A General Robust Approach to the Analysis of Income Distribution, Inequality and Poverty

Victoria-Feser, Maria-Pia

Robust Methods for the Analysis of Income Distribution, Inequality and Poverty

Maria-Pia Victoria-Feser
University of Geneva
CH-1211 Geneva 4

June 2000

Abstract

Income distribution embeds a large field of research subjects in economics. It is important to study how incomes are distributed among the members of a population in order for example to determine tax policies for redistribution to decrease inequality, or to implement social policies to reduce poverty. The available data come mostly from surveys (and not censuses as it is often believed) and are often subject to long debates about their reliability because the sources of errors are numerous. Moreover the forms in which the data are available is not always as one would expect, i.e. complete and continuous (micro data) but one also can only have data in a grouped form (in income classes) and/or truncated data where a portion of the original data has been omitted from the sample or simply not recorded.

Because of these data features, it is important to complement classical statistical procedures with robust ones. In this paper such methods are presented, especially for model selection, model fitting with several types of data, inequality and poverty analysis and ordering tools. The approach is based on the *Influence Function* (IF) developed by Hampel (1974) and further developed by Hampel, Ronchetti, Rousseeuw, and Stahel (1986). It is also shown through the analysis of real UK and Tunisian data, that robust techniques can give another picture of income distribution, inequality or poverty when compared to classical ones.

Key words: income distribution, inequality, poverty, robust statistics, influence function, model choice, grouped data, censored data, stochastic dominance.

Résumé

La distribution des revenus comporte une importante quantité de domaines de recherche en économie. Il est important de pouvoir étudier comment les revenus sont répartis au sein des membres d'une population pour pouvoir par exemple définir une politique de taxation et de redistribution afin de diminuer l'inégalité, ou implémenter des actions sociales pour diminuer la pauvreté. Les données à disposition proviennent essentiellement d'enquêtes (et non pas de rencensemment comme on pourrait le croire) et leur fiabilité soulève de grands débats car les sources d'erreur sont nombreuses. En plus, les données peuvent ne pas se présenter sous la forme habituelle de données continues et complètes, mais sous forme groupée (revenus par classe) et/ou sous forme censurée à savoir qu'une partie des revenus a été enlevée de l'échantillon ou simplement non enregistrée.

A cause de la particularité des données, il est important de compléter les analyses statistiques classiques au moyen d'analyses robustes. Dans cet articles de telles méthodes sont présentées, spécialement pour la sélection de modèle, l'estimation de modèle avec différents types de données, l'analyse de l'inégalité et de la pauvreté, et pour les outils de comparaison de distributions. L'approache est basée sur la *fonction d'influence* (IF) développée par Hampel (1974) et ensuite par Hampel, Ronchetti, Rousseeuw, and Stahel (1986). On montre aussi à travers l'analyse de données réelles Britanniques et Tunisiennes que les procédures robustes peuvent donner une autre représentation de la distribution des revenues, de l'inégalité et de la pauvreté lorsqu'elles sont comparées à des procédures classiques.

1 Introduction

In the field of welfare economics, income distribution plays a central role. The different possible analysis are for example the data fitting with appropriate models that describe the distribution of incomes, the choice of the appropriate model, testing its goodness of fit, the computation of inequality indices, comparisons of distributions using ordering tools such as the Lorenz curve, and the evaluation of poverty. Statistical inference is an essential issue, since one deals with data that mostly come from surveys and by the way it is done will influence the resulting economic policies, such as fiscal policies, investment policies, labour policies and redistribution policies.

Survey data are often subject to long debates about their reliability because the sources of errors are numerous (see e.g. Groves (1989). In the context of income distribution, it often happens that data providers will modify raw data to eliminate negative incomes or zeros (Jenkins (1997) or to censor high incomes for confidentiality reasons (Fichtenbaum and Shahidi (1988). Moreover, what constitutes an "income" is not always well perceived by people filling in the questionnaires. The time span during which the income is received is also a source of confusion (weekly, monthly, annual?) which might produce for example for some respondents monthly incomes when weekly incomes are in fact expected. The economic unit (person, nuclear family, household) doesn't include the same type of people depending on the country. The impact on estimates can be really serious (see e.g. Van Praag, Hagenaars, and Van Eck (1983).

Apart from the sources of "errors" for incomes, there might also be legitimate but extreme incomes that do not "fit" into the picture presented by the majority of the data. These incomes might be very small (from people whose beliefs make them choose to live with very little) or exceptionally very large, for example lottery winners (see e.g. Prieto Alaiz and Victoria-Feser (1996) or the famous German farmer who reported an income of 1000000 DM per month in the Luxembourg Income Study 1981 (German dataset). Though extreme incomes are clearly important from the point of view of the political authority, one can argue that they should have a limited influence when the goal of the study is to derive inequality or poverty measures or even other economic indicators that should reflect the economic and social situation of a given region as a whole. This is even more important when these indicators are fed (automatically) as parameters in more complex econometric models which are used to compare different countries possibly over time.

It should also be stressed that on top of the problem of reliability, the forms in which the data are available is not always as one would expect, i.e. complete and continuous (micro data) but one also can only have data in a grouped form (in income classes) and/or truncated data where a portion of the original data has been omitted from the sample or simply not recorded. This produces serious difficulties when it comes to estimation (see e.g. Ben Horim (1990, Nelson and Pope (1990).

Given the structure of the data and the possible "unusual features", the models that are supposed to describe them can only be thought of as approximations to the reality. Robust techniques appear therefore as natural candidates to try to partially alleviate the problems associated with income data. In this paper we review some aspects of the application of such techniques to the problem of the analysis of income distribution, inequality and poverty. More details can be found in Cowell and Victoria-Feser (1996a, 1996b, 1996c), Victoria-Feser (1993,1997) and Victoria-Feser and Ronchetti (1994,1997).

2 Model formulation

Let $F \in \mathfrak{F}$ be the true income distribution which belongs to the family $\mathfrak{F}$ of distribution functions. The income distribution might depend on a set of parameters θ and in that case we would write F_θ. What we observe is a sample of n incomes denoted by $x_1, \ldots, x_n$ generated independently by an approximation of F. In order to represent the impact of contamination on an income distribution we need a specific model of the contamination. Consider the elementary distribution $G^{(z)} \in \mathfrak{F}$ which has a unit point mass at an arbitrary point z and zero mass elsewhere:

$$G^{(z)}(x) = \iota(x \geq z). \tag{1}$$

where ι is the *indicator function* defined by

$$\iota(D) = \begin{cases} 1 & \text{if } D \text{ is true} \\ 0 & \text{if } D \text{ is false .} \end{cases}$$

We may use this to model an elementary form of contamination. Suppose that there is a small amount of undetectable contamination at point z in the income distribution. Then the sample which is actually observed will of course not be a realization of the true distribution F but of a mixture distribution $F_\varepsilon^{(z)}$ where

$$F_\varepsilon^{(z)}(x) := [1-\varepsilon]F(x) + \varepsilon G^{(z)}(x). \tag{2}$$

The parameter ε is used to capture the importance of the contamination relative to the true data. Hence, an observation drawn from $F_\varepsilon^{(z)}$ has probability $(1-\varepsilon)$ of being generated by F and probability ε of being equal to z. It should be noted that $F_\varepsilon^{(z)}$ is a particular case of a more general mixture distribution of the type $[1-\varepsilon]F(x) + \varepsilon G(x)$ where G is any distribution. However, for our purpose, this particular mixture distribution is not restrictive.

The central issue with which we are concerned can then be stated as follows. Suppose we wish to fit the model F to the data or choose among possible models in $\mathfrak{F}$ to fit the data, or compute an inequality index, or rank two distributions $F1$ and $F2 \in \mathfrak{F}$, or compare the poor in populations $F1$ and $F2 \in \mathfrak{F}$. Will the resulting analysis which will actually be based on $F1_\varepsilon^{(z)}$ (and $F2_\varepsilon^{(z)}$) give very misleading answers? If the amount of contamination is large relative to the true data then we might reasonably conclude that nothing much can be expected from the statistical analysis. However, if the amount of contamination is relatively small, we might reasonably expect that statistical inference should be robust under contamination, and might be concerned when this is not the case.

For any statistic T this idea can be made more precise by introducing the *influence function* IF. This is obtained by taking the derivative with respect to ε of the statistic at $F_\varepsilon^{(z)}$ when $\varepsilon \to 0$ thus:

$$IF(z; T, F) := \left. \frac{\partial}{\partial \varepsilon} T(F_\varepsilon^{(z)}) \right|_{\varepsilon \to 0}. \tag{3}$$

The IF for the statistic T measures the impact upon the estimate of an infinitesimal amount of contamination at the point z. It is a function of z, the point at which the contamination occurs. If the IF is unbounded for some value of z it means that the T-statistic may be catastrophically affected by data-contamination at income values close to z.

The IF was first introduced by Hampel (1968 (Hampel (1968, Hampel (1974) and can be thought of as a first-order approximation of the bias on the statistic due to the introduction of the contamination. In other words, if one plots the maximum (absolute) bias of the statistic as a function of ε, then the maximum (absolute) value of the IF is the slope of the tangent at $\varepsilon = 0$. This means that if the IF can be infinite, then so can be the bias of the statistic (see Hampel, Ronchetti, Rousseeuw, and Stahel (1986). Moreover, the IF describes the worst bias upon T when contamination of any kind (i.e. not restricted to $G^{(z)}$ but say G, where G is any distribution) is introduced in the model (see Hampel, Ronchetti, Rousseeuw, and Stahel (1986).

The IF has two main functions: it can be used to study robustness properties of statistics and to build robust estimators and robust test procedures. In the context of income distribution it has been used in particular to assess the robustness properties of inequality and poverty measures and ranking tools (see Cowell and Victoria-Feser 1996a, 1996b, 1996c). In the general context of parametric estimation, the IF has been used to define robust estimators. Indeed, the problem is first to find a general class of estimator and then within this class choose a robust one in an optimal way as it will be explained below. The class of M-estimators (Huber (1964) constitutes a good basis. It is given for a general parametric model F_θ and for a sample of n independent incomes $x_1, \ldots, x_n$ by the solution in θ of

$$\sum_{i=1}^{n} \psi(x_i; \theta) = 0$$

where ψ is a very general function. A particular case is given by the MLE when $\psi(x;\theta) = s(x;\theta)$. Another particular case is given by the optimal B-robust estimator (OBRE) of Hampel, Ronchetti, Rousseeuw, and Stahel (1986). The OBRE is optimal because it has the minimum (trace of the) asymptotic covariance matrix among the M-estimators with bounded (by a constant c) IF. There are actually several OBRE depending on the way the IF is bounded. The standardised OBRE is defined as the solution in θ of

$$\sum_{i=1}^{n} \left[s(x_i;\theta) - a \right] w_c(x_i; a, A) = 0 \tag{4}$$

where

$$w_c(x; a, A) = \min \left\{ 1; \frac{c}{\| A \left[s(x,\theta) - a \right] \|} \right\} \tag{5}$$

are the weights of the Huber function (Huber 1964, 1981), $s(\cdot;\theta)$ is the score function and the vector $a = a(\theta)$ and matrix $A = A(\theta)$ are defined implicitly by

$$E\left[[s(x;\theta) - a][s(x;\theta) - a]^T w_c(x;a,A)^2\right] = A^{-1}A^{-T} \tag{6}$$

$$E\left[s(x;\theta) w_c(x;a,A)\right] = a \tag{7}$$

A and a insure efficiency and consistency of the estimator. One can see in (4) that the robust estimator is relatively simple. It can be seen as a weighted MLE with weights given by (5) for data lying far from the bulk. It should be stressed that the OBRE is by construction robust not only to gross errors (outliers) but also to slight model misspecifications. Indeed, it has a bounded IF which measures the influence upon the estimator of any type of model misspecification. The degree of robustness is controlled by c. To gain robustness one should lower c and it is easy to see that when $c \to \infty$, $w_c(x;a,A) \to 1$, then $a \to 0$, $\psi(x;\theta) \to s(x;\theta)$ and one gets the MLE. However, there is a price to pay and this price is an efficiency loss. In practice one can choose c so that the OBRE has 95% efficiency compared to the MLE and still has good robustness properties.

When analysing income data, one doesn't always have continuous and complete data. In these cases, it was necessary to develop new robust estimators and a similar approach to Hampel, Ronchetti, Rousseeuw, and Stahel (1986) was used leading to robust estimators of the same type as in (4). The same can be said about testing procedures. In the following sections we review some of the recent developments in robust statistics for income distribution by type of data and/or type of analysis.

3 Fitting continuous and complete data

The problem here is how to estimate θ given a parametric model F_θ that supposedly describes the distribution of incomes. It is now widely known that the maximum likelihood estimator (MLE) is in general not a robust estimator. This is true for all the known models for income distribution (see Victoria-Feser (1993). With continuous and complete data, the OBRE is a suitable candidate for the parameters of income models such as the Gamma distribution which depends on two parameters and the Dagum type I (Dagum (1980) which depends on three parameters. The first parameter is usually for the scale, the second is for the shape of the distribution and a third parameter is sometimes added to model the thickness of the right tail.

As an example, we try to fit one data set to the income distributions mentioned above using the MLE and the standardized OBRE. Our data set is based upon a subsample ($n = 746$) of a standard data set of disposable income in the UK, 1979, where the income receiver is the household in receipt of social benefits (see Department of Social Security (1992). A Gamma distribution is first fitted. The histogram of the data and the estimated densities (MLE and OBRE, $c = 2$) are presented in Figure 1.

Figure 1 here

We can see that the MLE based on the Gamma model provides a very poor fit, whereas the OBRE captures the bulk of the data in the center of the distribution. It better estimates the mode of the distribution, but at the cost of underestimating the masses at the extremes (i.e. high incomes). These extreme incomes represent a small proportion of all incomes and while the MLE tries (not too successfully) to accommodate them, the OBRE concentrates on the more probable incomes, i.e. the ones around the mode of the distribution. One could argue that two parameters are not sufficient to describe this dataset, and that a third one is needed. The question is then: is it worthwhile to add a parameter to model a few extreme values and are we really safe with one more parameter? To give an element of answer to this problem, we choose to fit a Dagum type I distribution to the same data set. The histogram of the data and the estimated densities (MLE and OBRE, $c = 2$) are presented in Figure 2. We can see that with an extra parameter for the thickness of the right tail (to model extreme values), the MLE based on three parameters provides a better fit to these data as with a two parameters model, but it still underestimates the mode of the data, which is not the case with the OBRE. With a three parameters model however, the difference between the two estimators is reduced.

Figure 2 here

With this example we showed that the use of a robust estimator to fit a distribution can at least give another picture. This is of course possible provided that a correct (approximate) model can be specified. In the following section we discuss a testing procedure that allows one to select robustly a model among competing ones.

4 Choosing between two models

The problem here is that of choosing a model which, according to certain criteria, best represents the data. With income distribution models we can follow two approaches. One is to first fit a super model to the data, that is one with four parameters such as the generalized Beta distributions (see McDonald and Ransom (1979) which includes most of the models proposed to describe the distribution of incomes, and then test for the significance of the parameters. However, we found that in practice, estimating four parameters models is numerically very difficult even for the MLE, and becomes nearly impossible with the OBRE. Moreover, this procedure supposes that we have a super model at hand and this is not always the case in other types of analysis. We therefore follow the approach of Cox (1961, 1962) to test between two non-nested hypotheses.

In general, it is assumed under H_0 (the hypothesis under test) that the model is F^0_α and that under H_1 (the alternative hypothesis) the model is F^1_β with α and β being two sets of parameters. Atkinson (1970) and others have shown that the Cox test statistics can be interpreted as a Lagrange multiplier or score test on $H_0 : \lambda = 1$ (against $H_1 : \lambda \neq 1$) for the comprehensive model

$$f^c(x; \alpha, \beta) = f^0(x; \alpha)^\lambda \cdot f^1(x; \beta)^{1-\lambda} \left[\int f^0(y; \alpha)^\lambda \cdot f^1(y; \beta)^{1-\lambda} dy \right]^{-1}$$

where f^0 and f^1 are the densities of F^0_α and F^1_β. In practice one needs to estimate α and β which is done using the MLE (or pseudo MLE, i.e. the MLE under one of the hypothesis). The asymptotic null distribution of the test statistic is the Normal distribution.

Cox' test statistics have been often criticized for several reasons. One is the lack of accuracy of the approximation of the sample distribution of the statistic by its asymptotic distribution. Another reason but at least as important is the lack of robustness of the testing procedure. Robustness in testing has been introduced by Ronchetti (1982), Rousseeuw and Ronchetti (1979, 1981) who were the first to adapt Hampel's optimality problem for estimators to testing procedures. Hampel's optimality problem for testing procedures can be stated as: Under a bound on the influence of small contamination on the test's level and power (robustness requirement), the power of the test at the ideal model is maximized (efficiency requirement). Victoria-Feser

(1997) shows that Cox' test statistic is not robust in that an infinitesimal amount of contamination can lead to a significant test when the true model is the one under H_0, or in other words that the IF on the level of the test is unbounded. The interesting point is that it is not enough to use robust estimators for α and β in order to get a robust testing procedure, one also needs to downweight the influence of extreme data on the test statistic itself. Victoria-Feser (1997) therefore proposed a robust Cox-type test statistic based on the results for general parametric models of Heritier and Ronchetti (1994).

The asymptotically normal robust Cox-type or generalized Lagrange multiplier (GLM) test statistic is given by

$$U_{\mathrm{GLM}} = \frac{1}{n}\sum_{i=1}^{n}\left[A_{(21)}s^0(x_i;\hat{\alpha}) + A_{(22)}\left[\frac{\partial}{\partial\lambda}\log f^{\mathrm{c}}(x_i;\hat{\alpha},\beta)\bigg|_{\lambda=1} - a_{(2)}\right]\right]w_c(x_i;A,a)$$

where $\hat{\alpha}$ is the MLE of α,

$$w_c(x;A,a) = \min\left\{1;\frac{c}{\left|A_{(21)}s^0(x;\hat{\alpha}) + A_{(22)}\left[\frac{\partial}{\partial\lambda}\log f^{\mathrm{c}}(x;\hat{\alpha},\beta)\big|_{\lambda=1} - a_{(2)}\right]\right|}\right\}$$

and the vector $A_{(21)}$ $(1\times\dim(\alpha))$ and the scalars $A_{(22)}$ and $a_{(2)}$ are determined implicitly (see Victoria-Feser (1997). As with estimators, the robust test statistic can be seen as a weighted classical Cox-type test statistic. In fact, when $c \rightarrow \infty$, we get the classical Cox-type test statistic. Note also that in practice β needs to be estimated consistently like by means of the MLE.

A simulation study shows the performance of the U_{GLM} when compared to the classical Cox-type statistics. Contaminated and uncontaminated Pareto samples of size 200 were simulated (with shape parameter $\alpha = 3$ and fixed scale parameter γ) and the Pareto distribution against the truncated Exponential distribution was then tested using a classical and robust ($c = 2$) Cox-type test statistic (see also Victoria-Feser (1997)). It should be noted that the Pareto distribution is used to model zeromodal income distributions such as high incomes. The samples were contaminated by means of $(1-\varepsilon/\sqrt{200})F_{\alpha,\gamma} + \varepsilon/\sqrt{200}F_{\alpha,10\cdot\gamma}$, where for each amount of contamination 1000 samples were generated. The amount of contamination here is of the type $\varepsilon/\sqrt{n}$ which is appropriate in testing procedures (see also Heritier and Ronchetti (1994). Table 1 gives the actual levels of the classical and robust

Atkinson statistic, i.e. the ones obtained by the simulations. These levels are actually the proportion of times the test is significant when testing the Pareto against the truncated Exponential *and* the test is not significant when testing the truncated Exponential against the Pareto.

	Classical statistic				Robust statistic			
Amount of	Nominal levels				Nominal levels			
contamination ε	1%	3%	5%	10%	1%	3%	5%	10%
0%	2.1	3.1	3.5	5.2	1.3	3.5	5.5	10.2
3%	6.3	8.7	10.3	14.7	1.2	3.3	5.1	10.3
6%	13.1	18.5	22.5	27.6	1.4	3.6	5.4	10.7
10%	24.4	31.3	35.2	43.9	1.3	3.0	5.6	11.4
15%	35.6	44.6	49.9	58.1	1.4	4.1	7.9	14.5
20%	46.3	54.2	58.6	67.1	0.9	4.1	7.6	14.5

Table 1: **Actual levels of the classical and robust** ($c = 2$) **Cox-type test statistic (Pareto against Exponential)**

We can observe that the classical statistic has a very strange behaviour since when there is no contamination the null hypothesis is under-rejected and with only small amounts of contamination, the null hypothesis is over-rejected. The first phenomenon is probably due to the fact that the approximation of the actual distribution of the Cox-type statistic by means of its asymptotic distribution is not accurate and the second phenomenon is the lack of robustness. On the other hand, we find that with the robustified test statistic not only the asymptotic distribution is a good approximation of its sample distribution, but also that the small departures from the model under the null hypothesis do not influence the level of the test at least for amounts of contamination up to about $\varepsilon = 10\%$. With more contamination, the null hypothesis tends to be slightly over-rejected at the 5% and 10% levels, but this is not too drastic compared to the classical case. In other words, the robust test is very stable.

5 Fitting truncated data

By truncated data is meant that for a subset of the space in which the incomes are defined (typically the real line) no information is given about the nature of incomes that belong to that subset. The reasons why there is no information can be different, like for example reasons of confidentiality when high incomes (i.e. those exceeding a fixed level) are not disclosed by data providers, or for practical reasons mentioned in the introduction. With income data from surveys there might also be the problem of nonresponse which is a more complex problem than the one supposed here. Indeed, it is arguable that nonresponse rates may vary depending on the level of income so that an appropriate model that takes into account this relationship is needed: see for example Little and Rubin (1987).

To tackle the problem of truncated data, Victoria-Feser (1993) proposes to use either the marginal distribution or a generalization of the EM algorithm (Dempster, Laird, and Rubin (1977) for M-estimators (EMM algorithm). Let $\mathfrak{X} \subseteq \Re$ be the subset of values of income for which cases are observed and $\overline{\mathfrak{X}} \subseteq \Re$ the subset for which they are not. The EMM algorithm defines an OBRE-type estimator with truncated data and is given by the solution in θ of

$$\int_{\overline{\mathfrak{X}}} [s(x;\theta) - a]\, w_c(x;A,a) dF_\theta(x) + \int_{\mathfrak{X}} dF_\theta(x) \frac{1}{n} \sum_{i=1}^{n} [s(x_i;\theta) - a]\, w_c(x_i;A,a) = 0 \tag{8}$$

where w_c, A, and a are defined as in (5), (6) and (7). The first part of (8) estimates the missing part (expectation step) while the second part of (8) defines the maximization step.

Another approach would be to consider the marginal distribution, i.e.

$$\frac{F_\theta}{\int_{\mathfrak{X}} dF_\theta(x)} \tag{9}$$

and consider its score function and use the OBRE for continuous data. This approach and the EMM one are in general different (they are equivalent for the MLE), but their difference is small and due to the assumptions on the underlying model. Indeed, contrary to the marginal distribution approach, the EMM algorithm assumes that the underlying model is the complete one i.e. F_θ and not (9). This leads to a difference in the way the weights associated to each observation are computed; see Victoria-Feser (1993). To show the properties of the EMM approach with truncated data, we present a simulation study. Table 2 gives the bias and the MSE for the MLE and the OBRE (EMM) in the case of contaminated and truncated data (below a minimum value), that where generated by a Gamma distribution with scale parameter $\lambda = 1$ and shape parameter $\alpha = 3$. The contamination consists on 1% of randomly chosen data ten times their value.

Information loss (%)	Parameter	OBRE Bias	OBRE MSE	MLE Bias	MLE MSE
7.5	α	0.07	0.24	-2.39	5.71
	λ	0.02	0.03	-0.75	0.57
2.5	α	0.06	0.17	-2.39	5.71
	λ	0.02	0.0202	-0.78	0.39

Table 2: **Bias and MSE of the OBRE ($c = 2$) for truncated Gamma data**

The behaviour of the MLE is not satisfactory, whereas the OBRE has a small bias. In fact, the MLE is even more biased than in the case of complete data (see Victoria-Feser and Ronchetti (1994). This is not surprising, for truncating means incomplete information and robust estimators are constructed to deal with approximate models.

6 Fitting grouped data

Another feature of income data is that they can be available only in a grouped form, i.e. in the form of frequencies per class. In these cases the OBRE of (Hampel, Ronchetti, Rousseeuw, and Stahel (1986) cannot be used as such. The problem can be thought as a discrete data problem with a continuous underlying model. In addition to the usual effects of contamination, we have in this situation grouping effects where some observations may be shifted from one class to another because of rounding errors or class definition. A general class of estimators (Minimum Power Divergence Estimators or MPE) for the parameters of the underlying model based on grouped data was defined by Cressie and Read (1984). It includes the MLE, the minimum Hellinger distance estimator (MHDE), the estimator based on Pearson χ^2, etc. Though their *IF* is bounded, deviations from the underlying model can cause a large bias especially in classes with low probabilities. Victoria-Feser and Ronchetti (1997) propose a more general class of estimators containing the MPE, called MGP, which can be seen as M-estimators for grouped data. An optimal estimator (OBRE) that minimizes the asymptotic covariance matrix under a bound on its *IF* is derived, and it is shown that for a small efficiency loss, it is more robust than its classical (MPE) counterpart. Let p_j be the observed relative frequencies in J non overlapping classes $I_1, \ldots, I_J$ and $k_j(\theta) = \int_{I_j} dF_\theta(x)$ the corresponding true relative frequencies, the optimal MGP estimators are then given by the solution in θ of

$$\sum_{j=1}^{J} \left(\frac{p_j}{k_j(\theta)}\right)^{\gamma} A\left[\frac{\partial}{\partial\theta}k_j(\theta) - ak_j(\theta)\right] w_{c,j}(A,a) = 0 \tag{10}$$

where γ is an arbitrary constant (to be defined),

$$w_{c,j}(A,a) = \min\left\{1; \frac{c}{\left\|A\left[\frac{\partial}{\partial\theta}\log k_j(\theta) - a\right]\right\|}\right\}$$

and where the matrix $A = A(\theta)$ and vector $a = a(\theta)$ are determined implicitly by the equations

$$\sum_{j=1}^{J} A\left[\frac{\partial}{\partial\theta}k_j(\theta) - ak_j(\theta)\right] w_{c,j}(A,a) = 0$$

$$\sum_{j=1}^{J} A\left[\frac{\partial}{\partial\theta}k_j(\theta) - ak_j(\theta)\right] \frac{\partial}{\partial\theta^T}\log k_j(\theta) w_{c,j}(A,a) = I$$

Note that (10) defines a set of robust estimators. Each classical estimator has its robust counterpart. For example, the robust version of the MLE is obtained by setting $\gamma = 1$ and the robust MHDE by setting $\gamma = 0.5$. The bound c allows the control over the degree of robustness. Lowering c increases robustness but at the same time decreases efficiency. In practice we found however that a good degree of robustness can be achieved without loosing too much efficiency.

We illustrate the performance of MGP estimators with a small simulation from a Pareto distribution. The data are contaminated by taking ε proportion of randomly chosen observation and multiplying them by 10, and then grouped in 22 classes. The MPE and OBRE (for the MLE and MHDE) are computed and their Bias and MSE reported in Table 3 (the standard errors for the values of the bias are less than 0.02).

ε	MPE				OBRE			
	MLE		MHDE		Rob MLE		Rob MHDE	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0%	0.01	0.0101	0.05	0.013	0.02	0.0202	0.01	0.011
1%	-0.19	0.05	-0.06	0.015	-0.03	0.02	-0.01	0.015
3%	-0.55	0.27	-0.18	0.05	-0.13	0.04	-0.04	0.015
5%	-0.77	0.61	-0.29	0.11	-0.25	0.08	-0.06	0.02

Table 3: **Bias and MSE of classical and robust estimators for grouped data on Pareto simulated data**

We can see that although the *IF* of the MLE is bounded, when the underlying model is contaminated, the MLE has a large bias. On the other hand, with the corresponding robust estimator we can see that this bias and the overall MSE are considerably smaller. The MHDE has better robustness properties than the MLE, but it can be improved by using the corresponding robust version which has the best bias and MSE overall in the example.

The robust procedure is also applied to the UK data set presented above. The data are grouped in 58 equally sized classes. The first class is extended to 0 and the last class to ∞. The Gamma distribution is chosen as candidate to model the data and the MLE, MHDE and robust MHDE ($c = 175$ which corresponds to 95% efficiency) are computed. The histogram of the data and the estimated densities are presented in Figure 3.

Figure 3 here

We can see that while the MLE tries to accommodate the tails of the distribution, it misses the description of the bulk of the data in the center. The MHDE improves the fit of the majority of the data but it is its robust version which gives the best fit. The latter has an efficiency of 95% compared to the classical MPE. Finally, Victoria-Feser and Ronchetti (1997) show that the OBRE for grouped data has a bounded local shift sensitivity and therefore it controls the effects of grouping errors as well. This is not the case with classical MPE.

So far we have looked at statistical methods to describe the distribution of incomes. The parameters' estimates can be used to conduct further analyses such as inequality or poverty analysis. This will be developed in the following sections.

7 Estimating inequality

Inequality measures and other related tools of income distribution analysis are often used to summarize information about income distributions. They play an important part in political debates about economic and social trends and in welfare economics. At a theoretical level, inequality measures have been derived by requiring a number of essential properties. The most important (and most accepted) property is the principle of transfers (Dalton (1920), which states that the transfer of an arbitrary positive amount of income from a poorer income receiver to a richer one (such that the mean of the distribution is preserved) should increase the value of the inequality measure. On the other hand, inequality measures are estimated from income data. Therefore, it is important to understand the relationship between the economic properties that the inequality measure should fulfil and the statistical properties of the corresponding estimator. In this section we summarize some of the results obtained in Victoria-Feser (1993) and Cowell and Victoria-Feser (1996b) on the relationship between economic and robustness properties of inequality measures. The aim is to find a simple and convenient way to check if an inequality measure is resistant against data contamination and draw conclusion about general classes of inequality indices. It can be shown that the principle of transfers alone does not imply necessarily non resistance of an inequality measure. However, if another property is added, like the scale independence or decomposability of the index (see Cowell and Victoria-Feser (1996b) which restricts the class but still encompasses most of the widely used inequality indices, or if a more realistic specification of the estimation problem is considered (like the mean income has to be estimated rather than being specified a priori), then it can be shown that any inequality index satisfying these properties is not robust in that its IF is unbounded. In particular this is the case for the Kolm index (Kolm (1976a and Kolm (1976b), the generalized entropy family (Cowell (1980) which includes the Theil indices (Theil (1967), and the Gini index (Gini (1910).

To show the effect of data contamination on the estimator of the inequality, 100 samples of 200 observations from a Pareto distribution were generated and contaminated by multiplying by 10 a proportion of randomly chosen observations. Three indices are computed, the Gini index, the Theil index and the coefficient of variation index (CV) and the average values of these indices are given in Table 4. (The second row represents the true values of the indices and SD stands for standard deviation.)

Contamination	Gini (0.2000)	SD	Theil (0.0945)	SD	CV (0.1667)	SD
0%	0.1974	0.002	0.0908	0.003	0.1545	0.02
1%	0.2577	0.002	0.2068	0.005	0.4974	0.02
2%	0.2845	0.002	0.2584	0.006	0.6466	0.03
3%	0.3319	0.002	0.3485	0.007	0.8993	0.04
4%	0.3542	0.002	0.3920	0.007	1.0180	0.04
5%	0.3891	0.002	0.4500	0.007	1.1150	0.04

Table 4: **Average values of standard inequality indices from simulated uncontaminated and contaminated Pareto data**

We can see that the three indices are very sensitive to small contamination in the data. Cowell and Victoria-Feser (1996b) contains a more complete study involving other income distribution models. To avoid the effect of contamination on the estimated indices, one way is to estimate inequality from robust estimates of the parameters of income distribution models. Cowell and Victoria-Feser (1996b) show for example that for the generalized entropy family indices, the IF of the members of this class is proportional to the IF of the estimators of the parameters of the underlying income model. Therefore, if robust estimators for the parameters of the income model are used, then the resulting inequality measure will also be robust. These robust estimators of inequality convey a lot of information in that they provide a check against classical estimates; where discrepancies between the results emerge and are attributable to small deviations from the assumed model, this information should be taken into account in drawing conclusions about the "true" picture of inequality. For example, using the UK dataset, the three inequality indices presented above are estimated through the MLE and OBRE of the Gamma distribution and compared to direct (empirical) estimates on the whole sample and on a truncated sample in which 2% of the top incomes have been removed. This is to show to what extent the inequality measures considered here are sensitive to extreme values. The results are presented in Table 5. They reveal two interesting points. First we can see that the OBRE produces a distribution that exhibits uniformly lower inequality than that produced by the MLE (even by a factor 2!). Secondly, the same phenomenon can be observed by comparing the two direct estimates; the upper truncation of the sample has an impact upon inequality that is similar to switching from the MLE to the OBRE. This in not surprising when one recalls Figure 1.

	Gini	Theil	CV
MLE	0.1375	0.0299	0.0302
OBRE ($c = 2$)	0.0983	0.0152	0.0153
Direct estimates on whole sample	0.1287	0.0320	0.0358
Direct estimates on truncated sample	0.1120	0.0222	0.0228

Table 5: **Inequality indices on the UK data, using the Gamma distribution**

It should be stressed that the point here is not to say that the robust estimates are better than the classical ones in estimating inequality, they give however another picture. The assumption underlying the computation of the indexes is that the data follow a Gamma distribution. This is obviously not exactly the case, since there are differences between the MLE and the OBRE, but one can argue that for the majority of the data, the Gamma distribution is appropriate. Unfortunately, the data which are considered as "extreme" for the Gamma distribution are also those which have an important influence on the inequality measure, and also correspond to the truncated data, which altogether explain the result presented in Table 5. The use of a robust estimator in this particular case, reveals that a few incomes make a great difference in the inequality measure, but which measure is more appropriate is not an easy question to answer.

8 Evaluating poverty

Since the pioneering article by Sen (1976) poverty evaluation has become a prolific area of research in economics. The important theoretical issues are the determination of the poverty line below which an economic unit such as a household is considered as poor, the functions (indices) that should be used for aggregating the incomes of the poor, and how statistical inference can be conducted. All these issues are of course linked. We focus here on the robustness properties of poverty measures as developed in Cowell and Victoria-Feser (1996a).

A very important class of poverty indices (additively separable poverty indices) is defined at the distribution of income F by the functional

$$P(F;z) = \int p\left(z(F),x\right)dF(x)$$

where $z(F)$ is the poverty line which is determined by the income distribution itself (endogenous). Several important poverty measures emerge from this class. One simple but well known measure is the headcount ratio (proportion of poor) when $p(z(F),x) = \iota(x < z(F))$, another is the poverty gap (a modified proportion of poor which takes into account the relative size of poverty) when $p(z(F),x) = \iota(x < z(F))(z(F)-x)/z(F)$, and an more general one is the Foster-Greer-Thorbeke poverty index (Foster, Greer, and Thorbeke (1984) when $p(z(F),x) = \iota(x < z(F))\left[(z(F)-x)/z(F)\right]^{\gamma}$ (see also Cowell and Victoria-Feser (1994).

In practice poverty is estimated using the empirical distribution $F^{(n)}$ and one might wonder whether the effect of contaminated data can have an important influence on the estimated level of poverty. Cowell and Victoria-Feser (1996a) have tackled the issue and found that an infinitesimal amount of contamination can have a catastrophic effect on the poverty measure only through the evaluation of the poverty line. Therefore, if the latter is robustly estimated, then the analyst is quite safe.

Ayadi, Matoussi, and Victoria-Feser (1998) have implemented this result in a poverty analysis in Tunisia. The central issues of the research was to evaluate the poverty differential between urban and rural regions in different economic parts in Tunisia in a robust fashion. This lies heavily on a correct specification of the poverty line. The data at hand were on household consumption (survey of 1990) so that the approach of Ravallion and Bidani (1994) which permits to evaluate poverty with this type of data was followed. Tunisia is divided in 3 major economic regions, namely the Grand Tunis (around the capital), the littoral at the Mediterranean see, which has known since the independence an economic prosperity, and the interior (western part) which has several acute social and economic problems. The littoral and interior can also be divided into a rural and urban part. To determine poverty, the poverty line is divided into two parts, namely the amount of income to satisfy the food and the non food needs (such as housing, clothing, etc.). To take into account regional differences, consumption behaviour of the poor is estimated by determining the "average" food products or consumer's basket by using the data on quantity of different food products. This consists in finding the center of multivariate data. It is important to estimate this center in a robust fashion, because the consumer's basket is then evaluated in local market prices and constitutes an important part of the poverty line. The non food part is then estimated using a so-called AIDS model (Ravallion (1993) which is a regression model of income food share on total income and socio-economic variables related to the household. Here again, in order to avoid deviations due to a small number of extreme data, a robust approach is adopted (see Ayadi, Matoussi, and Victoria-Feser (1998).

To summarise the analysis, Table 6 presents three poverty indices calculated for each region of Tunisia following a robust approach to the evaluation of the poverty line compared to a classical one (they are the so-called lower poverty lines, see Ayadi, Matoussi, and Victoria-Feser (1998). The indices are the headcount ratio (HCR), the poverty gap (PG) and the Foster-Greer-Thorbeke (FGT) poverty index with parameter $\gamma = 2$. We can draw the following conclusions. In the more developed regions (Grand Tunis and Urban Littoral) not only poverty is less intense but also there is no real difference between the classical and robust approach. This is because the classical and robust estimators of the poverty lines give very similar values. In the other regions the differences between the classical and robust approaches are quite substantial (up to 45%), because the classical and robust estimators of the poverty lines differ substantially (see Ayadi, Matoussi, and Victoria-Feser (1998). When urban and littoral regions are compared, one can see that poverty is mainly rural which goes against some previous beliefs about the location of the poor. This difference is less important when the robust analysis is considered.

	Robust Analysis			Classical Analysis		
	HCR	PG	FGT	HCR	PG	FGT
Grand Tunis	4.4	0.81	0.24	4.3	0.79	0.24
Urban littoral	3.6	0.57	0.15	3.6	0.61	0.16
Urban interior	9.1	2.12	0.75	13.1	3.36	1.26
Rural littoral	8.4	1.63	0.54	10.9	2.42	0.82
Rural interior	11.9	3.15	1.18	16.0	4.28	1.66

Table 6: **Robust and classical poverty indices for the regions of Tunisia**

9 Ordering income distributions

The use of inequality measures is thought by many as being rather restrictive since the whole information about the distribution of incomes is summarized in a single number. To compare two or more income distributions one therefore often use ordering tools of which the Lorenz curve is certainly the most well known. These tools are mainly based on quantiles and cumulative distributions. For example, so-called first-order dominance criteria are simply defined for a distribution F by the quantile functional $Q = Q(F, q)$, i.e. by q^{th} quantiles of the distribution F. For example $Q(F, 0.5)$ is the median. Two (or more) distributions $F1$ and $F2$ are then compared at each level of q and if for $F1$ all the quantiles are larger or equal to those for $F2$, then it is said that $F1$ first-order dominates $F2$. The economic implication is that given a welfare function (with some rather mild conditions on it), the level of welfare in the population with income distribution $F1$ is higher. In practice, one can get three conclusions: dominance of one or the other distribution, and also no dominance of either because the order is reversed at some q's. A stronger concept of dominance (in terms of conditions on the welfare function) are the so-called second-order dominance tools based on the cumulative distribution functional C

$$C(F, q) = \int^{Q(F,q)} x dF(x)$$

In words, $C(F, q)$ is the total income earned by the q proportion of the poorest. The function $q \mapsto C(F, q)/\mu(F) = L_q$ defines the Lorenz curve which for each proportion q plots the proportion of total income of the population earned the q proportion of the poorest.

The issue which is of interest here is to investigate whether ordering tools can be influenced by data contamination as to lead to reversed orders when comparing empirical distributions. Cowell and Victoria-Feser (1996c) make this investigation using the IF and conclude that infinitesimal amounts of contaminations (in the tails of the distribution) can completely distort the picture given by second-order dominance criteria. How the problem is solved in practice is not obvious. The recourse to parametric models is not a good solution since one can prove that with some models, the resulting ordering curves never cross. Cowell and Victoria-Feser (2000a,b) propose two approaches, a pragmatic one based on trimmed samples and a semi-parametric one. The first one needs some more investigation into it, whereas the semi-parametric one seems very promising. The idea is to use a parametric model to fit the tail of the distribution and then construct an empirical ordering curve for the bulk of the distribution mixed with a parametric ordering tool for the tail of the distribution. The appropriate model is the Pareto model and the fit should be robust. One question that is left open is what proportion of the data in the tail should be considered?

To illustrate the point, Cowell and Victoria-Feser (2000a) made the following simulation exercise. Two samples of 10 000 observations where simulated from a Dagum I distribution which has the property that for large incomes the distribution converges to the Pareto distribution. The parameters were chosen in order to get two distributions such that one exactly dominates the other. Let $F1^{(n)}$ be the empirical distribution that dominates and $F2^{(n)}$ the one which is dominated. $F1^{(n)}$ was contaminated by multiplying 0.25% of the largest observations by 10 giving the empirical mixture distribution $F1_{\varepsilon}^{(n)}$. The three empirical distributions are given in Figure 4.

Figure 4 here

We can see that with only 0.25% of extreme data, the ordering of the distributions is completely reversed. When the upper tail is modelled using the Pareto distribution (5% of the upper tail), the situation changes. Figure 5 depicts the empirical Lorenz curve of $F2^{(n)}$ together with the semiparametric Lorenz curves of $F1_{\varepsilon}^{(n)}$, using the classical MLE and a robust estimator for the parameter of the Pareto model (see Cowell and Victoria-Feser (2000a)). We can see that with a robust semiparametric Lorenz curve the ordering is preserved. Actually there is no visual difference between the robust semiparametric Lorenz curve on contaminated data and the empirical Lorenz curve on non contaminated data for the same original distribution (see Cowell and Victoria-Feser (2000a)). The semiparametric Lorenz curve using the MLE is also distorted by the extreme data, although it is less distorted than the empirical Lorenz curve on contaminated data. This is not surprising since it is well known that the MLE is not robust.

Figure 5 here

10 Conclusions

We showed that robust techniques can play a useful role in income distribution analysis by providing more reliable fits and tests, more stable inequality measures and ordering tools. We don't advocate the unique use of robust techniques but we believe that they provide very useful information when combined together with a classical approach. They also have the advantage of being rather objective methods in that they do not rely on the judgment of the analyst in deciding often by eye which of the data could be considered as extreme. There are however still some developments to be made in this field of research, especially on the computational aspects. Indeed, although robust techniques are very important they are rather cumbersome to program and very computer intensive. A solution could be found in developing less efficient robust estimators which are easier to compute. But this another story.

References

Atkinson, A. C. (1970). A method for discriminating between models. *Journal of the Royal Statistical Society, Serie B 32*, 323–353.

Ayadi, M., M. S. Matoussi, and M.-P. Victoria-Feser (1998). Urban rural poverty comparisons in Tunisia: A robust statistical approach. Cahiers du département d'econométrie, University of Geneva.

Ben Horim, M. (1990). Stochastic dominance and truncated sample data. *Journal of Financial Research 13*, 105–116.

Cowell, F. A. (1980). Generalized entropy and the measurement of distributional change. *European Economic Review 13*, 147–159.

Cowell, F. A. and M.-P. Victoria-Feser (1994). Robustness properties of poverty indices. DARP discussion papers no 8, STICERD, London School of Economics, London WC2A 2AE.

Cowell, F. A. and M.-P. Victoria-Feser (1996a). Poverty measurement with contaminated data: A robust approach. *European Economic Revue 40*, 1761–1771.

Cowell, F. A. and M.-P. Victoria-Feser (1996b). Robustness properties of inequality measures. *Econometrica 64*, 77–101.

Cowell, F. A. and M.-P. Victoria-Feser (1996c). Welfare jugments in the presence of contaminated data. DARP discussion paper no 13, STICERD, London School of Economics, London WC2A 2AE.

Cowell, F. A. and M.-P. Victoria-Feser (2000a). Robust Semi-Parametric Lorenz Curves. DARP discussion paper no 50, STICERD, London School of Economics, London WC2A 2AE.

Cowell, F. A. and M.-P. Victoria-Feser (2000b). Distributional Dominance with Dirty Data. DARP discussion paper no 51, STICERD, London School of Economics, London WC2A 2AE.

Cox, D. R. (1961). Tests of separate families of hypotheses. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability 1*, Berkeley, pp. 105–123. University of California Press.

Cox, D. R. (1962). Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society, Serie B 24*, 406–424.

Cressie, N. A. C. and T. R. C. Read (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, Serie B 46*, 440–464.

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

Dalton, H. (1920). The measurement of the inequality of incomes. *Economic Journal 30*, 348–361.

Dempster, A. P., M. N. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Serie B 39*, 1–22.

Department of Social Security (1992). *Households below Average Income, 1979-1988/89*. London: HMSO.

Fichtenbaum, R. and H. Shahidi (1988). Truncation bias and the measurement of income inequality. *Journal of Business and Economic Statistics 6*, 335–337.

Foster, J., J. Greer, and E. Thorbeke (1984). A class of decomposable poverty measures. *Econometrica 52*, 761–765.

Gini, C. (1910). Indici di concentrazione e di dipendenza. In Gini 1955, p. 3–120. Atti della III Reunione della Sociatà Italiana per il Progresso delle Scienze.

Gini, C. (1955). *Memorie di Metodologia Statistica*. Roma: Pizetti e Salvemini Libreria Eredi Virgilio Veschi.

Groves, R. M. (1989). *Survey Errors and Survey Costs*. New York: John Wiley.

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

Heritier, S. and E. Ronchetti (1994). Robust bounded-influence tests in general parametric models. *Journal of the American Statistical Association 89*(427), 897–904.

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.

Jenkins, S. P. (1997). Trends in real income in Britain: A microeconomic analysis. *Empirical Economics 22*, 483–500.

Kolm, S. C. (1976a). Unequal inequalities. *Jounal of Economic Theory 12*, 416–442. Part I.

Kolm, S. C. (1976b). Unequal inequalities. *Jounal of Economic Theory 13*, 82–111. Part II.

Little, R. J. A. and D. B. Rubin (1987). *Statistical Analysis with Missing Data*. New York: Wiley.

McDonald, J. B. and M. R. Ransom (1979). Functional forms, estimation techniques and the distribution of income. *Econometrica 47*, 1513–1525.

Nelson, R. D. and R. D. Pope (1990). Imprecise tail estimation and the empirical failure of stochastic dominance. *ASA Proceedings, Business and Economic Statistics Section*, 374–379.

Prieto Alaiz, M. and M.-P. Victoria-Feser (1996). Modelling income distribution in Spain: A robust parametric approach. DARP discussion paper no 20, STICERD, London School of Economics, London WC2A 2AE.

Ravallion, M. (1993). *Poverty comparison*, Volume 56 of *Fundamentals of Pure and Applied Economics*. Chur, Switzerland: Harwood Academic Press.

Ravallion, M. and B. Bidani (1994). How robust is a poverty profile? *The World Bank Economic Review 8*, 75–102.

Ronchetti, E. (1982). *Robust Testing in Linear Models: The Infinitesimal Approach*. Ph. D. thesis, ETH, Zurich, Switzerland.

Rousseeuw, P. J. and E. Ronchetti (1979). The influence curve for tests. Research Report 21, ETH Zürich, Switzerland.

Rousseeuw, P. J. and E. Ronchetti (1981). Influence curves for general statistics. *Journal of Computational and Applied Mathematics 7*, 161–166.

Sen, A. K. (1976). Poverty: An ordinal approach to measurement. *Econometrica 44*, 219–231.

Theil, H. (1967). *Economics and Information Theory*. Amsterdam: North-Holland.

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. (1993). *Robust Methods for Personal Income Distribution Models*. Ph. D. thesis, University of Geneva, Switzerland. Thesis no 384.

Victoria-Feser, M.-P. (1997). Robust model choice test for non-nested hypothesis. *Journal of the Royal Statistical Society, Serie B 59*, 715–727.

Victoria-Feser, M.-P. and E. Ronchetti (1994). Robust methods for personal income distribution models. *The Canadian Journal of Statistics 22*, 247–258.

Victoria-Feser, M.-P. and E. Ronchetti (1997). Robust estimation for grouped data. *Journal of the American Statistical Association 92*, 333–340.

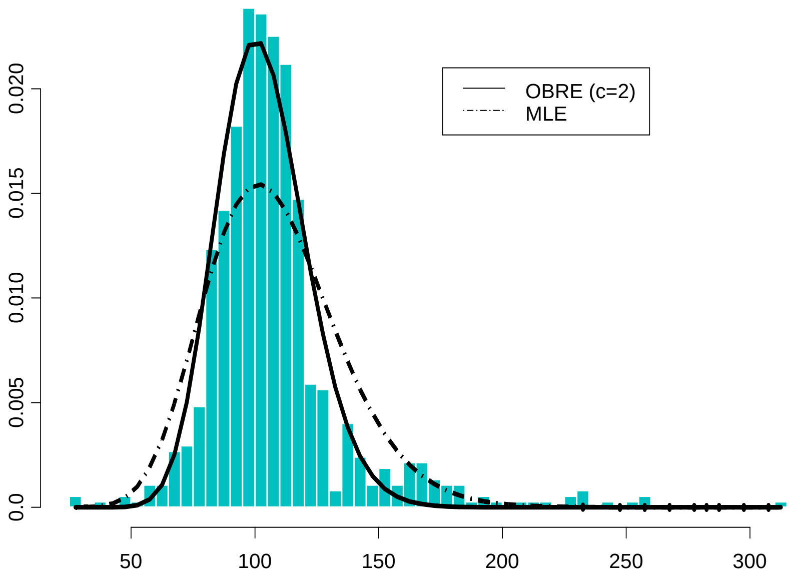

Figure 1: Gamma fitting of UK data

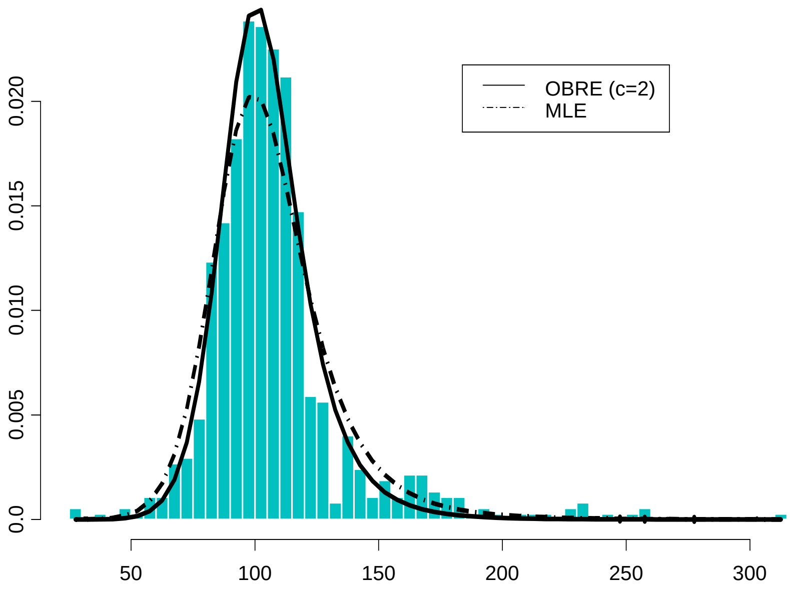

Figure 2: Dagum-I fitting of UK data

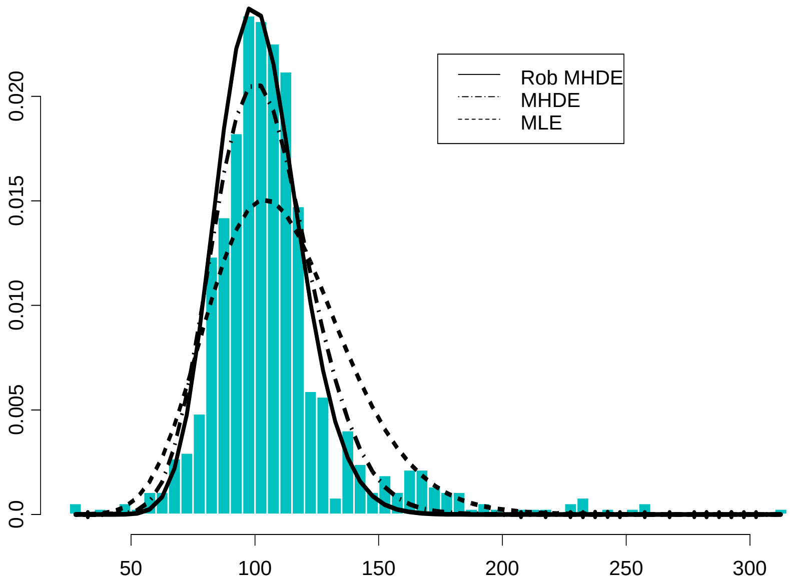

Figure 3: Gamma modelling of UK grouped data

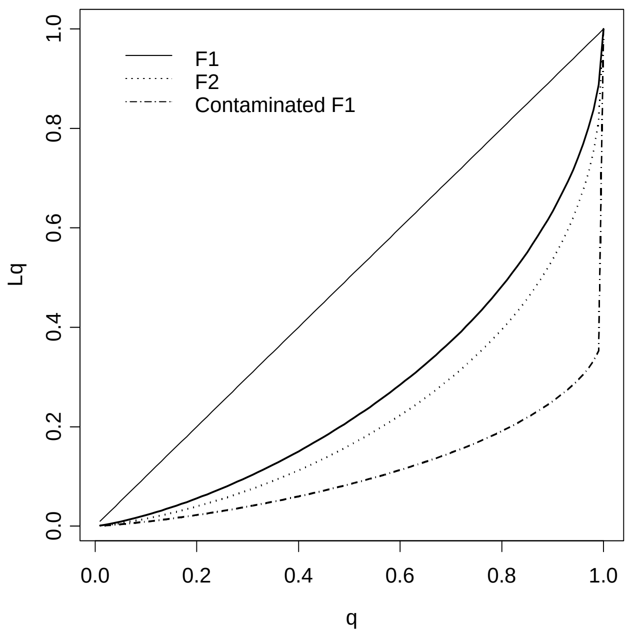

Figure 4: Lorenz curves comparisons of uncontaminated and contaminated emprirical distributions

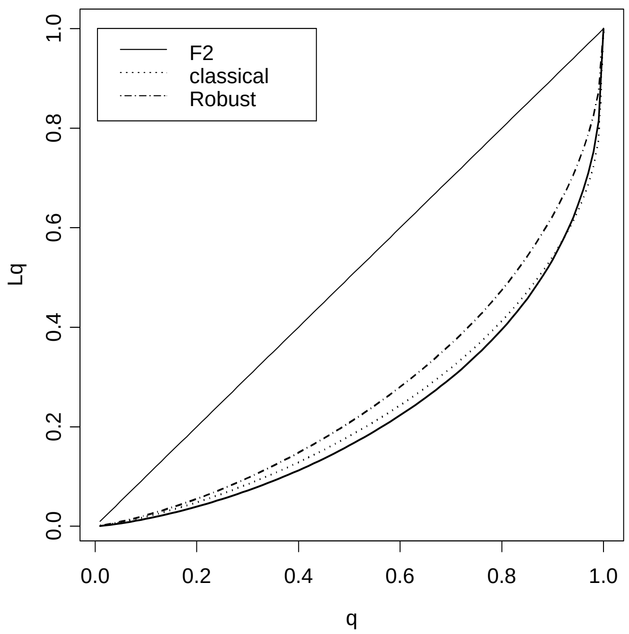

Figure 5: Semiparametric Lorenz curves: classical and robust