-b, b)$.

Appendix B

The ratio G/A for selected population groups 1973-1990

	Time								
	1973	1979	1982	1985	1986	1987	1988	1989	1990
Working Persons									
G	0.285	0.277	0.266	0.264	0.270	0.273	0.271	0.281	0.269
Standard deviation	0.004	0.004	0.006	0.007	0.004	0.006	0.004	0.009	0.005
G/A	0.72	0.73	0.73	0.73	0.73	0.73	0.73	0.74	0.73
Wage workers									
G	0.266	0.253	0.240	0.246	0.251	0.253	0.257	0.261	0.253
Standard deviation	0.003	0.004	0.003	0.007	0.003	0.005	0.005	0.010	0.005
G/A	0.70	0.71	0.71	0.71	0.71	0.72	0.73	0.73	0.73
Selfemployed other than farming and fishery									
G	0.369	0.366	0.405	0.408	0.360	0.365	0.347	0.364	0.367
Standard deviation	0.017	0.014	0.029	0.042	0.015	0.023	0.019	0.021	0.016
G/A	0.77	0.77	0.80	0.80	0.76	0.78	0.76	0.77	0.78
Married couples that are working									
G	0.180	0.166	0.167	0.170	0.174	0.178	0.184	0.204	0.184
Standard deviation	0.007	0.005	0.005	0.013	0.005	0.006	0.006	0.016	0.006
G/A	0.70	0.71	0.72	0.72	0.72	0.74	0.73	0.76	0.73