Projection Estimators for Generalized Linear Models

Andrea Bergesio and Victor J. Yohai

Andrea Bergesio is Teaching Assistant, Universidad Nacional del Litoral, IMAL, Güemes 3450, S3000GLN Santa Fe, Argentina (E–mail: anbergesio@gmail.com) and Víctor J. Yohai is Emeritus Professor, Departamento de Matemáticas, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón 1, 1428 Buenos Aires, Argentina (E–mail: vyohai@dm.uba.ar).
This work was supported by Grant X-018 from the Universidad de Buenos Aires, Grant PIP 5505 from CONICET, Argentina and Grant PICT 00899 from ANPCYT, Argentina. The authors thank Alejandro de Acosta for useful suggestions to prove Theorem 5 and to Pavel Čížek for sending us the programe to compute the AMSTLE. We also thank to the Associate Editor and two referees for their comments and suggestions which resulted in several important improvements on the first version of this paper.

ABSTRACT

We introduce a new class of robust estimators for generalized linear models which is an extension of the class of projection estimators for linear regression. These projection estimators are defined using an initial robust estimator for a generalized linear model with only one unknown parameter. We found a bound for the maximum asymptotic bias of the projection estimator caused by a fraction $\varepsilon$ of outlier contamination. For small $\varepsilon$, this bias is approximately twice the maximum bias of the initial estimator independently of the number of regressors. Since these projection estimators are not asymptotically normal, we define one-step weighted M-estimators starting at the projection estimators. These estimators have the same asymptotic normal distribution as the M-estimator and a degree of robustness close to the one of the projection estimator. We perform a Monte Carlo simulation for the case of binomial and Poisson regression with canonical links. This study shows that the proposed estimators compare favorably with respect to other robust estimators. Supplemental Material containing the proofs and the numerical algorithm used to compute the P-estimator is available on line.

KEY WORDS: logistic regression, robust estimators, maximum bias, one-step estimators.

# 1. INTRODUCTION

Let us consider a generalized linear model (GLM) where we observe a response $y \in R$ and a vector $\mathbf{x}= (x_1, \ldots, x_p)'$ of explanatory variables. It is assumed that

$$y|\mathbf{x} \sim F_{\theta}, \tag{1}$$

where $F_\theta$ is a discrete or continuous exponential family of distributions in $R$, $\theta \in \Theta \subset R$ with densities of the form

$$f(y, \theta) = \exp[m(\theta)y - J(\theta) - t(y)]I_D(y), \tag{2}$$

where $D$ is the support of $y$, $I_A$ denotes the indicator function of $A$ and

$$g(\theta) = \beta_0'\mathbf{x}, \tag{3}$$

where $\beta_0 \in R^p$ is unknown and $g : R \to R$ is a known link function. We will assume that $m(\theta)$ and $g(\theta)$ are both continuous and strictly increasing. The link $g(\theta) = m(\theta)$ is called canonical. The support $D$ is an open interval (whose extremes may be $-\infty$ or $\infty$) of the set of real numbers in the continuous case or of the non-negative integers in the discrete case.

One important case of GLM is binomial regression where $y$ takes the values 0 and 1, $\theta = P(y = 1)$ and the link function $g$ is generally a symmetric distribution function. The most popular binomial regression model is the logistic model with link function

$$g(\theta) = \log\left(\frac{\theta}{1-\theta}\right). \tag{4}$$

If the model is perfectly observed, an efficient estimator for GLM is the maximum likelihood estimator (MLE). However, for some models, like in the case that $F_\theta$ is the normal or Poisson family, a few outliers can make the MLE tend to infinity. Instead, in the case of binomial regression, Croux, Flandre and Haesbroeck (2002) showed that if the model has an intercept, not more than $2(p-1)$ badly classified leverage points can make that the MLE of all the coefficients different from the intercept tend to zero and not to infinity.

Robust estimators have the property that they are not much affected by a small fraction of outliers. Several robust estimators have been proposed for GLM.

Künsch, Stefanski and Carroll (1989) derived optimal conditional unbiased bounded influence estimators. However, Maronna, Bustos and Yohai (1979) have shown that in the case of a linear model, the breakdown point of these estimators tends to 0 when $p$ tends to infinity. Cantoni and Ronchetti (2001) defined robust estimators for GLM which can be considered a robustification of the quasi-likelihood estimators introduced by Wedderburn (1974). For binomial regression several other estimators have been proposed. Among them we can cite Pregibon (1981), Copas (1988), Carroll and Pederson (1993), Christmann (1994), Bianco and Yohai (1996), Kordzakhia, Mishra and Reiersølmoen (2001), Croux and Haesbroeck (2003), Müller and Neykov (2003), Bondell (2005), Gervini (2005) and Čížek (2008).

One way to measure the robustness of an estimator is using the concept of maximum asymptotic bias (MAB). A class of estimators that have very good MAB properties is the class of projection estimators. Maronna and Yohai (1993) introduced this class of estimators for linear regression; Maronna, Stahel and Yohai (1992) and Tyler (1994) employed them for estimating multivariate scatter and location, respectively. All these estimators are based on an initial robust estimator for a one–parameter model and the robustness of the initial estimator is inherited by the projection estimators of the parameters of the more complex model.

In this paper we extend the class of P-estimators for GLM and show that they have similar properties as the P-estimators for the linear model. We study with special detail the case of binomial and Poisson regression and a Monte Carlo study shows that they compare favorably to other robust proposals. One shortcoming of P-estimators is that they are not asymptotically normal, and as a consequence there is not a simple way to implement inference procedures based on them. To overcome this drawback we propose one–step weighted M-estimators based on the scoring algorithm starting from the P-estimators. These one-step estimators have the same normal asymptotic distribution as weighted M-estimators and keep the breakdown point of the initial estimators. Moreover, a Monte Carlo study shows that their behavior under outlier contamination remains very close to that of P-estimators.

In Section 2 we introduce P-estimators for GLM and prove their Fisher–consistency. In Section 3 we give a bound for the maximum bias of P-estimators. When the fraction of contamination is small this bound is approximately twice the maximum bias of the initial estimator. In Section 4 we study the order of consistency of the P-estimators. In Section 5 we introduce a class of initial estimators for the auxiliary model. In Section 6 we study the breakdown point of the P-estimators when the initial estimator is as in Section 5. In Section 7 we present one–step estimators weighted M-estimators that use a P-estimator as initial value. In Section 8 we show the results of a Monte Carlo study where we compare our P-estimators with other robust estimators for logistic and Poisson regression. In Section 9 we analyze a real dataset. Section 10 contains some conclusions. All the proofs and a description of the numerical algorithm to compute the P-estimators can be found in the Supplemental Material available on line.

## 2. PROJECTION ESTIMATES FOR GLM

Let $\mathcal{H}$ be the set of distribution functions on $R^{p+1}$. Let $\mathbf{T} : \mathcal{H} \rightarrow R^p$ be an estimating functional for $\beta_0$ in a GLM satisfying (1)-(3). Then given a sample $(y_1, \mathbf{x}_1), \ldots, (y_n, \mathbf{x}_n)$ we can define an estimator of $\beta_0$ by $\beta_n = \mathbf{T}(H_n)$, where $H_n$ is the empirical distribution. Since by the Glivenko-Cantelli Theorem $H_n \rightarrow H$ uniformly, if the functional $\mathbf{T}$ is continuous we have $\lim_{n\rightarrow\infty} \beta_n = \lim_{n\rightarrow\infty} \mathbf{T}(H_n) = \mathbf{T}(H)$. The estimating functional $\mathbf{T}$ is Fisher consistent for the GLM model if given a distribution $H_0$ of $(y, \mathbf{x})$ satisfying (1)-(3) we have

$$\mathbf{T}(H_0) = \beta_0. \tag{5}$$

We now introduce an auxiliary uniparametric model, where three random variables $(y, w, z)$ are observed, such that

$$y|(z, w) \sim F_\theta \tag{6}$$

and

$$g(\theta) = \alpha_0 w + z. \tag{7}$$

Then this auxiliary model is a GLM with one regular explanatory variable and one offset variable.

As we will see below, we will use this auxiliary model taking $z = \beta' \mathbf{x}$ where $\beta$ is a candidate for estimating $\beta_0$ and $w = \lambda' \mathbf{x}$, where $||\lambda|| = \mathbf{1}$. The vector $\lambda$ gives a tentative direction for modifying $\beta$ and $\alpha_0$ (that will depend on $\lambda$) is the size of the correction. The value $\beta$ is a good candidate for estimating $\beta_0$ if the size of the correction $\alpha_0$ is small for all $\lambda$. In general, it is easier to find robust estimators for this uniparametric model than for the original model with $p$ parameters. We will understand that an estimator for the auxiliary model is robust if a small change on the distribution of $(y, w, z)$ has a small influence on the estimator

Let $T_0(H^*)$ be a robust and Fisher consistent estimating functional of $\alpha_0$, where $H^*$ is the distribution of $(y, w, z)$. The Fisher consistency condition requires that for any $H_0^*$ satisfying (6)-(7) we have

$$T_0(H_0^*) = \alpha_0. \tag{8}$$

Examples of robust estimating functionals $T_0$ for the auxiliary model are given in Section 5.

We introduce the following notation. Let $\mathbf{v}$ be a random vector in $R^q$, $h_i : R^q \to R$, $1 \leq i \leq m$, then $\mathcal{L}((h_1(\mathbf{v}), ..., h_m(\mathbf{v})),F)$ is the distribution of $(h_1(\mathbf{v}), ..., h_j(\mathbf{v}))$ when $\mathbf{v}$ has distribution $F$. We will assume that $T_0$ satisfies the following equivariance properties

$$T_0(\mathcal{L}((y, \gamma w, z), H^*) = \frac{1}{\gamma} T_0(\mathcal{L}((y, w, z), H^*)) \text{ for all } \gamma \in R \tag{9}$$

and

$$T_0(\mathcal{L}((y, w, z + \gamma w), H^*)) = T_0(\mathcal{L}((y, w, z), H^*)) - \gamma \text{ for all } \gamma \in R \tag{10}$$

for any distribution $H^*$ of $(y, w, z)$.

Based on $T_0$ we define an estimating functional $\mathbf{T}(H)$ of $\beta_0$ for the $p-$dimensional GLM model (1)-(3). Note that if $(y, \mathbf{x})$ satisfies (1) and (3), then, for all $\lambda \in R^p$, $(y, \lambda' \mathbf{x}, \beta_0' \mathbf{x})$ satisfies (6) and (7) with $\alpha_0 = 0$. Therefore according to (8), if $H_0$ is the distribution of $(y, \mathbf{x})$ satisfying (1)-(3), we have

$$T_0(\mathcal{L}((y, \lambda' \mathbf{x}, \beta_0' \mathbf{x}), H_0)) = 0 \text{ for all } \lambda \in R^p. \tag{11}$$

Then, given a distribution $H$ of $(y, \mathbf{x})$ (not necessarily satisfying model (1)-(3)), it is natural to define the estimator $\beta = \mathbf{T}(H)$ so that for any $\lambda$ the value of $T_0(\mathcal{L}((y, \lambda' \mathbf{x}, \beta_0' \mathbf{x}), H))$ be as close to 0 as possible. Then the projection estimating functional of $\beta_0$ is defined as follows:

Let $S$ be a scale functional, i.e., $S$ is defined on all distributions $F$ on $R$ and satisfies $S(F) \geq 0$ for all $F$ and if $u$ is a random variable we have $S(\mathcal{L}(\gamma u, F)) = |\gamma| S(F)$. For example $S$ may be the MAD scale functional defined by

$$S(F) = \frac{1}{0.6745} \text{median}(\mathcal{L}(|u|, F)). \tag{12}$$

Another possibility is to use as $S$ an M-estimating functional of scale. See for example Maronna, Martin and Yohai (2006, Section 2.5.). Define for any $\beta$ and any distribution $H$ on $R^{p+1}$

$$A(\beta,H)=\max_{||\lambda||=\mathbf{1}} S\left(\mathcal{L}(\lambda'\mathbf{x},H)\right) \quad \left|T_0\left(\mathcal{L}((y,\lambda'\mathbf{x},\beta'\mathbf{x}),H)\right)\right|. \tag{13}$$

Then the projection estimating functional for the $p$-dimensional GLM is defined by

$$\mathbf{T}(H)=\arg\min_{\beta\in R^p} A(\beta,H). \tag{14}$$

The estimator associated to this functional, i.e., the functional applied to the empirical distribution, will be called projection estimator (P-estimator). The factor $S(\mathcal{L}(\lambda'\mathbf{x},H))$ is necessary for the affine equivariance of $\mathbf{T}$. Moreover, note that without this factor we would have $A(\beta,H)=\infty$. Note also that the maximum in (13) can be taken for $\lambda\in R^p$ instead that for $||\lambda||=1$.

**Theorem 1.** *Let* $\mathbf{C}$ *be a* $p\times p$ *non-singular matrix and* $H$ *a distribution on* $R^{p+1}$, *then* $\mathbf{T}(\mathcal{L}((y,\mathbf{C}\mathbf{x}),H))=\mathbf{C}'^{-1}\mathbf{T}(H)$.

The following Theorem proves the Fisher consistency of the P-estimating functionals.

**Theorem 2.** *Assume that* $H_0$ *is the distribution of* $(y,\mathbf{x})$ *satisfying model (1)-(3) and* $G_0$ *the corresponding marginal distribution of* $\mathbf{x}$. *Suppose that* $T_0$ *satisfies (8) and that for all* $\lambda\in R^p$, $S(\mathcal{L}(\lambda'\mathbf{x},G_0))>0$. *Then the P-estimating functional* $\mathbf{T}$ *is Fisher consistent for* $\beta_0$.

# 3. A BOUND FOR THE MAXIMUM BIAS OF P-ESTIMATES

Let $H_0$ be the distribution of $(y,\mathbf{x})$ under model (1)-(3) and let $G_0$ be the marginal distribution of $\mathbf{x}$. Note that $H_0$ is determined by $G_0$ and $\beta_0$. To study the robustness of an estimating functional $\mathbf{T}$, we start by defining for any $0<\varepsilon<1$

the contamination neighborhood of $H_0$ of size $\varepsilon$ as

$$\mathcal{V}(H_0,\varepsilon)=\{H:H=(1-\varepsilon)H_0+\varepsilon\widetilde{H},\ \widetilde{H}\text{ arbitrary}\}.$$

Then a measure of the robustness of an estimating functional $\mathbf{T}$ of $\beta_0$ is the maximum asymptotic bias (MAB) $B(\mathbf{T},\beta_0,H_0,\varepsilon)$, defined by

$$B(\mathbf{T},\beta_0,G_0,\varepsilon)=\sup_{H\in\mathcal{V}(H_0,\varepsilon)}\left[(\mathbf{T}(H)-\beta_0)'V(G_0)(\mathbf{T}(H)-\beta_0)\right]^{1/2},$$

where $V$ is an affine equivariant scatter estimating functional of a distribution $H$ in $R^p$. The reason why we standardize the bias using $V(G_0)$ is to make the MAB affine invariant. Note that this definition of MAB is similar the one given for linear models, see e.g., Martin, Yohai and Zamar (1989). By extension, we will call MAB of an estimator to the MAB of the corresponding estimating functional.

In this Section we derive a bound for the MAB similar to the one found in Maronna and Yohai (1993) of linear regression. Let

$$c^+(G_0,\varepsilon,\lambda)=\sup_{G\in\mathcal{V}(G_0,\varepsilon)}S(\mathcal{L}(\lambda'\mathbf{x},G)),\ d^+(G_0,\varepsilon)=\sup_{||\lambda||=1}c^+(G_0,\varepsilon,\lambda)\tag{15}$$

and

$$c^-(G_0,\varepsilon,\lambda)=\inf_{G\in\mathcal{V}(G_0,\varepsilon\}}S(\mathcal{L}(\lambda'\mathbf{x},G)),\ d^-(G_0,\varepsilon)=\inf_{||\lambda||=1}c^-(G_0,\varepsilon,\lambda).\tag{16}$$

We have already seen that for any $\lambda\in R^p$, $(y,\lambda'\mathbf{x},\beta_0'\mathbf{x})$ satisfies model (6)–(7) with $\alpha_0=0$. For any $\lambda\in R^p$, let $M_{\lambda,\beta_0}$ be the distribution of $(y,\lambda'\mathbf{x},\beta_0'\mathbf{x})$ when $(y,\mathbf{x})$ has distribution $H_0$. Then we define the maximum bias of $T_0$ at $M_{\lambda,\beta_0}$ by

$$B(T_0,\lambda,\beta_0,G_0,\varepsilon)=\sup_{M\in\mathcal{V}(M_{\lambda,\beta_0},\varepsilon)}\frac{|T_0(M))|}{(\lambda'V^{-1}(G_0)\lambda)^{1/2}}.\tag{17}$$

**Theorem 3.** *Suppose that* $(y, \mathbf{x})$ *satisfies model (1)-(3) and let* $G_0$ *be the marginal distribution of* $\mathbf{x}$. *Then we have*

$$B(\mathbf{T}, \varepsilon, \beta_0, G_0) \leq \left(\mathbf{1}+\frac{d^+(G_0,\varepsilon)}{d^-(G_0,\varepsilon)}\right) \sup_{||\lambda||=1} B(T_0, \lambda, \beta_0, G_0, \varepsilon). \tag{18}$$

**Remark.** Note that $\sup_{||\lambda||=1} B(T_0, \lambda, \beta_0, G_0, \varepsilon)$ depends only on the distributions of bidimensional projections of $\mathbf{x}$ and $d^+(G_0, \varepsilon)$ and $d^-(G_0, \varepsilon)$ on the distribution of one–dimensional projections. Then, in the case that $G_0$ is multivariate normal with covariance matrix equal to the identity, the bound given in Theorem 3 for the maximum bias of the projection estimating functionals depends only on $||\beta_0||$ and not on $p$.

**Remark.** Since

$$\lim_{\varepsilon \to 0} \frac{d^+(G_0,\varepsilon)}{d^-(G_0,\varepsilon)} = 1,$$

the bound for the maximum bias of $\mathbf{T}$ is for small $\varepsilon$ approximately $2 \sup_{||\lambda||=1} B(T_0,\lambda,\beta_0,G_0,\varepsilon)$. In Table 1 we show $d^+(G_0, \varepsilon)/d^-(G_0, \varepsilon)$ when $G_0$ is N(0, I) and $S$ is the MAD given by (12).

TABLE 1 ABOUT HERE

In Section 6 we will use the bound given by (18) to obtain a lower bound for the asymptotic breakdown point of the P-estimating functionals.

## 4. CONSISTENCY ORDER OF P-ESTIMATES

In this Section we prove that if the order of convergence of the estimator derived from $T_0$ is $n^{-1/2}$, then the order of convergence of the corresponding P-estimator is also $n^{-1/2}$.

**Theorem 4.** *Let* $(y_1, \mathbf{x}_1), \ldots, (y_n, \mathbf{x}_n)$ *be a random sample of model (1)-(3) and let* $H_n$ *be its empirical distribution. Assume that the initial estimating functional* $T_0$ *verifies*

$$n^{1/2} \sup_{||\lambda||=\mathbf{1}} |T_0(\mathcal{L}((y, \lambda'\mathbf{x}, \beta_0'\mathbf{x}), H_n))| = O_p(1). \tag{19}$$

*We also assume that the scale functional* $S$ *satisfies*

$$\text{limsup}_{n\to\infty} \sup_{||\lambda||=\mathbf{1}} S(\mathcal{L}(\lambda'\mathbf{x}, G_n)) < \infty \text{ a.s.} \tag{20}$$

and

$$\text{liminf}_{n\to\infty} \inf_{||\lambda||=\mathbf{1}} S(\mathcal{L}(\lambda'\mathbf{x}, G_n)) > 0 \text{ a.s.}, \tag{21}$$

*where* $G_n$ *is the empirical distribution of the* $\mathbf{x}_i$*'s. Let* $\mathbf{T}$ *be the corresponding P-estimating functional, then*

$$n^{1/2}||\mathbf{T}(H_n) - \beta_0|| = O_p(1)$$

*too.*

In the next Section we propose a family of initial estimating functionals satisfying (19). Assume that $P_{G_0}(\lambda'\mathbf{x} = 0) < 0.5$ for all $\lambda$, then is easy to prove that the MAD scale given by (12) and the M-scales with breakdown 0.5 satisfy (20) and (21).

We have not been able to find the asymptotic distribution of the projection estimators proposed in this paper. However Zuo (2003) shows that projection estimators for multivariate location have a non–normal asymptotic distribution. The fact that these estimators are analogous to the ones defined here, allows us to conjecture that our projection estimators have also non–normal limit distributions. This makes it difficult to use these estimators for statistical inference. However, in Section 7 we present one way to overcome this problem.

## 5. INITIAL ESTIMATES

Consider the auxiliary model (6)-(7) and let $L$ be a possible distribution of $(y, w, z)$. We shall define robust initial estimating functionals $T_0$ of the form

$$E_L\left[\eta(y - \delta(g^{-1}(T_0(L)w + z)))\kappa\left(\frac{w}{S(L)}\right)\right] = 0, \tag{22}$$

where the function $\delta$ is defined by

$$E_\theta\left[\eta(y - \delta(\theta))\right] = 0, \tag{23}$$

$\eta$ and $\kappa$ are odd and bounded functions and $S$ is a scale functional. We will assume to simplify the presentation that the functional $S$ is the same as the one used in the definition of the P-estimators, but this is not necessary.

If we use a canonical link function, i.e., $g(\theta) = m(\theta)$, the functional $T_0^{\text{MV}}$ corresponding to the maximum likelihood estimating functional is of the form (22)-(23). In fact $T_0^{\text{MV}}$ is defined by

$$E_L\left[(y - \delta(g^{-1}(T_0^{\text{MV}}(L)w + z)))w\right] = 0,$$

where $\delta(\theta) = E_\theta[y]$. However the projection estimators obtained using initial robust estimators of the form (22)-(23) have good robustness properties even if $g(\theta)$ is not the canonical link.

It is immediate to show that (23) implies the Fisher consistency of these estimators. It is also easy to verify that the equivariance conditions (9) and (10) hold.

Suppose that $g(\theta) = m(\theta)$ and that the median of $F_\theta$ is well defined. Then it may be shown that the estimator with $\eta(u) = \text{sign}(u)$, $\kappa(w) = \text{sign}(w)$ has minimum gross error sensitivity (GES) in a broad class of estimators: the class of conditionally unbiased general M-estimators (CUGM-estimators). We call this estimator *conditional unbiased minimum GES* (CUMGES) estimator. It is immediate to see that in this case $\delta(\theta)$ =median$_\theta(y)$. The concept of conditionally unbiased estimators was introduced in Künsch et al. (1989).

Assume now that $g(\theta) = m(\theta)$ and that the support $D$ of $F_\theta$ is the set of non-negative integers. Then, in this case the CUMGES estimator is given by $\eta(u) = \eta^H_{0.5}(u)$, where $\eta^H_k$ is the Huber family

$$\eta^H_k(u) = \text{sign}(u)\min(|u|, k) \tag{24}$$

and $\kappa(w)$ =sign$(w)$. It can be also proved that if $0 < k < 0.5$, all the estimators of the form (22) with $\eta = \eta^H_k$ and $0 < k \leq 0.5$ coincide. The definition of the class of CUGM-estimators for the auxiliary model and an heuristic derivation of the two CUMGES estimators can be found in Section 10 of the Supplemental Material.

When $F_\theta$ is Bernoulli with logistic link function and parameter $\theta = P_\theta(y = 1)$, it can be shown that

$$\eta^H_{0.5}(y - \delta(m(\theta))) = \frac{y - \theta}{2\max(\theta, 1 - \theta)}. \tag{25}$$

We will assume the following properties

**P1.** $\eta(y)$ is odd, bounded, non-decreasing and continuously differentiable. In the continuous case we assume that $\eta^{/}(u) > 0$ in a neighborhood of 0 and in the case where the support $D$ is the set of non-negative integers, $\eta'(u) > 0$ in the interval $[-0.5, 0.5]$.

**P2.** $\kappa$ is odd, continuous, bounded, non-decreasing and $\kappa(u) > 0$ for $u > 0$.

**P3.** There is an sphere $U \subset R^p$ such that all $\mathbf{x} \in U$ has a positive density.

**P4.** The link function $g(\theta)$ and $m(\theta)$ in (2) are continuously differentiable and strictly increasing. Note that this assumption implies that $f(\mathbf{x},\theta)$ is continuously differentiable in $\theta$ too.

The next Theorem shows that under assumptions P1-P4, the family of estimators of the form (22) satisfy condition (19) and therefore the corresponding P-estimators have order of consistency $n^{-1/2}$. Note that the functions $\eta$ and $\kappa$ corresponding to the CUMGES estimator do not satisfy the smoothness conditions required in P1 and P2 respectively. However they can be approximated by functions satisfying these assumptions.

**Theorem 5**. *Let $(y,\mathbf{x})$ be a random vector in $R^{p+1}$ satisfying that $y|\mathbf{x}$ has distribution $F_\theta$. Assume P1-P4, (20) and (21). Then the estimating functional $T_0$ given by (22) satisfies (19).*

The following Theorem gives a bound for $B(T_0,\lambda,\beta_0,G_0,\varepsilon)$ when $T_0$ is of the form (22). Consider the function

$$Q(\alpha,\lambda,s) = E_{H_0}\left[\eta(y-\delta^*(\alpha\lambda'\mathbf{x}+\beta_0'\mathbf{x}))\kappa\left(\frac{\lambda'\mathbf{x}}{s}\right)\right],$$

where $\delta^* = \delta(g^{-1})$ The function $Q(\alpha,\lambda,s)$ is continuous and monotone non-increasing in $\alpha$. Let $e(\varepsilon) = c_1c_2\varepsilon/(1-\varepsilon)$ where

$$c_1 = \sup\eta,\ \ c_2 = \sup\kappa \tag{26}$$

and define

$$\alpha_1(\varepsilon,\lambda) = \inf\{\alpha : Q(\alpha,\lambda,c^+(G_0,\varepsilon,\lambda)) \le e(\varepsilon)\},$$
$$\alpha_2(\varepsilon,\lambda) = \sup\{\alpha : Q(\alpha,\lambda,c^+(G_0,\varepsilon,\lambda)) \ge -e(\varepsilon)\},$$

where $c^+(G_0, \varepsilon, \lambda)$ is defined in (15).

**Theorem 6.** Suppose that $(y, \mathbf{x})$ satisfies a GLM and let $T_0$ be an estimating functional for the auxiliary model of the form.(22). Assume P1-P4, then

$$B(T_0, \lambda, \beta_0, G_0, \varepsilon) \leq \max\left(-\frac{\alpha_1(\varepsilon, \lambda)}{(\lambda' V^{-1}(G_0)\lambda)^{1/2}}, \frac{\alpha_2(\varepsilon, \lambda)}{(\lambda' V^{-1}(G_0)\lambda)^{1/2}}\right).$$

# 6. BREAKDOWN POINT OF PROJECTION ESTIMATES

The asymptotic breakdown point to infinity ($\mathrm{ABDP}_\infty$) of an estimating functional $\mathbf{T}$ for a GLM at $H_0$ is defined by

$$\varepsilon^*(\mathbf{T}, H_0) = \inf\{\varepsilon : M(\mathbf{T}, H_0, \varepsilon) = \infty\},$$

where

$$M(\mathbf{T}, H_0, \varepsilon) = \sup\{||\mathbf{T}(H)|| : H \in \mathcal{V}(H_0, \varepsilon)\}$$

and $H_0$ is the distribution of $(y, \mathbf{x})$. determined by $\beta_0$ and $G_0$. By extension, we call $\mathrm{ABDP}_\infty$ of an estimator to the $\mathrm{ABDP}_\infty$ of the corresponding estimating functional.

For the case that $F_\theta$ is the family of Bernoulli distributions, Croux et al. (2002) showed that the effect of large outliers is to take the estimators of $\beta_0$ to $\mathbf{0}$ instead to $\infty$. Then in this case a meaningful robustness measure is the BDP to zero which is similarly defined. However this BDP depends on $\beta_0$ and decreases to 0 when $||\beta_0||$ tends to 0. For this reason this measure is very difficult to compute and it will not be studied here.

In this Section we show that for some GLM models we can find P-estimating functionals with initial estimator $T_0$ of the form (22) with $\mathrm{ABDP}_\infty$ close to 0.5.

We will study the breakdown point of the projection estimating functionals in two cases where

$$\inf_{\theta} m(\theta) = \inf_{\theta} g(\theta) - \infty, \ \sup_{\theta} m(\theta) = \sup_{\theta} g(\theta) = \infty \tag{27}$$

hold.

**Case A.** Suppose that $F_\theta$ is a family of continuous distributions with support $R$ and satisfying (27). This is the case for example when $F_\theta$ is $\mathrm{N}(\theta, \sigma_0^2)$ where $\sigma_0$ is known and link function $g(\theta) = \theta$. Then (27) is satisfied.

**Case B.** Suppose that under $F_\theta$ the support of $y$ is the the set of non-negative integers and (27) is satisfied. This is the case for example when $F_\theta$ is the Poisson family of distributions with probability density $f(x, \theta) = \exp(\theta)\theta^x/x!$ and $g(\theta) = \log(\theta)$. Clearly (27) is satisfied for the Poisson family.

General lower bounds for the breakdown point of P-estimators with initial estimating functional of the form (22) can be obtained. However, since they are quite involved, they are not given here. In Theorem 7 we show that for these two cases we can found initial estimating functionals of the form (22) so that the corresponding projection estimating functionals have high breakdown point.

Let $\varepsilon^{+*}(S, G_0)$ and $\varepsilon^{-*}(S, G_0)$ be defined by

$$\varepsilon^{+*}(S, G_0) = \inf\{\varepsilon : d^+(G_0, \varepsilon) = \infty\},$$
$$\varepsilon^{-*}(S, G_0) = \inf\{\varepsilon : d^-(G_0, \varepsilon) = 0\},$$

where $d^+(S, G_0)$ and $d^-(G_0, \varepsilon)$ are defined in (15) and (16). Denote by

$$\xi_1(G_0) = \inf_{||\lambda||=1} P_{G_0}(\lambda'\mathbf{x} \neq 0), \xi_2(H_0) = \inf_{||\lambda||=1} P_{H_0}(\lambda'\mathbf{x} \neq 0, y \neq 0), \tag{28}$$

and

$$\vartheta_1(G_0) = \frac{\xi_1(G_0)}{1 + \xi_1(G_0)}, \vartheta_2(H_0) = \frac{\xi_2(H_0)}{1 + \xi_2(H_0)},$$

$$\varepsilon_{01}(G_0) = \min(\varepsilon^{+*}(S, G_0), \varepsilon^{-*}(S, G_0), \vartheta_1(G_0)),$$

$$\varepsilon_{02}(H_0) = \min(\varepsilon^{+*}(S, G_0), \varepsilon^{-*}(S, G_0), \vartheta_2(H_0)).$$

**Theorem 7.** *Consider a GLM model where (27) is satisfied and let* $\mathbf{T}$ *be a P-estimating functional with initial estimating functional* $T_0$ *as defined in (22). Then*

*(i) Suppose that* $F_\theta$ *is as in case A,* $\eta$ *satisfies P1 and P4 and assume* $\kappa(x)$ =sign$(x)$. *Then* $\varepsilon^*_\infty(\mathbf{T}, H_0) \geq \varepsilon_{01}(G_0)$.

*(ii) Suppose that* $F_\theta$ *is as in case B,* $\eta$ *satisfies P1 and P4. Assume also* $\eta(y) = c_1$ for $y \geq 1$ and $\kappa(x)$ =sign$(x)$. *Then* $\varepsilon^*_\infty(\mathbf{T}, H_0) \geq \varepsilon_{02}(H_0)$.

**Remark.** Note that according to Theorem 7 in the case A if $\varepsilon^{+*}(S, G_0) = \varepsilon^{-*}(S, G_0) = 0.5$ and $\xi_1(G_0) = 0$, we can obtain P-estimating functionals with breakdown point 0.5. In the case B, under the same conditions plus the condition that $P_{H_0}(y = 0)$ is small, we can find estimators with breakdown point close to 0.5 too. It can also be proved that if sign$(x)$ is approximated by a smooth function $\kappa(x)$ satisfying P2, the bounds for $\varepsilon^*_\infty(\mathbf{T}, H_0)$ given in Theorem 7 are still approximately correct. We also observe that in case B the breakdown point lower bound $\varepsilon_{02}$ tends to 0 when $P_{H_0}(y = 0)$ tends to 1. This may be explained by the the fact that in this case zero is an extreme value of the response $y$.

# 7. ONE–STEP WEIGHTED M–ESTIMATORS

## 7.1. Definition and asymptotic normality

We consider the class of weighted M-estimators $\widehat{\beta}_n^M$ of $\beta_0$ for the GLM defined by

$$\sum_{i=1}^{n} w(d(\mathbf{x}_i, \widehat{\mu}_n, \widehat{\mathbf{\Sigma}}_n))\psi(y_i, \beta'\mathbf{x}_i, s_n^*)\mathbf{x}_i = 0,$$

where $\psi : R^3 \to R$ satisfies the Fisher consistency condition

$$E_\theta \left[ \psi(y_i, g(\theta), s) \right] = 0 \text{ for all } \theta \text{ and for all } s,$$

$d(\mathbf{x}, \mu, \mathbf{\Sigma})$ is the square Mahalanobis distance

$$d(\mathbf{x}, \mu, \mathbf{\Sigma}) = (\mathbf{x} - \mu)' \mathbf{\Sigma}^{-1} (\mathbf{x} - \mu), \tag{29}$$

$\widehat{\mu}_n$ and $\widehat{\mathbf{\Sigma}}_n$ are robust estimators of multivariate location and scatter of $\mathbf{x}$ converging a.s. to values $\mu_0$ and $\mathbf{\Sigma}_0$, $w : R_{\geq 0} \to R_{\geq 0}$ ($R_{\geq 0}$ is the set of non-negative real numbers) is a non-increasing weight function that penalizes high leverage observations and $s_n^*$ is an adaptive tuning constant. Most often $s_n^*$ will be a robust scale estimator of the errors $y_i - q(g^{-1}(\beta_0' \mathbf{x}_i))$, $1 \leq i \leq n$, where $q(\theta) = E_\theta[y]$. In some cases, as when $F_\theta$ is Bernoulli, $s_n^*$ may not be necessary. In general denote by $N_p(\mu, \mathbf{\Sigma})$ the multivariate normal distribution of dimension $p$ with mean $\mu$ and covariance matrix $\mathbf{\Sigma}$ and by $\to^D$ convergence in distribution. Note that the maximum likelihood estimator is of this form with $\psi(y, \phi, s) = y - q(g^{-1}(\phi))$ and $w = 1$. Suppose that $s_n^* \to s_0^*$, then using general Theorems for the asymptotic normality of M-estimators (see for example Theorem 10.11 of Maronna et al. (2006)) it can be proved that under general conditions

$$n^{1/2}(\widehat{\beta}_n^M - \beta_0) \to^D N(0, V_0),$$

where

$$V_0 = \mathbf{A}_0^{-1} \mathbf{C}_0 \mathbf{A}_0^{-1\prime}, \tag{30}$$

$$\mathbf{C}_0 = E \left[ w(d(\mathbf{x}_i, \mu_0, \mathbf{\Sigma}_0))^2 \psi^2(y, \beta_0' \mathbf{x}_i, s_0^*) \mathbf{x}_i \mathbf{x}_i' \right]$$

and

$$\mathbf{A}_0 = E \left[ w(d(\mathbf{x}, \mu_0, \mathbf{\Sigma}_0)) \psi_2(y, \beta_0' \mathbf{x}, s_0^*) \mathbf{x}_i \mathbf{x}_i' \right], \tag{31}$$

where $\psi_2(y,\phi,s) = \partial\psi(y,\phi,s)/\partial\phi$.

If we have an initial estimator $\widehat{\beta}_{0n}$ of $\beta_0$, the one-step weighted M-estimator is defined by

$$\widehat{\beta}_{1n} = \widehat{\beta}_{0n} - \widehat{\mathbf{A}}_n^{-1}\frac{1}{n}\sum_{i=1}^{n} w(d(\mathbf{x}_i,\widehat{\mu}_n,\widehat{\mathbf{\Sigma}}_n))\psi(y_i,\widehat{\beta}'_{0n}\mathbf{x}_i,s_n^*)\mathbf{x}_i,$$

where $\widehat{\mathbf{A}}_n$ is a consistent estimator of $\mathbf{A}_0$. We consider two choices for $\widehat{\mathbf{A}}_n$ : Newton-Raphson, where

$$\widehat{\mathbf{A}}_n = \frac{1}{n}\sum_{i=1}^{n} w(d(\mathbf{x}_i,\widehat{\mu}_n,\widehat{\mathbf{\Sigma}}_n))\psi_2(y,\widehat{\beta}'_{0n}\mathbf{x}_i,s_n^*)\mathbf{x}_i\mathbf{x}_i', \tag{32}$$

and scoring, where

$$\widehat{\mathbf{A}}_n = \frac{1}{n}\sum_{i=1}^{n} w(d(\mathbf{x}_i,\widehat{\mu}_n,\widehat{\mathbf{\Sigma}}_n))\tau(g^{-1}(\widehat{\beta}'_{0n}\mathbf{x}_i),s_n^*))\mathbf{x}_i\mathbf{x}_i' \tag{33}$$

and

$$\tau(\theta,s) = E_\theta\left[\psi_2(y,g(\theta),s)\right].$$

As $s_n^*$ we can take a robust scale estimator, e.g., the MAD scale or an M-scale estimator of $y_i - q(g^{-1}(\widehat{\beta}'_0\mathbf{x}_i))$.

The idea of using one-step estimators to combine two different properties was used by Simpson et al. (1992). They obtained one-step estimators for the linear model that combine high breakdown point and bounded influence. In the Theorem 8 below we prove that for GLM models, under general conditions, the one–step estimator $\widehat{\beta}_{1n}$ is also asymptotically normal with covariance matrix given by (30). If $\widehat{\beta}_0$ is more robust than $\widehat{\beta}_n^M$, we can also expect that $\widehat{\beta}_{1n}$ be more robust than $\widehat{\beta}_n^M$ too. Then, one way to obtain a highly robust estimator for a generalized linear model which is asymptotically normal is to compute a one-step weighted M-estimator $\widehat{\beta}_{1n}$ taking as $\widehat{\beta}_{0n}$ the projection M-estimator. We have verified by simulation that –at least for the logistic and Poisson models– the one–step scoring estimator behaves more robustly than the one–step Newton-Raphson estimator. One possible explanation is that $y$ outliers have influence on $\widehat{\mathbf{A}}_n$ when this matrix is defined by (32), but $\widehat{\mathbf{A}}_n$ it is not affected when we use (33).

The following assumptions are required to prove the asymptotic normality of the one-step weighted M-estimators.

**A1.** $\psi(y, \phi, s)$ is bounded and continuous as a function of $(y, \phi, s)$ and continuously differentiable in $\phi$. Moreover $\psi_2(y, \phi, s)$ is bounded when $s$ is bounded away from 0.

**A2.** $E_\theta[\psi(y, g(\theta), s)] = 0$ for all $\theta$ and $s$.

**A3.** $w(z)$ is continuous, non-increasing and bounded. We also assume that $zw(z)$ is bounded.

**A4.** $n^{1/2}(\widehat{\beta}_{0n} - \beta_0) = O_p(1)$.

**A5.** $\widehat{\mathbf{A}}_n \rightarrow^p \mathbf{A}_0$.

**A6.** $n^{1/2}(\widehat{\mathbf{\Sigma}}_n - \Sigma_0) = O_p(1)$, $n^{1/2}(\widehat{\mu}_n - \mu_0) = O_p(1)$ and $n^{1/2}(s_n^* - s_0^*) = O_p(1)$.

**A7.** The matrix $\mathbf{A}_0$ is not singular

**Theorem 8.** *Let $(y_1, \mathbf{x}_1), ..., (y_n, \mathbf{x}_n)$ be i.i.d. observations satisfying a GLM given by (1), (2)and (3). Assume that A1-A7 and P4 hold, then $n^{1/2}(\widehat{\beta}_{1n} - \beta_0) \rightarrow^D N(0, V_0)$.*

The asymptotic covariance matrix $V_0$ of $\widehat{\beta}_{1n}$ can be estimated by $\widehat{V} = \widehat{\mathbf{A}}_n^{-1} \widehat{\mathbf{C}}_n \widehat{\mathbf{A}}_n^{-1\prime}$, where $\widehat{\mathbf{A}}_n$ is given by (32) or (33) and

$$\widehat{\mathbf{C}}_n = \frac{1}{n} \sum_{i=1}^{n} w^2(d(\mathbf{x}_i, \widehat{\mu}_n, \widehat{\mathbf{\Sigma}}_n)) \psi^2(y_i, \widehat{\beta}_{0n}' \mathbf{x}_i, s_n^*) \mathbf{x}_i \mathbf{x}_i'.$$

In the Monte Carlo study of Section 8 we include a one–step M-estimator of scoring type for the logistic and Poisson regression models starting from a projection estimator. The simulation results show that the degree of robustness of these estimators remains very close to the one of the projection estimator.

We can take as $\psi$ a function of the form

$$\psi(y, \phi, s) = \eta \left( \frac{y - \delta(g^{-1}(\phi), s)}{s} \right), \tag{34}$$

where $\delta(\theta, s)$ is chosen so that $E_\theta \left[ \eta \left( (y - \delta(\theta, s)) / s \right] = 0$ for all $\theta$ and $s$. Another possible family of $\psi$ functions, when $F_\theta$ is the Bernoulli family, is given in Bianco and Yohai (1996) and in Croux et al. (2002).

## 7.2 Asymptotic breakdown point

The estimating functional $\mathbf{T}_1(H)$ associated with the scores one-step weighted M-estimator can be written as

$$\mathbf{T}_1(H) = \mathbf{T}_0(H) - \mathbf{T}_{\mathbf{A}}(H)^{-1} E_H \left[ w(d(\mathbf{x}, \mathbf{T}_\mu(H), \mathbf{T}_{\mathbf{\Sigma}}(H))) \psi(y, \mathbf{T}_0(H)'\mathbf{x}, T_{S^*}(H)) \mathbf{x} \right],$$

where $\mathbf{T}_0, \mathbf{T}_\mu, \mathbf{T}_{\mathbf{\Sigma}}$, $\mathbf{T}_{\mathbf{A}}$ and $T_{S^*}$ are the functionals associated to the estimators $\widehat{\beta}_{0n}, \widehat{\mu}_n$, $\widehat{\mathbf{\Sigma}}_n$, $\widehat{\mathbf{A}}_n$ and $s_n^*$. The functional $\mathbf{T}_{\mathbf{A}}$ corresponding to the scores version of the one-step estimator is

$$\mathbf{T}_{\mathbf{A}}(H) = E_H \left[ w(d(\mathbf{x}, \mathbf{T}_\mu(H), \mathbf{T}_{\mathbf{\Sigma}}(H))) \tau(g^{-1}(\mathbf{T}_0(H)'\mathbf{x}), T_{S^*}(H)) \mathbf{x}\mathbf{x}' \right].$$

Given a symmetric matrix $\Omega$, we denote by $\gamma_1(\mathbf{\Omega}) \leq \gamma_2(\mathbf{\Omega}) \leq .... \leq \gamma_p(\mathbf{\Omega})$ the eigenvalues of $\mathbf{\Omega}$. Consider an estimating functional $\mathbf{T}_\Omega(H)$ of a $p \times p$ symmetric matrix $\Omega_0$, then we define the following asymptotic breakdown points of $\mathbf{T}_\Omega$ at $H_0$ as follows:

$$\varepsilon^{-*}(T_\Omega, H_0) = \inf \left\{ \varepsilon : \inf_{H \in \mathcal{V}(H_0, \varepsilon)} \inf_{1 \leq i \leq p} |\gamma_i(\mathbf{T}_A(H))| = 0 \right\},$$

$$\varepsilon^{+*}(T_\Omega, H_0) = \inf \left\{ \varepsilon : \sup_{H \in \mathcal{V}(H_0, \varepsilon)} \sup_{1 \leq i \leq p} |\gamma_i(\mathbf{T}_A(H))| = \infty \right\}$$

and

$$\varepsilon^*(T_\Omega, H_0) = \min(\varepsilon^{-*}(T_\Omega, H_0), \varepsilon^{+*}(T_\Omega, H_0)).$$

We also define the following asymptotic breakdown points of $T_{S^*}$ at $H_0$

$$\varepsilon^{-*}(T_{S^*}, H_0) = \inf \left\{ \varepsilon : \inf_{H \in \mathcal{V}(H_0,\varepsilon)} \mathbf{T}_{S^*}(H) = 0 \right\},$$

$$\varepsilon^{+*}(T_{S^*}, H_0) = \inf \left\{ \varepsilon : \sup_{H \in \mathcal{V}(H_0,\varepsilon)} \mathbf{T}_{S^*}(H) = \infty \right\}$$

and

$$\varepsilon^*(T_{S^*}, H_0) = \min(\varepsilon^{-*}(T_{S^*}, H_0), \varepsilon^{+*}(T_{S^*}, H_0)).$$

Finally let $\varepsilon^*(\mathbf{T}_0, H_0)$, $\varepsilon^*(\mathbf{T}_1, H_0)$ and $\varepsilon^*(\mathbf{T}_\mu, H_0)$ be the ABDP$_\infty$ s of $\mathbf{T}_0$, $\mathbf{T}_1$ and $\mathbf{T}_\mu$. We need the following additional assumptions:.

**A8**. The functional $\mathbf{T}_\Sigma$ is standardized, so that $\text{median}_H(d(\mathbf{x}, \mathbf{T}_\mu(H), \mathbf{T}_\Sigma(H))) = F^{-1}_{\chi^2_p}(0.5)$, where $F_{\chi^2_p}$ is the chi-squared distribution function with $p$ degrees of freedom.

**A9.** $w_0 = w(F^{-1}_{\chi^2_p}(0.5)) > 0$.

**A10.** For all $\theta$ and $s$ we have $\tau(\theta, s) < 0$.

Note that any scatter functional $\mathbf{T}_\Sigma$ can be standardized so that A8 is satisfied, and therefore this assumption is not restrictive. The assumption A10 is satisfied if $\psi(y, \theta, s)$ is monotone non-increasing in $\theta$, for example if $\psi(y, \theta, s)$ is given by (34) with $\eta$ monotone.

**Theorem 9.** *(a) Assume P4, A1, A2, A3 and A5. Then*

$$\varepsilon^*(\mathbf{T}_1, H_0) \geq \min(\varepsilon^*(\mathbf{T}_0, H_0), \varepsilon^*(\mathbf{T}_\mu, H_0), \varepsilon^*(\mathbf{T}_{\boldsymbol{\Sigma}}, H_0), \varepsilon^{-*}(\mathbf{T}_A, H_0)).$$

*(b) Assume also A8, A9, A10 and* $\xi_1(G_0) \geq 0,5$, *where* $\xi_1(G_0)$ *is defined in (28). Then*

$$\varepsilon^{-*}(\mathbf{T}_A, H_0) \geq \min\left(\varepsilon^*(\mathbf{T}_0, H_0),\ \varepsilon^*(\mathbf{T}_{S^*}, H_0), \frac{\xi_1(G_0) - 0.5}{\xi_1(G_0)}\right).$$

Note that Theorem 9 implies that if we take $\mathbf{T}_\Sigma, \mathbf{T}_\mu$ and $T_{S^*}$ with asymptotic breakdown point 0.5, $\psi(y,\phi,s)$ monotone non-increasing in $\phi$ and $\zeta_1(G_0)=1$ we obtain that $\varepsilon_1^*(\mathbf{T}_1, H_0) \geq \varepsilon_0^*(\mathbf{T}_0, H_0)$.

# 8. MONTE CARLO STUDY

To study the efficiency and robustness of the P-estimators we performed a Monte Carlo study for logistic and Poisson regression.

## 8.1 Logistic regression

We consider the logistic model where $\mathbf{x}_i = (1, x_{i1}, x_{i2})'$ and $x_{i1}$ and $x_{i2}$ are independent with distribution $\mathrm{N}_1(0{,}1)$ and $g(P(y_i = 1|\mathbf{x}_i)) = \beta'\mathbf{x}_i$, where $g(\theta)$ is given in (4) and $\beta = (\beta_1, \beta_2, \beta_3)' = (0, 1, -1)'$. This model was used to compare several estimators in the Monte Carlo study in Čížek (2008). We consider the following estimators.

**(i)** Maximum likelihood (**ML**), which is the solution of

$$\sum_{i=1}^{n} \left( y_i - \frac{\exp(\beta'\mathbf{x}_i)}{1+\exp(\beta'\mathbf{x}_i)} \right) \mathbf{x}_i = 0.$$

**(ii)** Weighted maximum likelihood (**WML**) estimator, defined by

$$\sum_{i=1}^{n} w(d(\mathbf{x}_i^*, \widehat{\mu}, \widehat{\mathbf{\Sigma}})) \left( y_i - \frac{\exp(\beta'\mathbf{x}_i)}{1+\exp(\beta'\mathbf{x}_i)} \right) \mathbf{x}_i = 0,$$

where $\mathbf{x}_i^* = (x_{i1}, x_{i2})'$, $d(\mathbf{x}, \mu, \mathbf{\Sigma})$ is defined in (29), $w(t) = I(t^2 \leq \chi^2_{2,\,0.975})$, $\chi^2_{p,\alpha}$ is the quantile $\alpha$ of a $\chi^2-$ distribution with $p$ degrees of freedom and $\widehat{\mu}$ and $\widehat{\mathbf{\Sigma}}$ are the location and scatter MCD estimators of the $\mathbf{x}_i^{*\prime}$s (Rousseeuw and Leroy, 1987, page 262) using 75% of the observations.

**(iii)** The weighted M-estimators (WM) proposed by Croux and Haesbroeck (2003) (also called weighted Bianco-Yohai estimators). These estimators where computed using the $\rho$-function with derivative

$$\rho'(t) = \begin{cases} \exp(-\sqrt{d}) & \text{if} \quad t \leq d \\ \exp(-\sqrt{t}) & \text{if} \quad t > d. \end{cases} \tag{35}$$

We took $d = 0.5$. The weights are the same as those used in the WML estimator. These estimators were computed using the program by Croux and Haesbroeck WBYlogreg downloaded from
http://www.econ.kuleuven.be/public/NDBAE06/programs.

**(iv)** A Projection estimator where the initial estimator $T_0$ is of the form (22 ) (PR). We took $\eta(y)$ as in (25) and $\kappa$ in the Huber family given in (24) with $k = 0.8$. The scale $S$ is the MAD scale given by (12).

**(v)** The scores one–step version described in Section 7 of the WM estimator (PR-WM). The starting estimator was PR and the WM estimator has function $\rho$ with derivative

$$\rho'(t) = \begin{cases} 1 & \text{if} \quad t \leq d \\ \exp(-a(t^{0.25} - d^{0.25})^2) & \text{if} \quad t > d. \end{cases}$$

We took $a = 6$ and $d = 0.5$. The weight function is $w(d) = \rho_B((d - Q_{\chi^2}(0.99, 5))/0.2)$, where $Q_{\chi 2}(\alpha, p)$ is the quantile $\alpha$ of the chi-squared distribution with $p$ degrees of freedom and $\rho_B$ is the bisquare loss function

$$\rho_B(u) = (1 - (1 - u^2)^3)I(|u| < 1) + I(|u| \geq 1.)$$

Note that $w(d)$ is a continuously differentiable function that takes value one if $d < Q_{\chi^2}(0.99, 5) - 0.2$ and zero if $d > Q_{\chi^2}(0.99, 5)$. The results with the

PR-WM that uses the function $\rho'$ in (35) are quite similar but slightly less robust under outlier contamination.

**(vi)** The adaptive maximum symmetrically trimmed likelihood estimator introduced by Čížek (2008) (AMSTLE). This estimator was computed with an R-code provided by P. Čížek.

We consider samples of size $100, 200$ and $400$ and we perform for each situation 500 replications. We study the behavior of the estimators under two situations:

- (i) No outliers.
- (ii) Approximately 5% of outliers. For this purpose we added 6 outliers to the samples of size 100, 11 to those of size 200 and 22 to those of size 400. The outliers are of the form $y = 0, \mathbf{x} = (1,x_1, x_2)'$ where $x_1$ and $x_2$ are independent with distribution $\mathrm{N}(z, 0.2)$ and $\mathrm{N}(-z, 0.2)$ The value of $z$ varies between 0.5 and 4.5 with steps of 0.1. As proved by Croux et al. (2002) this type of outliers make the MLE of $\beta_1$ and $\beta_2$ tend to 0 when $z \to \infty$.

For each estimator $\widehat{\beta}$ and each model we estimate the mean squared error MSE$=E\left[||\widehat{\beta}-\beta_0||^2\right]$ by $\widehat{\mathrm{MSE}} = \frac{1}{N}\sum_{i=1}^{N}\left\|\widehat{\beta}_i - \beta_0\right\|^2$, where $\widehat{\beta}_i$ is the estimate obtained with the $i$-th sample.

In Table 2 we show the MSE of ML and the MSE efficiencies of the other estimators with respect to ML when there is no contamination. These efficiencies are defined as the ratio between the MSEs of ML and that of the corresponding estimator, times 100. In Table 3 we report the maximum value of $\widehat{\mathrm{MSE}}$ under outlier contamination when $z$ varies between 0.5 and 5. In Figure 1 we plot the $\widehat{\mathrm{MSE}}$ of the robust estimators under 5% of contamination for $z$ between 0.5 and 2. All the estimators except ML take the maximum MSE in this range of values.

TABLE 2 ABOUT HERE

TABLE 3 ABOUT HERE

FIGURE 1 ABOUT HERE

Under the logistic model without outliers the MSE efficiencies of the PR-WM are about 86% which may be considered quite high for a highly robust estimator. We also observe that the efficiencies of the AMSTLE become close to 100% when the sample size increases. These high efficiencies are consistent with the fact that its asymptotic efficiency is 100%. Instead, under outlier contamination, the PR and PR-WM estimators have smaller maximum $\widehat{\text{MSE}}$ than the other estimators and therefore they may be considered the most robust ones. Figure 1 show that for small values of $z$ (smaller than 0.9), the AMSTLE is slightly better than the PR and PR-WM, for intermediate values of $z$ ($0.9 < z < 1.5$), the PR and PR-WM are better than the AMSTLE and for large values of $z$ ($z$ larger than 1.5), the AMSTLE is again better. Although the maximum bias of the PR estimator is slightly smaller than the one of the PR-WM, the PR-WM has the advantage, as we have seen in Section 7, of being asymptotically normal.

## 8.2 Poisson regression

Let $F_\theta$ be the Poisson family of distributions. Consider the GLM with link function $g(\theta) = m(\theta) = \log(\theta)$. The vector of covariables is $\mathbf{x} = (1, \mathbf{x}^*)'$, where $\mathbf{x}^* = (x_1, ...., x_5)'$ has distribution $\mathrm{N}_5(\mathbf{0}, \mathbf{I})$ and $y|\mathbf{x}$ has distribution $F_\theta$ with $\log(\theta) = \beta_0'\mathbf{x}$ and $\beta_0 = (0, c, c, c, c, c)'$ with $c = 0.4$. We will consider the following estimators:

**(i)** The maximum likelihood estimator (ML).

**(ii)** An estimator of the family of optimal conditionally unbiased bounded-influence estimators (CUBI) proposed in Künsch et al. (1989). We compute this estimator with the function qlmRob of the robust library of SPLUS and the default parameters.

**(iii)** A robust quasi likelihood estimator (RQL) proposed by Cantoni and Ronchetti (2001). We compute this estimator with the function qlmrob of the robustbase package of R. We used the option `weights.on.x="RobCov"`.

**(iv)** A projection estimator where the initial estimator $T_0$ for the auxiliary model is of the form (22 ) (PR). We took as $\eta(y)$ an odd and continuously differentiable function which is constant for $|y| \geq 1$. For $u \geq 0$, the function $\eta$ is defined by

$$\eta(u) = \begin{cases} u & \text{if } \; 0 \leq u \leq 0,5 \\ -u^2 + 2u - 0.25 & \text{if } \; 0.5 < u \leq 1 \\ 0.75 & \text{if } \quad u > 1 \end{cases}$$

and as $\kappa$ we took the Huber function $\eta_{0.8}^H$ given in (24). As $S$ we use the MAD scale given by (12). The number of subsamples was $N = 500$ and $h = 6$.

**(v)** The scores one–step version described in Section 7 of the WM estimator starting from PR (WM-PR). We took $\psi(y, \phi, s) = \eta((y - \delta(\exp(\phi), s))/s)$ where $\delta(\theta, s)$ is chosen so that $E_\theta\left[\eta((y - \delta(\theta, s))/s)\right] = 0$ and $\eta(u) = \tanh(u/1.5)$, where tanh stands for hyperbolic tangent. We used the same weight function $w$ as for the estimator PR-WM for the logistic model.

We consider samples of size 100, 200 and 400 and we perform for each situation 500 replications using the same samples for all the estimators. We study the behavior of the estimators under two situations:

- (i) Samples without outliers.
- (ii) Samples which contain 10% of outliers. For this purpose we replace 10 observations of each sample of size 100 for identical outliers. The outliers are of the form $(\mathbf{x}_0, y_0)$, where $\mathbf{x}_0 = (1, 2.5, 0, 0, 0, 0)$. The values of $y_0$ are 0, 1, 2, 3, 4, 5, 10, 15 and 20. The expected value of $y$ when $\mathbf{x} = \mathbf{x}_0$ and $\beta = \beta_0$ is 2.718.

In Table 4 we show the MSE of ML and the MSE efficiencies of the robust estimators with respect to ML when there is no contamination. In Figure 2 we plot the MSE of the four robust estimators as a function of $y_0$ when the fraction of outliers is 10%. The efficiency of the PR estimators is around 60 % which may be considered rather low, while the efficiencies of the other estimators are between 70 and 80%. Figure 2 shows that in this case the PR and the PR-WM are clearly more robust than the other estimators. The PR-WM which has a good behavior under the model and under outlier contamination, appears as the best option among the five estimators studied for this model.

TABLE 4 ABOUT HERE

FIGURE 2 ABOUT HERE

## 9. EXAMPLE: FOOD STAMPS DATA

The following example was used by Stefanski, Carroll and Ruppert (1986) and by Künsch et al. (1989). The response variable is participation on the Food Stamp Program, i.e., $y = 1$ denotes participation. The covariates include log(monthly income + 1) (LMI) and two dichotomous variables: tenancy (TEN) and supplemental income (SI). The model has also an intercept. The data consist of 150 observations.

The robust procedures considered are the same that were used in the Monte Carlo study for the logistic model. The only difference is that the weights used for the WML, WM and PR-WM estimators were computed using only the variable LMI. The reason is that any high breakdown covariance estimator of the set of variables that includes a dichotomous variable is singular. Table 5 shows the coefficients and standard errors of the estimators. We observe that the values of the robust estimates WML, WM, PR, AMSTLE and PR-WM are quite close and are far from the value of the ML estimator. The comparison between the fitted probabilities using the PR estimator and the observed $y$'s, reveals that observations 5 and 66 are outliers. In fact, their $y$ values are 0 and 1 and the corresponding fitted values of $P(y = 1)$ are 0.997 and 0.0187 respectively. In Figure 3 we show the Q-Q plots between theoretical deviance quantiles and observed deviances quantiles introduced by García Ben and Yohai (2004). Since the Q-Q plots of the WM, WML and AMSTLE are very similar, we omit the last two. We note that observation 5 sticks out more clearly as an outlier in the plot corresponding to the robust estimator and observation 66 appears as an outlier only in the plots corresponding to the robust estimators. We observe a cluster of 2 or 3 observations in the middle of the Q-Q plots that seems to behave differently. One possible explanation of this apparent misfit is the discontinuity of the theoretical distribution function of the deviances at 0. The reason of this discontinuity is the lack of observations$(y, \mathbf{x})$ such that $P(y|\mathbf{x})$ is close to one.

TABLE 5 ABOUT HERE

FIGURE 3 ABOUT HERE

# 10. CONCLUSIONS

We have introduced the class of P-estimators for GLM which are based on projections. These estimators have order of consistency $n^{1/2}$ and high breakdown point. One shortcoming of these estimators is that they are not asymptotically normal and as consequence inference based on these estimators is difficult to implement. To overcome this problem, we propose one-step scores weighted M-estimators that use a P-estimator as starting point. These estimators are asymptotically normal and inherit most of the robust properties of the P-estimators. A Monte Carlo simulation for the logistic and Poisson model confirm, that at least for this models the one step estimators, have a combination of efficiency and robustness that compares favorably to other estimators. Therefore we consider that they are a very good option for fitting GLM.

## 11. SUPPLEMENTAL MATERIAL

The Supplemental Material available on line contains all the proofs and a numerical algorithm for computing P-estimators.

## REFERENCES

Bianco, A. and Yohai, V. J. (1996), "Robust Estimation in the Logistic Regression Model", in *Robust Statistics, Data Analysis and Computer Intensive Methods,* Proceedings of the workshop in honor of Peter Huber, ed. H. Rieder, Lecture Notes in Statistics, 109, 7–34, New York, Springer-Verlag.

Bondell, H. D. (2005), "Minimum Distance Estimation for the Logistic Regression Model," *Biometrika*, 92, 724–731.

Cantoni, E. and Ronchetti, E. (2001), "Robust Inference for Generalized Linear Models," *Journal of the American Statistical Association*, 96, 1022–1030.

Carroll, R. J., and Pederson, S. (1993), "On Robustness in the Logistic Regression Model," *Journal of the Royal Statistical Society, Ser. B*, 55, 693–706.

Čížek, P. (2008), "Robust and Efficient Adaptive Estimation of Binary-Choice Regression Models", *Journal of the American Statistical Association*, 103, 687–696.

Christmann, A. (1994), "Least Median of Weighted Squares in Logistic Regression With Large Strata," *Biometrika*, 81, 413–417.

Copas, J. B. (1988), "Binary Regression Models for Contamination Data," *Journal of the Royal Statistical Society*, Ser. B, 50, 225–265.

Croux, C., Flandre, C. and Haesbroeck, G. (2002), "The Breakdown Behavior of the Maximum Likelihood Estimator in the Logistic Regression Model", *Statistics and Probability Letters*, 60, 377–386.

Croux, C. and Haesbroeck, G. (2003), "Implementing the Bianco and Yohai Estimator for Logistic Regression," *Computational Statistics and Data Analysis,* 44, 273–295.

García Ben, M. and Yohai, V. J. (2004), "Quantile-quantile Plot for Deviance Residuals in the Generalized Linear Model," *Journal of Computational and Graphical Statistics*, 13, 36–47.

Gervini, D. (2005), "Robust Adaptive Estimators for Binary Regression Models," *Journal of Statistical Planning and Inference*, 131, 297–311.

Kordzakhia, N., Mishra, G. D., and Reiersølmoen, L. (2001), "Robust Estimation in the Logistic Regression Model," *Journal of Statistical Planning and Inference*, 98, 211–223.

Künsch, H., Stefanski, L. and Carroll, R. (1989), "Conditionally Unbiased Bounded-Influence Estimation in General Regression Models, with Applications to Generalized Linear Models," *Journal of the American Statistical Association,* 84, 460–466.

Maronna, R, A., Bustos, O. and Yohai, V. J. (1979), "Bias and Efficiency Robustness of General M-estimators for Regression with Random Carriers," in *Smoothing techniques for curve estimation*, eds. T. Gasser and M. Rosenblatt, Lectures Notes in Mathematics No. 757, 91–111, New York, Springer Verlag.

Maronna, R. A., Stahel, W. and Yohai, V. J. (1992), "Robust-Bias Estimates for Multivariate Scatter Based on Projections," *Journal of Multivariate Analysis*, 42, 141–163.

Maronna, R. A. and Yohai, V. J. (1993), "Bias-Robust Estimates of Regression Based on Projections," *Annals of Statistics*, 21, 965–990.

Maronna, R. A., Martin, R. D. and Yohai, V. J. (2006), *Robust Statistics: Theory and Methods,* Chichister, Wiley.

Martin. R. A., Yohai, V. J. and Zamar. R. H. (1989), "Min-Max Bias Robust Regression," *Annals of Statistics,* **17**, 1608-1630.

Müller, C. H., and Neykov, N. M. (2003), "Breakdown Points of Trimmed Likelihood Estimators and Related Estimators in Generalized Linear Models," *Journal of Statistical Planning and Inference*, 116, 503–519.

Pregibon, D. (1981), "Logistic Regression Diagnostics," *Annals of Statistics*, 9, 705–724.

Rousseeuw, P. and Leroy, A. (1987), *Robust Regression and Outlier Detection*, New York, Wiley.

Simpson, D. G., Ruppert, D. and Carroll, R. J. (1992), "On One-Step GM Estimates and Stability of Inferences in Linear Regression," *Journal of the American Statistical Association*, 87, 439-450.

Stefanski, L. A., Carroll, R. J. and Ruppert, D. (1986), "Optimally Bounded Score Functions for Generalized Linear Models with Applications to Logistic Regression," *Biometrika*, 73, 413–424.

Tyler, D. E. (1994), "Finite Sample Breakdown Points of Projection Based Multivariate Location and Scatter Statistics," *Annals of Statistics*, 22, 1024–1044.

Wedderburn, R. W. M. (1974), "Quasi-likelihood Functions, Generalized Linear Models, and the Gauss–Newton Method," *Biometrika*, 61, 439–447.

Zuo, Y. (2003), "Projection based depth functions and associated medians," *The Annals of Statistics*, 31, 1460-1490

# TABLES

Table 1. Ratio between maximum and minimum value of S in $\mathcal{V}(M_{\lambda,\beta_0}, \varepsilon)$.

| $\varepsilon$ | 0.05 | 0.10 | 0.20 | 0.30 |
|---|---|---|---|---|
| $d^+/d^-$ | 1.13 | 1.29 | 1.81 | 2.91 |

Table 2. Efficiencies Without Outliers for the logistic model

| Sample Size | MSE | Efficiencies with respect to ML | | | | |
|---|---|---|---|---|---|---|
| | ML | WML | WM | PR | AMSTLE | PR-WM |
| 100 | 0.25 | 93.0 | 89.2 | 85.6 | 83.6 | 87.3 |
| 200 | 0.11 | 93.0 | 89.5 | 85.1 | 96.5 | 88.7 |
| 400 | 0.056 | 95.2 | 90.6 | 81.8 | 99.8 | 86.8 |

Table 3. Maximum Mean Squared Errors with **5%** of Outliers for logistic regression

| Sample Size | Estimates | | | | | |
|---|---|---|---|---|---|---|
| | ML | WML | WM | PR | AMSTLE | PR-WM |
| 100 | 1.93 | 0.68 | 0.54 | 0.35 | 0.59 | 0.39 |
| 200 | 1.84 | 0.61 | 0.48 | 0.27 | 0.43 | 0.32 |
| 400 | 1.82 | 0.60 | 0.46 | 0.22 | 0.42 | 0.28 |

Table 4. Efficiencies Without Outliers for Poisson Regression

| Sample Size | MSE | Efficiencies with respect to ML | | | |
|---|---|---|---|---|---|
| | ML | CUBI | RQL | PR | PR-WM |
| 100 | 0.051 | 74.2 | 80.0 | 64.3 | 81.4 |
| 200 | 0.024 | 70.6 | 81.1 | 59.1 | 80.8 |
| 400 | 0.012 | 71.6 | 80.0 | 60.1 | 82.9 |

Table 5. Estimated Coefficients for the Food Stamp Data

| Estimate | INT | TEN | SI | LMI |
|---|---|---|---|---|
| ML | 0.93 (1.63) | −1.85 (0.53) | 0.90 (0.50) | −0.33 (0.27) |
| WML | 5.27 (2.63) | −1.82 (0.53) | 0.68 (0.52) | −1.05 (0.45) |
| WM | 5.42 (1.48) | −1.88 (0.27) | 0.70 (0.27) | −1.08 (0.26) |
| PR | 5.79 | −1.84 | 0.53 | −1.12 |
| PR-WM | 5.67 (1.06) | −1.83 (0.19) | 0.60 (0.23) | −1.11 (0.17) |
| AMSTLE | 5.49 | −1.82 | 0.67 | −1.09 |

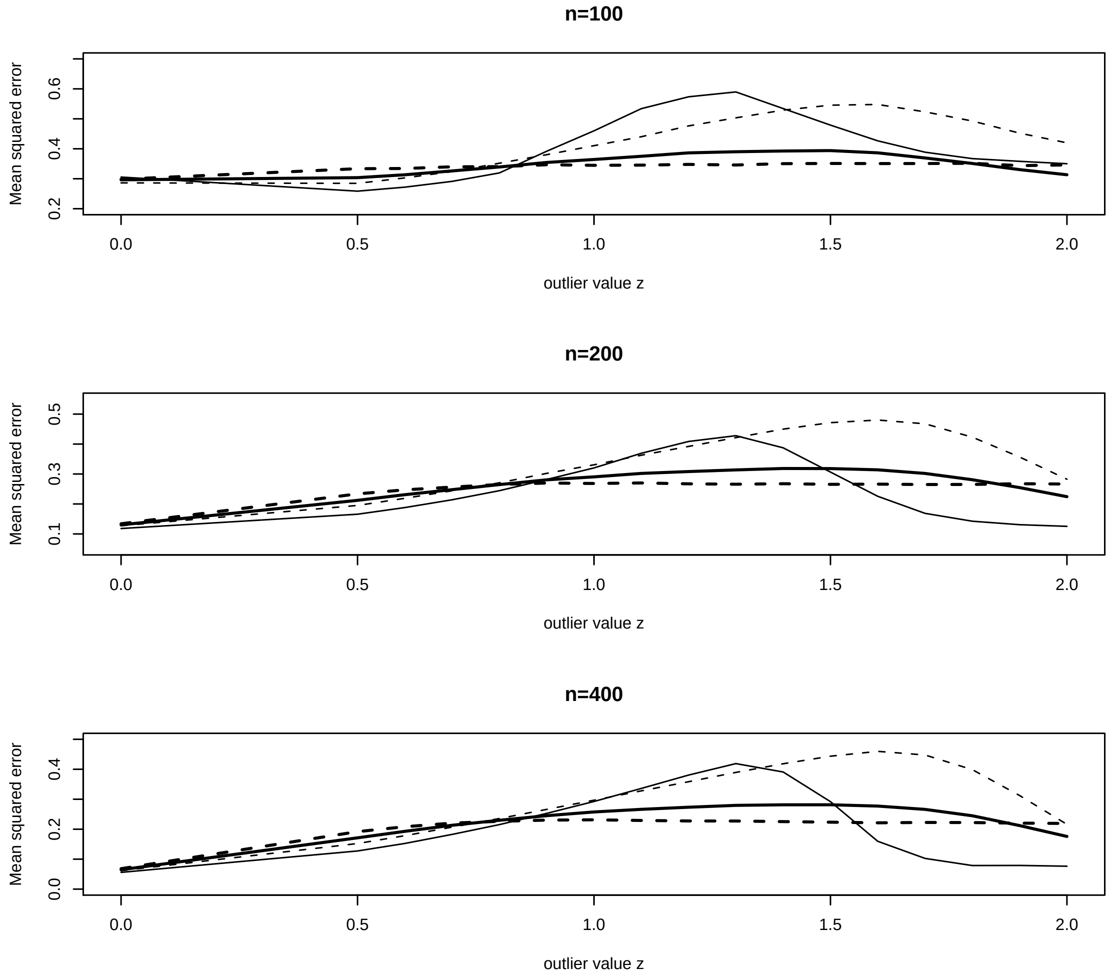


Figure 1: Mean Squared Errors Under Contamination for Logistic Regression. Solid Line : AMSTLE, Dashed Line: WM, Bold Solid Line: WM-PR and Bold Dashed Line: PR.

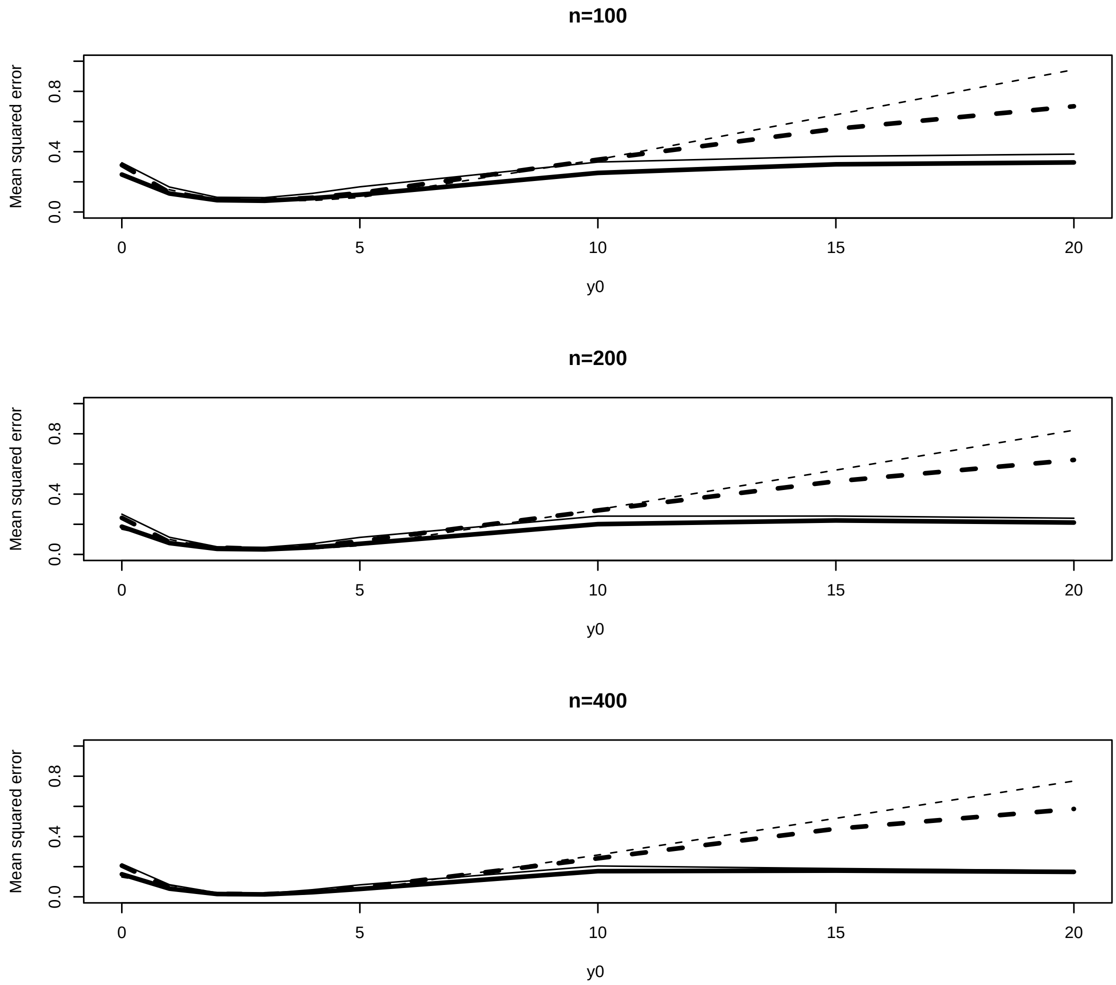


Figure 2: Mean Squared Errors Under Contamination for Poisson Regression Dashed Line: RQL, Bold Dashed line: CUBI, Solid Line: PR and Bold Solid Line: WM-PR.

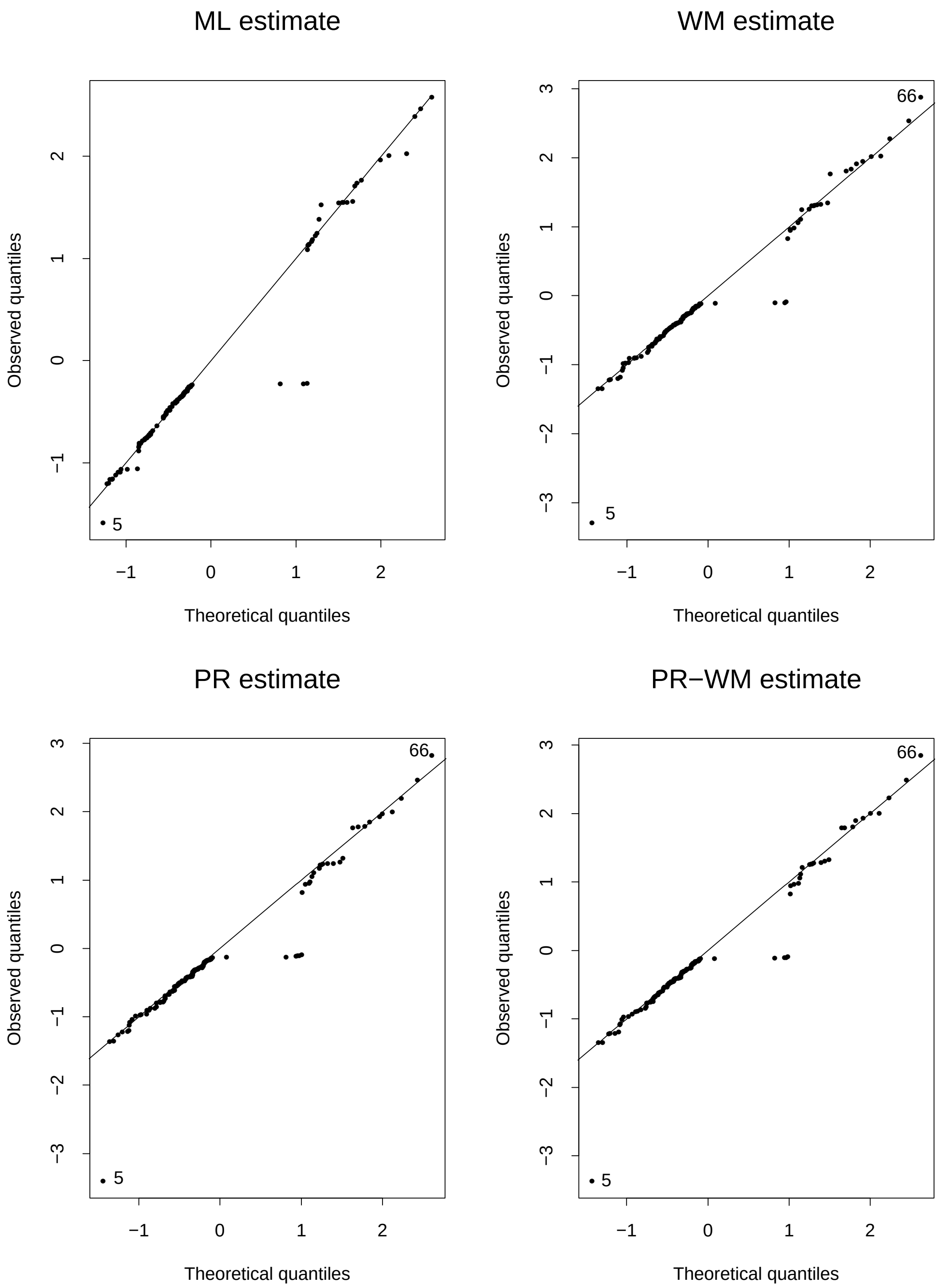


Figure 3: Q-Q Plots of Food Stamp Data