STICERD

Suntory and Toyota International Centres for Economics and Related Disciplines

Distributional Analysis Research Programme
Discussion Paper

Robust estimation of personal income distribution models

Maria-Pia Victoria-Feser

October 1993

LSE STICERD Research Paper No. DARP 04
This paper can be downloaded without charge from:
http://sticerd.lse.ac.uk/dps/darp/darp4.pdf

ROBUST ESTIMATION OF PERSONAL INCOME DISTRIBUTION MODELS

by

Maria-Pia Victoria-Feser
London School of Economics and Political Science

Discussion Paper
No.DARP/4
October 1993

The Toyota Centre
Suntory and Toyota International Centres for
Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Tel.: 020-7955 6678

Abstract

Statistical problems in modelling personal income distributions include estimation procedures, testing, and model choice. Typically, the parameters of a given model are estimated by classical procedures such as maximum likelihood and least-squares estimators. Unfortunately, the classical methods are very sensitive to model deviations such as gross errors in the data, grouping effects or model misspecifications. These deviations can ruin the values of the estimators and inequality measures and can produce false information about the distribution of the personal income in a given country. In this paper we discuss the use of robust techniques for the estimation of income distributions. These methods behave as the classical procedures at the model but are less influenced by model deviations and can be applied to general estimation problems.

Keywords: Personal income distributions, inequality measures, parametric models, influence function, M-estimator.

JEL Nos.: C13, D63.

Contact address: Mari-Pia Victoria-Feser, STICERD, London School of Economics and Political Science, Houghton Street, London WC2A 2AE.

1 Introduction

In this paper we discuss the robustness properties of estimators of parameters and derived inequality measures in models for personal income distributions (PID). These models and the associated inequality measures play a central role in the field of welfare economics and inference based on them can affect strongly the conclusions derived from the data. Here we focus on inference drawn from parametric models of PID.

Parametric models are only approximations to the reality and many studies on income data come to the conclusion that the model does not fit the data well. Common sources of deviations from assumed models include outliers (such as recording or definition errors), grouping effects (due for instance to a shift of some observations from a class to another), and other general misspecifications of the model; cf. Van Praag, Hagenaars, and Van Eck (1983) or Cowell and Victoria-Feser (1993). These deviations have drastic effects on the maximum likelihood estimator (MLE) for the parameters of the model becomes biased and inefficient when the model does not hold exactly. This in turn will affect the inequality measure computed from the estimated parameters. In particular, a few extreme observations can drive this quantity by themselves. The implication of this is for example an inequality measure which no longer represents the overall inequality structure of the population.

It should be stressed that our statistical approach to the study of PID is a complement to the problem of choosing the right model. It is important for PID models not only to fulfil a set of fundamental properties (see e.g. Dagum 1980) but also to be well estimated. Regularly, new models are developped and their authors show that they fit a set of data better than the others. We don't believe that a model that is perfectly adapted to every set of data exists, for the reasons argued above and for others (see e.g. Hampel, Ronchetti, Rousseeuw, and Stahel 1986). Therefore, the search for a better model fit can only be achieved by means of a combination of theoretical and practical considerations. Hence, estimation techniques which deal with model misspecification should be considered.

In PID, when the data are continuous, the deviations of the estimates can be caused by a few very high incomes. The MLE of two-parameters model, where one of the parameters is for the scale and the other for the shape, are very sensitive to high incomes. That means that because in such models no account is made for the heavy right tail, the MLE and other classical estimators can be biased and the fit of the distribution can be very bad. With a three or four-parameters model, the added parameters often deal with large observations by making the right tail more or less heavy. However, we will see that even less parsimonious models are non robust when estimated by means of classical estimators.

In order to illustrate the argument developed above, we fit a Gamma distribution to simulated data in Table 1. We assume a Gamma model with shape parameter $\alpha = 3$ and scale parameter $\lambda = 1$ and we consider the inequality measure given by the Theil's index (see section 3 for the definition) which has a true value of 0.16. We simulated samples of size 200 from a "3% contaminated Gamma model", i.e. the data were generated with probability .97 from the assumed Gamma(3,1) model and with probability .03 were outliers. (These contamination models will be discussed more in details in section 3.) Table 1 shows the bias (estimated by $\frac{1}{200}\sum_{i=1}^{2}00(\hat{\theta}_i - \theta)^2$ where $\hat{\theta}_i$ is the estimate for the i^{th} sample and θ is the true parameter) of the MLE and a robust estimator, the mean squared errors (MSE), and the derived Theil's index. All the values of the bias have standard errors smaller than 0.06.

It is clear from Table 1 that even a small amount of contamination (3%), i.e. a few outlying observations out of 200, has the effect to introduce a large bias in the MLE, to increase the MSE and to increase the Theil's index from 0.16 to 0.27. For a comparison we report the values of the robust estimator we will introduce in this paper. We see that it has very small bias and MSE and the Theil's index derived from it is practically not affected by the deviations in the data because its value is based on the overwhelming majority of the data. Similar effects can be observed with other models and other inequality measures. One can say that parametric models and inequality indices should take into account all the observations. However, we argue that the computation of classical *and* robust estimators gives a very important indication about the structure of the distribution. For instance, in this simulation we know that the true value of the Theil's index (i.e. at the Gamma(3,1) distribution) is of 0.16 and the robust estimate is equal to 0.17. Thus, the robust estimator estimates the Theil's index based on the majority (here 97%) of the data and is not influenced by the 3% of outliers. In a real case, by comparing the MLE and the robust estimator, one would draw the conclusion that the majority of the data leads to a Theil's index of 0.17 whereas by taking into account the outliers, one gets a Theil's index of 0.27. The outlying observations can be discovered immediately by simple inspection of the weights $W_c(x_i; \cdot)$ given by (18) below. They are those with weights close to zero.

Sometimes anomalous observations can be dealt with by a preliminary screening of the data. However, in view of the amount of data available nowadays and the automated procedures used to analyse them, robust techniques offer the advantage to take into account automatically possible deviations without a preliminary screening of the data. Moreover, the diagnostic information provided by these techniques can be used by the analyst to identify deviations from the model or from the data.

In this paper we will derive such robust estimators and associated inequality indices. In the past 20 years, robust statistics has been a central area of research in the statistical literature. Our results are based on the general theory developed in Huber (1981) and Hampel, Ronchetti, Rousseeuw, and Stahel (1986). Our goal is to show the usefulness of robust techniques for modelling PID data.

The paper is organized as follows. For the sake of completeness, we summarize in section 2 the basic ideas and techniques based on the general theory. In section 3 we apply these methods to PID models. We focus on Dagum's type I, and the Gamma and Pareto models but the same qualitative conclusions can be drawn for other models. Section 4 presents an application of robust methods to real data. Finally, some implications of these techniques with some concluding remarks are given in section 5.

2 Robustness concepts

Analysing and describing the PID of a given population of economic units involves typically the estimation of a parametric model. Robust methods deal with such models. In this section we summarize a few basic concepts of robust statistics for general parametric models which will be used in our application. The approach followed here is the one based on influence functions originated by Hampel (1968), Hampel (1974) and developed in Hampel, Ronchetti, Rousseeuw, and Stahel (1986).

Let $x_1, \ldots, x_n$ be n observations belonging to some sample space X and $\{F_\theta\}$ a parametric model with density f_θ, where the unknown parameter belongs to some parameter space $\Theta \subseteq \Re^p$. The empirical distribution $F^{(n)}$ is given by

$$F^{(n)}(x) = \frac{1}{n} \sum_{i=1}^{n} \Delta_x(x_i) \qquad (1)$$

where Δ_x is a point mass in x. As estimators of θ we consider statistics $T_n = T_n(x_1, \ldots, x_n)$ which can be represented (at least asymptotically) as functionals of the empirical distribution function, i.e. $T_n(x_1, \ldots, x_n) = T(F^{(n)})$.

One way of assessing the robustness of the estimator T is by means of the *influence function (IF)* which is defined at the model F_θ by

$$IF(x; T, F_\theta) = \lim_{\varepsilon \longrightarrow 0} \left[\frac{T((1-\varepsilon)F_\theta + \varepsilon \Delta_x) - T(F_\theta)}{\varepsilon} \right]. \qquad (2)$$

It describes the effect of a small contamination ($\varepsilon \Delta_x$) at the point x on the estimate, standardized by the mass of the contamination. Thus, the linear approximation $\varepsilon IF(x; T, F_\theta)$ measures the asymptotic bias of the estimator caused by the contamination. A desirable robustness property for an estimator is a bounded IF. Such an estimator is called $B-$ (or bias) robust.

The IF of a MLE is proportional to its scores function $s(x,\theta) = \frac{\partial}{\partial\theta}\log f_\theta(x)$. Unfortunatelly, most MLE for income models have an unbounded scores function and therefore an unbounded IF. We will see that this is the case for the Pareto, Gamma and Dagum's distributions.

The IF plays a central role in robustness considerations but can also be used to evaluate the asymptotic covariance matrix of an estimator. For an asymptotically normal estimator T, i.e.

$$\sqrt{n}(T_n - T(F_\theta)) \xrightarrow{D} N(0, V(T, F_\theta)), \tag{3}$$

its asymptotic covariance matrix can be written as

$$V(T, F_\theta) = \int IF(x; T, F_\theta) IF(x; T, F_\theta)^T dF_\theta(x). \tag{4}$$

Before stating the general optimality result, let us define the class of estimators we will be working with. It is the class of the M-estimators which are a generalization of the MLE (Huber 1964). An M-estimator is the solution T_n of the (system of) equation(s)

$$\sum_{i=1}^{n} \psi(x_i, T_n) = 0 \tag{5}$$

for some function $\psi : X \times \Re \to \Re^p$.

It is worth mentionning that this class is rich and includes a variety of well known estimators. For example, if ψ is the likelihood scores function, we obtain the MLE. Moreover, to any asymptotically normal estimator, there exists an asymptotically equivalent M-estimator. Hence there is no loss, at least asymptotically, in confining to the class of M-estimators.

The corresponding functional T is the solution of the equation

$$\int \psi(x, T) dF_\theta(x) = 0 \tag{6}$$

and Fisher consistency, i.e. $T(F_\theta) = \theta$ implies

$$\int \psi(x, \theta) dF_\theta(x) = 0 \tag{7}$$

for all θ.

The IF of an M-estimator defined by ψ at F_θ is given by

$$IF(x; \psi, F_\theta) = M(\psi, F_\theta)^{-1} \psi(x, \theta), \tag{8}$$

where

$$M(\psi, F_\theta) = -\int \frac{\partial}{\partial \theta} \psi(x, \theta) dF_\theta(x). \tag{9}$$

Under regularity conditions (see Huber 1981), an M-estimator is asymptotically normal with the asymptotic covariance matrix given by

$$V(T, F_\theta) = M(\psi, F_\theta)^{-1} Q(\psi, F_\theta) M(\psi, F_\theta)^{-T} \tag{10}$$

where $Q(\psi, F_\theta) = \int \psi(x, \theta) \psi(x, \theta)^T dF_\theta(x)$.

To build a B-robust estimator we have to put a bound on its

IF. However, doing that leads to an efficiency loss at the model. Hence, it is necessary to find a trade-off between robustness (the model holds only approximately) and efficiency. The best trade-off gives the *optimal B-robust estimator (OBRE)*. It is the M-estimator which minimizes the asymptotic covariance matrix[1] (10) under the constraint that it has a bounded IF (8).

Since the IF (8) is a p-vector, one can choose different norms to measure its maximum. The most natural way is to put an upper bound c on the Euclidian norm[2] of the IF, i.e.

$$\sup_x \|IF(x;\psi,F_\theta)\| \leq c \tag{11}$$

which leads to the unstandardized OBRE. Another way is to measure the IF in the metric given by the asymptotic covariance matrix of the estimator, hence to put an upper bound c on[3]

$$\sup_x \left\{IF(x;T,F_\theta)^T V(T,F_\theta)^{-1} IF(x;T,F_\theta)\right\}^{\frac{1}{2}} \tag{12}$$

which leads to the standardized OBRE. This estimator is invariant with respect to scale transformations.

The following theorem gives the OBRE in the standardized case for a general parametric model.

---

[1]In fact, a solution of this problem exists only if one minimizes the trace of the asymptotic covariance matrix; see Hampel, Ronchetti, Rousseeuw, and Stahel (1986).

[2]This robustness measure is called the *unstandardized gross-error sensitivity*.

[3]This robustness measure is called the *self-standardized gross-error sensitivity*.

Theorem

Assume a parametric model $\{F_\theta; \theta \in \Theta \subset \Re^p\}$ *with likelihood scores function* $s(x,\theta) = \frac{\partial}{\partial\theta}\log f_\theta(x)$, *a real constant* $c \geq \sqrt{p}$ *and denote by* $T^{(c)}$ *the solution for* θ *of*

$$\sum_{i=1}^{n} \psi_c^{A,a}(x_i,\theta) = 0, \tag{13}$$

where

$$\psi_c^{A,a}(x,\theta) = H_c(A[s(x,\theta) - a]), \tag{14}$$

$H_c(x) = x \cdot \min\left\{1; \frac{c}{||x||}\right\}$ *is the Huber function and* A, a *(respectively a* $p \times p$ *matrix and a* p*-dimentional vector) are determined by the equations:*

$$\int \psi_c^{A,a}(x,\theta)\psi_c^{A,a}(x,\theta)^T dF_\theta(x) = I \tag{15}$$

$$\int \psi_c^{A,a}(x,\theta) dF_\theta(x) = 0. \tag{16}$$

Then $T^{(c)}$ *is "admissible" B-robust in the sense that there is no M-estimator with the standardized gross-error sensitivity (12) bounded under c and a smaller*[4] *asymptotic covariance matrix.*

Proof: Hampel, Ronchetti, Rousseeuw, and Stahel (1986), p.245.

Notice that the matrix A and the vector a depend in general

[4] "Smaller" means that the difference of the matrices is positive semidefinite.

on the parameter θ and should be viewed as $A(\theta)$ and $a(\theta)$. The equation (13) can be written as

$$\sum_{i=1}^{n} \psi_c^{A,a}(x_i;\theta) = \sum_{i=1}^{n} A(\theta)\,[s(x_i;\theta) - a(\theta)]\cdot W_c^{A,a}(x_i;\theta) = 0 \quad (17)$$

where

$$W_c^{A,a}(x;\theta) = \min\left\{ 1\ ; \frac{c}{\|A(\theta)[s(x;\theta) - a(\theta)]\|} \right\} \qquad (18)$$

(18) defines the weights given by the OBRE to each observation, and can be used to detect the outlying observations.

Let us now interpret the result of the theorem. For efficiency reasons, the optimal estimator has to be as similar as possible to the MLE for the values of x in the bulk of data, i.e. at non-influential values of x. Therefore, its ψ-function equals the scores function s for those values. On the other hand, since the IF is proportional to the ψ-function, in order to obtain a bounded IF, one has to truncate the scores function where the bound c is exceeded. This is achieved by means of the Huber function. The matrix A and the vector a can be viewed as Lagrange multipliers for the constraints resulting from a bounded self-standardized gross-error sensitivity and Fisher consistency. Finally, the constant c is the bound on the IF and can be interpreted as the regulator between robustness and efficiency: for a lower c one gains robustness but looses efficiency and vice versa. The most robust estimator can be obtained by choosing the lower bound $c = \sqrt{p}$. On the other hand, $c = \infty$ gives the MLE (the most efficient but non robust). Typically, c is chosen as to achieve a 95% efficiency at the model. This depends in general on the model.

3 Application to personal income distributions

In this section we discuss the computation of the OBRE and present an application to three models of PID. It should be stressed that the results of section 2 can be applied to any income model.

To compute the OBRE, one requires solving (13) under the conditions (15) and (16). We propose here an algorithm based on the Newton-Raphson method. The main idea is to compute the matrix A and the vector a for a given θ by solving (15) and (16). This is followed by a Newton-Raphson step for (13) given these two matrices, and these steps are iterated until convergence.

More precisely, the algorithm can be defined by the following four steps:

Step 1: Fix a precision threshold η, an initial value for the parameter θ and initial values $a = 0$ and $A = J^{\frac{1}{2}}(\theta)^{-T}$ where

$$J(\theta) = \int s(x,\theta)s(x,\theta)^T dF_\theta(x)$$

is the Fisher information matrix.

Step 2: Solve the following equations with respect to a and A:

$$A^T A = M_2^{-1}$$

and

$$a = \frac{\int s(x,\theta)W_c(x,\theta)dF_\theta(x)}{\int W_c(x,\theta)dF_\theta(x)} \ ,$$

where

$$M_k = \int [s(x,\theta)-a][s(x,\theta)-a]^T W_c(x,\theta)^k dF_\theta(x) \ , \ k = 1,2.$$

The current values of θ, a and A are used as starting values to solve the given equations.

Step 3: Compute M_1 and $\Delta\theta = M_1^{-1}\{\frac{1}{n}\sum_{i=1}^n [s(x_i,\theta)-a]W_c(x_i,\theta)\}$.

Step 4: If $|\Delta\theta| > \eta$ then $\theta \rightarrow \theta + \Delta\theta$ and return to Step 2, else stop.

The algorithm is convergent provided the starting point is near to the solution. In the first step, we can take for instance the MLE as initial value for the parameter. However, it can be argued that, a more robust starting point like a trimmed moment estimator or a moment estimate based on the median and MAD[5] would be preferable. An alternative is to still use the MLE as the starting point but then compute an OBRE with a

[5]MAD denotes median absolute deviation and is often used as a robust estimator of the standard error in a normal model.

high value of the bound c and then use the estimate as starting point for another more robust (lower value of c) estimator.

The choice of the initial values for the matrices A and a in the second step is due to the fact that these values solve the equations for $c = \infty$ (corresponding to the MLE). Notice that integration can be avoided in Step 2 by replacing F_θ by its empirical distribution function. This means replacing the integrals with averages over the sample.

We now study the robustness properties of the MLE for the Dagum's, Gamma and Pareto model. The density (f_θ) and the scores functions (s) are

i) Pareto law:

$$f_\alpha(x) = \alpha x^{-(\alpha+1)} x_0^\alpha \ , \ 0 \leq x_0 \leq x < \infty$$

$$s(x;\alpha) = \left[\frac{1}{\alpha} - \log(x) + \log(x_0)\right]$$

ii) Gamma distribution:

$$f_{\alpha,\lambda}(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \ , \ 0 < x < \infty$$

$$s(x;\alpha,\lambda) = \begin{bmatrix} \log(\lambda) - \tilde{\Gamma}(\alpha) + \log(x) \\ \frac{\alpha}{\lambda} - x \end{bmatrix} ,$$

where $\alpha, \lambda > 0$, $\Gamma(\alpha) = \int_0^\infty e^{-u} u^{\alpha-1} du$ and $\tilde{\Gamma}(\alpha) = \frac{\partial}{\partial \alpha} \log \Gamma(\alpha)$.

iii) Dagum model type I:

$$f(x) = \frac{1}{B[1,\beta]}\lambda\delta x^{-\delta-1}\left(1+\lambda x^{-\delta}\right)^{-\beta-1} \quad , 0 \le x < \infty$$

where $\beta, \lambda > 0$, $\delta > 1$ and $B[p,q] = \int_0^1 t^{p-1}(1-t)^{q-1}dt$ is the Beta function.

$$s(x;\beta,\lambda,\delta) = \begin{bmatrix} -\tilde{B}[1,\beta] - \log(1+\lambda x^{-\delta}) \\ \frac{1}{\lambda} - (\beta+1)\frac{x^{-\delta}}{(1+\lambda x^{-\delta})} \\ \frac{1}{\delta} + (\beta+1)\lambda\frac{x^{-\delta}\log(x)}{(1+\lambda x^{-\delta})} - \log(x) \end{bmatrix} \tag{19}$$

where

$$\tilde{B}[1,\beta] = \frac{\frac{\partial}{\partial\beta}B[1,\beta]}{B[1,\beta]} = \frac{\int_0^1 t^{\beta-1}\log(t)dt}{\int_0^1 t^{\beta-1}dt}$$

It is easy to see that the MLE for these three models is not robust. Indeed, when we compute for example the gross-error sensitivity, we need the expression of the IF for the MLE of these models. But since the IF is proportional to s, we just have to study the behaviour of the scores function to see that the MLE is not robust (for the Dagum's model see Victoria-Feser 1993a). That means that one single point can carry the MLE arbitrarily far. This is also the case for most of other PID models.

In order to study the behaviour of the OBRE compared to the MLE when the data are contaminated, we performed a simulation study. We chose to study the Pareto and the Gamma distribution. To give a clearer picture of the effect of contamination on the analysis, we also compute an income inequality measure. We consider here Theil's index given by

$$I_{Theil} = E\left[\frac{x}{\mu}\log\left(\frac{x}{\mu}\right)\right]$$

where $\mu = E[x]$. It is computed by taking the expectation at the model with the estimated parameter.

The estimates of Theil's index are given by

$$I_{Theil}(\hat{\alpha}) = \frac{1}{\hat{\alpha}-1} - \log\left(\frac{\hat{\alpha}}{\hat{\alpha}-1}\right)$$

for the Pareto distribution, and

$$I_{Theil}(\hat{\alpha}) = \frac{1}{\hat{\alpha}} + \tilde{\Gamma}(\hat{\alpha}) - \log(\hat{\alpha})$$

for the Gamma distribution. Moreover, for any parametric model F_θ, it can be shown that the IF of the Theil's index is proportional to the IF of the estimators of the parameters (see Cowell and Victoria-Feser 1993). Therefore, an unbounded IF of the estimators of the parameters implies an unbounded IF of the Theil's index.

In order to compare numerically the MLE and the OBRE when the models do not hold exactly, we performed a simulation study by generating samples of size 200 according to the following distributions. Data were generated with the random number generator of Splus on SUN 3/60.

1. Gamma model 1: non contaminated model $\{F_{\alpha,\lambda}\}$, with $\alpha = 3$ and $\lambda = 1$.

2. Gamma model 2: model with 1% of very "bad" contamination; the upper 1% of the observations is multiplied by 10.

3. Gamma model 3: model with 5% of contamination given by $\{0.95F_{\alpha,\lambda} + 0.05F_{\alpha,0.1\lambda}\}$

4. Pareto model 1: non contaminated model $\{F_{\alpha,x_0}\}$, with $\alpha = 3$ (and $x_0 = 0.5$).

5. Pareto model 2: model with 5% of contamination given by $\{0.95F_{\alpha,x_0} + 0.05F_{\alpha,10x_0}\}$

In Tables 2 and 3, we give the bias ($E[\hat{\theta} - \theta]$), the MSE ($E[\hat{\theta} - \theta]^2$) and the corresponding Theil's index for the MLE and the OBRE (with different values of c). The true values of the Theil's index are 0.095 for the Pareto distribution and 0.16 for the Gamma distribution. All the computations were performed on a VAX 8700 (with VMS operating system). All the values of the bias in Table 2 have standard errors smaller than 0.06. For those in Table 3, the standard errors are smaller than 0.02.

As expected, the MLE is badly biased in the presence of a small amount of contamination introduced in the model. For instance the bias of $\hat{\alpha}$ in the Gamma model becomes substantial with only 1% of contamination and the Theil's index rises from 0.155 to 0.32. This is of course the worst type of contamination for the MLE. However, the picture is the same in the more realistic cases, namely the Gamma model 3 and the Pareto model 2. These situations represent the real case of a recording error of some percentage of the observations (comma error).

On the other hand, we can see that when the model holds exactly, the OBRE shows the same performance as the MLE. When an amount of contamination is introduced, the OBRE is stable and much less influenced than the MLE. Its MSE are much smaller than the one of the MLE (by a factor 10 in the Gamma model and by a factor 5 for the Pareto law). Consequently, the Theil's index based on OBRE is also stable.

The sensitivity of the OBRE to contamination depends on the choice of the bound c. As the results show, the lower the bound c, the less is the OBRE sensitive to contamination. However, lowering the bound c leads to an efficiency loss at the model. If we measure the efficiency at the model by the ratio between the traces of the asymptotic covariance matrices of the OBRE and the MLE i.e.,

$$\frac{tr\left[J(\theta)^{-1}\right]}{tr\left[\int IF(x,\theta,H_\theta)IF(x,\theta,H_\theta)^T dF_\theta(x)dx\right]}$$

the OBRE has 95% efficiency for a bound c of about 3.5, and

60% efficiency for $c = 1.5$ in the case of the Gamma model. For the Pareto distribution, the OBRE has 95% efficiency for $c = 3.0$ and 75% efficiency for $c = 1.5$. These results pratically do not change with respect to θ. In our simulations we have used different values for the bound c to show the behaviour of different OBRE. In an application, one can use a bound c such that 95% of efficiency is achieved.

4 Application to real data

In order to illustrate the usefulness of OBRE in the study of PID, we apply the techniques presented in the former sections to a real data set. Actually, we want to fit a Gamma model to the empirical distribution of total family income in 1981 in the USA using the Panel Study of Income Dynamics (PSID). We compare the Gamma distribution with the Dagum's model which has one more parameter. The total family income is defined as the sum of the total taxable income of head and spouse, the total transfer income of head and spouse, the total taxable income of all other members in the family unit and the total transfer income of all other members in the family unit. Since the Gamma variable is positive, we consider only the positive incomes. The sample size is 5199.

As with the simulations, in order to give a more realistic interpretation of the results, we also calculated an income inequality measure. This time we chose the Gini index. For the Gamma distribution (see Salem and Mount 1974) and the Dagum model (see e.g. Dagum 1985), the Gini index is given by respectively

$$I_{Gini} = 2\frac{B[0.5;\hat{\alpha},\hat{\alpha}+1]}{B[\hat{\alpha},\hat{\alpha}+1]} - 1 \tag{20}$$

and

$$I_{Gini} = \frac{\Gamma(\hat{\beta})\Gamma(2\hat{\beta}+1/\hat{\delta})}{\Gamma(2\hat{\beta})\Gamma(\hat{\beta}+1/\hat{\delta})} - 1 \tag{21}$$

where the incomplete Beta function is given by

$$B[t_0;\alpha,\beta] = \int_0^{t_0} t^{\alpha-1}(1-t)^{\beta-1}dt \ , \ 0 \leq t_0 < 1 \tag{22}$$

For the Gamma distribution, we first computed the MLE and obtained the estimates $\hat{\alpha} = 0.32$ and $\hat{\lambda} = 1.4 \cdot 10^{-5}$. Figure 1 gives the histogram of the empirical distribution and the plots of the estimated Gamma distribution by the MLE and the OBRE. MLE not only estimates a zeromodal distribution but also gives a very bad fit. The reason is that the MLE is almost completely determined by the highest observations that represent only a very small proportion of the data set. One could say that the Gamma model is not the best choice for this particular example and that a model with more than two parameters could deal with heavy tails. However, when we compute the OBRE ($c = 3.0$) we obtain the estimates $\hat{\alpha} = 1.67$ and $\hat{\lambda} = 7.8 \cdot 10^{-5}$ and this produces an excellent fit.

For the Dagum model, we found a small difference between the MLE and the OBRE ($c = 2.0$). Actually, the MLE of β, λ, δ are respectively 0.41, $1.18 \cdot 10^{14}$ and 3.15, whereas the OBRE are respectively 0.36, $1.22 \cdot 10^{14}$ and 3.13. If we look at the histogram and the plot of the estimated densities (see figure 2) we see no great difference in the fit: both estimators lead to a good fit.

If we compare the estimated Gini indexes, we remark that the ones computed from the MLE are radically different for the two models (0.320 for the Gamma and 4.120 for the Dagum distribution). If however we look at the Gini index computed from the OBRE, we find for the Gamma distribution a value of 0.4055 which is comparable the the value for the Dagum model.

So what can be concluded from this numerical example? First we showed that the MLE can lead a statistical analysis to false conclusions (especially, in our case, with the Gamma distribution). However, with the sample we have analysed, it seems that the estimated (MLE) Dagum model fits the data quite well. It should also be stressed that this model has one more parameter which permits the model to accomodate the large observations. But is it worthwhile to estimate one more parameter to accomodate a few observations? Perhaps not, especially when we know that more parsimonious models estimated robustly give the same results (at least with respect to the fit of the distribution and when comparing derived inequality measures). Moreover, as we have seen in section 3, the Dagum model is not robust when estimated by the MLE, so that, with some data samples, it could be that a robust estimator has to be preferred.

5 Conclusion

In this paper we show that robust methods can be used successfully in the estimation of income distribution models. These techniques still give reliable parameters estimates and inequality measures in the presence of deviations from the assumed model. We limited our discussion to continuous data. However, these methods can also be applied to truncated data and to grouped data where truncation and grouping effects play an important role in misspecification. These cases have been treated in Victoria-Feser (1993b) and are the subject of different papers.

Acknowledgments

The research was partially supported by the 'Fond National Suisse pour la Recherche Scientifique'.

References

Cowell, F. A. and M.-P. Victoria-Feser (1993). Robustness properties of inequality measures. Working paper, London School of Economics, UK. Submitted for publication.

Dagum, C. (1980). Generating systems and properties of income distribution models. *Metron 38*(3-4), 3–26.

Dagum, C. (1985). Analyses of income distribution and inequality by education and sex in Canada. *Advances in Econometrics 4*, 167–227.

Hampel, F. R. (1968). *Contribution to the Theory of Robust Estimation.* Ph. D. thesis, University of California, Berkeley.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association 69*, 383–393.

Hampel, F. R., E. Ronchetti, P. J. Rousseeuw, and W. A. Stahel (1986). *Robust Statistics: The Approach Based on Influence Functions.* New York: John Wiley.

Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics 35*, 73–101.

Huber, P. J. (1981). *Robust Statistics.* New York: John Wiley.

Salem, A. B. Z. and T. D. Mount (1974). A convenient descriptive model of income distribution: The Gamma density. *Econometrica 42*, 1115–1127.

Van Praag, B., A. Hagenaars, and W. Van Eck (1983). The influence of classification and observation errors on the measurement of income inequality. *Econometrica 51*, 1093–1108.

Victoria-Feser, M.-P. (1993a). Robust methods for personal income distribution models with application to Dagum's model. In C. Dagum and A. Lemmi (Eds.), *Income Distribution, Social Welfare, Inequality and Poverty*. JAI-Press of Greenwich. To appear.

Victoria-Feser, M.-P. (1993b). *Robust Methods for Personal Income Distribution Models*. Ph. D. thesis, University of Geneva. Thesis no 384.

	Parameter	Bias	MSE	Theil's Index
MLE	α	-1.33	1.89	0.27
	λ	-0.56	0.32	
Robust	α	-0.20	0.11	0.17
	λ	-0.09	0.02	

Table 1: **Effects of deviations from the model on MLE and a robust estimator**

		Bias(MLE)	Bias(OBRE)	Bound c	MSE	Theil's Index
(1) no contamination	α	0.05			0.07	0.155
	λ	0.01			0.01	
	α		0.06	$c = 4.0$	0.07	0.155
	λ		0.02		0.01	
(2) 1% contamination	α	-1.62			2.65	0.320
	λ	-0.64			0.41	
	α		0.01	$c = 5.0$	0.07	0.157
	λ		0.01		0.01	
	α		0.12	$c = 1.5$	0.17	0.152
	λ		0.04		0.02	
(3) 5% contamination	α	-1.72			3.09	0.342
	λ	-0.70			0.5	
	α		-0.63	$c = 4.0$	0.51	0.196
	λ		-0.28		0.09	
	α		-0.22	$c = 1.5$	0.16	0.169
	λ		-0.11		0.03	

Table 2: **Numerical comparison between the MLE and the OBRE for the Gamma model**

		Bias(MLE)	Bias(OBRE)	Bound c	MSE	Theil's Index
(4) no contamination	α	-0.01			0.05	0.095
	α		-0.01	$c = 2.0$	0.05	0.095
(5) 5% contamination	α	-0.75			0.61	0.212
	α		-0.47	$c = 3.0$	0.28	0.151
	α		-0.24	$c = 1.5$	0.11	0.118

Table 3: Numerical comparison between the MLE and the OBRE for the Pareto law

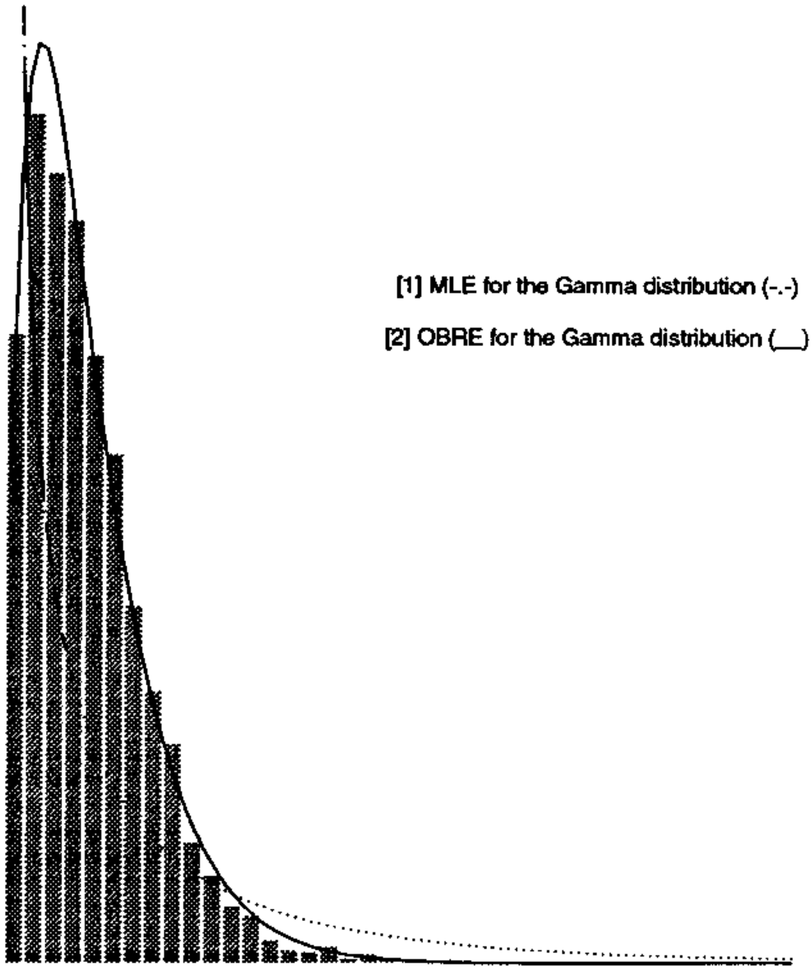

Figure 1: **MLE and OBRE of the Gamma distribution on PSID data**

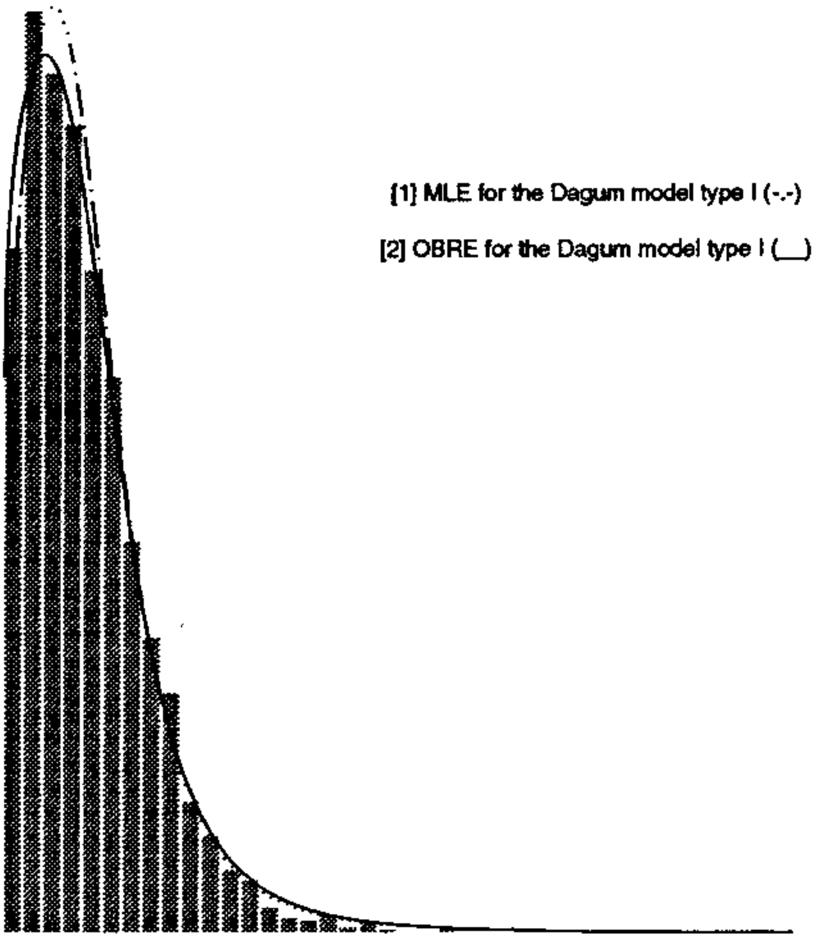

Figure 2: MLE and OBRE of the Dagum model on PSID data