Robust Income Distribution Estimation with Missing Data

by

Maria-Pia Victoria-Feser
University of Geneva

Discussion Paper
No. DARP 57
September 2000

Distributional Analysis Research Programme
The Toyota Centre
Suntory and Toyota International Centres for Economics and Related Disciplines
London School of Economics
Houghton Street
London WC2A 2AE

Distributional Analysis Research Programme

The Distributional Analysis Research Programme was established in 1993 with funding from the Economic and Social Research Council. It is located within the Suntory and Toyota International Centres for Economics and Related Disciplines (STICERD) at the London School of Economics and Political Science. The programme is directed by Frank Cowell. The Discussion Paper series is available free of charge and most papers are downloadable from the website. To subscribe to the DARP paper series, or for further information on the work of the Programme, please contact our Research Secretary, Sue Coles on:

Telephone: UK+20 7955 6678
Fax: UK+20 7955 6951
Email: s.coles@lse.ac.uk
Web site: http://sticerd.lse.ac.uk/DARP

ABSTRACT

With income distributions it is common to encounter the problem of missing data. When a parametric model is fitted to the data, the problem can be overcome by specifying the marginal distribution of the observed data. With classical methods of estimation such as the maximum likelihood (ML) an estimator of the parameters can be obtained in a straightforward manner. Unfortunately, it is well known that ML estimators are not robust estimators in the presence of contaminated data. In this paper, we propose a robust alternative to the ML estimator with truncated data, namely one based on M-estimators that we call the EMM estimator. We present an extensive simulation study where the EMM estimator based on optimal B-robust estimators (OBRE) is compared to a more conservative approach based on marginal density (MD) for truncated data, and show that the difference lies in the way the weights associated to each observation are computed. Finally, we also compare the EMM estimator based on the OBRE with the classical ML estimator when the data are contaminated, and show that contrary to the former, the latter can be seriously biased.

Keywords: M-estimators, influence function, EM algorithm, truncated data.

JEL classification: C13, D31, D63.

Correspondence to:
Professor Maria-Pia Victoria-Feser,
Faculty of Psychology and Education
Faculty of Economics and Social Sciences
University of Geneva
40, Bd du Pont d'Arve
CH-1211 Geneva 4
Tel: + 41 22 705 8218

Acknowledgements

The author was partially supported by the Swiss National Fund for Scientific Research. She would like to thank Professors E. Ronchetti and F. A. Cowell for helpful suggestions and comments.

1 Introduction

This paper deals with the problem of estimating parameters of a univariate distribution when some of the data are missing. The classical approach, ML estimation, specifies the marginal density of the random variable and the ML estimator is obtained by maximizing the log-likelihood of this marginal density. This is equivalent to a simple procedure in which sufficient statistics are adjusted and then used in the maximization procedure. When it comes to robust estimators, and we will focus on optimal B-robust estimators (OBRE) developed in Hampel, Ronchetti, Rousseeuw, and Stahel (1986), there are different approaches, depending on the way the weights associated with each observation are computed.

In the particular problem of income distribution analysis, it is well known that income data are not always as "clean" as one would like them to be; see Cowell and Victoria-Feser (1996). Moreover, income data are typically truncated: either in the low incomes because of taxation rules or in the high incomes for confidentiality reasons. For these reasons it is important to develop robust estimators that taken account of these common features of the data.

The paper is organized as follows. In section 2 the classical framework of ML estimation is presented and an interpretation of the estimation procedure is developped. This is important because it gives the motivation for the robust procedures that will be proposed in section 3. Section 4 presents a simulation study with the Gamma distribution, a commonly used model for income distribution. Finally, section 5 concludes.

2 Maximum likelihood estimation for truncated data

2.1 Specification

Let X be a (univariate) random variable, generally the income variable, defined in $\mathcal{X} \subseteq \mathbb{R}$ and let F_θ be a parametric model with corresponding density $f(\cdot;\theta)$ and score function $s(\cdot;\theta) = \frac{\partial}{\partial\theta}\log f(\cdot;\theta)$, such that $X \sim F_\theta$. Our aim is to estimate θ given a sample $y_1, \ldots, y_n$ only observable in a subset $\mathcal{Y}$ of $\mathcal{X}$, i.e. the data are not observed in $\mathcal{X}/\mathcal{Y} = \overline{\mathcal{Y}}$. Typically with income distributions, the missing data are actually truncated data in the lower and/or upper tail of the distribution. To estimate θ one specifies a model for the random variable Y which has generated the observed values given that $X \sim F_\theta$, and therefore $Y \sim \frac{F_\theta}{\int_{\mathcal{Y}} dF_\theta(x)}$, the marginal distribution. The score function for such a model is given by

$$\begin{aligned}\widetilde{s}(y;\theta) &= \frac{\partial}{\partial\theta}\left[\log f(y;\theta) - \log\int_{\mathcal{Y}} dF_\theta(x)\right] \\ &= s(y;\theta) - \frac{1}{\int_{\mathcal{Y}} dF_\theta(x)}\int_{\mathcal{Y}} s(x;\theta)dF_\theta(x) \qquad (1)\end{aligned}$$

The ML estimator is a consistent estimator over the marginal distribution, i.e.

$$\int_{\mathcal{Y}} \widetilde{s}(x;\theta)\frac{dF_\theta(x)}{\int_{\mathcal{Y}} dF_\theta(x)} = 0$$

but we also have that

$$\int_{\mathcal{X}} s(x;\theta)dF_\theta(x) = 0$$

Then the ML estimator can be written as the solution in θ of

$$\int_{\overline{\mathcal{Y}}} s(x;\theta)dF_\theta(x) + \int_{\mathcal{Y}} dF_\theta(x)\frac{1}{n}\sum_{i=1}^{n} s(y_i;\theta) = 0 \qquad (2)$$

2.2 Computation

In general one can use a Newton-Raphson method to solve (2), but in the case when F_θ belongs to the class of exponential families the procedure simplifies considerably.. Suppose first that $f(x;\theta)$ has the regular exponential-family form

$$f(x;\theta) = \frac{1}{d(\theta)}b(x)\exp\left[t(x)^T\theta\right]$$

where $t(x)$ denotes a $p \times 1$ vector of complete data sufficient statistics ($p = dim(\theta)$) and $d(\theta)$ is a constant which makes the integration of $f(x;\theta)$ equal to unity. Intuitively, when one has to estimate the parameters of a model with truncated data, one would try to complete the information through the estimation procedure. As a first step, one can compute the ML estimators ignoring the fact that the data set is truncated. In a second step, as these estimators depend on statistics calculated by means of the truncated data, one has to transform these statistics in order to push them near to the value they would have taken if the data set was complete. This intuition makes sense because for a parametric model from the class of exponential families we have

$$s(y;\theta) = -\frac{\partial}{\partial\theta}\log d(\theta) + t(y)$$

Putting this in (2) we get

$$-\frac{\partial}{\partial\theta}\log d(\theta) + \int_{\overline{\mathcal{Y}}} t(x)dF_\theta(x) + \int_{\mathcal{Y}} dF_\theta(x)\frac{1}{n}\sum_{i=1}^{n} t(y_i) = 0 \qquad (3)$$

Thus, in the framework of exponential families, one can use the following algorithm. Suppose that $\theta^{(k)}$ denotes the current value of θ after k cycles of the algorithm. The next cycle can be described in two steps, as follows:

- *E-step:* Estimate the complete-data sufficient statistics $t(x)$ by finding

$$t^{(k)} = \int_{\overline{\mathcal{Y}}} t(x) dF_{\theta^{(k)}}(x) + \int_{\mathcal{Y}} dF_{\theta^{(k)}}(x) \frac{1}{n} \sum_{i=1}^{n} t(y_i)$$

- *M-step:* Determine $\theta^{(k+1)}$ as the solution of the equation $\frac{\partial}{\partial \theta} \log d(\theta) = t^{(k)}$.

It should be stressed that this algorithm is nothing else but the EM algorithm proposed by Dempster et al. (1977).

2.3 Properties

With income data one often encounters the problem of extreme data or outliers. Extreme data can be seen as model misspecification in that the postulated model is not exact so that a few data don't fit into the assumption that $X \sim F_\theta$. If the extreme data represent a very small proportion of the observations, it is sensible to expect from the estimators not to be too much influenced by these data points. This is what is understood by the robustness property of estimators. To investigate this property, one uses the so-called *Influence Function* (IF) which is defined for a statistic written as a functional $T(F)$ as its first order approximation in a neighbourhood of F of infinitesimal size $F_\varepsilon = (1-\varepsilon)F + \varepsilon\Delta_z$ where Δ_z is the distribution with point mass 1 at an arbitrary point z. Formally, for differentiable functionals the IF is

$$IF(z; T, F) = \frac{\partial}{\partial \varepsilon} T(F_\varepsilon) \mid_{\varepsilon=0}$$

The IF was first introduced to assess the robustness properties of T by Hampel (1968, 1974). F_ε defines a neighbourhood of F of radius ε and can be seen as a contaminated version of the true model F. Data generated by F_ε are generated by F with probability $(1-\varepsilon)$ and by another arbitrary distribution Δ_z with probability ε. If ε is small (as it is supposed usually) it is desirable to have statistical procedures that are not influenced too much by these model deviations. A statistic with an unbounded IF is said to be non-robust because in this case an infinitesimal amount of contaminated data can drive the value of the statistic by itself (Hampel et al. 1986). From the delta method, the IF can be used to compute the asymptotic covariance matrix of T in that

$$n \operatorname{var}(T) = \int IF(z; T, F) IF(z; T, F)^T dF(z)$$

The IF of the ML estimator $\hat{\theta}$ for the parametric model F_θ when the data are truncated is given by

$$IF(z;\hat{\theta},F_\theta) = M^{-1}(s,\theta)\left[s(z;\theta) - \frac{1}{\int_{\mathcal{Y}} dF_\theta(x)}\int_{\mathcal{Y}} s(x;\theta)dF_\theta(x)\right] \tag{4}$$

where

$$\begin{aligned} M(s,\theta) \quad = \quad & \frac{1}{\left[\int_{\mathcal{Y}} dF_\theta(x)\right]^2}\left[\int_{\mathcal{Y}} dF_\theta(x)\int_{\mathcal{Y}} s(x;\theta)s^T(x;\theta)dF_\theta(x) - \right. \\ & \left. \int_{\mathcal{Y}} s(x;\theta)dF_\theta(x)\int_{\mathcal{Y}} s^T(x;\theta)dF_\theta(x)\right] \end{aligned}$$

The proof is given in Appendix A. The asymptotic covariance matrix of the ML estimator with truncated data is then given by

$$n\,\mathrm{var}(\hat{\theta}) = M^{-1}(s,\theta)Q(s,\theta)M^{-T}(s,\theta)$$

where

$$\begin{aligned} Q(s,\theta) \quad &= \quad \frac{1}{\int_{\mathcal{Y}} dF_\theta(x)}\int_{\mathcal{Y}} s(x;\theta)s^T(x;\theta)dF_\theta(x) - \\ & \quad \frac{1}{\left[\int_{\mathcal{Y}} dF_\theta(x)\right]^2}\int_{\mathcal{Y}} s(x;\theta)dF_\theta(x)\int_{\mathcal{Y}} s^T(x;\theta)dF_\theta(x) \\ &= \quad M(s,\theta) \end{aligned}$$

so that

$$n\,\mathrm{var}(\hat{\theta}) = M^{-1}(s,\theta)$$

which corresponds to the inverse of the Fisher information matrix.

The IF is especially useful for assessing the robustness properties of estimators. The IF of the ML estimator of θ when the data are truncated is given in (4). Dependent on the way data have been truncated and on the form of F_θ, the IF would become arbitrarily large and hence the ML estimator of θ would be biased. Typically with income distributions, F_θ is such that the score function takes large values in the tails of the distribution. The interesting feature is then that if the truncation occurs in both tails, the influence of contamination is limited, whereas if the truncation occurs in the lower tail (as it is the case for example with fiscal data), the IF is unbounded. It is therefore important to develop estimators which are robust to data contamination.

3 Robust estimators

3.1 General definition

We consider here the general class of M-estimators which are defined implicitly by the solution in θ of

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

with ψ being a function satisfying mild conditions (Huber 1964). In general, the ψ function defining the M-estimator is itself a function of the score function, i.e. $\psi(x;\theta) = K[s(x;\theta)]$. For example if

$$\psi(x;\theta) = s(x;\theta)$$

we have the ML estimator, and if

$$\psi(x;\theta) = H_c\left(A[s(x;\theta) - a]\right) \tag{6}$$

where

$$H_c(\mathbf{z}) = \min\left(1; \frac{c}{\|\mathbf{z}\|}\right)$$

are Huber weights depending on a tuning constant c, A is a $p \times p$ matrix and a is a $p \times 1$ vector which are determined implicitly by

$$\begin{aligned} E\left[A[s(x;\theta) - a][s(x;\theta) - a]^T A^T\right] &= I \\ E\left[A[s(x;\theta) - a]\right] &= 0 \end{aligned}$$

we have the OBRE in the standardized case (see Hampel et al. 1986). This estimator is the most efficient estimator in the class of M-estimators with a bounded (function of the) IF. The degree of robustness of the OBRE depends on the choice of c: the lower its value the more robust is the OBRE but it also looses efficiency compared to the ML estimator at the uncontaminated model. Typically c is chosen to achieve a given degree of efficiency. Note that when $c = \infty$, one gets the ML estimator.

With truncated data, there are at least two ways of choosing the ψ function. The first is to use the score function $\widetilde{s}(y;\theta)$ given in (1) and therefore define the M-estimator as

$$\sum_{i=1}^{n} \widetilde{\psi}(y_i;\theta) = \sum_{i=1}^{n} K\left[\widetilde{s}(y_i;\theta)\right] = 0 \tag{7}$$

with $\widetilde{\psi}$ satisfying

$$\int_{\mathcal{Y}} \widetilde{\psi}(x;\theta) dF_\theta(x) \Big/ \int_{\mathcal{Y}} dF_\theta(x) = 0 \tag{8}$$

for consistency. This is the approach based on the marginal distribution (MD approach). When K is such that one gets the OBRE, the algorithm proposed in Victoria-Feser and Ronchetti (1994) could be used.

Another approach would be to replace $s(\cdot;\theta)$ by $\psi(\cdot;\theta)$ in (2) and define the M-estimator as

$$\begin{aligned}\int_{\overline{\mathcal{Y}}} \psi(x;\theta)dF_\theta(x) + \int_{\mathcal{Y}} dF_\theta(x)\frac{1}{n}\sum_{i=1}^{n}\psi(y_i;\theta) &= \\ \int_{\overline{\mathcal{Y}}} K\left[s(x;\theta)\right]dF_\theta(x) + \int_{\mathcal{Y}} dF_\theta(x)\frac{1}{n}\sum_{i=1}^{n}K\left[s(y_i;\theta)\right] &= 0 \qquad (9)\end{aligned}$$

The idea behind this definition is the same as that on which the ML estimator is based: the value of the ψ function on the missing observations is replaced by its mathematical expectation. Note that by taking expectations in (9) one gets

$$\int_{\mathcal{X}} \psi(x;\theta)dF_\theta(x) = 0$$

We call this estimator the *EMM estimator* since it is based on the idea of the EM algorithm and extended to M-estimators. The two approaches are equivalent if and only if $K[s(\cdot;\theta)] = B\cdot s(\cdot;\theta)$, i.e. in the case of the ML estimator. In what follows, we will explore the two approaches in the case when K leads to the OBRE given by (6).

3.2 Properties

We derive here the IF for both estimators with general ψ functions. This will enable us to study the robustness properties of the estimators and derive their asymptotic covariance matrix. For the M-estimator based on the marginal distribution $\hat{\theta}_{MD}$, we have

$$IF(z,\hat{\theta}_{MD},F_\theta) = M^{-1}(\widetilde{\psi},F_\theta)\widetilde{\psi}(z;\theta)$$

where

$$M(\widetilde{\psi},F_\theta) = \int_{\mathcal{Y}} \widetilde{\psi}(x;\theta)s^T(x;\theta)d\frac{F_\theta(x)}{\int_{\mathcal{Y}} dF_\theta(x)}$$

(for details see Appendix B). For the EMM estimator $\hat{\theta}_{EMM}$ defined in (9) we have

$$IF(z,\hat{\theta}_{EMM},F_\theta) = M^{-1}(\psi,F_\theta)\left[\int_{\mathcal{Y}} dF_\theta(x)\psi(z;\theta) + \int_{\overline{\mathcal{Y}}}\psi(x;\theta)dF_\theta(x)\right]$$

where

$$\begin{aligned}M(\psi,F_\theta) &= \int_{\mathcal{Y}} \psi(x;\theta)s^T(x;\theta)dF_\theta(x) - \\ &\quad \int_{\mathcal{Y}} \psi(x;\theta)dF_\theta(x)\int_{\mathcal{Y}} s^T(x;\theta)dF_\theta(x)/\int_{\mathcal{Y}} dF_\theta(x)\end{aligned}$$

(see Appendix C). We note in both cases that if $\widetilde{\psi}$ or ψ is bounded, then the IF are bounded and therefore the resulting estimators are robust to small amounts of contamination.

The covariance matrices are respectively

$$\begin{aligned} n\,\mathrm{var}(\hat{\theta}_{MD}) &= M^{-1}(\widetilde{\psi},F_\theta)Q(\widetilde{\psi},F_\theta)M^{-T}(\widetilde{\psi},F_\theta) \\ Q(\widetilde{\psi},F_\theta) &= \int_{\mathcal{Y}} \widetilde{\psi}(x;\theta)\widetilde{\psi}^T(x;\theta)d\frac{F_\theta(x)}{\int_{\mathcal{Y}} dF_\theta(x)} \end{aligned}$$

and

$$\begin{aligned} n\,\mathrm{var}(\hat{\theta}_{EMM}) &= M^{-1}(\psi,F_\theta)Q(\psi,F_\theta)M^{-T}(\psi,F_\theta) \\ Q(\psi,F_\theta) &= \int_{\mathcal{Y}} dF_\theta(x)\int_{\mathcal{Y}} \psi(x;\theta)\psi^T(x;\theta)dF_\theta(x) - \\ & \quad \int_{\mathcal{Y}} \psi(x;\theta)dF_\theta(x)\int_{\mathcal{Y}} \psi^T(x;\theta)dF_\theta(x) \end{aligned}$$

3.3 OBRE with truncated data

As we have seen, the EMM estimator and the MD approach do not always give the same estimators. Under OBRE the two approaches are different because the statistical model foundation depends on the chosen approach. With MD estimation the underlying model is transformed: we consider the density the density on the incomplete data set rather than on the full data. With the EMM estimator, we keep the initial model (density over the complete data set) by replacing the truncated interval with its corresponding mathematical expectation. However, the empirical results will show that with OBRE, the two approaches give nearly the same estimates at least for small proportions of missing data. The difference between the two approaches is due to the computed weights on the observations which depend on the underlying model.

As we have seen before the EMM estimator is defined as the solution in θ of (9). In order to find the solution of this equation, we propose to use a Newton-Raphson step. The details are given in Appendix D.

4 Numerical comparisons

As an example, we will take the case of data generated by a Gamma distribution with shape parameter α and scale parameter λ:

$$F_{\alpha,\lambda}(x) = \int_0^x \frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}dx$$

where Γ denotes the Gamma function. We suppose without loss of generality that the data are truncated below a known minimum value x_m, an example commonly encountered in the study of income distributions.

4.1 EMM estimator and MD approach with OBRE

It is difficult to compare the two approaches by looking at the equations defining the estimators. However, we can notice that the EMM estimator is consistent on the complete distribution, whereas in the MD approach, the OBRE is consistent on the incomplete distribution. Thus the two solutions will be in general different and the reason for their difference can be seen in the weights (see below). This result is not surprising, because often equivalent classical procedures become different when they are robustified (see e.g. Heritier and Ronchetti 1994 with robust parametric tests). We can also observe that if we choose a bound c large enough the two approaches give the same result. In fact in this case we are using the ML estimator.

Empirically we tried to compute the difference in the weights calculated with the two approaches. We computed the matrix A and vector a when the distribution is a Gamma distribution with parameters $\alpha = 3$ and $\lambda = 1$. We chose $c = 2.0$.

In order to make this difference as clear as possible, we chose two different truncation points. In Figure 1, we can see that with a 10% of information loss in the lower tail, the difference in the computation of the weights by the two approaches is very small. In Figure 2, we made intentionally a high truncation point (30% of information loss) and observed that the difference between the two approaches is larger. In fact the weights associated to the EMM estimator are always the same for any amount of information loss. Thus, the difference is due to the weights calculated with the marginal distribution approach. This confirms the hypothesis we made in the introduction: the underlying model modifies the computation of the weights. The question which then arises is: which approach to choose? In income distribution models we know that the unreported lower incomes do really exist. Thus, it is more realistic to postulate that the true model is the complete one and choose the EMM estimator. What is more, since in certain cases the incomplete information does not only correspond to the case of truncated data, the EMM estimator is largely applicable in comparison to the approach with the marginal distribution. However, as we will see, the differences on the estimates between the two approaches are small at least in the case of a truncation corresponding to a information loss less than 10%.

In order to compare the standardized OBRE in the two approaches, we simulated 50 samples of 200 observations generated by a Gamma distribution with parameter values $\alpha = 3.0$ and $\lambda = 1.0$. In a second step, we contaminated the model with the mixture model $(1-\varepsilon)F_{\alpha,\lambda} + \varepsilon F_{\alpha,0.1\lambda}$ taking

$\varepsilon = 1\%$ and $\varepsilon = 3\%$. We also truncated the data under a minimum value x_m, corresponding to an information loss of between 0.8% and 8%.

Truncation	α, λ	OBRE (EMM) ($c = 2.0$)	MSE	OBRE (MD) ($c = 2,0$)	MSE
$x_m = 1.0$	α	3.12 (0.07)	0.24	3.12 (0.06)	0.22
(8%)	λ	1.03 (0.02)	0.03	1.03 (0.02)	0.02
$x_m = 0.8$	α	3.10 (0.06)	0.21	3.10 (0.06)	0.17
(4.7%)	λ	1.03 (0.02)	0.02	1.03 (0.02)	0.02
$x_m = 0.6$	α	3.09 (0.06)	0.17	3.09 (0.05)	0.14
(2.3%)	λ	1.02 (0.02)	0.02	1.03 (0.02)	0.02
$x_m = 0.4$	α	3.07 (0.05)	0.12	3.06 (0.04)	0.10
(0.8%)	λ	1.02 (0.02)	0.02	1.02 (0.02)	0.01

Table 1: **OBRE on non contaminated data, with the EMM algorithm and the marginal distribution approach (MD)**

As expected, in the non contaminated case, since the weights on the observations are nearly all equal to unity, the two approaches give the same estimators which are equivalent to the simulation values for the parameters (see Table 1). In the case of 1% of contamination (see Table 2), we cannot say that the estimators are different. In the case of 3% of contamination the difference is still small (see Table 3). The reason is because as we have seen in the previous section, the difference between the two approaches becomes clear when the loss of information is relatively large. In the contaminated cases the MSE have relatively the same values as in the non contaminated case, and moreover, they are the same when we compare the estimators between the two approaches.

4.2 EMM and ML estimators

In this section we compare the EMM to the ML estimator.

We took the same sample as before and contaminated 1% of the highest observations of each sample by multiplying them by 10. The results are very surprising. First, as we can see in Table 4, when the samples are not contaminated, the OBRE and the ML estimator give the expected results, with a comparable MSE. On the other hand, when we introduce the contamination (see Table 5) the behaviour of the ML estimator is catastrophic! This result si not surprising since we already know that the ML estimators are not robust to model contamination. The bias are even far worse than those in the complete information case (see Victoria-Feser and Ronchetti 1994). We

Truncation	α, λ	OBRE (EMM) ($c = 2.0$)	MSE	OBRE (MD) ($c = 2.0$)	MSE
$x_m = 1.0$	α	3.07 (0.07)	0.24	3.11 (0.07)	0.24
(8%)	λ	1.02 (0.02)	0.03	1.03 (0.02)	0.03
$x_m = 0.8$	α	3.07 (0.06)	0.21	3.08 (0.07)	0.22
(4.7%)	λ	1.02 (0.02)	0.02	1.03 (0.02)	0.03
$x_m = 0.6$	α	3.06 (0.06)	0.17	3.08 (0.06)	0.18
(2.3%)	λ	1.02 (0.02)	0.02	1.02 (0.02)	0.02
$x_m = 0.4$	α	3.06 (0.05)	0.12	3.05 (0.05)	0.12
(0.8%)	λ	1.02 (0.02)	0.02	1.02 (0.02)	0.02

Table 2: **OBRE on contaminated data at 1%, with the EMM algorithm and the marginal distribution approach (MD)**

Truncation	α, λ	OBRE (EMM) ($c = 2.0$)	MSE	OBRE (MD) ($c = 2.0$)	MSE
$x_m = 1.0$	α	2.59 (0.06)	0.31	2.64 (0.06)	0.30
(8%)	λ	0.85 (0.02)	0.04	0.87 (0.02)	0.04
$x_m = 0.8$	α	2.65 (0.05)	0.26	2.69 (0.05)	0.24
(4.7%)	λ	0.86 (0.02)	0.03	0.88 (0.02)	0.03
$x_m = 0.6$	α	2.71 (0.05)	0.20	2.73 (0.05)	0.19
(2.3%)	λ	0.88 (0.02)	0.03	0.89 (0.02)	0.03
$x_m = 0.4$	α	2.75 (0.04)	0.15	2.75 (0.04)	0.15
(0.8%)	λ	0.90 (0.02)	0.02	0.90 (0.01)	0.02

Table 3: **OBRE on contaminated data at 3%, with the EMM algorithm and the marginal distribution approach (MD)**

Truncation	α, λ	OBRE (EMM) ($c = 2.0$)	MSE	MLE	MSE
$x_m = 1.0$	α	3.12 (0.07)	0.24	3.13 (0.06)	0.18
(8%)	λ	1.03 (0.02)	0.03	1.04 (0.02)	0.02
$x_m = 0.8$	α	3.10 (0.06)	0.21	3.09 (0.05)	0.15
(4.7%)	λ	1.03 (0.02)	0.02	1.03 (0.02)	0.02
$x_m = 0.6$	α	3.09 (0.06)	0.17	3.08 (0.05)	0.13
(2.3%)	λ	1.02 (0.02)	0.02	1.02 (0.02)	0.01
$x_m = 0.4$	α	3.07 (0.05)	0.12	3.04 (0.04)	0.09
(0.8%)	λ	1.02 (0.02)	0.02	1.01 (0.02)	0.01

Table 4: **OBRE and MLE on non contaminated data, with the EMM algorithm**

chose here an extreme type of contamination to show our point, but a less extreme type of contamination (i.e. not choosing systematically the highest observation) would lead to similar conclusions.

Truncation	α, λ	OBRE (EMM) ($c = 2.0$)	MSE	MLE	MSE
$x_m = 1.0$	α	3.07 (0.07)	0.24	0.61 (0.00)	5.71
(8%)	λ	1.02 (0.02)	0.03	0.25 (0.00)	0.57
$x_m = 0.8$	α	3.07 (0.06)	0.21	0.61 (0.00)	5.51
(4.7%)	λ	1.02 (0.02)	0.02	0.23 (0.00)	0.59
$x_m = 0.6$	α	3.06 (0.06)	0.17	0.61 (0.00)	5.71
(2.3%)	λ	1.02 (0.02)	0.02	0.22 (0.00)	0.61
$x_m = 0.4$	α	3.06 (0.05)	0.12	0.65 (0.02)	5.60
(0.8%)	λ	1.02 (0.02)	0.02	0.21 (0.01)	0.62

Table 5: **OBRE and MLE on contaminated data at 1%, with the EMM algorithm**

5 Conclusion

Robust estimators with truncated data are particularly useful in the study of income distributions because of the features of the data one encounters. In this paper, we proposed the EMM estimator which is an extension of

M-estimators to the case of missing data. In the particular case of the ML estimator, the EMM estimator leads to the same estimator as when one considers the marginal distribution of the data. However, with robust estimators, the two approaches differ and their difference is due to the way in which the weights associated to each observations are computed.

We stressed our point by means of simulations study with the example of the Gamma distribution. We also showed that a robust approach is much safer than the ML estimator when the data are contaminated.

References

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

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.

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.

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

A IF of the ML estimator

The ML estimator $\hat{\theta}$ can be seen as a special case of (7) where K is the identity function. Using the results in (10), we have

$$IF(z;\hat{\theta},F_\theta) = M^{-1}(s,F_\theta)\left[s(z;\theta) - \frac{1}{\int_{\mathcal{Y}} dF_\theta(x)}\int_{\mathcal{Y}} s(x;\theta)dF_\theta(x)\right]$$

with

$$\begin{aligned} M(s,F_\theta) \quad = \quad & \frac{1}{\int_{\mathcal{Y}} dF_\theta(x)}\int_{\mathcal{Y}} s(x;\theta)s^T(x;\theta)dF_\theta(x) - \\ & \frac{1}{\left[\int_{\mathcal{Y}} dF_\theta(x)\right]^2}\int_{\mathcal{Y}} s(x;\theta)dF_\theta(x)\int_{\mathcal{Y}} s^T(x;\theta)dF_\theta(x) \end{aligned}$$

B IF of M-estimators with missing data

BVery generally, the IF for M-estimators $\hat{\theta}$ as defined in (5) is given by

$$IF(z;\hat{\theta},F_\theta) = M^{-1}(\psi,F_\theta)\psi(z;\theta)$$

where

$$M(\psi,F_\theta) = -\int \frac{\partial}{\partial\theta^T}\psi(x;\theta)dF_\theta(x)$$

(see Hampel et al. 1986). The IF for M-estimators $\hat{\theta}_{MD}$ as defined in (7) is deduced from this result, i.e.

$$IF(z;\hat{\theta}_{MD},F_\theta) = M^{-1}(\widetilde{\psi},F_\theta)\widetilde{\psi}(z;\theta) \tag{10}$$

where

$$M(\widetilde{\psi},F_\theta) = -\int_{\mathcal{Y}} \frac{\partial}{\partial\theta^T}\psi(x;\theta)dF_\theta(x)/\int_{\mathcal{Y}} dF_\theta(x)$$

From (8) we get

$$-\int_{\mathcal{Y}} \frac{\partial}{\partial\theta^T}\psi(x;\theta)dF_\theta(x) = \int_{\mathcal{Y}} \psi(x;\theta)s^T(x;\theta)dF_\theta(x)$$

and

$$M(\widetilde{\psi},F_\theta) = \int_{\mathcal{Y}} \psi(x;\theta)s^T(x;\theta)dF_\theta(x)/\int_{\mathcal{Y}} dF_\theta(x)$$

C IF of EMM estimators

We also use the general results on the IF for M-estimators. Let

$$\psi^*(y;\theta) = \int_{\overline{\mathcal{Y}}} \psi(x;\theta) dF_\theta(x) + \int_{\mathcal{Y}} dF_\theta(x) \psi(y;\theta)$$

so that the EMM estimator $\hat{\theta}_{EMM}$ is defined implicitly by

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

The IF is then given by

$$IF(z;\hat{\theta}_{EMM}, F_\theta) = M^{-1}(\psi^*, F_\theta)\psi^*(z;\theta)$$

with

$$\begin{aligned}
M(\psi^*, F_\theta) &= \int_{\mathcal{Y}} \psi^*(x;\theta) s^T(x;\theta) dF_\theta(x) / \int_{\mathcal{Y}} dF_\theta(x) \\
&= \int_{\overline{\mathcal{Y}}} \psi(x;\theta) dF_\theta(x) \int_{\mathcal{Y}} s^T(x;\theta) dF_\theta(x) / \int_{\mathcal{Y}} dF_\theta(x) + \\
&\quad \int_{\mathcal{Y}} \psi(x;\theta) s^T(x;\theta) dF_\theta(x) \\
&= \int_{\mathcal{Y}} \psi(x;\theta) s^T(x;\theta) dF_\theta(x) - \\
&\quad \int_{\mathcal{Y}} \psi(x;\theta) dF_\theta(x) \int_{\mathcal{Y}} s^T(x;\theta) dF_\theta(x) / \int_{\mathcal{Y}} dF_\theta(x) \\
&= M(\psi, \theta)
\end{aligned}$$

D Algorithm for EMM estimator

By making the following approximation

$$\frac{\partial}{\partial\theta} \left\{ \int_{\overline{\mathcal{Y}}} \psi(x;\theta) dF_\theta(x) + \int_{\mathcal{Y}} dF_\theta(x) \frac{1}{n} \sum_{i=1}^{n} \psi(y_i;\theta) \right\} \cong - \int_{\mathcal{X}} \psi(x;\theta) s^T(x;\theta) dF_\theta(x) \tag{11}$$

the Newton-Raphson step $\Delta\theta$ is given by

$$\begin{aligned}
\Delta\theta &= \left\{ \int_{\mathcal{X}} \psi(x;\theta) s^T(x;\theta) dF_\theta(x) \right\}^{-1} \left[\int_{\overline{\mathcal{Y}}} \psi(x;\theta) dF_\theta(x) + \right. \\
&\quad \left. \int_{\mathcal{Y}} dF_\theta(x) \frac{1}{n} \sum_{i=1}^{n} \psi(y_i;\theta) \right]
\end{aligned} \tag{12}$$

The complete algorithm to find the EMM estimator based on the standardized optimal OBRE is given by the following four steps:

- *Step 1:* Fix a precision threshold η, an initial value for θ and initial values for the vector a and the matrix A. For example, the ML estimator of θ, $a = 0$ and $A = J^{1/2}(\theta)^{-T}$ where

$$J(\theta) = \int_{\mathcal{X}} s(x;\theta)s(x;\theta)^T dF_\theta(x) \tag{13}$$

- *Step 2:* Solve the following equations for A and a

$$A^T A = M_2^{-1}(\theta) \tag{14}$$

$$a = \frac{\int_{\mathcal{X}} s(x;\theta)W_c(x;\theta)dF_\theta(x)}{\int_{\mathcal{X}} W_c(x;\theta)dF_\theta(x)} \tag{15}$$

where

$$M_k(\theta) = \int_{\mathcal{X}} [s(x;\theta) - a][s(x;\theta) - a]^T W_c^k(x;\theta)dF_\theta(x)$$

and

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

- *Step 3:* Compute $M_1(\theta)$ and

$$\Delta\theta = M_1^{-1}(\theta)A^{-1}\left[\int_{\overline{\mathcal{Y}}} \psi(x;\theta)dF_\theta(x) + \int_{\mathcal{Y}} dF_\theta(x)\frac{1}{n}\sum_{i=1}^{n}\psi(y_i;\theta)\right]$$

- *Step 4:* If $|\Delta\theta| \geq \eta$ then $\theta^{(k+1)} = \theta^{(k)} + \Delta\theta$ and return to step 2, else stop.

The function ψ is defined for the OBRE in the standardized case by equation (6) and the following ones. In step 2, one can use an iterative process in which, given current values for θ, a and A, the right hand sides of (14) and (15) are computed and then the equations are solved for a and A. In our experience however, a one step improvement using current values of θ, a and A is enough. The algorithm is convergent provided the starting point is near the solution. We propose here the ML estimator, and to insure convergence, one can first compute an OBRE with a high value for the bound c and then use the estimate as starting point for another more robust (lower value for c) estimator.

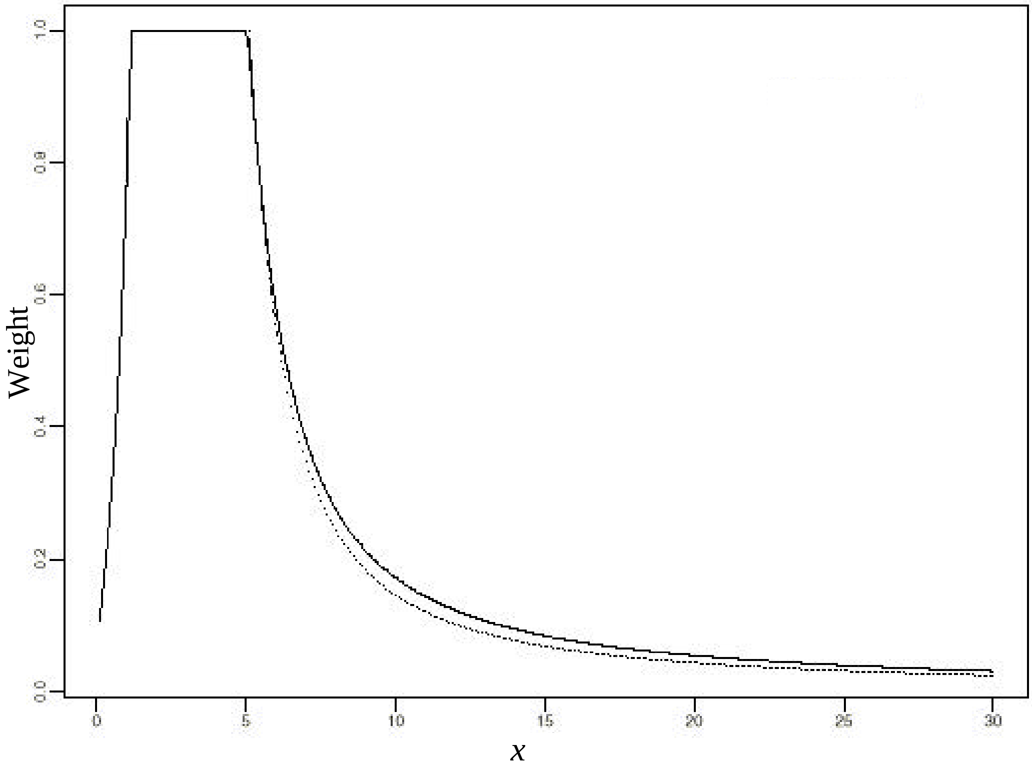

Figure 1: Weights of the OBRE ($c = 2.0$) with the EMM estimator (solid line) and the MD approach (dotted line) with 10% of truncated data

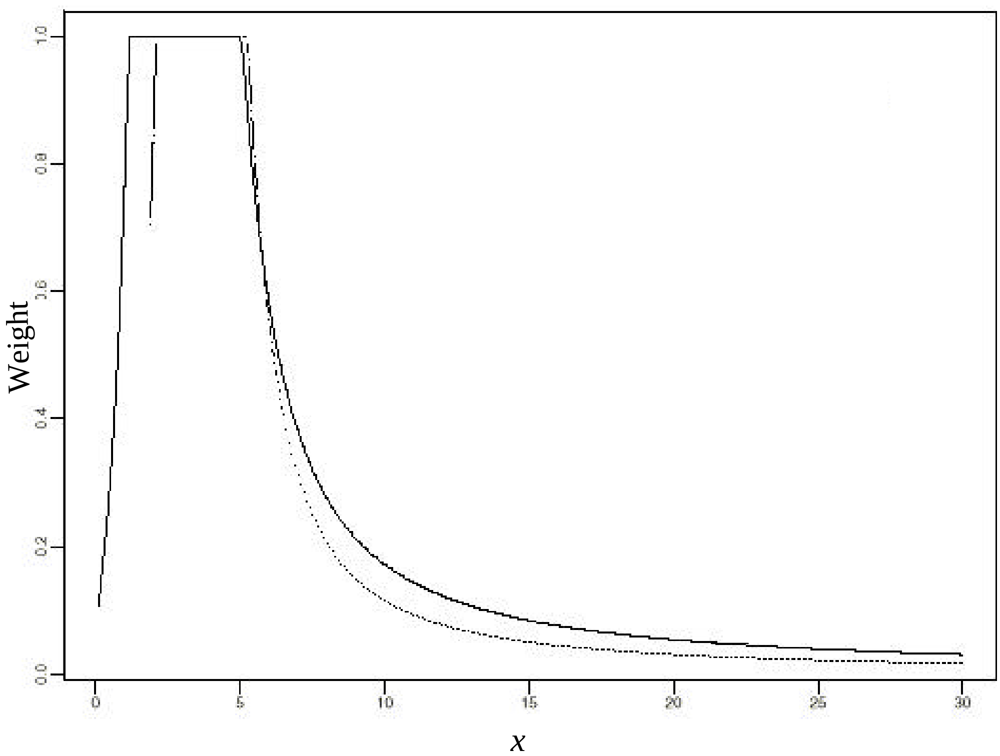

Figure 2: Weights of the OBRE ($c = 2.0$) with the EMM estimator (solid line) and the MD appraoch (dotted line) with 30% of truncated data