Robust Estimating Equations in the Presence of a Nuisance Parameter

L. Greco, L. Ventura

2003.8

Dipartimento di Scienze Statistiche
Università degli Studi
Via C. Battisti 241-243
35121 Padova

Maggio 2003

Robust Estimating Equations in the Presence of a Nuisance Parameter

Luca Greco and Laura Ventura
Department of Statistics, University of Padova
Via C. Battisti 241, 35121 Padova, Italy

16th May 2003

1 Introduction

Consider a sample $y = (y_1, \ldots, y_n)$ of n independent observations with distribution function $F_\theta = F(y; \theta)$ depending on an unknown parameter $\theta \in \Theta \subseteq \mathbb{R}^p$, $p > 1$. Suppose that θ is partitioned as $\theta = (\tau, \lambda)$, i.e. into a scalar parameter of interest τ and a $(p-1)$–dimensional nuisance parameter λ. Consider inference about τ in the presence of the nuisance parameter λ, based on the observation of y. For models with a nuisance parameter, classical inference is often based on a pseudo-likelihood function, i.e. a function of the data and τ having properties similar to those of a likelihood function when there is no nuisance parameter. The most commonly used pseudo-loglikelihood function is the profile loglikelihood function, given by

$$\ell_p(\tau) = \ell(\tau, \hat{\lambda}_\tau) = \sum_{i=1}^{n} \ell(\tau, \hat{\lambda}_\tau; y_i) \ , \tag{1}$$

where $\ell(\theta) = \ell(\tau, \lambda)$ denotes the usual loglikelihood function for θ and $\hat{\lambda}_\tau$ is the maximum likelihood estimate (MLE) of λ for fixed τ. Other commonly used pseudo-likelihoods include the adjusted profile, the conditional and the marginal likelihood functions (see e.g. Pace and Salvan, 1997, and Severini, 2000).

In many situations of practical interest, there is no certainty that the data y come from the specified model F_θ and may in fact come from some neighborhood of the model. It is well-known that standard likelihood procedures are not robust and the need for robust statistical procedures for estimation and testing has been stressed by many authors in the statistical literature; cfr. for instance, Huber (1981), Hampel et al. (1986) and Markatou and Ronchetti (1997). However, while robust literature offers many solutions for inference on the whole parameter θ of the model, the situation with a nuisance parameter has been somewhat neglected. In this paper, we discuss the general problem of robust inference about τ, in the presence of the nuisance parameter λ. In particular, it is of interest to construct a robust estimating equation for the scalar parameter of interest in the presence of the nuisance parameter λ. The approach followed in this paper is esentially based on the results of Yuan and Jennrich (2000), which discuss a general approach to estimating equations with nuisance parameters, mixed with a basic truncation argument of the theory of robust statistics to robustify estimating equations (see Hampel et al, 1986, chap. 2, and Carroll and Ruppert, 1988, chap. 6).

The paper is organized as follows. Section 2 presents the profile loglikelihood and to the generalised profile loglikelihood, obtained by substituting the nuisance parameter with a suitable estimate of λ other than the MLE of λ for fixed τ. The latter function is the starting point for the proposed robust estimating equations. The approach to obtain robust estimating equations is developed in Section 3. Finally, Section 4 discusses some final remarks.

2 Profile and generalised profile likelihoods

Consider the profile loglikelihood function (1). This function is obtained from $\ell(\tau, \lambda)$ by replacing λ with its MLE for fixed τ, i.e. with $\hat{\lambda}_\tau$. The MLE of λ for fixed τ is given as the solution $\hat{\lambda}_\tau$ of the partial loglikelihood equation

$$\ell_\lambda(\tau, \hat{\lambda}_\tau) = \sum_{i=1}^{n} \ell_\lambda(\tau, \hat{\lambda}_\tau; y_i) = 0 \ , \tag{2}$$

where $\ell_\lambda = \partial \ell(\tau, \lambda)/\partial\lambda$. To set up notation, in the following, we will denote first partial derivatives of $\ell(\tau, \lambda)$ by $\ell_\lambda = \partial \ell(\tau, \lambda)/\partial\lambda$ and $\ell_\tau = \partial \ell(\tau, \lambda)/\partial\tau$ and, with an obvious notation, second partial derivatives by $\ell_{\lambda\lambda}$, $\ell_{\lambda\tau}$ and $\ell_{\tau\tau}$, and so on.

The profile loglikelihood $\ell_p(\tau)$ is not a genuine likelihood function. In particular, the profile score function $\ell_p'(\tau) = \partial \ell_p(\tau)/\partial\tau$ has not zero mean, but its expectation is of order $O(1)$. However, it has many properties as a genuine loglikelihood for τ (see e.g. Pace and Salvan, 1997, and Severini, 2000). First of all, the profile MLE of τ, $\hat{\tau}$, given as the solution of the profile score function

$$\ell_p'(\tau) = \sum_{i=1}^{n} \ell_\tau(\tau, \hat{\lambda}_\tau; y_i) = 0 \ , \tag{3}$$

coincides to the MLE for τ. Moreover, the profile loglikelihood ratio statistic for testing $\tau = \tau_0$ versus $\tau \neq \tau_0$ may be written as $W_p(\tau_0) = 2(\ell_p(\hat{\tau}) - \ell_p(\tau_0))$ and has a standard χ^2_1 distribution. Similarly, the directed likelihood $r_p(\tau_0) = \text{sgn}(\hat{\tau} - \tau_0)\sqrt{W_p(\tau_0)}$ has standard normal distribution. In view of this, for setting confidence regions or for testing hypothesis, $W_p(\tau)$ or $r_p(\tau)$ may be used in a standard way.

Let $\tilde{\lambda}_\tau$ denote an alternative estimate of λ for fixed τ. Following Severini (1998), a generalised profile loglikelihood $\tilde{\ell}_p(\tau) = \ell(\tau, \tilde{\lambda}_\tau)$ can be used for inference about τ, provided that some week conditions are satisfied. In particular, it is assumed that

$$\tilde{\lambda}_{\tau_0} = \lambda_{\tau_0} + O_p(n^{-1/2}) \ , \tag{4}$$

where λ_{τ_0} denotes the value of λ that maximizes $E_0\{\ell(\tau, \lambda)\}$ holding τ fixed, and that

$$\tilde{\lambda}'_{\tau_0} = \hat{\lambda}'_{\tau_0} + O_p(n^{-1/2}) \ , \tag{5}$$

where $\tilde{\lambda}'_\tau$ and $\hat{\lambda}'_\tau$ denote, respectively, the derivatives of $\tilde{\lambda}_\tau$ and $\hat{\lambda}_\tau$ with respect to τ evaluated in τ_0, i.e. $\tilde{\lambda}'_{\tau_0} = \partial\tilde{\lambda}_\tau/\partial\tau|_{\tau=\tau_0}$ and $\hat{\lambda}'_{\tau_0} = \partial\hat{\lambda}_\tau/\partial\tau|_{\tau=\tau_0}$. Condition (5) can be expressed as $\tilde{\lambda}'_{\tau_0} = \lambda'_{\tau_0} + O_p(n^{-1/2})$, where $\lambda'_{\tau_0} = \partial\lambda_\tau/\partial\tau|_{\tau=\tau_0}$. Quantity λ'_{τ_0} may be expressed in terms of the expected information matrix $i(\theta_0) = i(\tau_0, \lambda_0)$ by noting that λ_τ must satisfy $E_{\theta_0}[\ell_\lambda(\tau, \lambda_\tau)] = 0$, for all τ, and hence $E_{\theta_0}[\ell_{\tau\lambda}(\theta_0) + \ell_{\lambda\lambda}(\theta_0)\lambda'_{\tau_0}] = 0$, so that

$$\lambda'_{\tau_0} = -i_{\lambda\lambda}(\theta_0)^{-1} i_{\lambda\tau}(\theta_0) \ .$$

If conditions (4) and (5) are satisfied, it follows that $\tilde{\ell}_p(\tau)$ has the same first-order properties as $\ell_p(\tau)$. Following Severini and Wong (1992), in this case $\tilde{\ell}_p(\tau)$ is a first-order generalised profile loglikelihood function and $\tilde{\lambda}_\tau$ is a first-order estimate of λ for fixed τ.

Another approach, which will be investigated in this paper, is to use estimates of λ of the form $\tilde{\lambda}$, i.e. which do not depend on τ. Gong and Samaniego (1981) discuss this general procedure and consider the first order asymptotic theory of the generalised MLE (GMLE) for τ based on $\ell(\tau, \tilde{\lambda})$; see also Pierce (1982), Parke (1986), Yuan and Jennrich (2000)

2.1 Estimation of the nuisance parameter

The added flexibility in the choice of $\tilde{\lambda}_\tau$ or $\tilde{\lambda}$ allows for the possibility of constructing a generalised profile likelihood function with properties similar to those of $\ell_p(\tau)$.

We first briefly consider the problem of choosing estimates $\tilde{\lambda}_\tau$ that can be used to construct a generalised profile loglikelihood function $\tilde{\ell}_p(\tau)$. One approach is to consider approximations to $\hat{\lambda}_\tau$ along the lines suggested by Cox and Wermuth (1990), when the exact computation of $\hat{\lambda}_\tau$ is difficult. Let $\hat{\tau}$, $\hat{\lambda}$ denote the MLEs of τ, λ. Then an approximation to $\hat{\lambda}_\tau$ is given by

$$\tilde{\lambda}_\tau = \hat{\lambda} - \ell_{\lambda\lambda}(\hat{\tau}, \hat{\lambda})^{-1} \ell_{\tau\lambda}(\hat{\tau}, \hat{\lambda})(\tau - \hat{\tau}) \ , \tag{6}$$

which is a second-order estimate. A closely related second-order estimate is given by

$$\tilde{\lambda}_\tau = \bar{\lambda} + \ell_{\lambda\lambda}(\bar{\tau}, \bar{\lambda})^{-1} \ell_\lambda(\tau, \bar{\lambda}) \ , \tag{7}$$

which only requires that $\bar{\tau}$, $\bar{\lambda}$ be $\sqrt{n}$-consistent. If τ and λ are orthogonal, then $\lambda'_{\tau_0} = 0$. It follows that any $\sqrt{n}$-consistent estimate of λ is a first-order estimate and any estimate $\tilde{\lambda}$ satisfying $\tilde{\lambda} = \hat{\lambda} + O_p(n^{-1})$ is a second-order estimate; this was noted by Cox and Reid (1987). Estimates (6) and (7) can be viewed as approximations to $\hat{\lambda}_\tau$. When robustness is required, $\tilde{\lambda}_\tau$ could be computed from (7) with $(\bar{\tau}, \bar{\lambda})$ suitable robust estimators. One possibility could be the OBRE for (τ, λ), which represnts the solution of the problem of finding the most efficient estimator under the condition of B-robustness. However, this choice does not assure that the etimator $\tilde{\lambda}_\tau$ is B-robust.

When inference about τ is based on a generalized profile likelihood of the form $\tilde{\ell}_p(\tau) = \ell(\tau, \tilde{\lambda}_\tau)$, the estimator for τ is given as the solution of the estimating equation

$$\begin{aligned} \tilde{\ell}'_p(\tau) &= \frac{\partial}{\partial\tau}\ell(\tau, \tilde{\lambda}_\tau) = \ell_\tau(\tau, \tilde{\lambda}_\tau) + \ell_\lambda(\tau, \tilde{\lambda}_\tau)\frac{\partial\tilde{\lambda}_\tau}{\partial\tau} = \\ &= \sum_{i=1}^n \ell_\tau(\tau, \tilde{\lambda}_\tau; y_i) + \frac{\partial\tilde{\lambda}_\tau}{\partial\tau}\sum_{i=1}^n \ell_\lambda(\tau, \tilde{\lambda}_\tau; y_i) = 0 \ , \end{aligned} \tag{8}$$

which does not coincide with the profile loglikelihood equation (3) for τ, since the partial derivative $\ell_\lambda(\tau, \lambda)$ evaluated in $\tilde{\lambda}_\tau$ does not vanish. The GMLE $\tilde{\tau}$, solution of (8), is not B-robust since its influence function still depends on the partial derivatives of the likelihood that are tipically

Distribution	MLE	OBRE	GMLE
$Ga(2,1)$	2.06 (0.279)	2.03 (0.297)	2.07 (0.283)
$0.95Ga(2,1) + 0.05Ga(2,5)$	1.55 (0.242)	1.83 (0.265)	1.97 (0.458)

Table 1: Simulation results for MLE, OBRE, GMLE of the shape parameter for the gamma distribution. The parameter of interest is the shape parameter.

unbounded. However, it could be argued that $\tilde{\tau}$ exhibits a more stable performance in terms of bias (not in terms of variability) in a neighborhood of the assumed model than the MLE $\hat{\tau}$. To investigate this performance a Monte Carlo study has been performed. In particular, suppose it is of interest to make inference on the shape parameter τ of a gamma distribution $Ga(\tau, \lambda)$ in the presence of the nuisance scale parameter λ. Table 1 gives the results of a Monte Carlo experiment (10000 trials) comparing the bias of $\hat{\tau}$ (MLE),$\tilde{\tau}$ (OBRE) and $\tilde{\tau}$ (GMLE) with their standard errors, for a sample size $n = 100$.

The main problem in using an estimate of the form $\tilde{\lambda}_\tau$, to eliminate the nuisance parameter from the likelihood equation for τ, is that its influence function is not in general bounded. In fact, it can be easily shown that the influence function of $\tilde{\lambda}_\tau$ depends on the partial loglikelihood derivative $\ell_\tau(\tau, \lambda; y)$, which is not in general bounded. On the contrary, if attention is restricted to an estimate of the form $\tilde{\lambda}$, i.e. which does not depend on τ, then it can be shown that its influence function is bounded if $\tilde{\tau}$ is choosen in the class of B-robust estimators.

In view of this, when robustness is required, we consider estimators $\tilde{\lambda}$ belonging the general class of M-estimators with bounded influence functions (cfr. Hampel et al., 1986) in order to introduce B-robustness. M-estimators are a generalization of the MLE. In general, an M-estimator for θ is defined as the root $\hat{\theta}_M$ of the unbiased estimating equation

$$\Psi_\theta(\theta) = \sum_{i=1}^{n} \psi(y_i; \theta) = 0 \ , \tag{9}$$

for some function $\psi(\cdot) : \mathcal{Y} \times \mathbb{R}^p \to \mathbb{R}^p$. Under broad conditions, which we will assume throughout this paper (cfr. Hampel et al., 1986), it can be shown that $\hat{\theta}_M$ is consistent and asymptotically normal, with mean θ and variance $V(\theta) = M(\theta)^{-1}\Omega(\theta)(M(\theta)^{-1})^{\mathsf{T}}$, where $M(\theta) = -\int \partial\psi(y;\theta)/\partial\theta^{\mathsf{T}} dF_\theta$ and $\Omega(\theta) = \int \psi(y;\theta)\psi(y;\theta)^{\mathsf{T}} dF_\theta$. Large-sample tests and confidence regions for θ can be constructed in a standard way using an estimate of the asymptotic covariance matrix $V(\theta)$. Since the influence function of $\hat{\theta}_M$ at a point x is given by

$$\mathrm{IF}_{\hat{\theta}_M}(x) = M(\theta)^{-1}\psi(x;\theta) \ ,$$

the M-estimator is B-robust at the assumed model F_θ if and only if $\psi(x;\theta)$ is bounded. The influence function describes the effect of a small contamination at the point x on the estimate. By bounding the influence function, we are able to ensure that small deviations from the model distribution do not cause large changes in the estimates.

Let us consider a generalised profile loglikelihood $\tilde{\ell}_p(\tau) = \ell(\tau, \tilde{\lambda})$, where $\tilde{\lambda}$ is an appropriate estimate of λ, obtained as the solution of an estimating equation of the form

$$\Psi_\lambda(\tilde{\lambda}) = \sum_{i=1}^{n} \psi_\lambda(y_i; \tilde{\lambda}) = 0 \ . \tag{10}$$

An estimating equation of the form (10) is still the estimating equation of a M-estimator of λ. In view of this, the estimator $\tilde{\lambda}$ can be viewed as a functional of the empirical distribution function and its influence function can be computed in a usual way from (10). In particular, it can be shown that for the estimator solution of (10), the influence function is proportional to $\psi_\lambda(x;\lambda)$. In this case, if $\tilde{\lambda}$ is chosen in the class of robust M-estimators, then its influence function is bounded. Moreover, $\tilde{\lambda}$ is consistent and asymptotically normal, with mean λ and variance $V_{\lambda\lambda} = M_{\lambda\lambda}^{-1}\Omega_{\lambda\lambda}(M_{\lambda\lambda}^{-1})^\top$, where $M_{\lambda\lambda} = -\int \partial\psi_\lambda(y;\lambda)/\partial\lambda^\top dF_\theta$ and $\Omega_{\lambda\lambda} = \int \psi_\lambda(y;\lambda)\psi_\lambda(y;\lambda)^\top dF_\theta$.

Let us focus now on the interest parameter τ, when the estimator $\tilde{\lambda}$ is used to eliminate the nuisance parameter. When inference about τ is based on a generalized profile likelihood of the form $\tilde{\ell}_p(\tau) = \ell(\tau, \tilde{\lambda})$, the GMLE for τ is given as the solution of the estimating equation

$$\tilde{\ell}'_p(\tau) = \frac{\partial}{\partial\tau}\ell(\tau,\tilde{\lambda}) = \ell_\tau(\tau,\tilde{\lambda}) = \sum_{i=1}^{n} \ell_\tau(\tau,\tilde{\lambda};y_i) \ . \quad (11)$$

The influence function of $\tilde{\tau}$, solution of (11) can be computed in a standard way. In particular, it can be easily shown that the influence function of $\tilde{\tau}$ is given by

$$IF_{\tilde{\tau}}(x) = M_{\tau\tau}^{-1}\ell_\tau(\tau,\lambda;x) - M_{\tau\tau}^{-1}M_{\tau\lambda}M_{\lambda\lambda}^{-1}\psi_\lambda(x;\lambda) \ , \quad (12)$$

where $M_{\tau\tau} = -\int \partial\ell_\tau(\theta;y)/\partial\tau dF_\theta$ and $M_{\tau\lambda} = -\int \partial\ell_\tau(\theta;y)/\partial\lambda^\top dF_\theta$. Since the influence function depends on $\ell_\tau(\tau,\lambda;x)$, in general the estimator for τ, given as the solution of (11), is not robust against outliers and influential observations. In view of this, when robustness is required, the estimating equation (11) must be modified. In the next section, a procedure to obtain a robust estimator for the interest parameter τ will be discussed.

3 A robust estimator for the interest parameter

A general approach to find robust estimates is based on the idea of starting with a parametric model and modifying the usual loglikelihood equations to achieve robust estimates. This idea has been used successfully by many authors (see Hampel at al., 1986). The appropriate downweighting of the maximum likelihood score function is the technique used to obtain robust estimates with high efficiency, i.e. the optimal estimators of Hampel et al. (1986). However, the downweighting can be carried out also on the probability scale for the model under consideration (see Field and Smith, 1994).

The aim of this section is to robustify the estimating equation (11) by applying a basic truncation argument of the theory of robust statistics (see e.g. Carroll and Ruppert, 1988, chap. 6). This seems a natural way to construct a robust profile estimating equation, since (11) defines a loglikelihood equation.

Consider the general situation of a parametric family F_θ depending on a scalar parameter θ. In order to bound the score function ℓ_θ for θ it is necessary to downweight observations for extreme values of y. This robustification is achieved by a truncation of the score function $\ell_\theta(\theta;y)$. Formally, the optimal robust estimator of Hampel for θ (see Hampel et al. 1986, chap. 2) is the solution of the estimating equation

$$\sum_{i=1}^{n}\psi(y_i;\theta) = \sum_{i=1}^{n}[\ell_\theta(\theta;y_i) - a]_{-b}^{b} = 0 \ , \quad (13)$$

where a is a such that $\int \psi(y;\theta)dF_\theta = 0$, and $b > 0$ is some constant related to the bound on the influence function. The notation $[h(x)]_{-b}^{b}$ for an arbitrary function $h(x)$ means truncation at levels b and $-b$. To solve (13) in general a numerical algorithm must be used. The basic idea is to start with the score function and to modify it so that the influence function is bounded and that it is Fisher-consistent.

Consider now the application on the above procedure to obtain a robust estimate to the situation with a nuisance parameter. The idea developed here consists in downeighting the estimating equation (11) using a technique similar to that of Hampel et al. (1986). This seems the natural way to find a robust estimate for τ since equation (11) represents a generalized profile loglikelihood equation. The proposed robust estimator for τ, that is $\tilde{\tau}_r$ (RGMLE), is the solution of the estimating equation

$$\tilde{\Psi}(\tau) = \sum_{i=1}^{n} \tilde{\psi}(\tau; y_i) = \sum_{i=1}^{n} \left[\ell_\tau(\tau, \tilde{\lambda}; y_i) - a \right]_{-b}^{b} = 0 \ , \tag{14}$$

where, similarly to (13), a is such that $\int \tilde{\psi}(\tau;y)dF_\theta = 0$, and $b > 0$ is a constant related to the bound on the influence function (12). The notation $[h(x)]_{-b}^{b}$ for an arbitrary function $h(x)$ means truncation at levels b and $-b$.

Some properties of the estimate $\tilde{\tau}_r$ are easy to evaluate. It is clear that the estimate is Fisher consistent from the inclusion of a. Moreover, for the presence of b the estimator is B-robust. To study the asymptotic variance and normality of the estimator, the most useful references are Pierce (1982), Clarke (1983, 1986), Parke (1986), Yuan and Jennrich (2000). In particular, it can be shown that the asymptotic variance of $\tilde{\tau}_r$ is (see the Appendix for more details)

$$V_{\tilde{\tau}_r} = M_{\tau\tau}^{-1} \Omega_{\tau\tau} (M_{\lambda\lambda}^{-1})^{\mathsf{T}} + M^{\tau\lambda} \Omega_{\lambda\tau} (M_{\tau\tau}^{-1})^{\mathsf{T}} \ , \tag{15}$$

where $\Omega_{\tau\tau} = \int \tilde{\psi}_\tau(\tau;y)^2 dF_\theta$, $\Omega_{\lambda\tau} = \int \psi_\lambda(y;\lambda)\tilde{\psi}_\tau(\tau;y)dF_\theta$, and $M^{\tau\lambda}$ is defined in the Appendix. The first term on the right hand side of (15) is the asymptotic variance corresponding to τ if λ were known, whereas the second term reflects the cost of estimating the nuisance parameter.

An example Consider a regression-scale and shape model with Exponential Power (EP) distributed errors. The interest parameter is the shape parameter κ, while $\lambda = (\beta_0, \beta_1, \sigma)$ are treated as nuisance parameters. Robust estimates for the nuisance parameter are obatined by applying the Huber Proposal 2 for regression-scale. To compare the performance of the proposed RGMLE for κ in terms of efficiency when the model is correctly specified, but also under small arbitrary departures from the assumed model, we simulate data from two different scenarios: (1) from an EP model with the parameter choice $(\beta_0, \beta_1, \sigma, \kappa) = (13, -1, 1, 1.5)$ and (2) from a contamination model of the form $F_\epsilon = (1-\epsilon)\text{EP} + \epsilon\ \text{EP2}$, where EP2 is an EP model with greater scale parameter. The results of the Monte Carlo study (based on 10000 trials) for MLE, the optimal B-robust estimator (OBRE) and RGMLE for κ are given in Table 2 for $n = 100$. Boxplots of MLE and RGMLE are showed in Figure 1 and Figure 2. Some features for the RGMLE emerge. First of all, the RGMLE exhibits its robustness properties under the contaminated model and appears more robust than the OBRE. Under the central model, of course the MLE performs better than robust estimators, but the RGMLE seems preferable to the OBRE. Moreover, the RGMLE appears right-skewed distributed with a range of variability larger then the MLE in both scenarios.

Distribution	MLE	RGMLE	OBRE
true model	1.49 (0.221)	1.63 (0.250)	1.75 (0.403)
contaminated model	1.06 (0.219)	1.54 (0.322)	1.65 (0.363)

Table 2: Simulation results for MLE, RGMLE and OBRE of the shape parameter for the EP distribution.

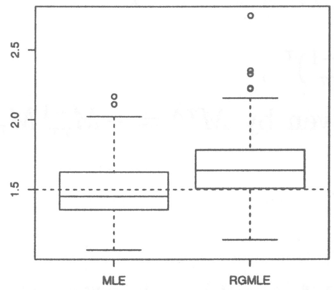

Figure 1: true model

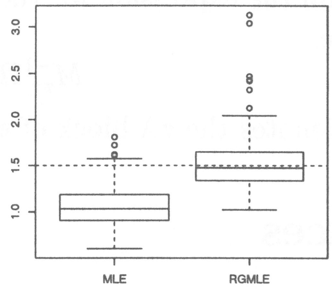

Figure 2: contaminated model

4 Final Remarks

The aim of this paper is to present a general procedure to obtain a robust estimator for a scalar parameter of interest in the presence of nuisance parameters. The proposed estimator is obtained by applying a basic truncation argument of the theory of robust statistics to the estimating equation for the interest parameter derived from a generalised profile likelihood function. The method discussed in this paper is still under investigation. In particular, it seems important to :
(1) better investigate the behaviour of the RGMLE under the true model and under model contamination by varying the tuning costant b and the amount of contamination;
(2) compare the RGMLE to the OBRE;
(3) obtain a robust quasi profile likelihood (Adimari and Ventura, 2002), starting from (14);
(4) define a quasi-likelihood ratio statistic for setting confidence regions or for testing hypothesis, with the classical asymptotic distribution.

5 Appendix

The RGMLE $\tilde{\tau}_r$ is obtained by solving for the interest parameter τ the following system of equations:

$$\Psi(y;\tau,\lambda) = \begin{pmatrix} \sum_{i=1}^n \psi_\tau(y_i;\tau,\lambda) = 0 \\ \sum_{i=1}^n \psi_\lambda(y_i;\lambda) = 0 \end{pmatrix} .$$

By applying standard robust arguments, it is possible to shown that the corresponding influence function is $M(\tau,\lambda)^{-1}\Psi(y;\tau,\lambda)$, where

$$M(\tau,\lambda) = \begin{pmatrix} M_{\tau\tau} \; M_{\tau\lambda} \\ M_{\lambda\tau} \; M_{\lambda\lambda} \end{pmatrix} ,$$

with $M_{\tau\tau} = -\int \partial\psi_\tau/\partial\tau dF_\theta$, $M_{\tau\lambda} = -\int \partial\psi_\tau/\partial\lambda^{\mathsf{T}} dF_\theta$, and so on.

The asymptotic variance matrix of $(\tilde{\tau}_r, \tilde{\lambda})$ is

$$V(\tau,\lambda) = M(\tau,\lambda)^{-1}\Omega(\tau,\lambda)(M(\tau,\lambda)^{\mathsf{T}})^{-1} ,$$

where

$$\Omega(\tau,\lambda) = \begin{pmatrix} \Omega_{\tau\tau} & \Omega_{\tau\lambda} \\ \Omega_{\lambda\tau} & \Omega_{\lambda\lambda} \end{pmatrix} ,$$

with $\Omega_{\tau,\tau} = \int \psi_\tau(y;\tau,\lambda)^2 dF_\theta$, $\Omega_{\tau,\lambda} = \int \psi_\tau(y;\tau,\lambda)\psi_\lambda(y;\lambda)^{\mathsf{T}} dF_\theta$, and so on. It is then straightforward to obtain the expression for the asymptotic variance of $\tilde{\tau}_r$, that corresponds to the $\tau\tau$ block of $V(\tau,\lambda)$, i.e.

$$M_{\tau\tau}^{-1}\Omega_{\tau\tau}(M_{\lambda\lambda}^{-1})^{\mathsf{T}} + M^{\tau\lambda}\Omega_{\lambda\tau}(M_{\tau\tau}^{-1})^{\mathsf{T}} ,$$

where $M^{\tau\lambda}$ denotes the $\tau\lambda$ block of the inverse of $M(\tau,\lambda)$, given by $M^{\tau\lambda} = -M_{\tau\tau}^{-1}M_{\tau\lambda}M_{\lambda\lambda}^{-1}$.

References

Adimari, G., Ventura, L. (2002), Quasi-profile loglikelihood for unbiased estimating functions, *Ann. Inst. Statist. Math.*, **54**, 235–244.

Carroll, R.J., Ruppert, D. (1988), *Transformation and Weighting in Regression*, Chapman and Hall, London.

Clarke, B.R. (1983), Uniqueness and Frechet differentiability of functional solutions to maximum likelihood type equations, *Ann. Statist.*, **11**, 1196-1205.

Clarke, B.R. (1986), Nonsmooth analysis and Frechet differentiability of M-functionals, *Probab. Th. Rel. Fields*, **73**, 197–209.

Cox, D.R., Reid, N. (1987), Parameter orthogonality and approximate conditional inference (with Discussion), *J. R. Statist. Soc.* B, **49**, 1–39.

Cox, D.R., Wermuth, N. (1990), An approximation to maximum likelihood estimates in reduced models, *Biometrika*, **77**, 747–761.

Dupuis, D.J., Field, C.A. (1998), Robust estimation of extremes, *Can. J. Statist.*, **26**, 199–215.

Field, C., Smith, B. (1994), Robust estimation - A weighted maximum likelihood approach, *Internat. Statist. Rev.*, **62**, 405–424.

Gong, G., Samaniego, F.J. (1981), Pseudo maximum likelihood estimation: theory and applications, *Ann. Statist.*, **9**, 861–869.

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

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

Markatou, M., Ronchetti, E. (1997), Robust inference: The approach based on influence functions, In: *Handbook of Statistics*, Maddala, G.S., Rao, C.R. (Eds.), Vol. 15, North-Holland, Amsterdam, 49–75.

Pace, L., Salvan, A. (1997), *Principles of Statistical Inference*, World Scientific, Singapore.

Parke, W.R. (1986), Pseudo maximum likelihood estimation: the asymptotic distribution, *Ann. Statist.*, **14**, 355 – 357.

Pierce, D.A. (1982), The asymptotic effect of substituting estimators for parameters in certain types of statistics, *The Annals of Statistics*, **10**, 475–478.

Severini, T.A.. (1998), Likelihood functions for inference in the presence of a nuisance parameter, *Biometrika*, **85**, 507–522.

Severini, T.A.. (2000), *Likelihood Methods in Statistics*, Oxford University Press, New York.

Severini, T.A., Wong, W.H. (1992), Profile likelihood and conditionally parametric models, *Ann. Statist.*, **20**, 1768–1802.

Victoria-Feser, M.P., Ronchetti, E. (1994), Robust Methods for Personal-income Distribution Models, *Can. Journ. Statist.*, **22**, 247–258.

Yuan, K.H., Jennrich, R.I. (2000), Estimating equations with nuisance parameters: theory and applications, *Ann. Inst. Statist. Math.*, **52**, 343 – 350.